

Psicometría

Conceptos básicos y aplicaciones prácticas con Rcommander

Paula Elosua

ARGITALPEN ZERBITZUA
SERVICIO EDITORIAL

eman ta zabal zazu



Universidad del País Vasco Euskal Herriko Unibertsitatea

www.ehu.es
ISBN: 978-84-694-5566-1

Prólogo.....1

PARTE I. R y Rcommander

1 INTRODUCCIÓN 2

1.1 Instalación de R..... 6

1.2 Primer contacto..... 8

1.3 Opciones de trabajo con R..... 12

1.4 Obtención de ayuda sobre R..... 14

1.5 Paquetes 19

1.6 R a modo de calculadora 20

1.7 Notas básicas sobre R 21

2 RCOMMANDER 23

2.1 Instalación de Rcommander 23

2.2 Descripción de Rcommander 26

2.3 Opciones generales de la barra de menús..... 29

2.4 Lectura e importación de datos 33

 2.4.1 Opciones de configuración 42

3 MANIPULACION DE DATOS Y TRANSFORMACION DE VARIABLES..... 44

3.1 Manipulación de datos 44

 3.1.1 Seleccionar conjunto de datos activo..... 45

 3.1.2 Actualizar conjunto de datos activo..... 45

 3.1.3 Ayuda sobre el conjunto de datos activo 46

 3.1.4 Variables del conjunto de datos activo 46

 3.1.5 Establecer nombres de casos 47

 3.1.6 Filtrar el conjunto de datos activo 47

 3.1.7 Eliminar los casos con valores omitidos..... 49

 3.1.8 Guardar el conjunto de datos activo 50

3.2 Manipulación de variables 51

 3.2.1 Recodificar variables 52

 3.2.2 Calcular una nueva variable 54

 3.2.3 Añadir número de observaciones al conjunto de datos 55

 3.2.4 Tipificar variables 56

 3.2.5 Convertir variable numérica en factor 57

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

3.2.6	Segmentar una variable numérica.....	58
3.2.7	Reordenar niveles de factor	60
3.2.8	Renombrar variables.....	60
3.2.9	Eliminar variables del conjunto de datos.....	61

PARTE II. Construcción y análisis de escalas

4	ANÁLISIS DE ESCALAS	62
4.1	Presentación del ejemplo.....	63
4.2	Lectura de datos.....	65
4.3	Corrección de la prueba. Recodificación de los ítems	69
4.3.1	Ítems inversos	71
4.4	Corrección de la prueba. Obtención de la puntuación.....	76
4.5	Análisis de ítems.....	80
4.6	Análisis de ítems. Índices de forma	82
4.6.1	Media aritmética	82
4.6.2	Índice de atracción.....	83
4.6.3	Gráfico de la distribución de respuestas	86
5	FIABILIDAD. ÍNDICE DE DISCRIMINACIÓN Y COEFICIENTE ALFA	89
5.1	Índices de relación. Matriz de correlaciones	89
5.2	Índices de relación. Índice de discriminación	93
5.3	Fiabilidad.....	94
5.3.1	Modelo lineal clásico.....	95
5.3.2	Coefficiente de fiabilidad.....	98
5.3.3	Procedimientos empíricos para la estimación del coeficiente de fiabilidad	99
5.3.4	Estimación de la puntuación verdadera	101
5.3.5	Coefficiente alfa.....	106
5.3.6	Índices derivados del modelo factorial.....	108
5.4	Rcommander y la estimación de la fiabilidad	111
5.4.1	Matriz de correlación.....	111
5.4.2	Índice de discriminación y coeficiente alfa	113
5.4.3	Coefficiente Theta.....	116
5.4.4	Error estándar de medida y estimación de la puntuación verdadera	116
6	VALIDEZ.....	119
6.1	Sesgo.....	122
6.2	Fuentes de evidencia	124
6.2.1	Contenido.....	124

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

6.2.2	Proceso de respuesta.....	128
6.2.3	Estructura interna.....	129
6.2.4	Relaciones con otras variables.....	134
6.2.5	Generalización de la validez.....	136
6.2.6	Consecuencias	137
7	ESTRUCTURA INTERNA. DIMENSIONALIDAD	140
7.1	Modelo de análisis factorial	141
7.1.1	Supuestos y consecuencias	143
7.1.2	Reproducción de la matriz de correlación	145
7.1.3	Representación matricial	146
7.1.4	Extracción factorial.....	148
7.1.5	Modelo de componentes principales	151
7.1.6	Rotación factorial	153
7.2	Etapas en la ejecución de un análisis factorial.....	156
7.2.1	Interpretación de la solución factorial	157
7.3	Análisis factorial y Rcommander	163
7.3.1	Modelo unidimensional	163
7.3.2	Modelo bidimensional	168
7.3.3	Representación gráfica de la solución factorial.....	171
7.3.4	Estimación de las puntuaciones factoriales	172
7.3.5	Análisis de componentes principales.....	173
8	FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM	176
8.1	Conceptos básicos	177
8.2	Estandarización.....	181
8.3	Regresión logística	182
8.3.1	Consideraciones sobre el funcionamiento diferencial del ítem	184
8.4	Funcionamiento diferencial del ítem y Rcommander	186
8.4.1	Regresión logística	186
8.4.2	Diferencias entre medias ponderadas	188
9	COMPARACIÓN ENTRE GRUPOS. (ALICIA LÓPEZ-JÁUREGUD).....	192
9.1	Comparación de dos grupos. Pruebas de comparación de medias	193
9.1.1	Prueba para una media.....	194
9.1.2	Comparación de dos medias	197
9.1.3	Pruebas para las varianzas	204
9.2	Pruebas para las proporciones	206
9.2.1	Tablas de doble entrada. Inferencia.....	211
9.3	Comparación entre más de dos grupos (ANOVA)	215

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

9.3.1	ANOVA de un factor.....	215
9.4	Variación entre grupos y variación intra grupos.....	217
9.4.2	ANOVA de un factor mediante Rcommander	225
9.4.3	Análisis de varianza de dos factores.....	228
9.5	Efectos principales e interacción	230
9.5.2	Análisis de varianza con Rcommander.....	235
10	TEST-CRITERIO. REGRESIÓN LINEAL	238
10.1	Regresión simple	239
10.1.1	Ecuación de la recta.....	239
10.1.2	Relaciones entre variables en las ciencias sociales	240
10.1.3	Estimación de la recta de regresión	241
10.1.4	Error típico de estimación.....	244
10.1.5	Estimación del criterio.....	244
10.1.6	Descomposición de la varianza y coeficiente de determinación	245
10.1.7	Modelo lineal.....	249
10.2	Regresión simple con Rcommander	250
10.3	Regresión múltiple	254
10.3.1	Modelo.....	254
10.3.2	Supuestos	257
10.3.3	Estimación de los parámetros	258
10.3.4	Descomposición de la varianza	259
10.3.5	Coficiente de determinación múltiple.....	261
10.3.6	Inferencias sobre el modelo.....	263
10.4	Correlación múltiple con Rcommander	265
10.4.1	Diagnósticos del modelo de regresión.....	269
10.4.2	Representación gráfica de distribuciones multivariadas	271
11	INTERPRETACIÓN NORMATIVA DE PUNTUACIONES. BAREMOS	274
11.1	Transformaciones lineales.....	274
11.2	Transformaciones no lineales	276
11.3	Baremos con Rcommander	278
12	REFERENCIAS BIBLIOGRÁFICAS	284

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Prólogo

El estudio de una escala de actitudes, un inventario de personalidad o una prueba de aptitud numérica comprende un conjunto de análisis cuya finalidad es garantizar la fiabilidad de los datos y la validez de las inferencias derivadas de ellos. El corpus que da sustento teórico, aplicado y ético a tales menesteres se cimienta sobre la disciplina denominada psicometría. La psicometría se ocupa del proceso de construcción y validación de escalas cuyo objetivo es medir variables de naturaleza psicológica.

El libro trata en profundidad y de forma actualizada los dos grandes temas de la medición empírica, la fiabilidad y la validez, así como la tecnología implicada en su análisis: regresión simple y múltiple, análisis factorial, equiparación de puntuaciones, funcionamiento diferencial del ítem, análisis de varianza. Todo ello se aborda desde una doble perspectiva teórica y aplicada. La primera aporta las bases necesarias para su comprensión y la segunda contribuye a su asimilación por medio de un ejemplo real que se ilustra con la ayuda del mejor software disponible para el análisis de datos y la psicometría, R.

R es un entorno de trabajo que se distribuye bajo licencia libre, y que ofrece al estudiante, profesor o investigador todas las herramientas necesarias para el análisis de datos y de escalas. En este manual, el autor introduce al lector en R Commander, un paquete gratuito, fiable y fácil de usar que incluye las funciones precisas para analizar datos y escalas.

El libro se estructura en dos partes; en la primera de ellas se introducen el entorno R y R Commander y se muestran las funciones para el manejo de datos y modificación de variables. En la segunda parte, se incluyen los conceptos psicométricos básicos y se ejemplifican aplicándolos sobre una escala de datos real.

1 Introducción

En el ámbito de las ciencias sociales se dispone de una variada gama de programas para el análisis de datos, que bien bajo licencia privativa o bajo licencia libre cumplen las exigencias del usuario independientemente del tipo de explotación o de la sencillez o complejidad de los modelos que quiera evaluar. Entre los primeros, se podrían citar programas tan conocidos como PASW (antes SPSS), SAS, STATISTICA, Systat, Stata o GenStat que han copado durante años el mercado, y han sido, y son utilizados como herramienta para la docencia por prácticamente todas las universidades tanto españolas, europeas como americanas. Desde sus primeras versiones hasta las actuales, estos programas han ido incorporando modelos de análisis cada vez más complejos a la par que han ido mejorando sus interfaces gráficas (GUI Graphical User Interfaces). El desarrollo de los menús contextuales basados en la gestión de ventanas los han convertido en programas de fácil manejo, y al mismo tiempo ha permitido una socialización en el uso de modelos de análisis complejos que son ahora accesibles al usuario medio.

Sin embargo, el carácter comercial y la licencia privativa de todos ellos es una limitación que valoro negativamente. Desde el punto de vista del profesor que asume la responsabilidad de ofrecer y facilitar al alumno las bases para una formación continua que no finaliza con el abandono de la universidad, la accesibilidad es una característica substancial. La disponibilidad de una herramienta, siempre asequible a coste 0, es un factor que debiera considerarse en la docencia, por ofrecer tanto al alumno como al futuro profesional una autonomía de trabajo efectiva que facilita el proceso de enseñanza-aprendizaje.

Añádase, que los programas comerciales comparten otra característica referida a su funcionamiento. Son herramientas cerradas, prediseñadas, y en consecuencia difícilmente manipulables, que operan a través de algoritmos

Conceptos básicos y aplicaciones prácticas con Rcommander

empaquetados cuya adaptación a la situaciones concretas de análisis se torna ardua y a veces imposible.

Sin embargo ambas situaciones pueden cambiar de una forma atractiva tanto para el alumno no especializado en análisis estadístico, como para el profesor que imparte docencia en análisis de datos. El cambio se produce a través de la adopción de una herramienta que cumple formalmente todos los criterios de calidad exigibles a un programa para el tratamiento de datos y que además de ser de libre distribución es moldeable en función de los intereses del usuario. La herramienta que cumple este triple criterio, calidad, accesibilidad y adaptabilidad es R.

R viene avalado técnicamente por ser el entorno de mayor implementación entre la población estadística, lo cual garantiza su validez; es la implementación GNU de S, lo cual garantiza la gratuidad de su distribución. Además R ofrece la capacidad de combinar análisis prediseñados con análisis específicos adaptados a cada situación concreta, lo cual garantiza su versatilidad. Validez, gratuidad y versatilidad son las cartas de presentación de R frente al resto de software para el análisis de datos.

El entorno R, que era utilizado hasta hace poco tiempo, prácticamente sólo por estadísticos o personal altamente especializado, se convierte con la ayuda de una interfaz gráfica (por ejemplo, Rcommander de John Fox) en una herramienta para el análisis de datos con un potencial inusitado, a la vez que cumple los criterios de facilidad de uso y gratuidad. Estas características convierten a R en un poderoso aliado para la enseñanza del análisis de datos, pero además, ofrece al alumno o profesional una herramienta continuamente actualizada que le brinda la posibilidad de una autonomía de trabajo no disponible bajo cualquier ningún otro entorno (Elosua, 2009).

R es un entorno de programación y análisis estadístico y gráfico derivado del lenguaje de programación S (Becker, Chambers y Wilks, 1988;

Conceptos básicos y aplicaciones prácticas con Rcommander

Chambers y Hastie, 1992; Venables y Ripley, 2000). Existe una versión de este lenguaje distribuida por *Insightful Corporation* bajo el nombre comercial de S-Plus, y una versión libre con código abierto conocida como R. Esta última fue desarrollada por Ross Ihaka y Robert Gentleman (esta es una de las razones del nombre R; Ihaka y Gentleman, 1996; otra de las razones deriva de la fonética de R “our” –nuestro- que lo enlaza con la filosofía del software libre) del Departamento de Estadística de la Universidad de Auckland (Nueva Zelanda). La primera versión de R se difundió rápidamente y la expansión es hoy irrefrenable. Desde su creación, R se alimenta y crece con los trabajos de investigadores provenientes de prácticamente todas las ramas del conocimiento. Las aportaciones desinteresadas de funciones y librerías de propósito tanto general como específico hacen de R un entorno dinámico formado por una comunidad en movimiento continuo y acelerado que se inscribe dentro de la filosofía del software libre.

R, en tanto en cuanto software libre, se inscribe dentro del proyecto GNU *General Public Licence*. (Licencia Pública General, GNU). Se trata de una licencia creada por *Free Software Foundation* (Fundación para el software libre) organización fundada por Richard Matthew Stallman (rms) en el año 1985. El principal propósito de la licencia GNU es declarar la libertad del uso, modificación y distribución del software y protegerlo de intentos de privatización que puedan de algún modo restringir su uso (el contenido de la licencia puede consultarse en el sitio <http://www.gnu.org/copyleft/gpl.html>). Dentro de esta licencia se distribuyen un sin fin de programas, muchos de los cuales son versiones libres del software informático generalista más utilizado. De entre ellos tal vez los más extendidos sean la suite ofimática *OpenOffice*, el navegador *Mozilla*, los artículos de *wikipedia*, el sistema operativo *GNU/Linux*, o el editor de textos *Emacs*.

Parte de la vasta información disponible sobre R es accesible a través de la web CRAN (*Comprehensive R Archive Network*; <http://cran.r-project.org/>), sitio oficial de R. Es la página base del proyecto R, desde la cual se puede descargar la última versión del programa (un equipo formado por unas doce personas, R

Conceptos básicos y aplicaciones prácticas con Rcommander

Development Core Team; asumió en 1997 las labores de actualización semestral del código de R), consultar manuales sobre R, obtener ayuda sobre su funcionamiento a través de un sistema de ayuda online, y en definitiva, estar al corriente de los movimientos en este entorno de trabajo. La bibliografía sobre R es amplia, porque a la propiamente desarrollada para R se añade la bibliografía sobre S o S-plus. Entre los libros más importantes es imprescindible citar el manual de referencia que desarrolla y actualiza con cada versión de R el *R Development Core Team* (2008) y el ya clásico y excelente volumen de Venables y Ripley (2002) que se ha convertido en el libro de cabecera de los usuarios de R y S “*Modern Applied Statistics with S-plus*”. Como obras introductorias, las más aconsejables serían el sencillo trabajo de Venables, Smith y the R Development Core Team (2007) publicado bajo el título “*An Introduction to R*” y el volumen de Paradis (2005) titulado “*R for Beginners*”. Existen traducciones al castellano de ambas en el sitio <http://cran.es.r-project.org/>, donde también pueden consultarse los textos originales. La obra de Dalgaard (2002), “*Introductory Statistics with R*”, es un volumen compacto que cubre el contenido de un curso básico de estadística, y al mismo tiempo introduce al lector en los conceptos elementales de la programación en R. Son más recientes el libro de Braun y Murdoch (2007) que ofrece un primer curso en estadística utilizando R, y la obra publicada por Crawley (2008) que en sus 900 páginas ilustra el proceso de modelado estadístico con R. También es altamente aconsejable el libro de Fox (2002) titulado “*An R and S-plus Companion to Applied Regression*”. Para usuarios iniciados la última actualización de la obra de Chambers (2007), uno de los artífices del lenguaje S, es una excelente opción. La transición entre el SPSS o SAS al R puede acompañarse de la lectura del trabajo de Muenchen que con el título “*R for SAS and SPSS Users*” ofrece dos productos; un corto y accesible en red ([http://rforsasandspssusers.com /](http://rforsasandspssusers.com/)), y otro de 470 páginas publicado recientemente por Springer (Muenchen, 2009).

En R confluyen características que convierten a este entorno en algo único: es libre, tiene versiones para distintas plataformas (Windows, Linux, Unix, o Mac), está continuamente actualizado y siempre a la vanguardia de los

Conceptos básicos y aplicaciones prácticas con Rcommander

más avanzados modelos estadísticos. R es un software libre que se inscribe dentro del proyecto *GNU* (acrónimo recursivo de *GNU in not UNIX*), lo que significa que los usuarios pueden utilizarlo, modificarlo y actualizarlo en función de sus necesidades. La información disponible sobre R es accesible a través de la web CRAN (*Comprehensive R Archive Network*; <http://cran.r-project.org/>).

R es más que un software para el análisis de datos; es un entorno de programación en el que se integran multitud de paquetes que continuamente van incrementando sus capacidades. Incluye desde funciones básicas relacionadas con los análisis descriptivos de datos, hasta los modelos más complejos concernientes con los últimos avances en el campo de la estadística, la psicometría y el análisis de datos en entornos como la psicología, biología, medicina o informática.

Aparte de las capacidades de análisis estadístico, R es un potentísimo generador de gráficos que cuenta con numerosas y variadas funciones y librerías diseñadas con esta finalidad. Es posible componer un simple plot, definir figuras extremadamente complejas e incluso crear animaciones. Puede disfrutarse de una muestra de gráficos generados con R en el sitio <http://addictedtor.free.fr/graphiques/>. Todos ellos evidencian la versatilidad y posibilidades de R.

1.1 Instalación de R

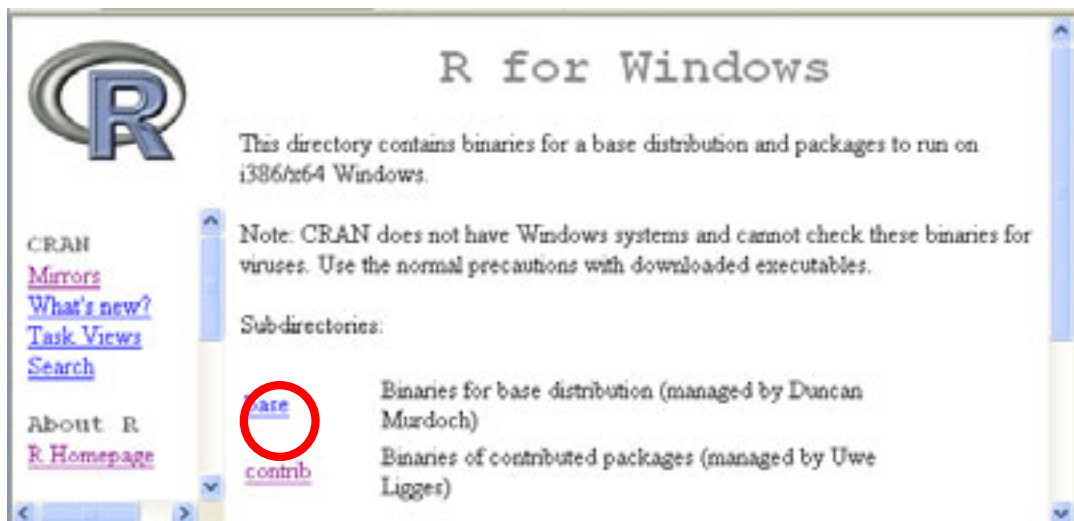
El primer paso en la utilización de R comienza con su instalación. El proceso, aunque similar para las distintas plataformas, será brevemente descrito para Windows (aunque parezca una contradicción GNU-Windows). La instalación de R bajo Windows se lleva a cabo directamente desde la página de CRAN (*Comprehensive R Archive Network*) a la que se puede acceder a través de la dirección. <http://cran.r-project.org/>.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



1.1. Figura. Página CRAN

En esta página, una vez seleccionada la opción Windows, aparecerá una pantalla como la que se muestra en la siguiente figura,



1.2. Figura. Bajar R

En ella se seleccionará la opción `base` que permitirá la instalación del paquete `base` o paquete básico para la ejecución de R. Las funciones incluidas en `base` son fácilmente ampliables a través de los más de 2100 paquetes o contribuciones añadidas al entorno R; cantidad que va incrementándose día a día.

Conceptos básicos y aplicaciones prácticas con Rcommander



1.3. Figura. Archivos R

El archivo de instalación de R es el archivo con extensión `.exe` (R2.10.1.exe; esta es la versión activa en el momento de redacción de este manual). Una vez seleccionado el archivo es posible guardarlo o ejecutarlo directamente. El tamaño del archivo (*32 megabytes*) y el tipo de conexión a Internet hacen aconsejable seleccionar un repositorio (mirror; espejo de CRAN) cercano al lugar donde habitamos para agilizar la transferencia del fichero.

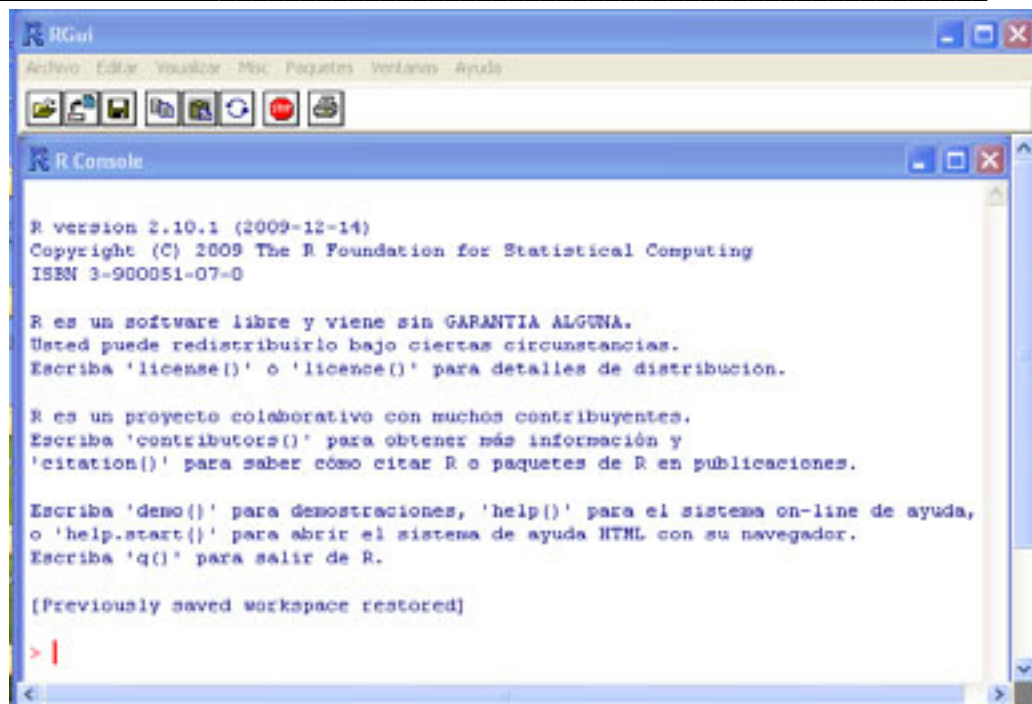
Una vez ejecutado el archivo de instalación, éste preguntará por el idioma deseado, así como por el directorio en el que se quiera instalar R. Por defecto, el programa se instalará en `C:\Archivos de programa\R\R2.10.1`, opción que puede modificarse a conveniencia.

1.2 Primer contacto

Una vez instalado el programa es posible ejecutarlo a través del icono de acceso directo que aparecerá en el escritorio, o bien a través de las ventanas Inicio>Todos los programas>R>R2.10.1. La primera pantalla que aparece cuando se ejecuta R recibe el nombre de consola de R (*R console*). En ella puede comprobarse la versión instalada (R versión 2.10.1).

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander



1.4. Figura. Consola R

La presencia del símbolo del sistema o `prompt`, en color rojo, (`>`) indica que R está listo para recibir comandos. Por ejemplo, tecleando `2 + 3` (pruebe el lector cualquier otra operación) y pulsando `<ENTER>`, R devolverá el resultado de la operación aritmética.

```
> 2+3
[1] 5
>
```

El `[1]` indica el orden de aparición de los resultados. En este caso sólo se ha solicitado un resultado. Si la salida hubiera sido más compleja R reconocería cada uno de los elementos por un número correlativo entre corchetes que indica la posición del resultado. El símbolo de sistema “`>`” indica que R está listo para recibir otra instrucción.

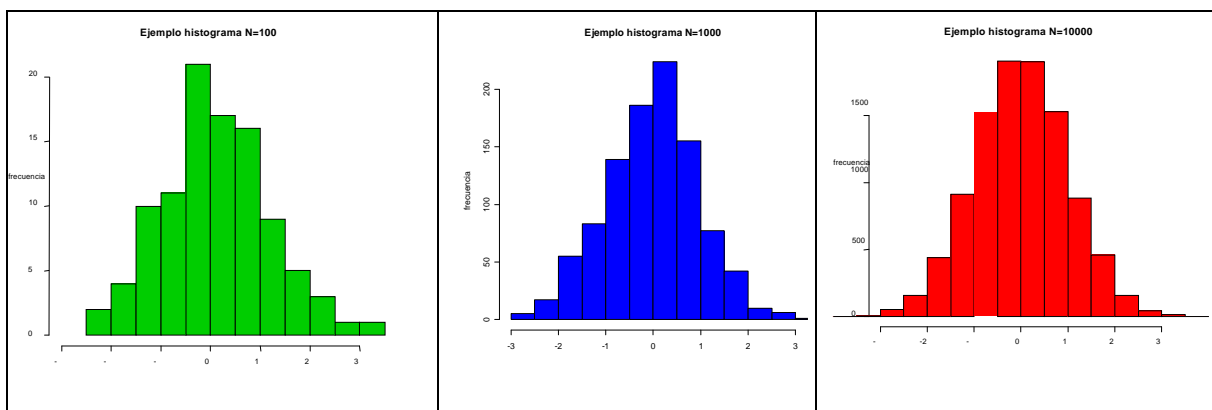
Si se desea generar una serie aleatoria de números que se distribuyan según la ley normal, y después representar su distribución por medio de un

Conceptos básicos y aplicaciones prácticas con Rcommander

histograma el comando a emplear sería el siguiente (más adelante se explicarán éste y otros comandos).

```
hist (rnorm(100), col=3, xlim=c(-3,3),main="Ejemplo histograma N=100",xlab=" ", ylab="frecuencia")
```

Se ha ejecutado este comando en tres ocasiones diferentes variando el número de elementos generados. La sucesión de histogramas muestra la aproximación a la distribución normal de la muestra a medida que aumenta el número de elementos generados (N=100, N=1000; N=10000). La representación gráfica obtenida sólo quiere ser un ejemplo de la facilidad y versatilidad de R, propiedades ambas que se irán descubriendo a lo largo del presente manual.



1.5. Figura. Ejemplo de representaciones gráficas

Además del símbolo del sistema (>), la consola de R puede presentar un símbolo “+” (también en color rojo) para indicar que la instrucción dada a R es incompleta y que no puede ejecutarse sin antes finalizar correctamente la secuencia de comandos. Por ejemplo si se teclea “3+5+” (“la instrucción no está completa y por tanto R lo advierte con la presencia del símbolo “+”. Si ante el símbolo + se teclea algo que pueda concluir la instrucción a ejecutar (por ejemplo “7*3”) R devolverá el resultado de la operación, 29.


Conceptos básicos y aplicaciones prácticas con Rcommander

Otro símbolo importante en el entorno de programación R es el símbolo “#”, utilizado para añadir comentarios no ejecutables durante la construcción de un código.

Es posible escribir más de una función en la misma línea de comandos, para ello basta utilizar un carácter delimitador entre ellas“;”.

```
> 3+5+(  
+  
  
> 3+5+(  
+ 7*3)  
[1] 29  
> 2+3;4*7  
[1] 5  
[1] 28
```

Para abandonar R tras una sesión de trabajo son varias las opciones disponibles:

- 1.- Teclear directamente en la consola `q` ().
- 2.- Acceder a la opción `Salir` a través de la barra de menús `Archivo>Salir`.
- 3.- Pulsar directamente sobre el icono de salida .


Antes de abandonar la sesión de trabajo el sistema preguntará si se desea guardar una copia (la imagen) del área de trabajo. Si la respuesta es afirmativa se salvarán todos los “objetos” creados en el área de trabajo durante la última sesión; en caso contrario no se podrán recuperar los objetos creados en una sesión posterior. También es posible salvar sólo parte de los objetos; para ello el comando a utilizar sería `save(nombre objeto, file="nombre archivo")`. Por defecto, la extensión utilizada para almacenar un área de trabajo es “.Rdata”. Los comandos que se han ido generando se almacenan en

Conceptos básicos y aplicaciones prácticas con Rcommander

el fichero `.Rhistory` del cual son fácilmente recuperables desde cualquier editor de texto.

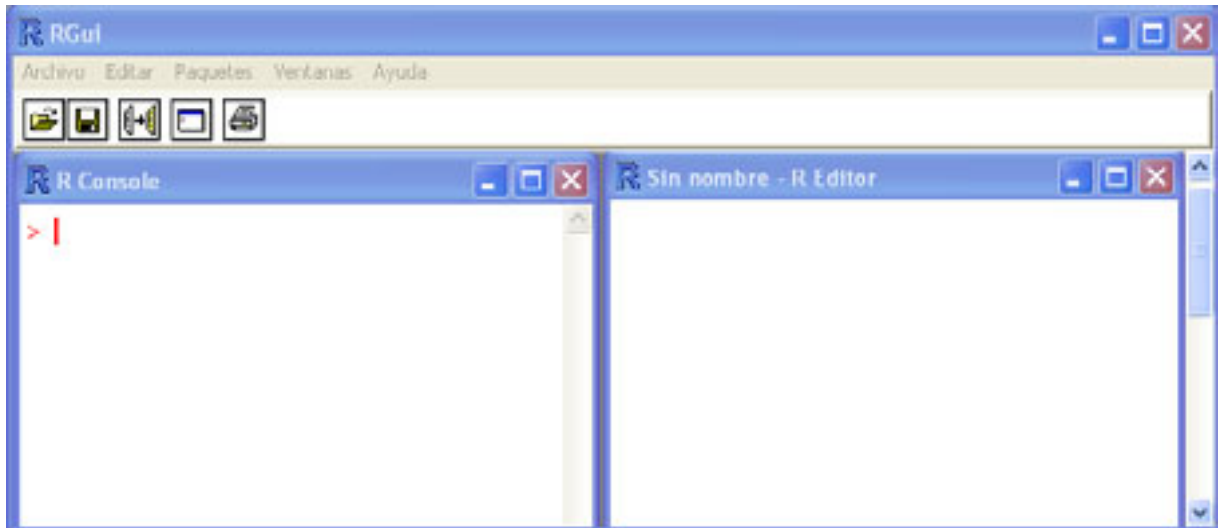
1.3 Opciones de trabajo con R

Bajo el entorno Windows (que es la plataforma que describiremos a lo largo de esta introducción a R) existen varias opciones de trabajo, es decir, diferentes vías para la introducción y ejecución de instrucciones. Es posible utilizar sólo una de ellas, o utilizar varias simultáneamente. Las necesidades del usuario determinarán en cada caso la elección del modo de operar. Básicamente se dispone de 4 modos diferentes para trabajar con R:

1. **Trabajar directamente sobre la ventana de comandos (R console)**, tecleando los comandos en la línea de “símbolo del sistema” “>”. Es un modo de trabajo interactivo que permite ejecutar acciones simples porque sólo puede ejecutarse un comando cada vez. Las flechas de dirección facilitan el modo de operar con la consola. La flecha hacia arriba (↑) permite acceder a comandos previos y la flecha hacia abajo (↓) da acceso a comandos posteriores al actual. Las flechas hacia la izquierda (←) y hacia la derecha (→) mueven el cursor en dichos sentidos dentro de una línea. Los comandos pueden copiarse y pegarse, como en cualquier editor de textos, con las funciones `Ctrl+C` y `Ctrl+V`.
2. **Abrir y trabajar en una ventana de escritura (*Script window*)** a la que se accede seleccionando la opción Nuevo Script de la barra de menús, `Archivo>Nuevo Script`. Esta opción permite crear códigos más complejos que pueden ejecutarse por bloques. Los códigos se teclearán directamente en esta ventana o podrán ser copiados de un archivo ya existente en formato ASCII. Para ejecutar un subconjunto de comandos, se debe de marcar la sección a ejecutar y pulsar `CTRL+R` o en su defecto el icono de ejecución . Cuando

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

se utiliza la ventana de escritura, las salidas de R se muestran en la consola, por lo tanto es recomendable mantener simultáneamente abiertas ambas ventanas: consola y ventana de comandos.



1.6. Figura. RConsole y ventana de escritur

- Utilizar programas específicos de edición de códigos.** Entre los más utilizados se pueden citar *Tinn-R*, (<http://www.sciviews.org/Tinn-R/>), *WinEdit* (<http://www.winedt.com/>) o *Emacs* (<http://www.gnu.org/software/emacs/>), software creado inicialmente para el entorno Unix bajo la filosofía GNU. Son programas de edición especializados que facilitan enormemente el trabajo con códigos o programas (conjuntos de instrucciones, comandos necesarios para la ejecución de determinados trabajos) más o menos complejos. Es la opción preferida de los usuarios avanzados de R.
- Utilizar interfaces gráficas.** Existen varias GUI (*Graphical user interfaz*) que facilitan el trabajo con R a personas desconocedoras de este entorno de programación y análisis. Entre las GUI disponibles mencionaremos, *R.NET* (<http://www.u.arizona.edu/~ryckman/RNet.php>), *Poor Man's GUI* (<http://www.math.csi.cuny.edu/pmg>), *Rkward*

Conceptos básicos y aplicaciones prácticas con Rcommander

(<http://es.wikipedia.org/wiki/RKward>) todavía no disponible para Windows, *Rcommander* (Fox, 2005), o *RExcel* (<http://www.statconn.com>) que se distribuye bajo licencia privativa. De entre ellas la más extendida es *Rcommander*; entre sus cualidades podrían destacarse la adecuación hacia los contenidos metodológicos impartidos en las ciencias sociales y la simplicidad de uso. *Rcommander* es un paquete adicional de R (*Rcmdr*) creado por John Fox que añade una interfaz gráfica con un sistema de ventanas y menús desplegable que convierten a R en un entorno amigable. Es el intermediario perfecto para acercar al usuario habitual de paquetes comerciales al entorno de programación R, permitiendo una transición sencilla hacia esta nueva filosofía de trabajo.

1.4 Obtención de ayuda sobre R

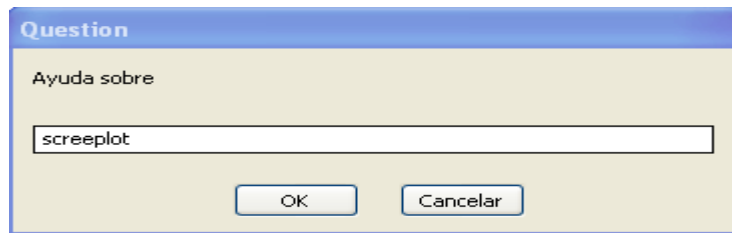
El entorno R es extraordinariamente amplio e incorpora diversas fuentes de ayuda sobre procedimientos, comandos, paquetes o sobre la aplicación de determinados modelos matemáticos que resultan de gran utilidad ante cuestiones o dudas que surgen durante una sesión de trabajo.

El modo básico de obtención de ayuda en R es a través de la opción *Ayuda* accesible desde la barra de menús, desde ella se pueden seleccionar varias alternativas

1.- La opción *Ayuda>Funciones R(texto)* abre una ventana interactiva en la que se tecleará el tópico sobre el que se busca información.

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander



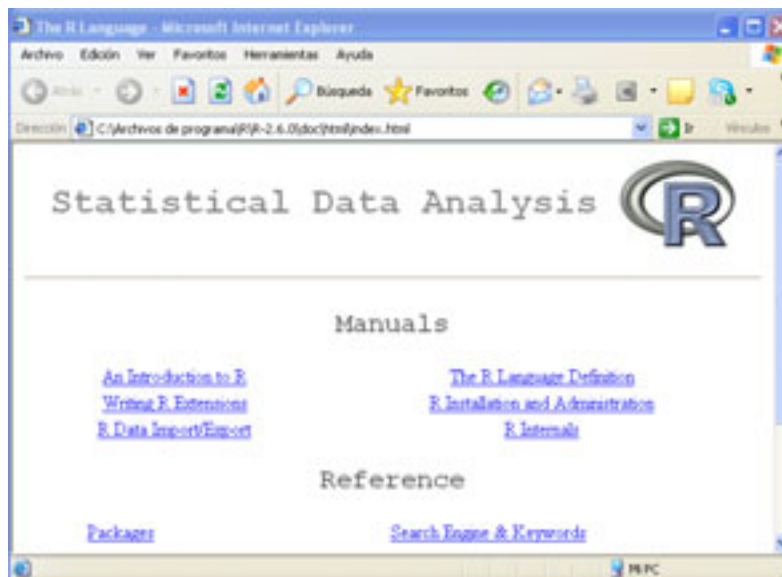
1.7. Figura. Menú de ayuda

Este modo de obtener información sobre un tópico concreto, es equivalente a teclear `help(screeplot)` sobre la consola R, o incluso `?screeplot`.

```
> help(screeplot)
> ?screeplot
```

Ambos modos dan acceso a una página de información sobre el tópico indicado.

2.- La opción `Ayuda>Ayuda Html`, acciona el explorador definido por defecto y muestra un pantalla a través de la cual se accede a la información almacenada en la memoria.



1.8. Figura. Ayuda Html

Conceptos básicos y aplicaciones prácticas con Rcommander

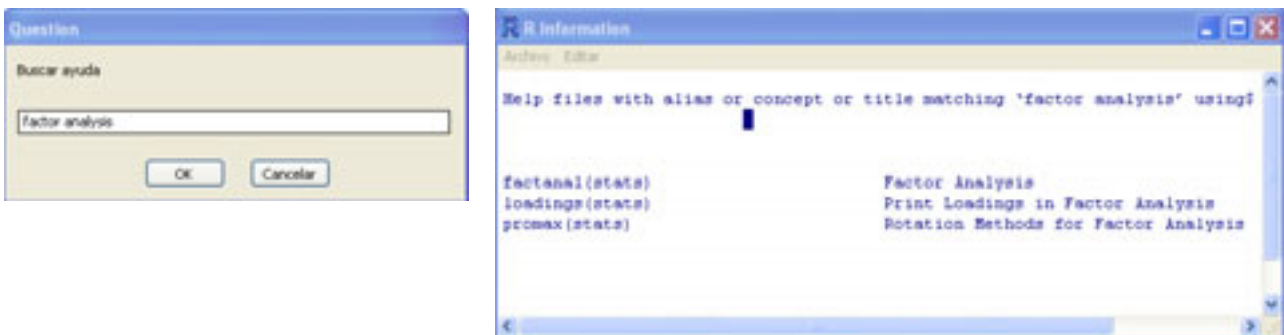
Podemos acceder a la misma página a través del comando `help.start()` que se tecleará directamente sobre la consola R.

3.- Es posible ampliar la base de la exploración y optar por una búsqueda en la web de R. Para ello es suficiente con seleccionar la opción de búsqueda `Ayuda>search.r.project.org`.

4.- Existen otras dos funciones de búsqueda `help.search()` y `apropos()`, que pueden ser útiles cuando se desea encontrar información sobre una función o modelo determinado que no se sabe exactamente donde localizar. Por ejemplo: supongamos que se quiere obtener información sobre el análisis factorial “*factor analysis*”; A través de la barra de menús se seleccionará la opción `Ayuda>Buscar Ayuda...` por medio de la cual se accede a una página en la que se tecleará el texto que describe lo que se desea buscar. Se puede obtener la misma información tecleando directamente sobre la consola lo siguiente:

```
help.search("factor analysis")
```

Este comando ofrece como salida una lista de todas las funciones cuyas páginas de ayuda contienen la palabra situada entre comillas. En este caso “*factor analysis*”.



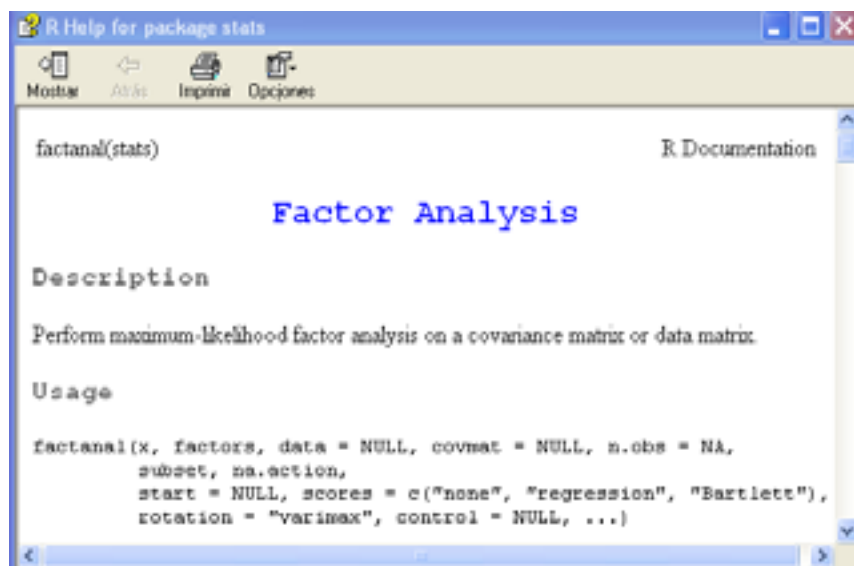
1.9. Figura. Help.search()

Conceptos básicos y aplicaciones prácticas con Rcommander

La salida indica que existe una función llamada `factanal` en la librería `stats`, en la que también se encuentran las funciones `loadings` y `promax`. Si se desea acceder al contenido de estas funciones, a su descripción, bastaría con teclear en la consola el nombre de la función, y la librería que la contiene:

```
help(factanal, package=stats)
```

Este comando abriría una página de ayuda sobre la función `factanal` en la que se incluye una descripción de la misma, su uso, el modo de especificar sus argumentos y un conjunto de ejemplos.



1.10. Figura. Página de información sobre comando

Si tecleamos directamente sobre la consola

```
> apropos("vector")
```

Se accede a una lista de todas las funciones que incluyen el texto entrecomillado; en este caso “vector”.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



1.11. Figura.apropos()

Todas estas funciones de ayuda ofrecerán información sobre paquetes que tengamos instalados. Sin embargo, en muchas ocasiones el interés se centra en saber si R dispone de funciones para llevar a cabo determinado tipo de análisis, y esta información va más allá de las librerías instaladas. La función `RSiteSearch()` escrita directamente sobre la consola R ejecutará una búsqueda en la Web de R.

Existen además de las descritas varias fuentes de información sobre R disponibles en la red; R dispone de una lista de distribución (<https://stat.ethz.ch/mailman/listinfo/r-help>) y de un buscador específico sobre funciones, comandos o cuestiones relacionadas con R <http://www.rseek.org/>.



1.12. Figura. Rseek

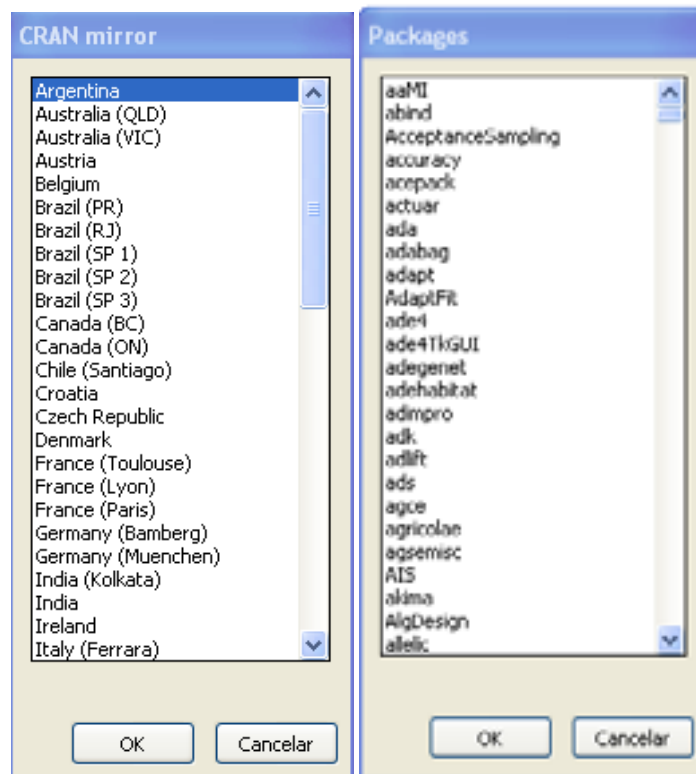
La página principal de CRAN almacena una gran cantidad de información sobre R, que incluye manuales de uso en varios idiomas, información sobre paquetes o listas sobre FAQ (*Frequently Asked Questions*) que pueden solucionar más de una duda al usuario. Entre ellas resulta especialmente atractiva la *wiki* (<http://wiki.r-project.org/rwiki/doku.php>).

Conceptos básicos y aplicaciones prácticas con Rcommander

1.5 Paquetes

R es un sistema dinámico en el que las contribuciones de autores individuales se ponen a disposición de los usuarios en forma de librerías o paquetes (*packages*). Cuando se instala R por primera vez se instala el paquete BASE, que puede ampliarse fácilmente por medio de paquetes adicionales que irán configurando R a la medida de las necesidades de cada usuario. Un paquete no es más que un conjunto de funciones y datos que una vez instalados pueden activarse a través del comando `library()`. La instalación de una nueva librería es una tarea sencilla. Para ello es posible utilizar el menú desplegable de la ventana de comandos de R y seleccionar la opción `instalar paquete(s)...`

Esta opción abrirá una ventana con un repertorio de espejos (*mirrors*) accesibles desde los cuales es posible transferir los paquetes deseados.



1.13. Figura.Espejos CRAN y paquetes

Conceptos básicos y aplicaciones prácticas con Rcommander

Es posible acceder a este listado de centros imagen a través del comando `CRAN.packages()` que se tecleará directamente sobre la consola. Una vez seleccionado el sitio se visualizará/n y seleccionará/n la/s librería/s a instalar.

La utilización de un paquete exige además de instalarlo, cargarlo. La instalación solo es necesaria llevarla a cabo una vez, pero deberemos de cargar el paquete en cada sesión de trabajo que lo necesitemos. Es una tarea muy sencilla; basta con seleccionar de la barra de menús de la consola la opción Paquetes>Cargar Paquete o teclear directamente sobre la consola `library(nombre del paquete)`.

Es sencillo conocer las librerías instaladas en memoria; para obtener un listado basta con teclear el comando `installed.packages()`. R devolverá información sobre los paquetes instalados, su versión y el nombre y ruta de las librerías que los contienen.

1.6 R a modo de calculadora

El modo de trabajar más simple o primitivo con R sería utilizarlo a modo de calculadora. En este sentido R evalúa y devuelve los resultados de cualquier expresión introducida en la línea de comandos. Dado que se trata de funciones básicas que pueden diferir ligeramente de cómo son definidas en otros entornos la tabla siguiente ofrece una breve descripción de las funciones algebraicas más comunes

Función	Operación
<code>+, -, *, /</code>	Suma, Resta, Multiplicación, División
<code>abs</code>	Valor absoluto
<code>asin acos atan</code>	Inversas de las funciones trigonométricas
<code>exp, log</code>	Exponencial y logaritmo natural
<code>Round</code>	Redondeo
<code>sin con tan</code>	Funciones trigonométricas

Conceptos básicos y aplicaciones prácticas con Rcommander

<code>sqrt ()</code>	<code>^</code>	Raíz cuadrada, Potencia
<code>%/ %</code>		División entera
<code>%%</code>		Resto de la división

1.1. Tabla. Funciones algebraicas

Como ejemplo, que puede servir al lector de primer ejercicio con R se muestra la ejecución de algunas de ellas.

```
> 2+3
[1] 5
> sqrt(20)
[1] 4.472136
> 5%/ %3 # genera la parte entera de la división 5/3
[1] 1
> 5%%3 # devuelve el resto de la división 5/3
[1] 2
> pi*5^2 # área de un círculo de radio 5
[1] 78.53982
> 1000*(1+0.05)^3-100
[1] 1057.625
> sqrt(c(10,100,1000))
[1] 3.162278 10.000000 31.622777
```

1.7 Notas básicas sobre R

R es un entorno de programación orientado a objetos. Esto significa que dependiendo del tipo de objeto al que se aplique una función, es decir, dependiendo de las características de los objetos, de sus atributos, el mismo comando actuará de diferente manera. En este primer capítulo, no vamos a describir los tipos de objetos, ni las características de cada uno; esto se hará en la parte final de este manual introductorio. En este punto, baste mencionar que el concepto de fichero de datos con el que se trabaja bajo SPSS o BMDP (por citar dos programas comerciales) se flexibiliza bajo el entorno R, donde la unidad de análisis u objeto más común es una *data frame* o marco de datos. Por el

Conceptos básicos y aplicaciones prácticas con Rcommander

momento, un marco de datos es una matriz de datos similar a los ficheros de datos que manipula SPSS o BMDP.

Como todo paquete estadístico, R utiliza por defecto un directorio de trabajo donde guardar las sesiones de trabajo, las salidas o los objetos que se desee almacenar. Puede comprobarse cual es el directorio de trabajo activo con la función `getwd()`, y éste puede modificarse con la opción `setwd()`. La fijación del directorio de trabajo es una opción recomendable. Es importante apuntar que en R los *path* utilizados para fijar directorios se definen a través del símbolo “/” o “\” en lugar del símbolo habitual bajo Windows “\”.

```
> setwd ("C:\\Documents and Settings\\Rmanual") ## CORRECTO
> setwd ("C:/Documents and Settings/Rmanual") ## CORRECTO
> setwd ("C\Documents and Settings\Rmanual") ## NO NO NO
CORRECTO
```

En R los nombres de los objetos son combinaciones de letras, números y el carácter “.”. No es posible definir el nombre de un objeto utilizando como primer carácter un número, sin embargo, sí es posible utilizar el “.”. R además es caso sensitivo; una letra mayúscula es diferente de su homónima minúscula

```
> 1.data #nombre incorrecto
> .data #nombre correcto
> Data #distinto de data
```

2 Rcommander

2.1 Instalación de Rcommander

Rcommander es un paquete adicional de R concebido como una interfaz gráfica (Graphical User Interface – GUI) que incorpora funciones para el análisis estadístico y generación de gráficos. Consigue, a través de un sistema de ventanas, convertir a R en un entorno amigable que facilita enormemente su utilización a aquellos usuarios no familiarizados con este entorno de programación. Trabajar en R con la ayuda de Rcommander se convierte en una experiencia similar a analizar datos bajo cualquier paquete estadístico al uso (SPSS, SAS, S-plus). Los menús desplegables que ofrece Rcommander, permiten al usuario que apenas conoce el entorno R, realizar las tareas de importar/exportar ficheros, manipular variables (recodificar, computar), seleccionar casos, describir datos, generar gráficos o ajustar modelos. Pero además, al mismo tiempo, Rcommander abre la posibilidad de modificar los comandos básicos o añadir nuevos comandos para realizar tareas que se ajusten a necesidades específicas.

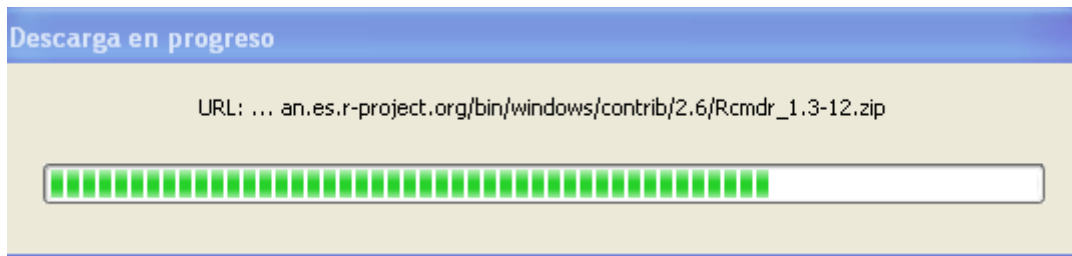
Las capacidades del Rcommander (como las del entorno R) han ido y van continuamente ampliándose. La versión sobre la que trabajamos fue construida para dar cobertura a un curso básico de análisis de datos. Esto significa que sólo una pequeña fracción del potencial estadístico y de gestión de datos disponibles bajo R (actualmente son más de 2100 los paquetes disponibles) es accesible por medio de Rcommander.

En definitiva, consideramos que Rcommander (Rcmdr) es el medio ideal para acercarse a R. Constituye un entorno de trabajo que muchos usuarios no necesitarán ampliar. Sin embargo, familiarizarse con Rcommander permite

Conceptos básicos y aplicaciones prácticas con Rcommander

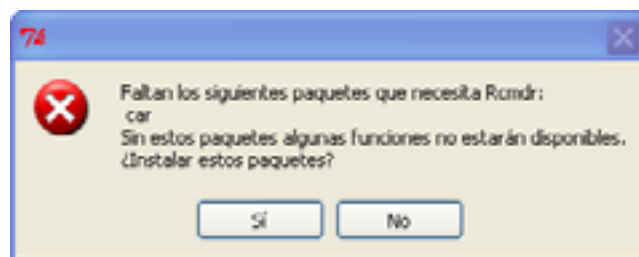
ir profundizando en este entorno de programación de modo que puede ampliarse y adaptarse a necesidades particulares (Fox, 2007).

Antes de comenzar a trabajar con Rcommander (Rcmdr) es necesario instalar y cargar el paquete Rcmdr. Como se ha indicado, la instalación y carga se llevan a cabo a través de las opciones Paquetes>Instalar.



2.1. Figura. Ventana de descarga de Rcommander

Una vez instalado Rcmdr es necesario “cargarlo”. Para ello, cuando se ejecute R se seleccionará la opción Paquetes > Cargar y de entre el listado de librerías disponibles se elegirá Rcmdr. También es posible teclear directamente en la consola `library(Rcmdr)`. La primera vez que se solicite cargar el paquete Rcmdr, R avisará que para el correcto funcionamiento de Rcmdr es necesario instalar algunos paquetes adicionales. El procedimiento de instalación busca las librerías necesarias para el correcto funcionamiento de las funciones disponibles en Rcommander, y solicita al usuario permiso para instalar aquellos paquetes que no encontrándose previamente instalados son necesarios. El usuario sólo necesitará asentir a los requerimientos del proceso de instalación.



2.2. Figura. Solicitud de permiso instalación de paquetes

Conceptos básicos y aplicaciones prácticas con Rcommander

R abrirá una nueva ventana en la que se seleccionará el sitio del que se desee instalar los paquetes faltantes. Es posible optar por la opción por defecto CRAN.



2.3. Figura. Instalación de paquetes adicionales

Una vez instalado y cargado Rcmdr en la consola R irán apareciendo los paquetes que ha sido necesario instalar, y la versión de Rcommander , 1.5-4., mostrando esta apariencia:



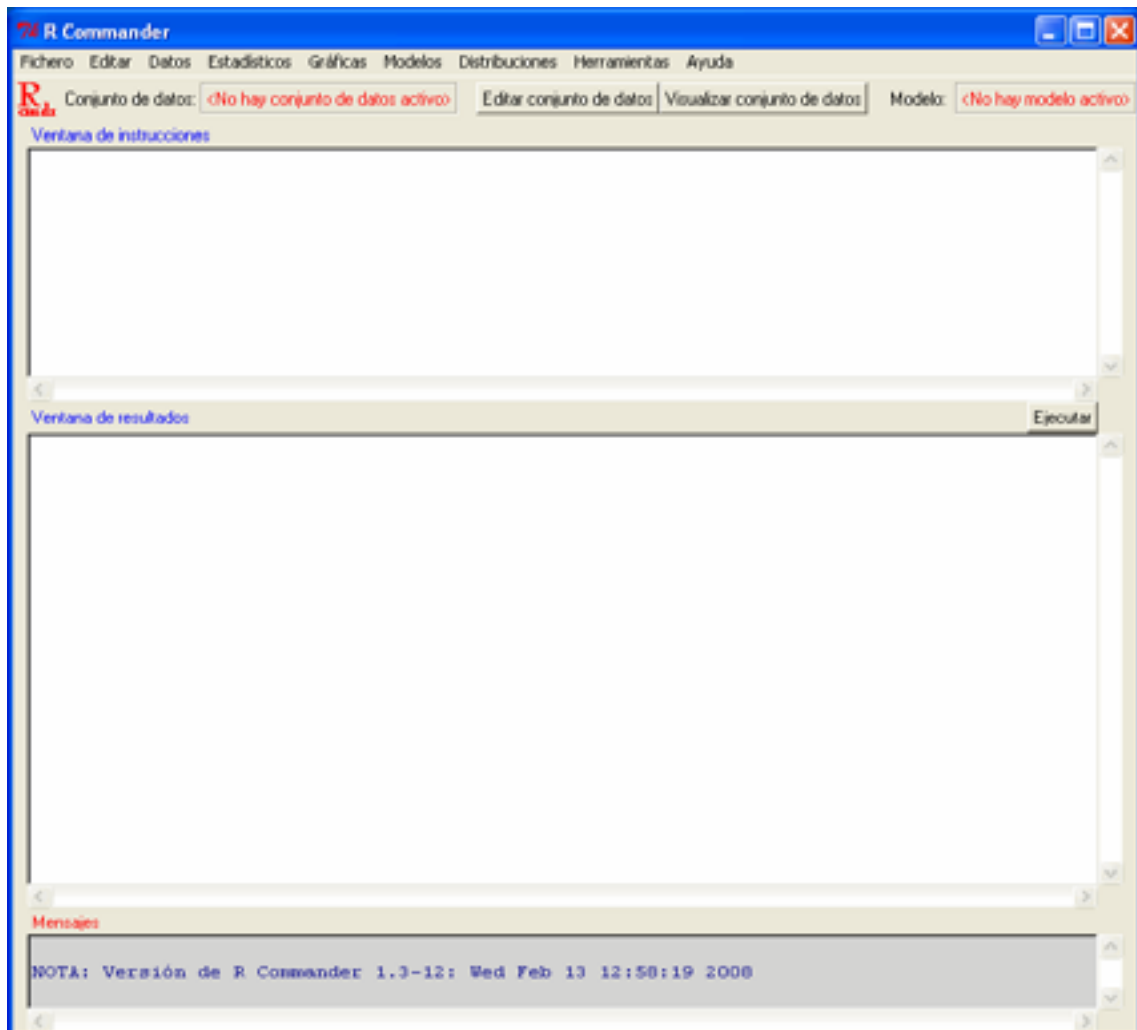
2.4. Figura. Consola R, Rcommander

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

2.2 Descripción de Rcommander

La pantalla inicial de Rcommander Graphical user Interface a cuya descripción se dedica este capítulo, presenta el siguiente aspecto.



2.5. Figura. Ventana Rcommander

La ventana básica de *Rcommander* aparece dividida en cinco secciones:

1.- Barra de menús.

Conceptos básicos y aplicaciones prácticas con Rcommander

Permite acceder por medio del ratón a las operaciones relacionadas con el manejo de ficheros, manipulación/descripción de datos o ajuste de modelos estadísticos, generación de gráficos o varias opciones de configuración de Rcommander. Se presentarán de forma detallada cada una de las opciones disponibles.



2.6. Figura. Barra de menús de Rcommander

2.- Conjunto de datos.

La segunda línea de la pantalla principal de Rcommander informa sobre los conjuntos de datos que se encuentran disponibles en una sesión de trabajo. Aunque solamente es posible ejecutar comandos o funciones sobre un conjunto de datos activo, es posible disponer en la misma sesión de trabajo de varios conjuntos de datos que pueden ir intercambiándose simplemente con seleccionarlos en la ventana en la que se indica cual es el conjunto de datos activo. Rcommander incorpora además un editor de datos y la posibilidad de visualizarlos. El editor de datos de Rcommander podría utilizarse para introducir un pequeño conjunto de datos; sin embargo sus capacidades como editor son muy limitadas. Por ello recomendamos la utilización de editores más potentes dado que los datos podrán ser importados con facilidad. Por último se muestra el modelo activo, esto es, el modelo matemático que se está aplicando sobre los datos.



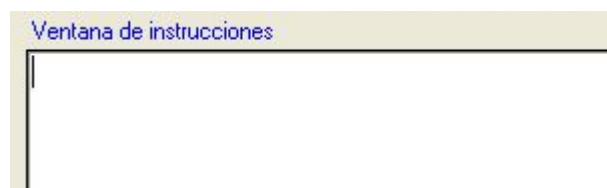
2.7. Figura. Barra de datos de Rcommander

La opción de visualizar datos está fijada por defecto para poder acceder a un fichero que contiene un número de variables igual o menor a 100. A partir de este umbral (que puede ser modificado) se utiliza el editor de datos.

3.- Ventana de instrucciones

Todas las operaciones ejecutadas por medio de los menús de Rcommander generan de forma automática comandos en R que se visualizan en la ventana de instrucciones (*script window*). Esta ventana cumple además las funciones de una consola de edición de códigos en la que el usuario puede escribir directamente el comando u operación que desee ejecutar. Una vez escrita una función basta seleccionarla con el ratón y pulsar la opción **Ejecutar** o las teclas `Ctrl+r` (run) para su ejecución. Si el usuario desea escribir un comando que ocupa más de una línea deberá comenzar la segunda línea y siguientes con uno o más espacios en blanco o tabuladores. La combinación de teclas `Ctrl+a` seleccionará todas las líneas de la ventana de instrucciones, y `Ctrl+s` abrirá una ventana para la grabación del fichero de instrucciones.

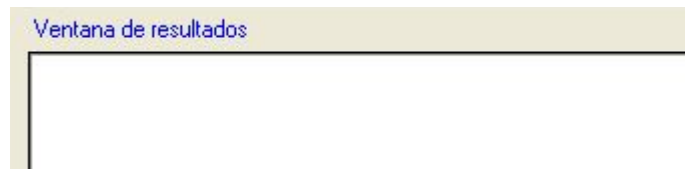
El hecho de que todas las acciones ejecutadas en Rcommander aparezcan en la ventana de instrucciones permite familiarizarse con el lenguaje de programación R de modo que con un poco de dedicación e interés se podrán reconocer y por tanto adaptar y manipular los comandos básicos que aparecen en esta ventana.



2.8. Figura. Ventana de Instrucciones de Rcommander

4.- Ventana de salidas (*output window*)

Es una ventana de edición en la que irán apareciendo tanto los comandos a ejecutar (en color rojo) como los resultados generados por ellos (en color azul oscuro).



2.9. Figura. Ventana de Resultados de Rcommander

Es importante hacer notar que los gráficos producidos por Rcommander generan por defecto otra ventana (dispositivo gráfico) de salida.

5- Ventana de mensajes

La ventana de mensajes ofrece al usuario información sobre los errores cometidos (mensajes en rojo), avisos (*warnings*) surgidos durante la ejecución de los comandos (mensajes en verde) o información sobre el conjunto de datos que se está utilizando (mensajes en azul oscuro).

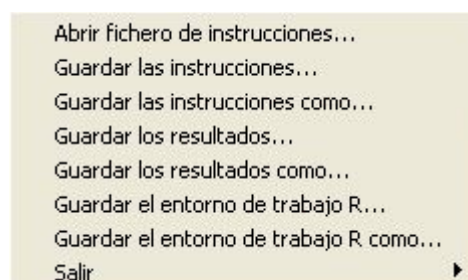


2.10. Figura. Ventana de mensajes de Rcommander

2.3 Opciones generales de la barra de menús

La barra de menús de Rcommander ofrece las siguientes opciones generales:

Fichero. Permite gestionar la lectura y grabación de ficheros de instrucciones o de ficheros de salidas generados por R.

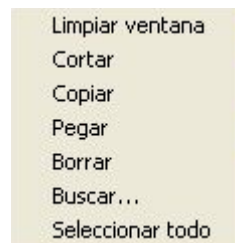


Conceptos básicos y aplicaciones prácticas con Rcommander

2.11. Figura. Opciones de Fichero de Rcommander

Las opciones que encierra este menú se refieren a la apertura y grabación de las sesiones de trabajo bajo Rcommander. Los ficheros de instrucciones se graban por defecto con la extensión “.R”, mientras que las salidas generadas se almacenan en formato texto “.txt”. Las opciones relacionadas con el entorno de trabajo permiten guardar todos los objetos (vectores, matrices, marcos de datos...) generados durante una sesión de trabajo con el fin de poder acceder a ellos en sesiones posteriores. Estos objetos podrán ser cargados a través de la opción *Cargar conjunto de datos* que se encuentra en el menú desplegable dependiente de datos (*Datos> Cargar conjunto de datos*). Aunque es posible guardar la sesión de trabajo bajo cualquier extensión, la opción de cargar los datos utiliza por defecto la extensión “.Rda” (Rdatos).

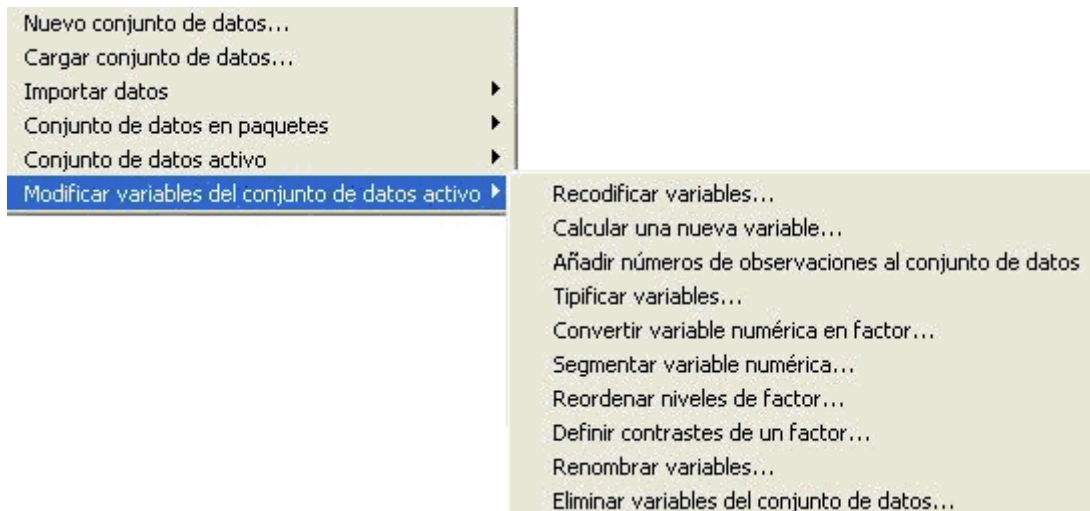
Edición. Las opciones que abre el menú de edición son las propias de cualquier editor al uso (cortar, copiar, pegar, buscar...) que podrán ser utilizadas tanto en la ventana de instrucciones como en la ventana de salidas.



2.12. Figura. Opciones de edición de Rcommander

Datos Esta opción da paso a una serie de submenús relacionados con la lectura e importación de conjuntos de datos y la manipulación de variables contenidas en ellos.

Conceptos básicos y aplicaciones prácticas con Rcommander



2.13. Figura. Opciones de Datos de Rcommander

Estadísticos. Accede a los submenús asociados con la descripción de datos, análisis estadísticos básicos, tests no paramétricos, estudio de la fiabilidad y dimensionalidad o ajuste de modelos lineales.



2.14. Figura. Opciones de Estadísticos de Rcommander

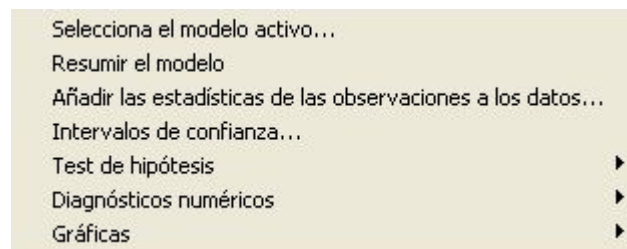
Gráficos. Por medio de esta opción se accede a un submenú para la generación de gráficos básicos. La gama que ofrece Rcommander es amplia, aunque al igual que todas las funciones que utiliza Rcommander es posible adaptarla a situaciones y necesidades específicas.

Conceptos básicos y aplicaciones prácticas con Rcommander



2.15. Figura. Opciones de Gráficos de Rcommander

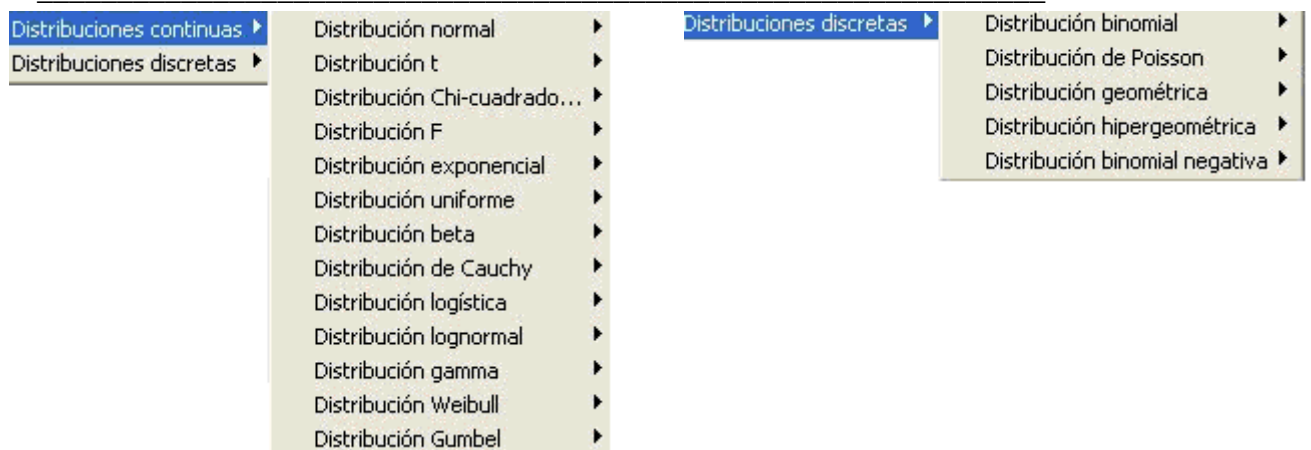
Modelos. Una vez ajustado un modelo Rcommander ofrece la posibilidad de profundizar en él por medio de información diagnóstica sobre su ajuste que puede acompañarse además con información gráfica.



2.16. Figura. Opciones de Modelos de Rcommander

Distribuciones. Esta ventana ofrece un acceso rápido y eficiente a las distribuciones estadísticas más comunes, tanto continuas como discretas que permite generar datos, dibujar gráficos, estimar cuantiles o probabilidades. Es un buen apoyo a la docencia a la par que evita el uso de tablas estadísticas.

Conceptos básicos y aplicaciones prácticas con Rcommander



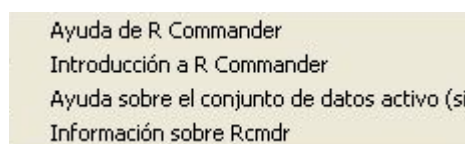
2.17. Figura. Opciones de Distribuciones de Rcommander

Herramientas. Permite cargar distintos paquetes de R (que tienen que estar ya instalados) así como fijar varios parámetros relacionados con la configuración de Rcommander.



2.18. Figura. Opciones de Herramientas de Rcommander

Ayuda. Por medio de esta opción se accede a un fichero que contiene información sobre el manejo de Rcommander.



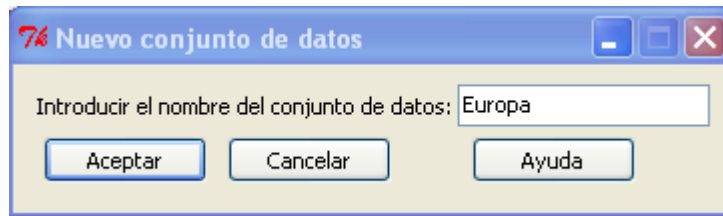
2.19. Figura. Opciones de Ayuda de Rcommander

2.4 Lectura e importación de datos

Los datos a analizar bajo R han de definirse como el conjunto de datos activo. En el entorno R no se utiliza el término fichero; en su lugar se habla de objetos (vectores, matrices, factores, listas, marcos de datos, funciones). Prácticamente todo es un objeto en el entorno R. Un fichero de datos habitual en

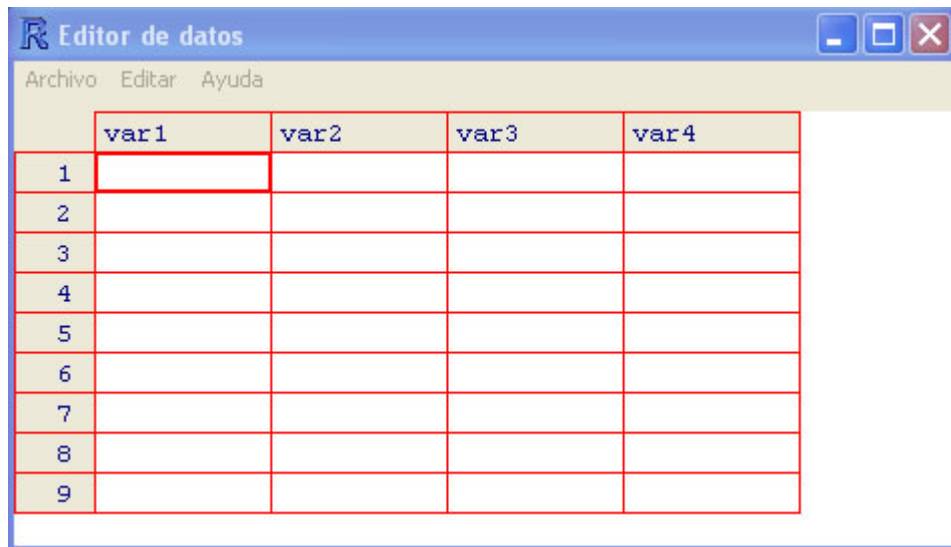
Conceptos básicos y aplicaciones prácticas con Rcommander

Esta opción requiere introducir los datos directamente a través del editor de Rcommander. La selección de esta opción abre una ventana en la cual se dará un nombre al conjunto de datos que se vaya a crear.



2.20. Figura. Nuevo conjunto de datos

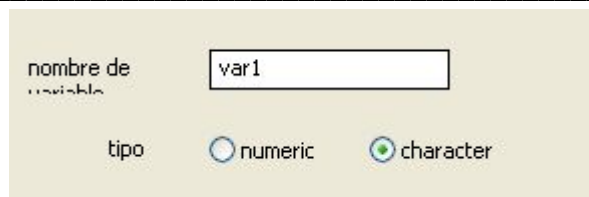
En este ejemplo se ha creado el objeto “Europa”. El editor de datos de Rcommander es bastante limitado, por lo que este modo de introducir datos es adecuado solamente cuando el número de datos no es cuantioso.



2.21. Figura. Editor de datos de Rcommander.

La ventana de edición permite asignar un nombre a cada una de las variables que se vaya a generar, y definir si la variable es de tipo numérico o de tipo carácter. Para ello basta con hacer *click* en cualquiera de las columnas (var1, var2...) de la primera fila del editor.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



nombre de variable:

tipo: numeric character

2.22. Figura. Definición del tipo de variables.

	Pais	Habitantes	var3	var4	var5
1	Alemania	82900000			
2	Austria	8e+06			
3	Belgica	1.03e+08			
4	Chipre	790000			
5	Dinamarca	5300000			
6	Eslovaquia	5400000			
7	Eslovenia	2e+06			
8	España	42700000			
9	Estonia	1400000			
10	Finlandia	5800000			
11	Francia	59600000			
12	Grecia	1.1e+07			
13	Holanda	16100000			
14	Hungria	10100000			
15	Irlanda	3900000			
16	Italia	57300000			
17	Letonia	2300000			
18	Lituania	3700000			
19	Luxemburgo	448000			

2.23. Figura. Editor de datos.

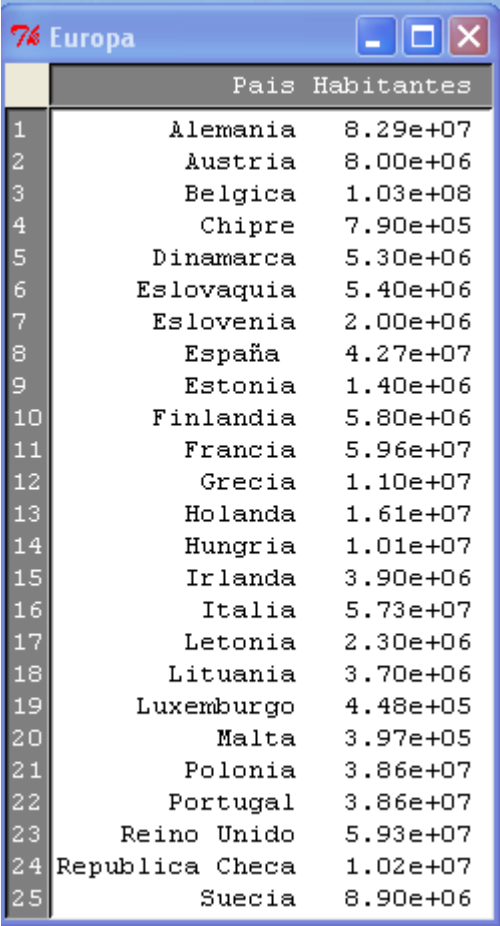
Las teclas de Avance Página (AvPag) y Retroceso Página (RePag) permiten acceder a las pantallas posterior y anterior a la pantalla visible. La tecla *Home* sitúa el cursor en la primera celda de la matriz (Fila1 Columna1). La tecla *End* lleva al cursor a la última fila de la matriz. Puede redimensionarse el tamaño de las columnas situándose en cualquiera de las celdas y pulsando la tecla derecha del ratón. Una vez finalizada la introducción de datos la opción Cerrar del menú Archivo (Archivo>Cerrar) nos hará abandonar la ventana de edición. La ventana de instrucciones mostrará el siguiente comando, que indica la creación de un marco de datos (un archivo de datos) de nombre

Conceptos básicos y aplicaciones prácticas con Rcommander

Europa. En la ventana *Conjunto de Datos* aparecerá ahora *Europa*. Disponemos de un conjunto de datos activo sobre el que se ejecutarán las funciones que indiquemos.

```
Europa <- edit(as.data.frame(NULL))
```

La opción de visualizar datos mostrará la siguiente pantalla, en la que aparecen los nombres de las variables (País, Habitantes) y los datos correspondientes a los 25 países de la Unión Europea que conforman este conjunto de datos.



	País	Habitantes
1	Alemania	8.29e+07
2	Austria	8.00e+06
3	Belgica	1.03e+08
4	Chipre	7.90e+05
5	Dinamarca	5.30e+06
6	Eslovaquia	5.40e+06
7	Eslovenia	2.00e+06
8	España	4.27e+07
9	Estonia	1.40e+06
10	Finlandia	5.80e+06
11	Francia	5.96e+07
12	Grecia	1.10e+07
13	Holanda	1.61e+07
14	Hungria	1.01e+07
15	Irlanda	3.90e+06
16	Italia	5.73e+07
17	Letonia	2.30e+06
18	Lituania	3.70e+06
19	Luxemburgo	4.48e+05
20	Malta	3.97e+05
21	Polonia	3.86e+07
22	Portugal	3.86e+07
23	Reino Unido	5.93e+07
24	Republica Checa	1.02e+07
25	Suecia	8.90e+06

2.24. Figura. Visualizador de datos.

Es importante recordar que la creación del marco de datos no implica que los datos nombrados como “Europa” hayan sido grabados; si se desea trabajar con ellos en sesiones posteriores tiene que ejecutarse la opción de guardar datos

Conceptos básicos y aplicaciones prácticas con Rcommander

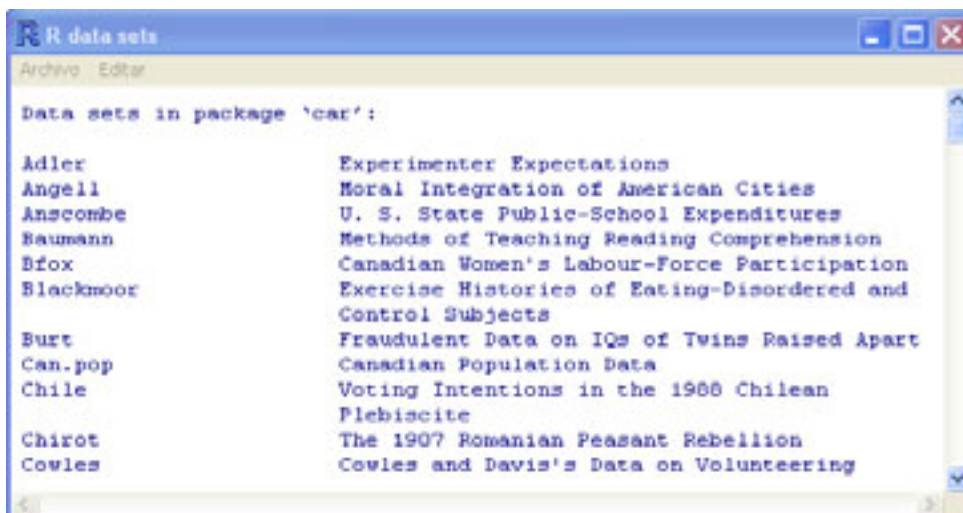
(Datos>Conjunto de datos activo>Guardar conjunto de datos activos)

2.- Cargar conjunto de datos

Esta opción brinda el acceso a los datos incluidos en los paquetes de R instalados en memoria, y a los objetos generados y almacenados en sesiones anteriores de trabajo. Son dos las opciones de Rcommander para acceder a datos que se encuentran en paquetes: Datos>Cargar conjunto de datos, y Datos>Conjunto de datos en paquetes.

Utilizando la primera se accede a los objetos con extensión “.RDA” que han sido guardados en sesiones de trabajo anteriores. Por ejemplo para recuperar los datos “Europa.rda” esta es la opción a utilizar.

La segunda opción, “Datos>Conjunto de datos en paquetes”, permite explorar los datos que habitualmente incorporan los paquetes de R. La opción ofrece como salida un listado de los conjuntos de datos incluidos en los paquetes.

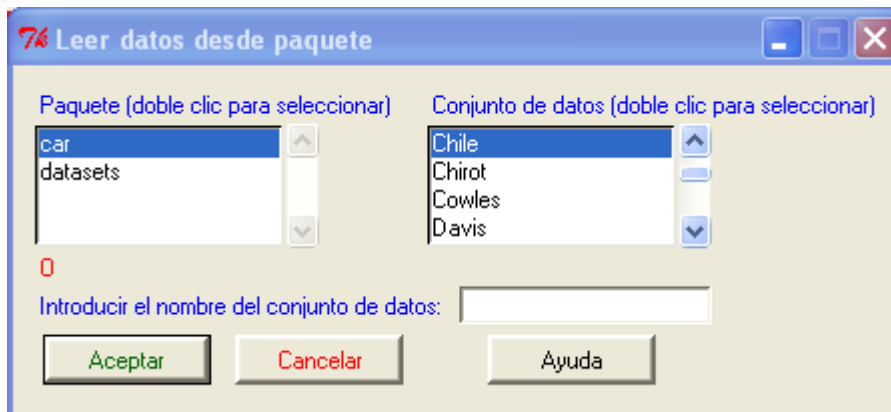


2.25. Figura. Listado de conjuntos de datos en el paquete “car”.

Para utilizar alguno de ellos se selecciona la opción “Conjunto de datos en paquetes>Leer conjunto de datos desde paquete

Conceptos básicos y aplicaciones prácticas con Rcommander

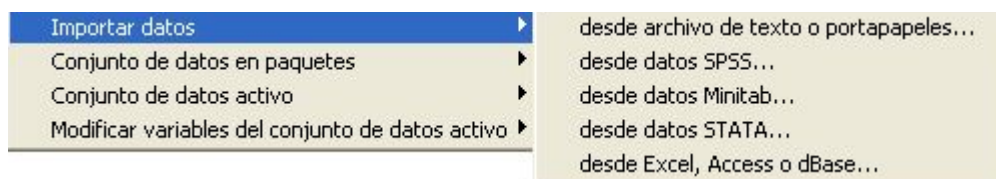
adjunto...”, que da acceso a una ventana en la que además de elegir el paquete y el conjunto de datos, éste último puede ser renombrado.



2.26. Figura. Selección de un conjunto de datos “car”.

3.- Importar datos

Esta opción permite trabajar con datos almacenados en formato ASCII, con datos creados con software estadístico (STATA, Minitab, SPSS...) o con datos provenientes de programas como *Excel* o *Acces*. El menú accesible por medio de esta opción muestra los formatos importables desde Rcommander.

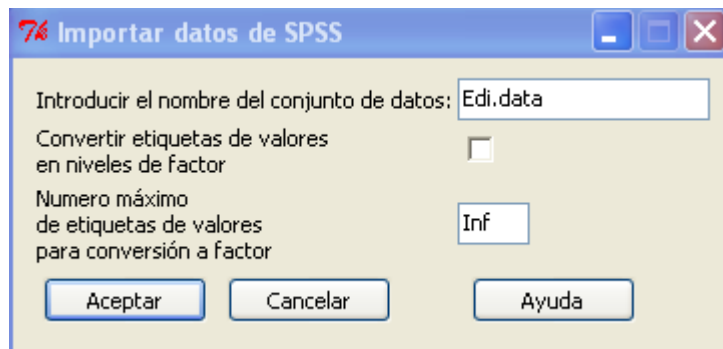


2.27. Figura. Formatos de ficheros importables desde R.

En el ejemplo mostrado a continuación y que guiará a partir de ahora los análisis realizados en esta introducción a R, se importará un archivo de datos generado con SPSS, cuya extensión es .sav. Los comentarios al respecto son fácilmente generalizables para prácticamente todos los formatos de archivo especificados.

Conceptos básicos y aplicaciones prácticas con Rcommander

Para importar un fichero de datos desde SPSS se procede a través de la opción, Datos>Importar Datos>desde datos SPSS..., que abrirá una ventana de diálogo en la que se especificarán las opciones de importación.



2.28. Figura. Importar archivo desde SPSS.

En primer lugar, se ha de especificar un nombre para el conjunto de datos, que no tendrá que coincidir con el nombre del archivo de datos que se va a importar. La opción de la ventana “Convertir etiquetas de valores en niveles de factor” es especialmente importante en el caso de variables categóricas ordenadas; por ejemplo, respuestas a ítems con formato Likert en las que las opciones de respuesta se corresponden con una escala graduada. R, diferencia entre variables numéricas y factores; se trata de tipos de objetos diferentes y que por lo tanto reciben un tratamiento diferente. Dentro de un análisis de fiabilidad por ejemplo, es necesario considerar estas variables, ítems, como variables numéricas por lo tanto no interesaría convertirlas en factores. Si seleccionamos la opción de conversión, convertiremos esas variables numéricas en variables categóricas o factores, limitando de este modo su tratamiento. Por ello, dependiendo del tipo de dato y del tipo de tratamiento que se les vaya a procurar será el investigador el que determine la conveniencia o no de convertir en factor los valores de la variable a importar.

Una vez pulsada la opción *Aceptar*, Rcommander permite abrir el fichero de datos a importar. El comando que genera esta opción, utiliza la

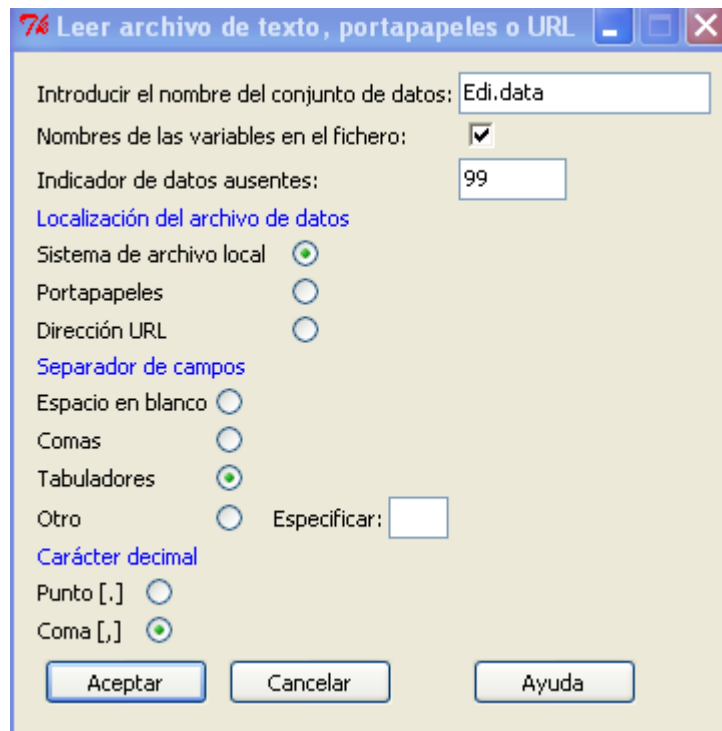
Conceptos básicos y aplicaciones prácticas con Rcommander

función `read.spss` para acceder al fichero de datos “Insatisfacción corporal.sav”, que en este ejemplo se almacenará bajo el nombre de `Edi.data`.

```
Edi.data <- read.spss("C:/Insatisfaccion corporal.sav",  
se.value.labels=FALSE, max.value.labels=Inf, to.data.frame=TRUE)
```

Tras leer el archivo podrá leerse en la ventana de mensajes: “NOTA: El conjunto de datos `Edi.data` tiene 976 filas y 24 columnas.”

En el caso de que se fuera a importar un fichero en formato ASCII, formato básico para el intercambio de ficheros entre programas, la selección de la opción Datos>Importar Datos>Desde archivo de texto o portapapeles, abrirá a una ventana en la que se definirán cuestiones generales de formato que determinarán la calidad de la importación.



2.29. Figura. Importar archivo desde formato ASCII.

En este sentido son especialmente importantes el indicador de datos ausentes y el carácter decimal. En R los datos ausentes son indicados al sistema por los caracteres NA (*Non Available*; No disponible). Si no se especifica el

Conceptos básicos y aplicaciones prácticas con Rcommander

modo en que se han codificado los valores ausentes todos los valores diferentes a NA serán tratados como valores presentes, lo cual originará resultados cuanto menos extraños. Además de los datos ausentes, es importante especificar el carácter decimal utilizado en el fichero ASCII. Si se ha utilizado la coma como carácter decimal y no se especifica de este modo, todas las variables con caracteres decimales serán tratadas como variables categóricas.

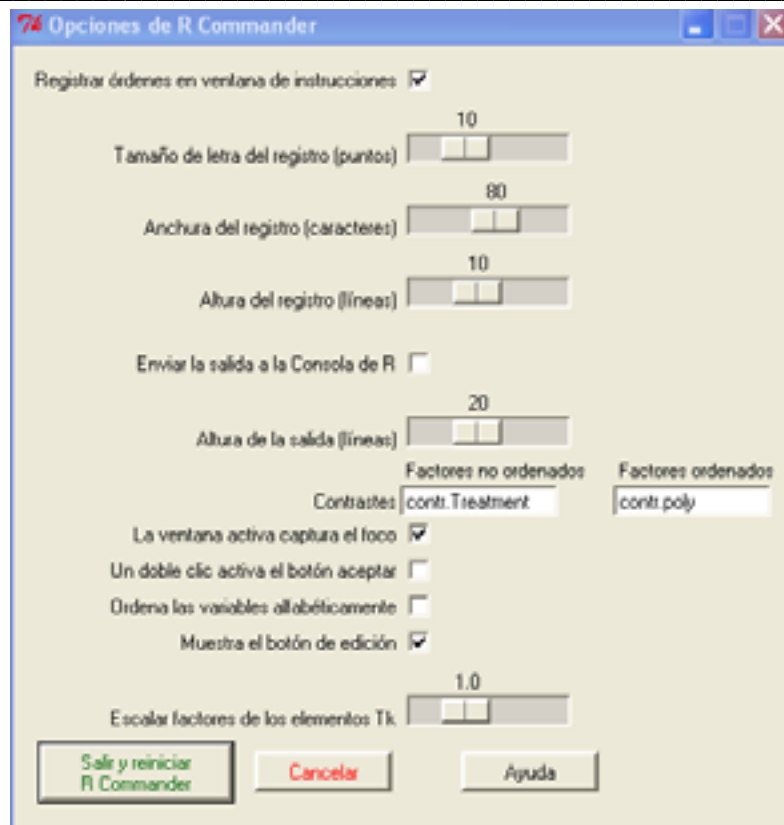
El comando utilizado por R para la lectura de ficheros ASCII aparecerá en la ventana de instrucciones,

```
Edi.data <- read.table("C:/Insatisfaccion corporal.txt",  
header=TRUE, sep=" ", na.strings="99", dec="," , strip.white=TRUE)
```

2.4.1 Opciones de configuración

En la opción Herramientas de la barra de menús es posible modificar algunas configuraciones básicas de *Rcommander* referidas tanto a su aspecto, como a especificaciones que puede ser interesante adaptar a nuestros hábitos de trabajo.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



2.30. Figura. Configuración de Rcommander.

Entre ellas, tal vez la más práctica, sea la relacionada con la visualización de las variables que forman el conjunto de datos a analizar. El investigador utiliza normalmente un orden preestablecido en la generación de las variables, sin embargo, Rcommander por defecto ordena las variables alfabéticamente. Aunque esta opción no alterará en absoluto los resultados de ningún análisis, es conveniente fijarla a los hábitos propios.

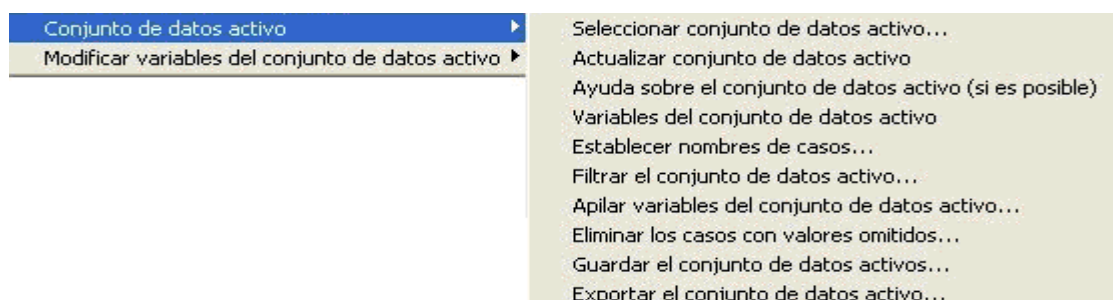
La ventana de instrucciones o comandos de Rcommander no es una consola R aunque resulta útil para la modificación/generación de pequeños códigos de programación. Para la generación de códigos más o menos complicados siempre es recomendable utilizar el editor de comandos del propio R, o editores de programación tales como el *TinnR*, *R-Winedit*, o *Emacs*.

3 Manipulación de datos y transformación de variables

Una vez fijado el conjunto de datos activo, los menús desplegables de Rcommander permiten manipularlo y transformar sus variables. El primer tipo de acciones que se encuentran encerradas en el menú “Datos >Conjunto de datos activo” están relacionadas con la definición y gestión del conjunto de datos activo; por ejemplo, la selección o eliminación de casos. El segundo tipo de funciones se relaciona con la manipulación de variables (creación de nuevas variables, recodificación de valores o conversión). En este capítulo se describen brevemente las opciones que presenta Rcommander para ejecutar estas funciones.

3.1 Manipulación de datos

Este apartado se relaciona directamente con las operaciones aplicadas sobre el conjunto de datos activo referidas a aspectos como la eliminación de valores faltantes, la selección de casos o la exportación del conjunto de datos. Las opciones disponibles que se muestran en la siguiente figura serán descritas con brevedad:

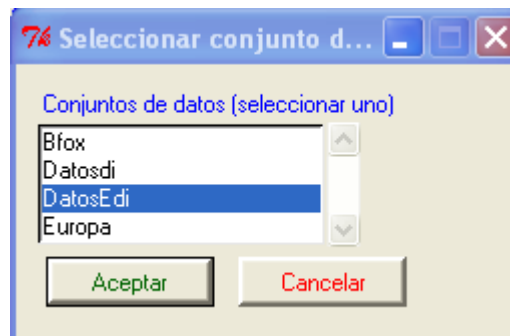


3.1. Figura. Opciones de manipulación de datos

Conceptos básicos y aplicaciones prácticas con Rcommander

3.1.1 Seleccionar conjunto de datos activo.

Los comandos que se ejecutan en R se aplican sobre un conjunto de datos activo (un objeto de R) lo cual no significa que no podamos disponer simultáneamente de más de un conjunto de datos. Por ejemplo, hasta ahora hemos generado un conjunto de datos “Europa” y hemos importado un fichero de datos de SPSS al que se ha denominado “Edi .data”, ambos aparecerán en la ventana de Rcommander bajo el epígrafe *Conjunto de Datos*. Sin embargo, siempre es necesario seleccionar aquél sobre el que se va a trabajar. La opción “Datos> Conjunto de datos activo>Seleccionar conjunto de datos activo...” accede a una ventana de diálogo en la que se selecciona el objeto que se convertirá en el conjunto activo.



3.2. Figura. Seleccionar conjunto de datos activo

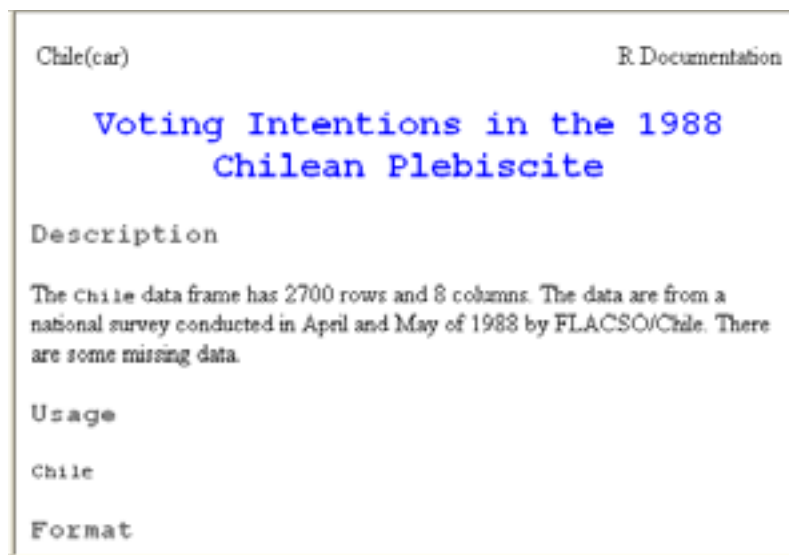
3.1.2 Actualizar conjunto de datos activo

El trabajo con el conjunto de datos puede dirigirse desde Rcommander o desde la consola de R. Si se trabaja con esta última conviene actualizar el conjunto de datos después de un proceso de manipulación de variables en el que se han recodificado o modificado sus valores para que incorpore las transformaciones efectuadas. La opción Datos> Conjunto de datos activo>Actualizar conjunto de datos activo tiene como finalidad ejecutar este reajuste.

3.1.3 Ayuda sobre el conjunto de datos activo

Los paquetes instalados en R además de funciones y comandos incorporan bases de datos de libre acceso y disposición. Es posible, además, que algunos de los paquetes incorporen información sobre los datos que aportan; la opción `Datos>Conjunto de datos activo>Ayuda sobre el conjunto de datos activo` accede a ella.

Por ejemplo en uno de los paquetes que utiliza Rcommander (`car`) existe un conjunto de datos denominado “Chile”, si se ejecuta la opción sobre información referida a este paquete se accede a una página como la que muestra la siguiente ilustración.



3.3. Figura. Información sobre un conjunto de datos perteneciente a un paquete.

3.1.4 Variables del conjunto de datos activo

Ofrece un listado con los nombres de las variables incluidas en el conjunto de datos activo. Con esta opción es posible comprobar las variables que forman “`Edi.data`”. El comando en R que permite obtener este listado es `names`, y su ejecución genera en la ventana de salidas el siguiente listado. El

Conceptos básicos y aplicaciones prácticas con Rcommander

número [1] indica la posición de “Edad” en el conjunto de datos. La variable “Pesideal” ocupa el lugar [4], la variable “IC_7” ocupa el lugar [19] .

```
> names(Edi.data)
[1] "Edad"      "Sexo"      "Peso"      "Pesideal"  "Altura"    "OD_1"
[7] "OD_2"      "OD_3"      "OD_4"      "OD_5"      "OD_6"      "OD_7"
[13] "IC_1"      "IC_2"      "IC_3"      "IC_4"      "IC_5"      "IC_6"
[19] "IC_7"      "IC_8"      "IC_9"      "IC_10"     "BA"        "B"
```

3.1.5 Establecer nombres de casos

Al igual que es posible asignar nombres de variables dentro de un objeto (fichero de datos) es posible asignar nombres a cada una de las filas. Para que la asignación sea posible es necesario disponer de una variable que incluya estos nombres de fila (rowname).

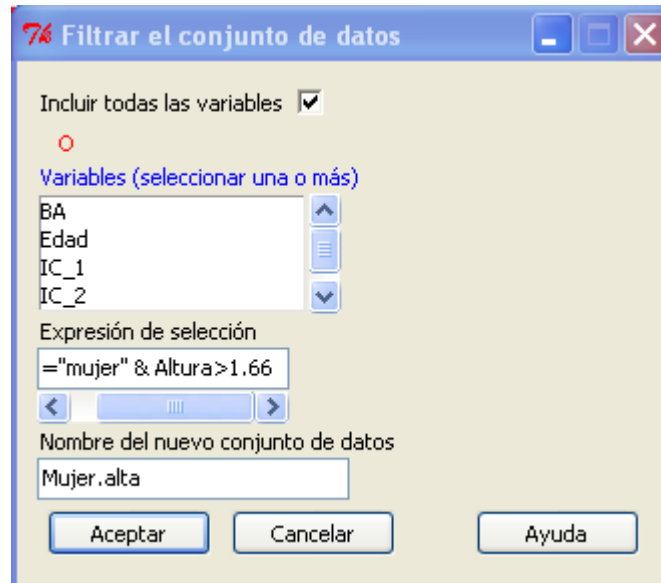
Si bien en el conjunto de datos actual pudiera carecer de sentido asignar nombres a las filas, esta opción resulta interesante en la definición de determinadas salidas de resultados pues permitirá dar un formato específico al marco de datos. Por ejemplo, si se lleva a cabo un análisis de ítems, se podría crear como salida un objeto (marco de datos) en el que las columnas estarían formadas por varios índices estadísticos, y se podría definir para cada fila el nombre del ítem al que ésta asociado.

3.1.6 Filtrar el conjunto de datos activo

Esta opción del menú Datos tiene como finalidad seleccionar un subconjunto de casos. La opción Datos>Conjunto de datos activo>Filtrar el conjunto de datos activo da acceso a una ventana que permitirá llevar a cabo la selección. La filtración generará un nuevo

Conceptos básicos y aplicaciones prácticas con Rcommander

conjunto de datos, “solo.mujer” en el que es posible incluir todas las variables originales o solamente un subconjunto de ellas.



3.4. Figura. Filtrar datos.

La ventana “Expresión de selección” es utilizada para definir la/s condición/es que deben de cumplir los datos que se desea filtrar. En este ejemplo, se ha seleccionado una submuestra que cumple dos condiciones “ser mujer” y “tener una altura mayor a 1,66 cm”. La variable `sexo` es categórica y por tanto sus valores habrán de especificarse entre comillas. El nuevo conjunto de datos se llama “Mujer.alta” y se convierte de forma automática en el conjunto de datos activo. En la ventana de mensajes es posible leer “NOTA: El conjunto de datos `mujer.alta` tiene 139 filas y 24 columnas.”

La especificación de las condiciones que han de cumplir las variables en una selección ha de regirse por los operadores lógicos que utiliza R. La siguiente tabla ofrece un resumen de la sintaxis de las expresiones lógicas disponibles en este entorno

Operaciones de comparación

Conceptos básicos y aplicaciones prácticas con Rcommander

<	Menor
<=	Menor o igual
>	Mayor
>=	Mayor o igual
= =	Igualdad lógica
!=	Distinto
Operadores lógicos	
&	Conjunción (y)
	Disyunción lógica (o)
!	Negación
xor	Disyunción exclusiva

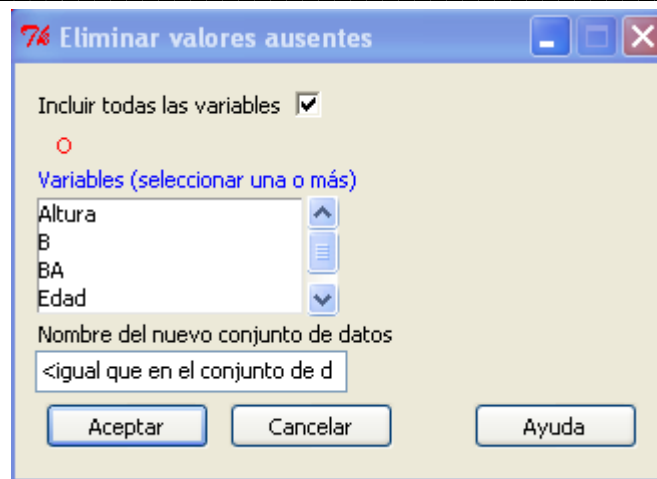
3.1 Tabla. Operadores lógicos

El comando R generador de esta selección se muestra en la ventana de instrucciones:

```
Mujer.alt <- subset(Edi.data, subset=Sexo=="mujer" & Altura>1.66)
```

3.1.7 Eliminar los casos con valores omitidos

Esta opción (Datos>Conjunto de datos activo> Eliminar los casos con valores omitidos) permite crear un archivo en el que se omiten todos los casos que presenten valores faltantes. Conviene recordar que en el entorno R los valores ausentes son codificados como NA (*Non available*). Es posible eliminar del conjunto de datos activos aquellos casos que presenten valores ausentes en determinadas variables; al mismo tiempo Rcommander permite crear un nuevo objeto sin NAs o escribir sobre el objeto antiguo el nuevo objeto sin NAs.



3.5. Figura. Eliminar valores ausentes.

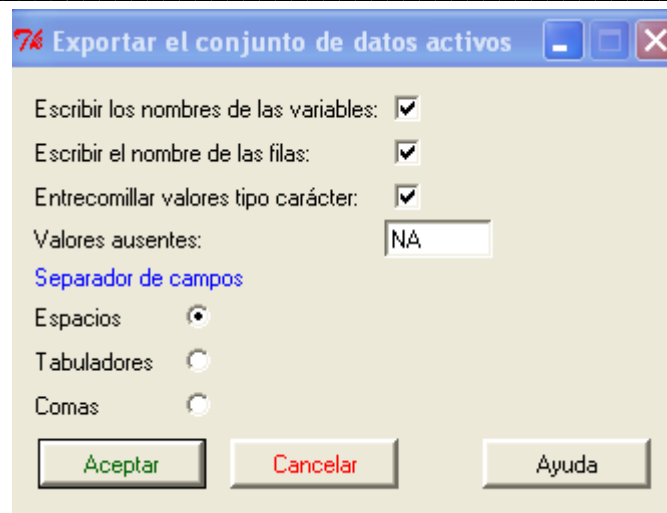
El comando que permite gestionar los valores ausentes es `na.omit`.

```
sin.mis <- na.omit(mujer.alt)
```

3.1.8 Guardar el conjunto de datos activo

La opción Datos>Conjunto de datos activo>Guardar conjunto de datos activos permite guardar el objeto de trabajo activo con extensión por defecto `.Rda` (datos R). Es importante recordar que independientemente de cómo se renombre el objeto de trabajo, esta opción guarda el conjunto de datos activo que se muestra en la ventana plana “Conjunto de Datos”.

También es posible exportar los datos como archivo de texto (extensión `.txt`, o `.dat`). La opción Datos>Conjunto de datos activo>Exportar el conjunto de datos activo abre una ventana de diálogo en la que se fijan varias opciones de exportación. La selección de una u otra dependerá de los intereses del investigador.



3.6. Figura. Exportar el conjunto de datos.

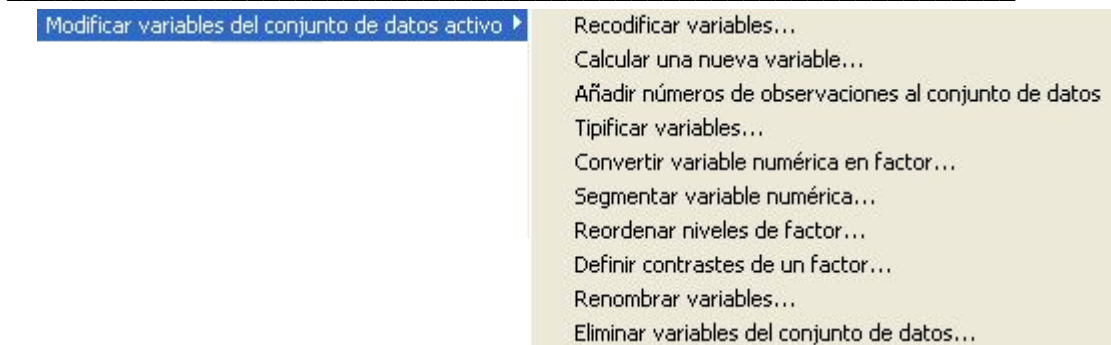
Rcommander utiliza la función `save` para exportar objetos. La exportación de datos generará el siguiente comando en la ventana de instrucciones.

```
save("Europa", file="C:/Europa.txt")
```

3.2 Manipulación de variables

Las opciones disponibles en Rcommander para la transformación o creación de variables sobre el conjunto de datos activo se encuentran en el menú dependiente de `Datos>Modificar variables del conjunto de datos activo`. La figura siguiente muestra las posibilidades incorporadas por Rcommander.

Conceptos básicos y aplicaciones prácticas con Rcommander



3.7. Figura. Opciones de manipulación de variables.

3.2.1 Recodificar variables

Permite asignar nuevos valores a las variables existentes en el conjunto de datos activo, y crear nuevas variables partiendo de valores previos. Por ejemplo, en el conjunto de datos activo `Edi.data` la variable `Edad` posee un rango de valores entre 10-27; es posible recodificar estos los valores y crear una nueva variable en la que los sujetos reciban un valor en función de la franja de edad en la que se encuentren. Se trataría de crear una variable con 3 niveles (por ejemplo) de modo que los sujetos que tengan menos de 14 años reciban el valor 1, las personas con edades entre los 14 y 16 años tengan un valor de 2, y al resto se les asigna el valor 3. La opción `recodificar variable` abre una ventana de diálogo en el que selecciona la variable a modificar, se nombra la variable que se desee crear y se introducen las directrices de recodificación, las reglas para la nueva asignación.

Conceptos básicos y aplicaciones prácticas con Rcommander



3.8. Figura. Recodificar variables.

En este ejemplo se ha recodificado una sola variable, pero es posible recodificar simultáneamente un conjunto de variables (piénsese en un subconjunto de ítems de un cuestionario o escala). Es importante especificar si se desea convertir la nueva variable en un factor (útil para los análisis de varianza), o se desea mantener su carácter numérico. Una vez seleccionadas las variables a modificar se podrá asignar un nuevo nombre a esta variable, o en su defecto mantener el nombre anterior. Es especialmente útil la opción de poder incluir un prefijo que se añade al nombre de variable viejo para poder diferenciar las nuevas variables recodificadas.

En la especificación de las reglas de asignación es suficiente indicar “valor antiguo= valor nuevo” en las situaciones en las que se recodifique una sola variable. Para una secuencia de valores consecutivos, se podrían indicar todos los valores separados por medio de comas, o utilizar dos puntos (“:”) para indicar a R que recodifique todos los valores comprendidos entre ellos. Cuando se han de recodificar varios valores es posible especificar sólo algunos de ellos y recodificar el resto utilizando “else” para indicar que se asigne un valor al resto de condiciones no especificadas.

18 = 1

Conceptos básicos y aplicaciones prácticas con Rcommander

```
18:23 = 1
18:23, 25, 27:30=1
ELSE = 2
```

El comando que genera esta opción en la ventana de instrucciones es `recode`. En él se indica el nombre de la nueva variable (`Edi.data$Edad_Re`), la variable a recodificar (`Edad` perteneciente al marco de datos `Edi.data`) y los valores antiguos junto a sus nuevas asignaciones.

```
Edi.data$Edad_Re <- recode(Edi.data$Edad, '10:13=1; 14:16=2;
17:27=3;as.factor.result=TRUE)
Edi.data$sexo_ber <- recode(Edi.data$sexo, '"mujer"="1";
"varón"="2"; ', as.factor.result=TRUE)
```

En el caso de que las variables a recodificar fueran categóricas es necesario indicar los valores antiguos entre comillas; para los valores nuevos se deberán utilizar las comillas si se desea mantener la variable como alfanumérica; en caso de que quiera reconvertirse una variable carácter en una variable numérica basta con especificar los nuevos valores sin el uso de comillas.

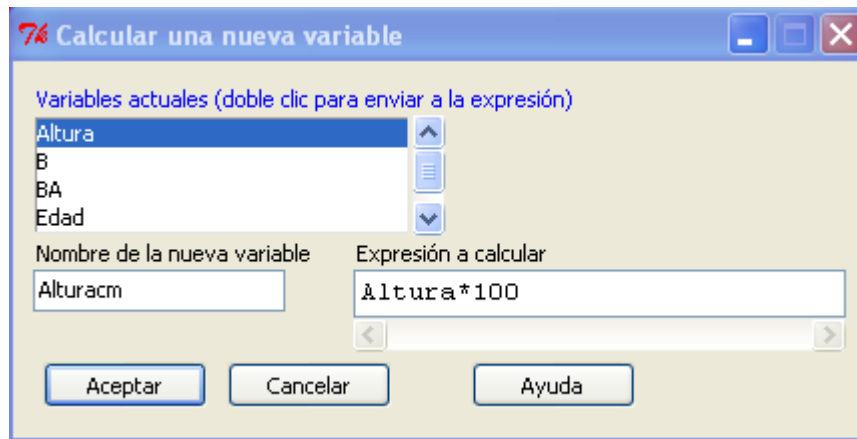
3.2.2 Calcular una nueva variable

Esta opción permite crear nuevas variables mediante transformaciones o combinaciones de variables ya existentes en el conjunto de datos activo o por medio de operaciones algebraicas. La nueva variable será añadida como columna al conjunto de datos activo.

En el ejemplo siguiente se ha creado una nueva variable a partir de la variable `altura`. La variable `altura` está expresada en metros y por medio de esta opción se creará una variable cuya métrica sean centímetros. Se trata de una simple transformación algebraica consistente en multiplicar la variable antigua

Conceptos básicos y aplicaciones prácticas con Rcommander

por 100. El comando que aparecerá en la ventana de escritura indica el nombre de la nueva variable altura_cm que pertenece (símbolo \$) al marco de datos Edi.data.



3.9. Figura. Calcular una nueva variable.

```
Edi.data$Alturacm <- with(Edi.data, Altura*100)
```

Las funciones aritméticas más comunes utilizadas por Rcommander se resumen en la siguiente tabla.

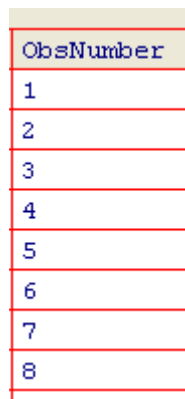
Función	Operación
+, -, *, /	Suma, Resta, Multiplicación, División
abs	Valor absoluto
asin acos	Inversas de las funciones trigonométricas
atan	
exp, log	Exponencial y logaritmo natural
round	Redondea
sin con tan	Funciones trigonométricas
sqrt () , ^	Raíz cuadrada, Potencia
%/%	División entera
%%	Resto de la división

3.2 Tabla. Funciones algebraicas

3.2.3 Añadir número de observaciones al conjunto de datos

Conceptos básicos y aplicaciones prácticas con Rcommander

Esta opción, Datos>Modificar variables del conjunto de datos activo>Añadir números de observaciones al conjunto de datos genera una variable que se añade al conjunto de datos activo y que asigna un valor secuencial a cada uno de los casos (filas) en el conjunto de datos activo. El nombre que por defecto recibe esta nueva variable es `ObsNumber` (número de observación).



ObsNumber
1
2
3
4
5
6
7
8

3.10. Figura. ObsNumber.

```
Edi.data$ObsNumber <- 1:976
```

3.2.4 Tipificar variables

Esta opción genera una variable tipificada que se añade como nueva variable en el conjunto de datos activo. El nombre por defecto de esta nueva variable está formado por el nombre de la variable antigua sin tipificar antepuesta por “Z.”. Si la variable a tipificar es `Altura`, la variable tipificada será `Z.Altura`.

En la ventana de instrucciones se puede comprobar el comando utilizado por Rcommander para realizar la tipificación; el comando `scale` genera una variable tipificada que es posteriormente añadida al marco de datos `Edi.data`. El proceso consiste en generar una variables de nombre “Z”, añadirla al

Conceptos básicos y aplicaciones prácticas con Rcommander

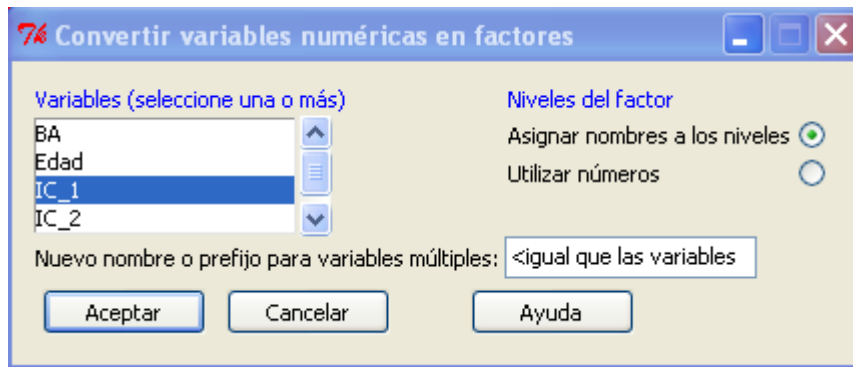
conjunto de datos con el nombre `Z.Altura`, y una vez incorporada al conjunto de datos activo finalmente eliminarla `remove(.Z)`.

```
.Z <- scale(Edi.data[,c("Altura")])
Edi.data$Z.Altura <- .Z[,1]
remove(.Z)
```

3.2.5 Convertir variable numérica en factor

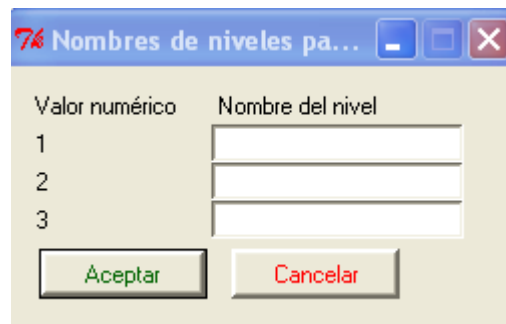
Un factor es una variable categórica utilizada normalmente con el fin de clasificar y reorganizar datos. Es una variable discreta con un número de niveles finito. Los factores en R son un modo (el más eficiente) de almacenar variables categóricas. Por defecto los niveles del factor se almacenan siguiendo el orden alfanumérico, por lo tanto en el caso de la variable `sexo` cuyos niveles pueden ser “Varón” y “Mujer”, mujer precederá a varón. El orden de los niveles del factor es importante porque determinará el orden en que aparecerán en las tablas o gráficos en los que vayan a utilizarse.

Rcommander permite transformar cualquier variable en factor; es importante sin embargo tener en cuenta que las variables que vayan a convertirse en factores no han de presentar demasiadas categorías. Por ejemplo, no tendría sentido convertir en factor la variable `altura`. Sin embargo, sí podríamos convertir en factor una recodificación de esta variable en varios niveles. La conversión de una variable en factor comienza con el siguiente cuadro de diálogo:



3.11. Figura. Conversión en factores.

En este cuadro de diálogo se selecciona la variable que se desea convertir en factor, se le asigna un nuevo nombre (o se mantiene el antiguo), y se definen los nombres de cada uno de sus niveles o se numeran éstos. La pantalla que permite asignar nombres a cada uno de los niveles del factor es la siguiente:



3.12. Figura. Niveles de los factores.

El comando que permite tratar una variable como factor aparece en la ventana de instrucciones:

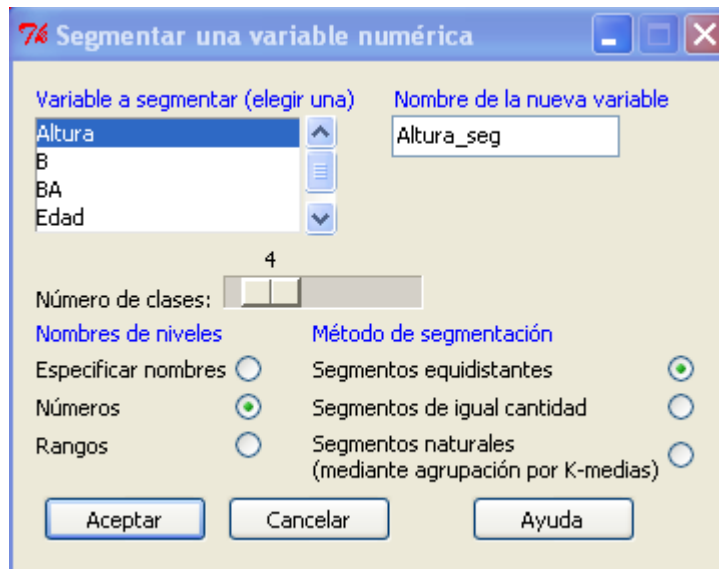
```
Edi.data$IC_1 <- as.factor(Edi.data$IC_1)
```

3.2.6 Segmentar una variable numérica

La opción de segmentación permite generar categorías a partir de los valores de una variable existente. Segmentar una variable numérica abre una ventana de diálogo en la que se seleccionará la variable que queremos

Conceptos básicos y aplicaciones prácticas con Rcommander

segmentar, se le asignará un nuevo nombre, y se definirán los criterios para ejecutar la categorización en un número de clases que definirá el investigador. La variable resultante es un factor. En el ejemplo se categoriza la variable altura en 4 clases equidistantes que serán nombradas utilizando números.



3.13. Figura. Segmentación de una variable.

El resultado es una variable con 4 valores equidistantes que muestra la siguiente tabla de frecuencias.

1	2	3	4
47	433	352	31

Si se hubiera definido un método de segmentación que mantenga segmentos con un número de casos similar, se hubiera obtenido una distribución de frecuencias que se aproxima a la distribución uniforme.

1	2	3	4
218	235	211	199

La opción Rangos disponible en esta ventana generaría una tabla de frecuencias para rangos de valores,

Conceptos básicos y aplicaciones prácticas con Rcommander

(1.32,1.49]	(1.49,1.66]	(1.66,1.83]	(1.83,2]
47	433	352	31

```
Edi.data$altura_seg <- bin.var(Edi.data$Altura, bins=4,
method='intervals', labels=c('1','2','3','4'))
```

3.2.7 Reordenar niveles de factor

Por defecto los niveles de un factor se almacenan siguiendo un orden alfanumérico, por lo tanto “Mujer” precedería a “Varón” o “1” precedería a “2”. El orden de los niveles del factor es importante porque determinará el orden en que aparecerán en las tablas o gráficos en los que vayan a utilizarse. Para imponer un orden predeterminado se utilizará la opción Reordenar niveles de factor. A partir de ella se accede a una ventana en la cual se podrá renombrar o mantener el nombre de la variable original. Es importante la opción “Factor de tipo ordenado” porque permite considerar los factores como variables categóricas ordenadas, o simplemente categóricas.

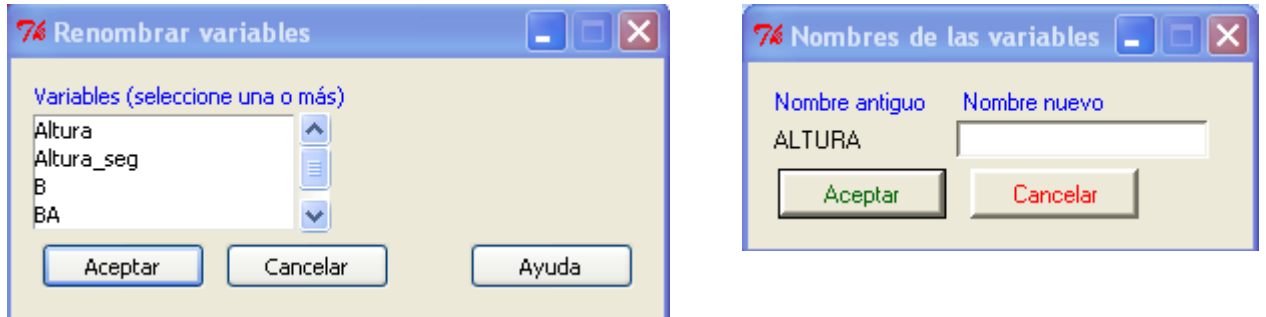


3.14. Figura. Reordenar niveles de factor.

3.2.8 Renombrar variables

Conceptos básicos y aplicaciones prácticas con Rcommander

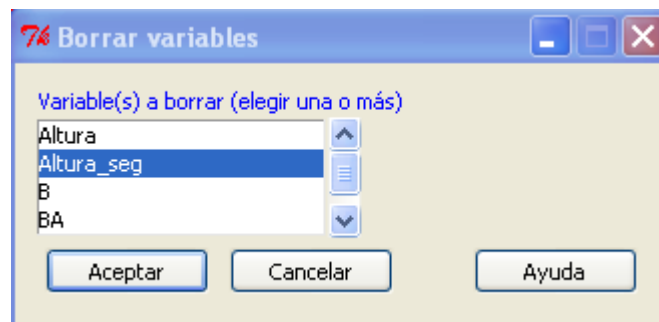
Esta opción permite asignar un nuevo nombre de variable a otra ya existente.



3.15. Figura. Renombrar variables.

3.2.9 Eliminar variables del conjunto de datos

Esta opción es útil para eliminar variables del conjunto de datos activo. Basta para ello seleccionar del siguiente cuadro de diálogo la/s variable/s a eliminar:



3.16. Figura. Eliminar variables.

4 Análisis de escalas

El estudio de una escala de actitudes, un inventario de personalidad o una prueba de aptitud numérica comprende un conjunto de análisis cuya finalidad es garantizar la fiabilidad de los datos y la validez de las inferencias que se derivan de ellos. La fiabilidad es un concepto relacionado con la precisión o consistencia interna de las medidas, y la validez se centra en el estudio de evidencias que avalen el uso de las puntuaciones obtenidas (Martínez Arias, 1995; Elosua, 2005). La fiabilidad y la validez son conceptos clave que han de ser estudiados, analizados y verificados en aras a un correcto uso de los tests.

Entre los acercamiento posibles al estudio de la fiabilidad de las medidas cabe diferenciar entre aquellos cuyo objetivo es analizar su estabilidad temporal y aquellos que focalizan su interés en la consistencia interna. Los primeros estiman la correlación entre las medidas obtenidas por el mismo test en dos momentos temporales. Los segundos ofrecen un indicador del grado de covariación entre los elementos del test. De entre los métodos desarrollados para la evaluación de la consistencia, el más utilizado en la investigación aplicada es el coeficiente alfa de Cronbach (Elosua y Zumbo, 2008).

El estudio de la validez se torna un asunto más complejo en tanto en cuanto es imposible reducir la validez de una inferencia a un indicador numérico. En la búsqueda de evidencias que avalen las inferencias construidas a partir de un test es posible diferenciar entre evidencias de tipo interno y evidencias externas (Elosua, 2003). Las primeras, incluirían un estudio de la estructura interna o dimensionalidad del test. Básicamente, son abordadas desde el modelo del factor común (son posibles acercamientos no lineales al tema, McDonald (1999)) a través de análisis factoriales que pueden ser exploratorios o confirmatorios, y a través del análisis de componentes principales. Su objetivo es reducir la información contenida en un conjunto (n) de variables observadas (ítems, escalas parciales, escalas) a un número de factores o dimensiones ($m < n$)

Conceptos básicos y aplicaciones prácticas con Rcommander

de modo que sea posible explicar la variabilidad observada en los datos a través de ellas.

Las segundas se basan en la recopilación de evidencias que analizan la relación entre las puntuaciones obtenidas por el test y criterios externos. Entre las evidencias utilizadas para garantizar la validez se encuentran la evidencia convergente/discriminante basada en el análisis de la matriz multirasgo/multimétodo (Campbell y Fiske, 1959) y las relaciones test/criterio que pueden establecerse por medio del coeficiente de validez, o correlación entre las puntuaciones obtenidas en el test y el criterio que se quiere predecir.

Rcommander permite estudiar la fiabilidad y estudiar fuentes de evidencia de validez interna y externa a través de funciones específicas diseñadas para el estudio de escalas o por medio de la aplicación de modelos estadísticos de regresión.

4.1 Presentación del ejemplo

La utilización de Rcommander en el estudio de las características de un inventario, tests o escala se presentará con el apoyo de los análisis ejecutados sobre un conjunto de datos provenientes de una investigación que tuvo como finalidad la adaptación al español del “*Eating “Disorder Inventory-3”* (Garner, 2004; Elosua, López-Jáuregui y Sánchez, 2010); los datos pertenecen a dos de las escalas parciales de este inventario; las dos están relacionadas con escalas para el estudio de trastornos de la conducta alimentaria: Insatisfacción corporal (IC) y Obsesión por la delgadez (OD). La escala Insatisfacción corporal está compuesta por 10 ítems, y la escala Obsesión por la delgadez por 7. Se analizará con mayor detalle la escala Insatisfacción Corporal. Todos los ítems tienen 6 categorías de respuesta (Siempre, Casi siempre, A menudo, A veces, Pocas veces, Nunca) que se responden en la dirección sintomática en una escala graduada con mínimo 0 y máximo 5 (0-1-2-3-4-5). Es decir a una mayor

Conceptos básicos y aplicaciones prácticas con Rcommander

puntuación se le asocia un nivel mayor de insatisfacción corporal. El contenido de los 10 ítems de la escala Insatisfacción Corporal es el siguiente:

- 1.- *Creo que mi estómago es demasiado grande.*
- 2.- *Pienso que mis muslos son demasiado gruesos.*
- 3.- *Creo que mi estómago tiene el tamaño adecuado.*
- 4.- *Me siento satisfecho con mi figura.*
- 5.- *Me gusta la forma de mi trasero.*
- 6.- *Creo que mis caderas son demasiado anchas.*
- 7.- *Me siento hinchado después de una comida normal.*
- 8.- *Creo que el tamaño de mis muslos es adecuado.*
- 9.- *Creo que mi trasero es demasiado grande.*
- 10.- *Creo que mis caderas tienen el tamaño adecuado.*

Los ítems de la escala Obsesión por la Delgadez son:

- 1.- *Como dulces y carbohidratos sin preocuparme.*
- 2.- *Pienso en ponerme a dieta.*
- 3.- *Me siento muy culpable cuando como en exceso.*
- 4.- *Me aterroriza ganar peso.*
- 5.- *Exagero o doy demasiada importancia al peso.*
- 6.- *Estoy preocupada porque querría ser una persona más delgada.*
- 7.- *Creo que puedo conseguir mis objetivos.*

El conjunto de datos a analizar está compuesto por las siguientes variables:

- 1.- *Edad: variable numérica*
- 2.- *Sexo: Factor con dos niveles Varón/Mujer*
- 3.- *Peso: Peso en kilogramos de cada participante*
- 4.- *Peso ideal. Peso considerado ideal por cada participante*
- 5.- *Altura. Altura en metros de cada participante*
- 6:12.- *OD_1 : OD_7.- Ítems perteneciente a la escala Obsesión por la delgadez*
- 13:22.- *IC_1 : IC_10:- Ítems pertenecientes a la escala Insatisfacción Corporal*

Conceptos básicos y aplicaciones prácticas con Rcommander

- 23.- *BA. Valores de cada uno de los participantes en una escala de Baja Autoestima*
- 24.- *B. Valores de cada uno de los participantes en una escala de comportamiento bulímicos; Bulimia*

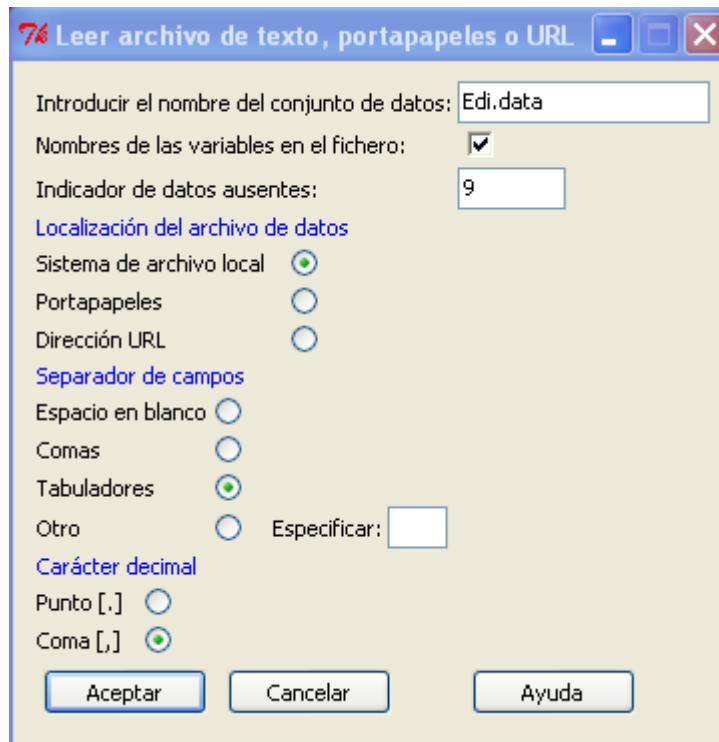
El análisis psicométrico de los datos seguirá un orden secuencial similar a cualquier estudio de este tipo. Básicamente se ajustará a las siguientes pautas.

- 1.- *Lectura de datos*
- 2.- *Corrección de la prueba.*
- 3.- *Análisis descriptivo de los ítems*
- 4.- *Fiabilidad. Índice de discriminación y coeficiente alfa.*
- 5.- *Validez. Estructura interna. Análisis factorial exploratorio*
- 6.- *Validez. Comparación de grupos*
- 7.- *Validez. Funcionamiento diferencial del ítem*
- 7.- *Construcción de baremos*

4.2 Lectura de datos

Los datos se encuentran en el fichero con formato ASCII `Edi0.dat` en el que los diferentes campos (variables) han sido delimitados por tabuladores. Tenemos que importar los datos, y para ello utilizamos la opción `Datos > Importar Datos > Desde archivo de texto o portapapeles`. La ventana abierta permitirá asignar un nombre al conjunto de datos en R que no tiene por qué coincidir con el nombre de fichero que está siendo importado. Es necesario especificar la información necesaria para la correcta lectura de los datos. En nuestro caso, tal y como puede verse en la imagen siguiente, hemos llamado al conjunto de datos `Edi.data`, indicamos a R que los nombres de las variables están incluidas en el fichero, que utilizamos como separador entre variables el tabulador, y que las variables numéricas utilizan la coma como signo decimal.

Conceptos básicos y aplicaciones prácticas con Rcommander



4.1. Figura. Importación de datos

Cuando pulsemos **Aceptar** aparecerá el comando de lectura tanto en la ventana de instrucciones como en la ventana de salida, y al mismo tiempo en la ventana de mensajes podemos comprobar el siguiente mensaje:

```
Edi.data <-read.table("C:/Edi0.dat", header=TRUE, sep="\t",
na.strings="NA", dec=",", strip.white=TRUE)
```

NOTA: El conjunto de datos Edi.data tiene 976 filas y 24 columnas.

El primer comando de R asigna el fichero Edi01.dat a un objeto R, un marco de datos (ver capítulo siguiente), que se convertirá a partir de ahora en el conjunto de datos activo. R ha leído un fichero de datos compuesto por 976 filas (casos) y 24 columnas (variables). La apariencia del fichero podremos comprobarla por medio de las opciones de **Editar conjunto de datos** o **Visualizar conjunto de datos**.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

	Edad	Sexo	Peso	Pesideal	Altura	OD_1	OD_2	OD_3	OD_4	OD_5	OD_6	OD_7	IC_1	IC_2	IC_3
1	16	mujer	58	52	1.67	0	5	5	5	5	5	5	4	5	
2	18	mujer	NA	52	1.70	3	2	0	5	2	2	2	0	1	
3	17	mujer	46	50	1.62	5	0	2	0	1	0	0	0	0	
4	17	mujer	55	50	1.60	0	5	3	5	4	4	5	4	2	
5	17	mujer	55	53	1.70	2	2	0	3	2	3	0	0	0	
6	18	mujer	NA	60	1.75	1	5	3	5	3	3	4	3	4	
7	17	mujer	NA	NA	NA	5	0	0	0	3	0	0	0	0	
8	17	mujer	NA	49	NA	4	5	5	5	5	5	5	5	5	
9	18	mujer	52	52	1.59	4	1	0	1	0	0	0	1	2	
10	18	mujer	NA	59	1.65	1	2	3	3	3	2	2	1	3	
11	17	mujer	68	65	1.66	5	2	1	2	2	2	3	2	3	
12	17	mujer	64	NA	1.52	2	2	0	1	1	1	0	1	1	
13	18	mujer	49	48	1.58	5	2	1	2	0	2	0	1	1	
14	17	mujer	59	42	1.45	2	5	5	3	4	3	5	5	5	
15	16	mujer	51	50	1.63	2	4	5	5	2	2	4	4	5	
16	17	mujer	NA	53	1.74	2	4	2	5	2	5	5	5	5	
17	14	mujer	NA	49	1.65	3	1	1	2	2	2	2	1	3	
18	13	mujer	58	52	1.66	2	2	3	2	2	2	0	3	3	
19	14	mujer	NA	52	1.69	3	0	0	0	0	0	0	0	0	
20	14	mujer	60	57	1.55	1	2	0	1	1	2	1	2	1	

4.2. Figura. Visualizar datos.

El contenido y características de las variables que componen el marco de datos (data frame) se obtiene fácilmente a través de la opción Estadísticos>Resúmenes>Conjunto de datos activo.

Edad	Sexo	Peso	Pesideal
Min. :10.00	mujer:517	Min. : 26.00	Min. : 6.00
1st Qu.:13.00	varón:459	1st Qu.: 50.00	1st Qu.: 50.00
Median :14.00		Median : 56.00	Median : 55.00
Mean :14.80		Mean : 56.93	Mean : 56.79
3rd Qu.:16.00		3rd Qu.: 64.00	3rd Qu.: 62.00
Max. :27.00		Max. : 98.00	Max. : 85.00
NA's : 8.00		NA's :204.00	NA's :540.00
Altura	OD_1	OD_2	OD_3
Min. : 1.320	Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.: 1.590	1st Qu.:2.000	1st Qu.:0.000	1st Qu.:0.000
Median : 1.650	Median :3.000	Median :1.000	Median :1.000
Mean : 1.651	Mean :2.756	Mean :1.490	Mean :1.363
3rd Qu.: 1.720	3rd Qu.:4.000	3rd Qu.:3.000	3rd Qu.:2.000
Max. : 2.000	Max. :5.000	Max. :5.000	Max. :5.000
NA's :113.000	NA's :3.000	NA's :8.000	NA's :9.000

Conceptos básicos y aplicaciones prácticas con Rcommander

<i>OD_4</i>	<i>OD_5</i>	<i>OD_6</i>	<i>OD_7</i>
<i>Min. :0.000</i>	<i>Min. :0.000</i>	<i>Min. :0.000</i>	<i>Min. :0.000</i>
<i>1st Qu.:1.000</i>	<i>1st Qu.:0.000</i>	<i>1st Qu.:0.000</i>	<i>1st Qu.:0.000</i>
<i>Median :2.000</i>	<i>Median :1.000</i>	<i>Median :1.000</i>	<i>Median :1.000</i>
<i>Mean :2.263</i>	<i>Mean :1.643</i>	<i>Mean :1.746</i>	<i>Mean :1.846</i>
<i>3rd Qu.:4.000</i>	<i>3rd Qu.:3.000</i>	<i>3rd Qu.:3.000</i>	<i>3rd Qu.:3.000</i>
<i>Max. :5.000</i>	<i>Max. :5.000</i>	<i>Max. :5.000</i>	<i>Max. :5.000</i>
<i>NA's :9.000</i>	<i>NA's :3.000</i>	<i>NA's :7.000</i>	<i>NA's :7.000</i>
<i>IC_1</i>	<i>IC_2</i>	<i>IC_3</i>	<i>IC_4</i>
<i>Min. :0.000</i>	<i>Min. :0.000</i>	<i>Min. : 0.000</i>	<i>Min. : 0.000</i>
<i>1st Qu.:0.000</i>	<i>1st Qu.:0.000</i>	<i>1st Qu.: 1.000</i>	<i>1st Qu.: 2.000</i>
<i>Median :1.000</i>	<i>Median :1.000</i>	<i>Median : 3.000</i>	<i>Median : 3.000</i>
<i>Mean :1.596</i>	<i>Mean :1.708</i>	<i>Mean : 2.731</i>	<i>Mean : 2.939</i>
<i>3rd Qu.:2.500</i>	<i>3rd Qu.:3.000</i>	<i>3rd Qu.: 4.000</i>	<i>3rd Qu.: 4.000</i>
<i>Max. :5.000</i>	<i>Max. :5.000</i>	<i>Max. : 5.000</i>	<i>Max. : 5.000</i>
<i>NA's :1.000</i>	<i>NA's :2.000</i>	<i>NA's :10.000</i>	<i>NA's :11.000</i>
<i>IC_5</i>	<i>IC_6</i>	<i>IC_7</i>	<i>IC_8</i>
<i>Min. : 0.000</i>	<i>Min. : 0.000</i>	<i>Min. :0.000</i>	<i>Min. : 0.000</i>
<i>1st Qu.: 1.000</i>	<i>1st Qu.: 0.000</i>	<i>1st Qu.:1.000</i>	<i>1st Qu.: 1.000</i>
<i>Median : 2.000</i>	<i>Median : 1.000</i>	<i>Median :2.000</i>	<i>Median : 2.000</i>
<i>Mean : 2.552</i>	<i>Mean : 1.379</i>	<i>Mean :1.757</i>	<i>Mean : 2.546</i>
<i>3rd Qu.: 4.000</i>	<i>3rd Qu.: 2.000</i>	<i>3rd Qu.:2.000</i>	<i>3rd Qu.: 4.000</i>
<i>Max. : 5.000</i>	<i>Max. : 5.000</i>	<i>Max. :5.000</i>	<i>Max. : 5.000</i>
<i>NA's :28.000</i>	<i>NA's :10.000</i>	<i>NA's :5.000</i>	<i>NA's :10.000</i>
<i>IC_9</i>	<i>IC_10</i>	<i>BA</i>	<i>B</i>
<i>Min. : 0.000</i>	<i>Min. : 0.000</i>	<i>Min. : 0.000</i>	<i>Min. : 0.000</i>
<i>1st Qu.: 0.000</i>	<i>1st Qu.: 1.000</i>	<i>1st Qu.: 1.000</i>	<i>1st Qu.: 1.000</i>
<i>Median : 1.000</i>	<i>Median : 3.000</i>	<i>Median : 4.000</i>	<i>Median : 2.000</i>
<i>Mean : 1.368</i>	<i>Mean : 2.754</i>	<i>Mean : 4.303</i>	<i>Mean : 3.598</i>
<i>3rd Qu.: 2.000</i>	<i>3rd Qu.: 4.000</i>	<i>3rd Qu.: 6.000</i>	<i>3rd Qu.: 5.000</i>
<i>Max. : 5.000</i>	<i>Max. : 5.000</i>	<i>Max. :22.000</i>	<i>Max. :31.000</i>
<i>NA's :12.000</i>	<i>NA's :12.000</i>	<i>NA's :88.000</i>	<i>NA's :74.000</i>

De este modo se logra una breve descripción numérica de las variables que conforman el conjunto de datos en: valor mínimo (Min), valor máximo (Max), mediana (Median) y media aritmética (Mean). También nos ofrece los valores del primer y tercer cuartil; valores que dejan bajo ellos el 25% y el 75%

Conceptos básicos y aplicaciones prácticas con Rcommander

de la distribución. Desde el punto de vista del análisis de ítems con este comando comprobamos que para los ítems analizados el valor mínimo es 0 y el valor máximo 5. No existen valores aberrantes en nuestro conjunto de datos. El comando que bajo R ofrece esta salida es:

```
summary(Edi.data)
```

4.3 Corrección de la prueba. Recodificación de los ítems

La puntuación obtenida en un test o escala es una variable compuesta formada por combinación lineal de variables simples o de variables compuestas. Piénsese en una escala parcial compuesta por varios ítems, y en la puntuación total obtenida por medio de las puntuaciones parciales. La variable compuesta (X) es un indicador empírico del constructo que se pretende medir. La combinación lineal que define una variable compuesta se representa por medio de la siguiente ecuación en la que n representa el número de elementos a combinar.

$$X_j = w_1X_{1j} + w_2X_{2j} + w_3X_{3j} + \dots + w_nX_{nj}$$

Dónde X_j es la puntuación obtenida por el sujeto j ,
 w_i , i es el coeficiente de ponderación del ítem i ,
 X_{ij} , es la puntuación obtenida por el sujeto j en el ítem i .

Dado que el peso de cada uno de los elementos sobre la puntuación total puede diferir, la ponderación (w) es necesaria para reflejar el peso asignado a cada uno de los elementos. Por ejemplo, supongamos que la nota final en una asignatura está compuesta por los resultados obtenidos en dos exámenes diferentes, y que el peso del primer examen en la nota final es el doble que el que tendrá el segundo. En este caso la variable compuesta formada por dos variables simples podría representarse del siguiente modo:

$$X_{total} = 2X_1 + X_2$$

Conceptos básicos y aplicaciones prácticas con Rcommander

En las situaciones de test, sin embargo, la fórmula habitual asigna el mismo peso a cada uno de los ítems que componen la prueba. En estas circunstancias el coeficiente de ponderación es 1 ($w_i=1$) y la expresión anterior se simplifica y adquiere la siguiente forma:

$$X = \sum_{i=1}^n X_i$$

Donde n es el número de elementos o ítems
 X_i , la puntuación obtenida en el ítem i .

La puntuación obtenida, X ; es la representación empírica del constructo que se desea medir. La exigencia demandada a la puntuación total respecto a su representatividad del constructo que quiere medir se hace extensible a cada uno de sus componentes. La puntuación final no será representativa ni significativa del constructo medido si sus componentes no alcanzan esas características (relevancia y representatividad).

Por ello, el análisis de una escala exige el análisis de sus componentes individuales a fin de explorar y comprobar su representatividad y consistencia. De hecho, si las variables simples no cumplen estos requisitos no podrá lograrse una variable compuesta que sea un indicador empírico relevante.

Existen más factores además de la ponderación, relacionados con la codificación de los ítems que afectan a la puntuación empírica; entre ellos, la presencia de ítems inversos y el azar.

En general, y utilizando como criterio la existencia de criterios objetivos para la corrección de las respuestas dadas a un ítem y, aunque existen más formatos que los expuestos, los ítems pueden clasificarse en dos grupos; aquellos que puede puntuarse en función de criterios objetivos (correcto, incorrecto, parcialmente correcto), y aquellos que carecen de criterios objetivos para su corrección (actitudes, intereses, personalidad...). Los ítems pertenecientes al primer grupo se conocen como ítems de ejecución máxima, y

Conceptos básicos y aplicaciones prácticas con Rcommander

los del segundo grupo son los ítems de ejecución típica. El azar puede afectar a los ítems de ejecución máxima, pues es posible responder correctamente aún sin conocer cuál es la respuesta correcta. En los ítems de ejecución típica el azar no es habitualmente un factor a considerar. En este tipo de ítems y con el fin de controlar los efectos de los estilos de respuesta es habitual utilizar ítems inversos.

4.3.1 Ítems inversos

Antes de la obtención de la puntuación total en una escala, es importante determinar cuál es el sentido de la medida o la valencia del ítem, es decir, cual es la dirección en que los ítems reflejan el constructo a medir.

- | |
|--|
| <p>1.- <i>Creo que mi estómago es demasiado grande</i></p> <ul style="list-style-type: none">1.- <i>Siempre</i>2.- <i>Casi siempre</i>3.- <i>Muchas veces</i>4.- <i>Algunas veces</i>5.- <i>Casi nunca</i>6.- <i>Nunca</i> <p>2.- <i>Creo que mi estómago tiene el tamaño adecuado</i></p> <ul style="list-style-type: none">1.- <i>Siempre</i>2.- <i>Casi siempre</i>3.- <i>Muchas veces</i>4.- <i>Algunas veces</i>5.- <i>Casi nunca</i>6.- <i>Nunca</i> |
|--|

Ambos ítems se relacionan con el mismo constructo, insatisfacción corporal, pero en sentido contrario. La inferencia en cuanto a la insatisfacción corporal de una persona que selecciona la opción “Siempre” en el primer ítem es

Conceptos básicos y aplicaciones prácticas con Rcommander

opuesta a las conclusiones sobre una persona que elige la opción “Siempre” en el segundo ítem. En el primer caso la puntuación asignada sería 5 y reflejaría un nivel de insatisfacción corporal mayor que el de aquella persona que hubiera obtenido un 1. Sin embargo aquél que responde “Siempre” al segundo ítem poseerá un nivel de insatisfacción corporal menor que la persona que responde al mismo ítem con la opción “Casi nunca”. Dado que la dirección de la escala es sintomática, a mayor puntuación mayor insatisfacción corporal, los ítems que representan el constructo en la dirección asintomática son ítems inversos.



4.3. Figura. Ítems inversos

Antes de proceder al análisis sobre conjuntos de datos que contienen ítems inversos, es necesario redireccionarlos situando todos los ítems de la escala en el mismo sentido. Los manuales de los tests incorporan la información necesaria para acometer esta tarea. En general, se aplicará la siguiente regla de reconversión:

$$X' = (X_{Ma} + X_{mi}) - X$$

dónde X' es la puntuación recodificada,
 X_{Ma} y X_{mi} , los valores máximo y mínimo de la variable a recodificar
 X , la opción seleccionada.

De este modo los valores 0, 1, 2, 3, 4 y 5 se convierten en 5, 4, 3, 2, 1 y 0.

Conceptos básicos y aplicaciones prácticas con Rcommander

$$X'_0 = (5 + 0) - 0 = 5$$

$$X'_1 = (5 + 0) - 1 = 4$$

$$X'_2 = (5 + 0) - 2 = 3$$

$$X'_3 = (5 + 0) - 3 = 2$$

$$X'_4 = (5 + 0) - 4 = 1$$

$$X'_5 = (5 + 0) - 5 = 0$$

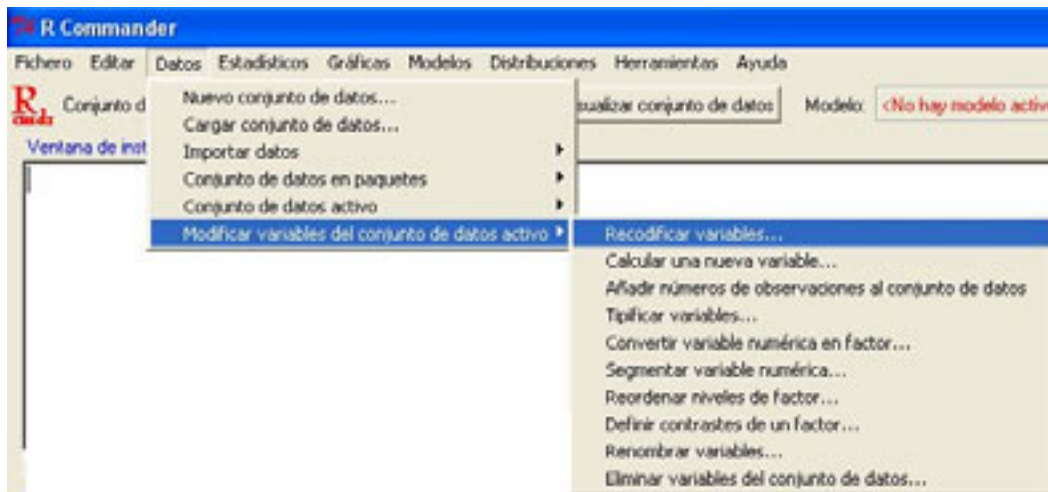
La utilización de ítems inversos ofrece al investigador la posibilidad de controlar dos factores relacionados con el estilo de respuesta que pueden contaminar la puntuación total. Por un lado la deseabilidad social o tendencia a responder en función de lo socialmente aceptado, y la aquiescencia o tendencia a manifestar un acuerdo sistemático con el enunciado del ítem independientemente de su contenido.

Además del control de la aquiescencia o deseabilidad social pueden existir razones de tipo teórico que justifiquen la utilización de alguna regla de codificación en la corrección de los ítems. Un ejemplo de esta situación es la corrección del “*Eating Disorder Inventory-3*” (EDI-3). En el manual de este cuestionario su autor recomienda utilizar una regla de corrección según la cual las dos categorías menos sintomáticas se funden en una sola. En el caso de los ítems de insatisfacción corporal las reglas de asignación a aplicar quedan recogidas en la siguiente tabla:

Ítem	Valores iniciales	Valores recodificados
1-2-6-7-9	5-4-3-2-1-0	4-3-2-1-0-0
3-4-5-8-10	5-4-3-2-1-0	0-0-1-2-3-4

4.1 Tabla. Recodificación ítems Insatisfacción Corporal

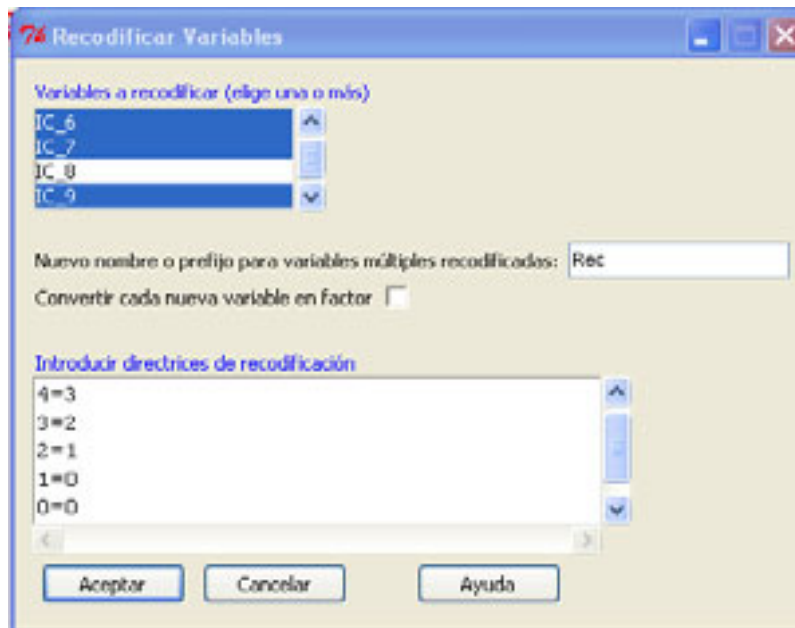
Para ejecutar esta reasignación de valores en Rcommander la opción a seleccionar es Recodificar Variables (Datos>Modificar variables del conjunto de datos activo>Recodificar variables).



4.4. Figura. Recodificación de variables.

La ventana de Rcommander permite seleccionar simultáneamente todas aquellas variables cuya recodificación se ajusta al mismo patrón (mantener pulsada la tecla Ctrl). En el proceso de recodificación podemos utilizar las mismas variables del conjunto de datos activo y reasignarles los nuevos valores, o podemos crear nuevas variables que contengan los valores recodificados. Esta es la opción por la que hemos optado y que recomendamos al lector. Téngase en cuenta que si se utilizan las mismas variables sería imposible invertir el proceso. Como resultado de la recodificación obtenemos nuevas variables cuyo nombre es igual al de las variables ya existentes precedidas por el prefijo Rec.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



4.5. Figura. Asignación de nuevos valores.

Una vez recodificadas las variables, podemos comprobar que R las ha añadido al conjunto de datos activo.

	RecIC_1	RecIC_2	RecIC_6	RecIC_7	RecIC_9	RecIC_3	RecIC_4
1	3	4	4	0	4	4	4
2	0	0	0	1	0	0	0
3	0	0	1	0	0	1	2
4	3	1	1	1	3	4	3
5	0	0	0	2	0	1	2
6	2	3	0	0	1	3	4
7	0	0	0	0	0	3	4
8	4	4	4	0	4	4	4
9	0	1	0	0	0	0	0
10	0	2	1	1	2	0	2
11	1	2	1	2	1	3	2
12	0	0	1	1	0	2	0
13	0	0	1	0	0	1	0
14	4	4	NA	1	4	3	4
15	3	4	4	3	3	3	3
16	4	4	4	0	4	4	4
17	0	2	0	0	3	2	1
18	2	2	1	0	0	2	1
19	0	0	1	2	0	0	1

4.6. Figura. Visualización del conjunto de datos activo.

Conceptos básicos y aplicaciones prácticas con Rcommander

Los comandos utilizados por R para la recodificación aparecen en la ventana de instrucciones, en donde podemos comprobar que la función utilizada es `recode`. El comando crea una nueva variable (\$) en el conjunto de datos activo (`Edi.data`) denominada `RecIC_1`, a partir de la recodificación de los valores de la variable `IC_1` perteneciente (\$) al conjunto de datos activo (`Edi.data`).

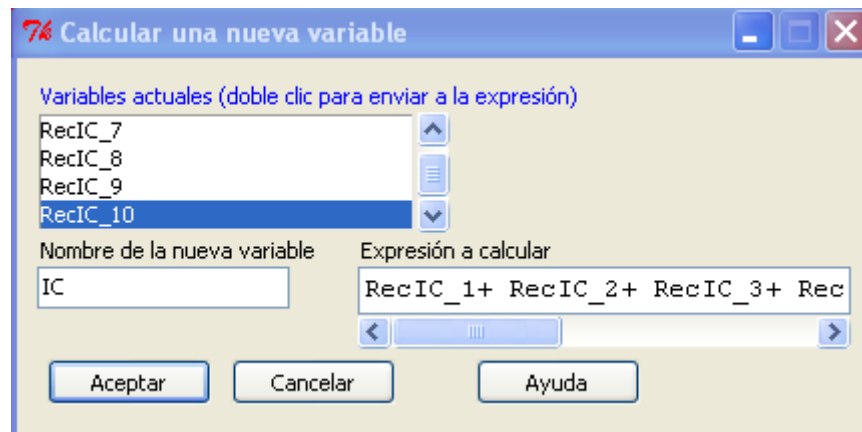
```
Edi.data$RecIC_1 <- recode(Edi.data$IC_1, '5=4; 4=3; 3=2; 2=1; 1=0;
0=0; ', as.factor.result=FALSE)
Edi.data$RecIC_2 <- recode(Edi.data$IC_2, '5=4; 4=3; 3=2; 2=1; 1=0;
0=0; ', as.factor.result=FALSE)
Edi.data$RecIC_6 <- recode(Edi.data$IC_6, '5=4; 4=3; 3=2; 2=1; 1=0;
0=0; ', as.factor.result=FALSE)
Edi.data$RecIC_7 <- recode(Edi.data$IC_7, '5=4; 4=3; 3=2; 2=1; 1=0;
0=0; ', as.factor.result=FALSE)
Edi.data$RecIC_9 <- recode(Edi.data$IC_9, '5=4; 4=3; 3=2; 2=1; 1=0;
0=0; ', as.factor.result=FALSE)
Edi.data$RecIC_3 <- recode(Edi.data$IC_3, '0=0; 1=0; 2=1; 3=2; 4=3;
5=4; ', as.factor.result=FALSE)
Edi.data$RecIC_4 <- recode(Edi.data$IC_4, '0=0; 1=0; 2=1; 3=2; 4=3;
5=4; ', as.factor.result=FALSE)
Edi.data$RecIC_5 <- recode(Edi.data$IC_5, '0=0; 1=0; 2=1; 3=2; 4=3;
5=4; ', as.factor.result=FALSE)
Edi.data$RecIC_8 <- recode(Edi.data$IC_8, '0=0; 1=0; 2=1; 3=2; 4=3;
5=4; ', as.factor.result=FALSE)
Edi.data$RecIC_10 <- recode(Edi.data$IC_10, '0=0; 1=0; 2=1; 3=2;
4=3; 5=4; ', as.factor.result=FALSE)
```

4.4 Corrección de la prueba. Obtención de la puntuación

Una vez recodificados los ítems de la escala Insatisfacción Corporal es posible obtener la puntuación para esta escala. Para ello es suficiente sumar los valores en cada uno de los ítems. Esta operación generará una nueva variable (IC) que se añadirá al conjunto de datos activo.

Conceptos básicos y aplicaciones prácticas con Rcommander

Para construir una nueva variable la opción de la barra de menús de Rcommander sería, Datos>Modificar variables del conjunto de datos activo>Calcular una nueva variable... Esta opción abre la siguiente ventana en la que se definirán el nombre de la variable a crear y la operación algebraica necesaria para ello.

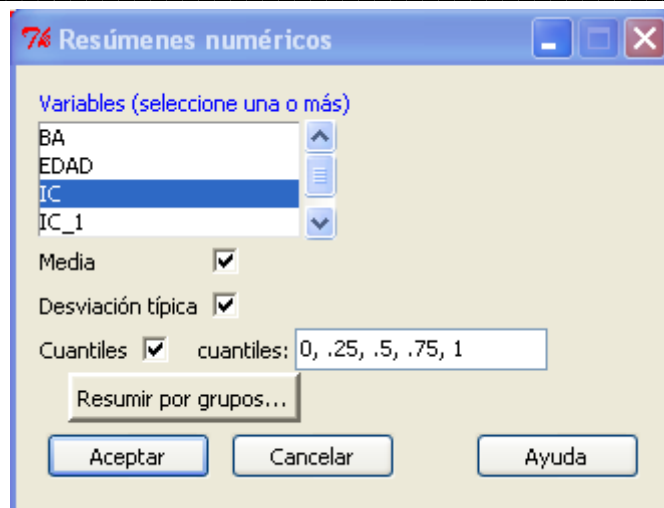


4.7. Figura. Cálculo de la puntuación total.

En el cómputo de la puntuación total es importante recordar que en el caso de que se hubiera recodificado ítems son éstos los que hay que incluir y no los originales.

```
> Edi.data$IC <- with(Edi.data, RecIC_1+ RecIC_2+ RecIC_3+ RecIC_4+  
RecIC_5+ + RecIC_6+ RecIC_7+ RecIC_8+ RecIC_9+ RecIC_10)
```

Una vez creada la nueva variable, para obtener información descriptiva sobre ella basta con seleccionar de la barra de menús la opción Estadísticos>Resúmenes>Resúmenes numéricos y elegir la variable IC del conjunto de variables.

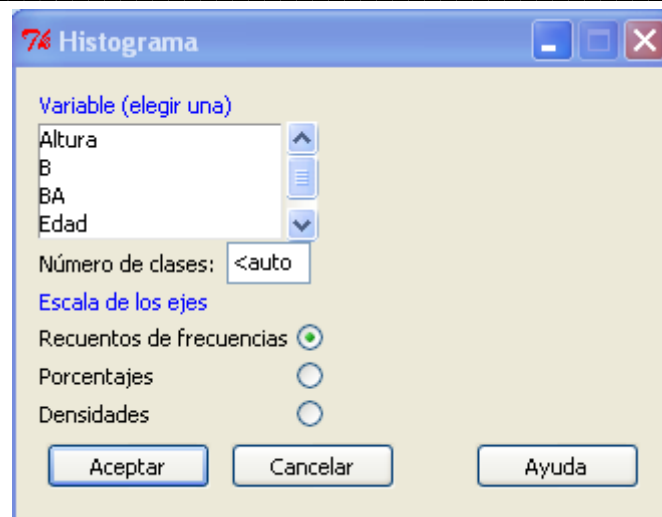


4.8. Figura. Resumen numérico.

La salida de resultados muestra que la media aritmética de la variable IC es 12,20 (mean), su desviación estándar (sd) es 10,22 y que el valor de la mediana es 11 (50%).

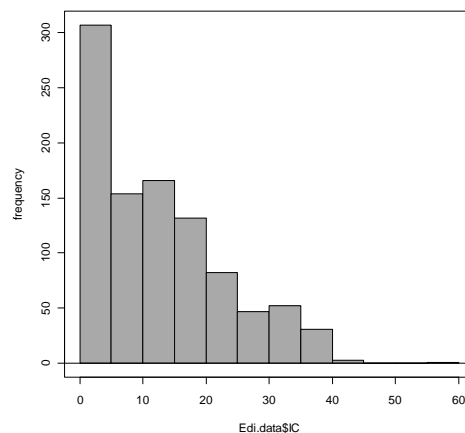
<i>mean</i>	<i>sd</i>	0%	25%	50%	75%	100%	<i>n</i>	<i>NA</i>
12.20705	10.22606	0	3	11	18	40	908	68

Para obtener una representación gráfica de la distribución de estas variables Rcommander ofrece varias posibilidades accesibles en la opción *Graficas* de la barra de menús. En este caso optamos por el histograma porque estamos interesados en una sola variable, la naturaleza de la variable es continua, y queremos información sobre la distribución de esta variable en el total del grupo de participantes. La opción *Graficas>Histograma* abre la siguiente ventana en la que se definirá la variable de interés y se fijarán varios parámetros de representación.



4.9. Figura. Ventana Histograma

Este comando y el resto de comandos relacionados con la ejecución de gráficos abren una nueva ventana en R (dispositivo gráfico); los gráficos no aparecen en la ventana de salidas de Rcommander. Una vez generado el gráfico es sencillo copiarlo o guardarlo bajo diferentes formatos gráficos; para ello es suficiente utilizar el botón derecho del ratón (Ctrl+C, Ctrl+V).



4.10. Figura. Insatisfacción corporal. Histograma

En el eje X del histograma se rotula el nombre de la variable (`Edi.data$IC`), y en el eje Y sus frecuencias absolutas (*frequency*). La apariencia de este gráfico es muy básica y algo tosca, pero es sencillo modificarlo y adaptarlo a los gustos y necesidades del investigador.

El histograma refleja una distribución sesgada de la variable insatisfacción corporal, en el que existe una acumulación de sujetos en el lado izquierdo de la distribución, donde los valores son más bajos. Hay muy pocos sujetos en la cola derecha de la distribución. ¿Qué significa esto? Para una correcta interpretación de cualquier distribución es importante conocer la variable representada. En nuestro caso la variable de estudio es la insatisfacción corporal, y está siendo analizada en una muestra no-clínica. Por lo tanto, el perfil que muestra el histograma es el perfil que esperábamos. La mayor parte de la población tiene niveles bajos de insatisfacción corporal, y existe un mínimo de la muestra con niveles de insatisfacción corporal elevados. ¿Cuál hubiera sido el perfil del histograma en una muestra clínica afectada por problemas de conducta alimentaria?

4.5 Análisis de ítems

El análisis de ítems trata de estudiar las características sustantivas y formales de un conjunto de ítems pertenecientes a una escala determinada. El análisis ha de ser cualitativo y cuantitativo y siempre ha de estar conectado con los objetivos del test. El análisis cualitativo de ítems estudiará aspectos lingüísticos y de contenido con el fin de analizar la adecuación lingüística, cultural y sustantiva en relación al objeto medido y a la población diana, que garantice tanto su relevancia como su pertinencia (Elosua, 2003; Elosua, Hambleton y Zenisky, 2006).

Sólo tras una primera fase, ineludible, en la que un grupo de expertos evalúa cada uno de los ítems comenzaría el análisis cuantitativo. Con el fin de que la puntuación obtenida en el test sea significativa y representativa del dominio sobre el cual se basa la inferencia, el análisis de ítems estudiará el modo en el que éstos alcanzan los objetivos marcados.

Conceptos básicos y aplicaciones prácticas con Rcommander

Los índices numéricos más habituales incluidos en el análisis de ítems pueden clasificarse en dos grandes bloques; aquellos relacionados con el modelo formal sobre el cual se ha construido el test, y aquellos índices relacionados con la validez del test. Dentro de la teoría clásica de test pertenecen al primer grupo el índice de dificultad, el índice de discriminación y el índice de fiabilidad (algo menos utilizado que los anteriores). El índice de dificultad se halla estrechamente ligado a la distribución de las respuestas, y el resto de índices se construye sobre el estudio de las relaciones entre variables. Entre los índices relacionados con la validez el índice de validez y el funcionamiento diferencial del ítem son los más destacados. El primero analiza la relación del ítem con un criterio externo y el segundo es una herramienta útil en el estudio de la estructura interna del test.

Excepto el funcionamiento diferencial del ítem, el resto son índices clásicos de la literatura psicométrica que son tratados prácticamente en todos los manuales de psicometría, y son acometidos por todo aquel interesado en la construcción o adaptación de test. El estudio del funcionamiento diferencial del ítem es algo más reciente respecto a los demás índices; sin embargo, ha de reconocerse que tiene ya una amplia trayectoria dentro de la psicometría y que existen ya procedimientos de estimación sencilla accesibles a todo investigador novel.

Modelo formal	Índice de dificultad Índice de atracción Índice de discriminación Índice de fiabilidad
Validez	Índice de validez Funcionamiento diferencial del ítem

4.2 Tabla. Análisis de ítems. Índices

4.6 Análisis de ítems. Índices de forma

4.6.1 Media aritmética

Cuando la naturaleza de la variable a analizar es dicotómica o el ítem es de ejecución máxima, la media aritmética del ítem informa sobre el nivel de dificultad del elemento; de hecho, en condiciones de dicotomía (correcto/incorrecto) la media aritmética del ítem se conoce como índice de dificultad del ítem (p_i) y no es más que la proporción de sujetos que responden correctamente al ítem. Para formatos de respuesta no dicotómicos la interpretación de la media aritmética requiere asociarla con la desviación estándar del ítem.

$$p_i = \frac{1}{N} \sum_{j=1}^N X_{ij}$$

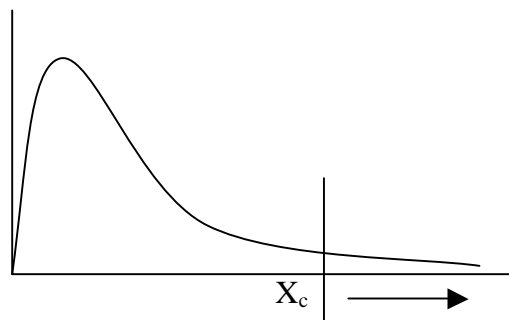
Donde p_i es el índice de dificultad del ítem i ,
 X_{ij} , es la respuesta dada al ítem i por el sujeto j
 N , es el número de sujetos.

A la pregunta común acerca del valor óptimo del índice de dificultad de un ítem o del valor recomendable para la media aritmética de un ítem es menester responder que el análisis de ítems ha de enmarcarse siempre en el contexto de los objetivos perseguidos. Sólo ellos podrán definir el nivel de discriminación necesario y sólo ellos permitirán delimitar si la información sobre los sujetos ha de ser más precisa en algún punto o intervalo del continuo de medida. En función de los objetivos del test las medias aritméticas de las respuestas a los ítems podrán ser diferentes (altas, medias o bajas). Por ejemplo, en los contextos en que la distribución buscada sea asimétrica y se precise una discriminación fina en los valores altos de la escala las medias aritméticas de los ítems que permitan tal objetivo tendrían que ser evidentemente bajas (ítems difíciles).

Supongamos que se desee seleccionar un grupo de alumnos para la concesión de becas de alto nivel. Ese objetivo exige discriminar entre un grupo

Conceptos básicos y aplicaciones prácticas con Rcommander

de alumnos aquellos que podrán acceder a la beca; es decir, la selección de candidatos se llevaría a cabo en un área de la distribución limitada por los valores más altos de la escala. Para ello la distribución debería ser sesgada hacia el lado derecho y ello implicaría utilizar índices de dificultad bajos (ítems difíciles). En esta situación, que en el contexto de la teoría clásica de test está siempre supeditada a las características del grupo normativo, serían pocos los que respondieran a los ítems correctamente, y la mayoría de los miembros del grupo obtendría puntuaciones bajas. Es decir, la mayoría de las personas del grupo obtendría puntuaciones por debajo del punto de corte, X_c y sobre este punto habría un grupo de candidatos menor.



4.11. Figura. Distribución sesgada

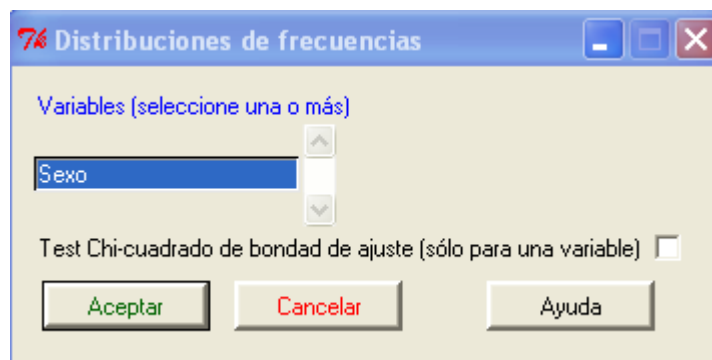
4.6.2 Índice de atracción

En el caso de escalas graduadas, escalas Likert, o ítems de ejecución óptima, el concepto de índice de dificultad carece de sentido, y habría de hablarse de la media aritmética del ítem. Sin embargo, resulta más informativo que la media aritmética el estudio de la distribución de frecuencias asociada a las respuestas. De este modo obtendríamos información sobre la “atracción” que ejerce cada una de las opciones de respuesta en la muestra de interés. El número (o en su caso proporción o porcentaje) de sujetos que elige cada una de las categorías del ítem se conoce como índice de atracción del ítem.

En este ejemplo obtendremos la distribución de frecuencias referida a cada uno de los ítems de la escala que estamos analizando. La opción para la

Conceptos básicos y aplicaciones prácticas con Rcommander

obtención de la tabla de frecuencias es: Estadísticos>Resúmenes>Distribución de frecuencias. Ahora bien, si nos fijamos en la ventana a la que accedemos, comprobaremos que sólo aparecen en ella variables categóricas, es decir, factores. Dado que el tratamiento que están recibiendo los ítems es de tipo numérico no aparecen en la ventana de selección de variables.



4.12. Figura. Selección de variables.

Para salvar esta restricción podemos hacer uso de dos opciones.

- Convertir las variables a analizar en factores a través de la opción Datos>Modificar variables del conjunto de datos activo> Convertir variable numérica en factor y optar posteriormente al análisis de distribución de frecuencias.
- Teclar directamente en la ventana de instrucciones el siguiente comando.

```
tabla.freq <-as.data.frame( apply(Edi.data[, 25:34], 2, table) )  
tabla.freq
```

Aconsejamos la segunda opción. En ella se especifica el conjunto de datos activo (`Edi.data`), y las columnas que ocupan las variables que queremos analizar; los ítems recodificados ocupan las posiciones 25-34 (la función `apply` será introducida más adelante).

Conceptos básicos y aplicaciones prácticas con Rcommander

Rcommander ofrece la siguiente salida: nombre de cada uno de los ítems analizados (RecIC_1...RecIC_10) y la distribución de respuestas correspondiente.

	RecIC_1	RecIC_2	RecIC_6	RecIC_7	RecIC_9	RecIC_3	RecIC_4	RecIC_5	RecIC_8	RecIC_10
0	556	535	612	476	617	382	433	308	338	374
1	175	145	135	270	136	128	129	137	124	126
2	82	102	75	104	73	172	200	222	177	190
3	76	67	65	72	54	171	115	152	190	169
4	86	125	79	49	84	113	88	129	137	105

La interpretación de la tabla es sencilla; Respecto al primer ítem de la escala Insatisfacción corporal se concluye que a 556 personas les ha sido asignada la puntuación 0, mientras que 175 han logrado una puntuación de 1; el valor 2 ha sido obtenido por 82 personas, el 3 por 76 participantes, y finalmente son 86 personas las que recibieron una puntuación de 4 en este ítem.

Si en lugar de frecuencias absolutas el interés se centrara en la obtención de proporciones, será suficiente escribir en la ventana de comandos la siguiente función:

```
prop.table (tabla.freq,2)
```

RecIC_1	RecIC_2	RecIC_6	RecIC_7	RecIC_9	RecIC_3	RecIC_4	RecIC_5	RecIC_8	RecIC_10	
0	0.57	0.55	0.63	0.49	0.64	0.40	0.45	0.32	0.35	0.39
1	0.18	0.15	0.14	0.28	0.14	0.13	0.13	0.14	0.13	0.13
2	0.08	0.10	0.08	0.11	0.08	0.18	0.21	0.23	0.18	0.20
3	0.08	0.07	0.07	0.07	0.06	0.18	0.12	0.16	0.20	0.18
4	0.09	0.13	0.08	0.05	0.09	0.12	0.09	0.14	0.14	0.11

Los estadísticos descriptivos de estas distribuciones, la media aritmética y la varianza, se obtienen a través de la opción Estadísticos > Resúmenes > Resúmenes numéricos. En ella se seleccionarán las variables de interés dentro del conjunto de datos activo. En nuestro caso, las variables son los 10 ítems recodificados. El comando que generará tal información es el siguiente.

Conceptos básicos y aplicaciones prácticas con Rcommander

```
numSummary(Edi.data[,c("RecIC_1", "RecIC_2", "RecIC_3", "RecIC_4",  
"RecIC_5", "RecIC_6", "RecIC_7", "RecIC_8", "RecIC_9",  
RecIC_10")], statistics=c("mean", "sd", "quantiles"),  
quantiles=c(0,.25,.5,.75,1))
```

La función a aplicar es numSummary sobre las variables del conjunto de datos activo (nótese que en esta ocasión en lugar de utilizar el símbolo \$ de pertenencia, se ha optado por crear un vector (c) para acceder a las variables pertenecientes al conjunto de datos activo). La ventana de salida de Rcommander ofrece una tabla con el siguiente aspecto:

	mean	sd	0%	25%	50%	75%	100%	n	NA
RecIC_1	1.0225641	1.542520	0	0	0	1.5	5	975	1
RecIC_2	1.2063655	1.715512	0	0	0	2.0	5	974	2
RecIC_3	1.4875776	1.448727	0	0	1	3.0	4	966	10
RecIC_4	1.2704663	1.371921	0	0	1	2.0	4	965	11
RecIC_5	1.6381857	1.421251	0	0	2	3.0	4	948	28
RecIC_6	0.9057971	1.510731	0	0	0	1.0	5	966	10
RecIC_7	0.9670443	1.304609	0	0	1	1.0	5	971	5
RecIC_8	1.6521739	1.474175	0	0	2	3.0	4	966	10
RecIC_9	0.8962656	1.525017	0	0	0	1.0	5	964	12
RecIC_10	1.4865145	1.425486	0	0	1	3.0	4	964	12

En ella aparecen la media aritmética de cada ítem (mean), la desviación estándar (sd), el primer cuartil (25%), el segundo cuartil o mediana de la distribución (50%), tercer cuartil (75%), número de sujetos (n) y el número de valores faltantes en cada uno de los ítems (NA).

4.6.3 Gráfico de la distribución de respuestas

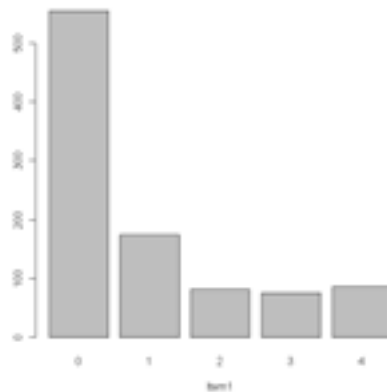
Otro modo de conseguir información representativa y clara sobre la distribución de frecuencias nos lo ofrece el gráfico de barras. Su obtención es sencilla en Rcommander; basta seleccionar en la barra de menús la opción

Conceptos básicos y aplicaciones prácticas con Rcommander

Graficas>Gráfica de barras, donde podemos seleccionar la/s variable/s cuya distribución queremos conocer.

La gráfica generada tiene un aspecto básico y sencillo que puede ser fácilmente modificado y mejorado por medio de la especificación de algunos parámetros.

```
barplot ( table(Edi.data[,25]), xlab="Item1" )  
ó  
barplot ( table(Edi.data$RecIC_1), xlab= "Item1" )
```



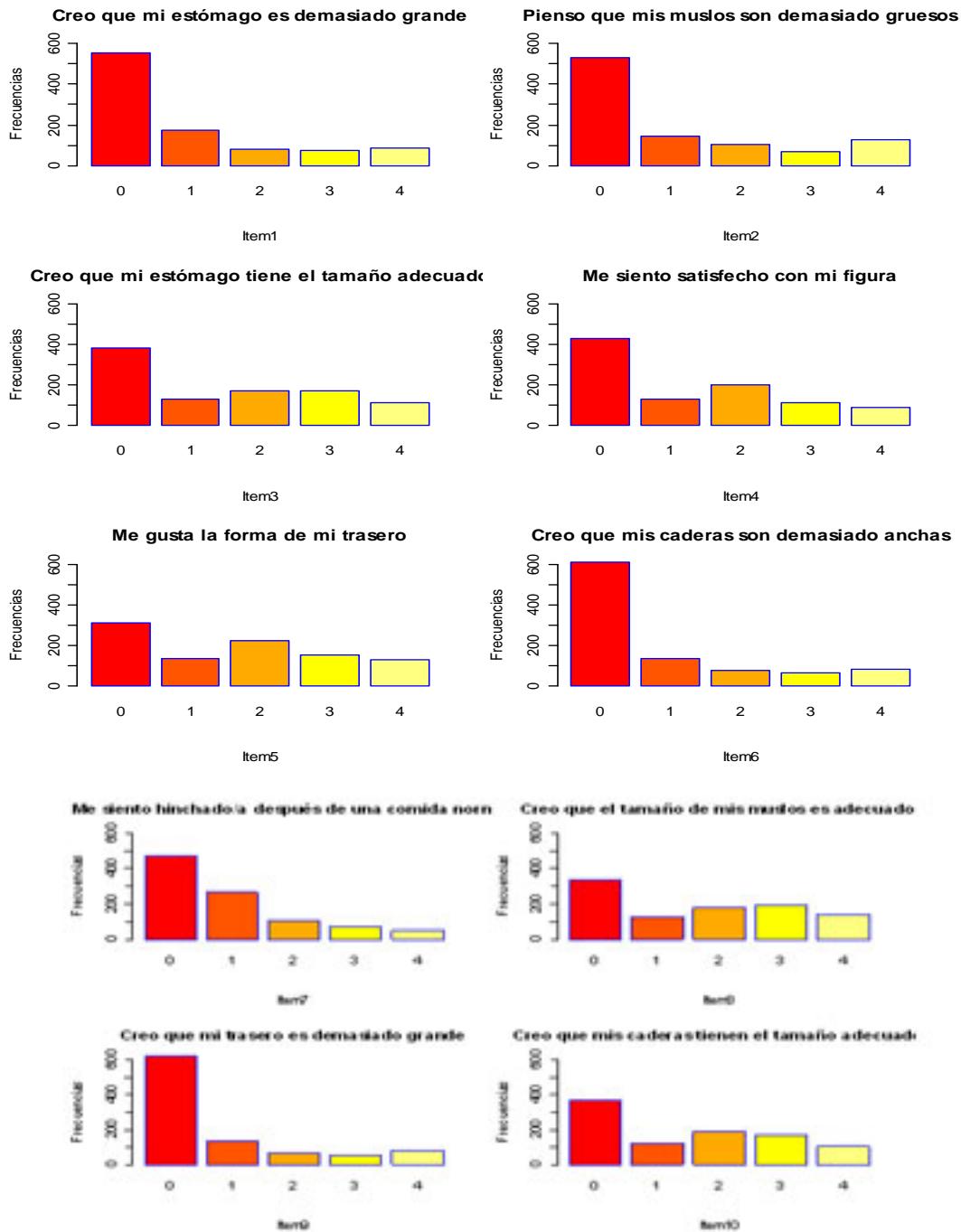
4.13. Figura. Diagrama de barras

Es posible modificar el aspecto del diagrama fijando los límites del eje de ordenadas (`ylim=range(0,100)`), añadiendo un título al mismo `ylab="Frecuencias"`, especificando el color de las barras (`col=heat.colors(5)`), o incorporando un título, que en este caso identificamos con el contenido del ítem (`main= "Creo que mi estómago es demasiado grande"`), y un indicativo del lugar que ocupa en la escala (`sub="Item1"`).

```
barplot(table(Edi.data$Rec_item1),ylim=range(0,100),  
ylab="Frecuencias", border="blue", col=heat.colors(5),main=  
"Creo que mi estómago es demasiado grande", sub="Item1")
```

Conceptos básicos y aplicaciones prácticas con Rcommander

Automatizando este procedimiento de generación de gráficos se crearon las siguientes gráficas de barras para el conjunto de ítems que forman la escala insatisfacción corporal.



4.14. Figura. Diagrama de barras de los ítems.

5 Fiabilidad. Índice de discriminación y coeficiente alfa

5.1 Índices de relación. Matriz de correlaciones

El estudio de las relaciones entre ítems ocupa un lugar destacado en el análisis de ítems; su estudio es abordado por los índices de relación, que básicamente analizan las correlaciones entre los ítems o entre conjuntos de ítems. Una correlación es un índice del grado de relación entre dos variables, si alguna de las variables es un ítem la información obtenida se refiere a éste último. El procedimiento de estimación de la correlación depende de la

Conceptos básicos y aplicaciones prácticas con Rcommander

naturaleza (discreta/continua) de las variables. Es posible estimar una correlación biserial, biserial-puntual, tetracórica, policórica o producto-momento de Pearson.

Correlación biserial puntual

Cuando una de las variables es dicotómica (p.e. ítem) y la otra variable es continua (p.e. la puntuación total) la correlación puede estimarse aplicando la siguiente fórmula:

$$\rho_{bp} = \frac{\mu_p - \mu_x}{\sigma_x} \sqrt{p/q} = \frac{\mu_p - \mu_q}{\sigma_x} \sqrt{pq}$$

Donde μ_p media aritmética en el test de los sujetos que han respondido al ítem correctamente,

μ_x , media aritmética del test,

μ_q , media aritmética en el test de los sujetos que han respondido al ítem incorrectamente,

σ_x , desviación estándar del test,

p, proporción de sujetos que han respondido correctamente

q, proporción de sujetos que han respondido incorrectamente.

Correlación biserial (poliserial)

Cuando se asume que a la variable dicotómica (o categorizada) subyace una variable latente que se distribuye según la ley normal, la correlación a estimar sería la correlación biserial. En este contexto, una de las variables es continua y la otra es una variable supuestamente dicotomizada o politomizada:

$$\rho_b = \frac{\mu_p - \mu_x}{\sigma_x} \frac{p}{y}$$

Donde μ_p media aritmética en el test de los sujetos que han respondido al ítem correctamente,

μ_x , media aritmética del test,

σ_x , desviación estándar del test,

p, proporción de sujetos que han respondido correctamente

y, la ordenada en el punto z que se corresponde al área bajo la curva normal asociada a p.

Conceptos básicos y aplicaciones prácticas con Rcommander

Lord y Novick (1968) formulan de este modo la relación entre la correlación biserial y la correlación biserial puntual.

$$\rho_{bp} = \frac{\rho_b \sqrt{pq}}{y}$$

Para los mismos datos, la estimación ofrecida por la correlación biserial es mayor que la obtenida por la correlación biserial puntual. La diferencia entre ambos índices es mínima cuando el índice de dificultad del ítem es medio, y su diferencia crece en la medida en que el índice de dificultad aumenta o disminuye, hasta el punto de que la correlación biserial puede incluso ser cuatro veces mayor que la correlación biserial-puntual (Magnuson, 1967).

Correlación Phi

Se utiliza cuando ambas variables son dicotómicas.

$$\rho_{ij} = \frac{P_{ij} - P_i P_j}{\sqrt{P_i Q_i P_j Q_j}}$$

Donde p_{ij} proporción de sujetos que responden correctamente a los ítems i y j

p_i proporción de sujetos que responden correctamente al ítem i

p_j proporción de sujetos que responden correctamente al ítem j

q_i proporción de sujetos que responden incorrectamente al ítem i

q_j proporción de sujetos que responden incorrectamente al ítem j .

Correlación tetracórica (policórica)

Cuando se trabaja con variables dicotomizadas (o categorizadas) se pueden estimar correlaciones tetracóricas (policóricas). En su aplicación se asume que bajo las variables dicotomizadas (categorizadas) subyacen dos variables que se distribuyen según la ley normal. No existe una fórmula simple para la obtención de la correlación tetracórica; R posee entre otros, el paquete `polycor` a través del cual se pueden estimar correlaciones policóricas y poliseriales.

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Correlación de Pearson

Cuando las dos variables son continuas y se asume una relación lineal entre ellas la correlación a utilizar es la correlación de Pearson. La correlación es una covarianza normalizada, de modo que el intervalo de valores posibles es -1,1.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Donde ρ_{XY} es la correlación entre X e Y,
 σ_X y σ_Y , son las desviaciones estándar de las variables X e Y
 $Cov(X,Y)$, es la covarianza entre las variables X e Y.

La expresión anterior es una formulación general de la correlación lineal entre dos variables observadas de las que la correlación biserial-puntual y phi no son más que casos especiales.

Si las correlaciones se estiman entre todos los ítems que componen una escala se obtiene la matriz de correlaciones. Esta matriz aporta información relevante sobre la estructura interna del test. La matriz de correlaciones es una matriz simétrica ($\rho_{12} = \rho_{21}$.) con unos en la diagonal.

Los valores esperables en la matriz de correlaciones entre ítems son por lo general, positivos. La aparición de valores negativos se debe habitualmente a la presencia de ítems inversos no recodificados. En el caso de que se analicen ítems pertenecientes a distintas escalas parciales, las correlaciones entre ítems pertenecientes a la misma escala parcial serán mayores que las estimadas entre ítems pertenecientes a distintas escalas. La inspección de la matriz de correlaciones ofrece información relevante sobre la estructura interna del test. Sin embargo, a medida que aumenta al número de ítems es difícil extraer información significativa y clara; habría que aplicar sobre ella procedimientos y modelos más adecuados entre los que destaca el modelo de análisis factorial.

5.2 Índices de relación. Índice de discriminación

La relación entre cada uno de los ítems que componen una escala y la escala misma se analiza habitualmente por medio del índice de discriminación (r_{iX}); la información aportada por este índice está ligada a la consistencia interna del test. El índice se obtiene por medio del coeficiente de correlación entre el ítem objeto de estudio y la puntuación total. Aunque teóricamente y al tratarse de una correlación los valores del índice de discriminación varían en un rango de -1,1, en el contexto del análisis de ítems no tendrían sentido los índices de discriminación negativos.

A pesar de que la fijación de puntos de corte referidos a la calidad de un índice es siempre interpretable en función del contexto, entre la comunidad psicométrica se acepta la siguiente clasificación referida al índice de discriminación (Ebel, 1965; Elosua, 2005).

r_{iX}	Nivel de discriminación
$0,40 < a_i$	Muy bueno
$0,30 < a_i < 0,39$	Bueno
$0,20 < a_i < 0,29$	Deficiente
$a_i < 0,19$	Insuficiente

4.3 Tabla. Índice de discriminación.

La relación ítem-test estimada por medio de la correlación ítem-puntuación total, presenta una característica que es preceptivo corregir cuando el número de ítems que componen la escala es reducido. La correlación ítem-test obtenida de este modo está inflada espuriamente por la inclusión del ítem cuyo índice de discriminación se desea estimar en la puntuación total de la escala. Este aumento espurio puede ajustarse fácilmente eliminando de la puntuación total la influencia del ítem y estimando el índice de discriminación corregido, entre el ítem y un compuesto formado por el resto de ítems de la escala. Este índice habitualmente recibe el nombre de “Índice de discriminación corregido”.

Conceptos básicos y aplicaciones prácticas con Rcommander

Ambos índices, el índice de discriminación y el índice de discriminación corregido, son ofrecidos por Rcommander.

Se pueden utilizar dos acercamientos para la obtención del índice de discriminación corregido; estimar la correlación entre el ítem y la escala de la que previamente se ha eliminado el mismo, o aplicar directamente la siguiente fórmula:

$$\rho_{i(X-i)} = \frac{\rho_{iX}\sigma_X - \sigma_i}{\sqrt{\sigma_i^2 + \sigma_X^2 - 2\rho_{iX}\sigma_i\sigma_X}}$$

Donde $\rho_{i(X-i)}$ es la correlación corregida entre el ítem y la escala,
 ρ_{iX} , es la correlación sin corregir entre el ítem y la escala,
 σ_X , es la desviación estándar de la escala
 σ_i , es la desviación estándar del ítem.

5.3 Fiabilidad

En todo proceso de medida por el cual se asignan valores numéricos a ciertas propiedades de los objetos interfieren los errores de medida. Los errores de medida afectan a todas las medidas, sean físicas, médicas, químicas o las medidas utilizadas en el marco de las ciencias sociales. Su influencia sobre el resultado final es diferente en función del grado de desarrollo métrico de la ciencia en cuestión. No serían comparables los errores de medida que afectan a la medición de una magnitud física con los errores de medida que acompaña toda medición psicológica. Una prueba del grado de fiabilidad de la primera respecto a la segunda podría obtenerse analizando sucesivas mediciones de la misma distancia y evaluando las discrepancias entre ellas. A menor discrepancia mayor fiabilidad. En el caso de la medicina por ejemplo, piénsese en las continuas mediciones de tensión arterial; en este caso -y en ausencia de variables externas que la modifiquen, las discrepancias entre mediciones realizadas son habituales; y la discrepancia aumenta si los sanitarios que efectúan la medición

Conceptos básicos y aplicaciones prácticas con Rcommander

son diferentes y utilizan distintos tensiómetros. En el contexto psicológico podría hipotetizarse las repetidas administraciones de un test y registrar sus distintos valores para el mismo sujeto.

Asumir la existencia de errores de medida es fundamental, y estimar su efecto sobre las puntuaciones es uno de los objetivos que persiguen los modelos psicométricos. Dentro de la teoría clásica de tests el concepto que aborda la estimación del efecto de los errores de medida es la fiabilidad.

La fiabilidad se refiere a la precisión de las puntuaciones, y su objetivo es estimar hasta qué punto los errores interfieren en el proceso de medida. Estos, son por definición aleatorios, no controlados y de su estimación se deriva el grado de fiabilidad de las puntuaciones del test.

5.3.1 Modelo lineal clásico

La teoría clásica de tests se construye en torno al concepto de fiabilidad. Los orígenes del modelo se sitúan en los trabajos de Spearman (1904, 1907, 1913), que Gulliksen sistematiza (1950) y se reformulan en la obra de Lord y Novick (1968). El objetivo del modelo es estimar la influencia de los errores de medida con el fin de obtener información sobre el grado de certidumbre (fiabilidad) de los datos obtenidos por el test.

MODELO	$X = V + E$
SUPUESTOS	1. $V = E[X]$ 2. $\rho_{V,E} = 0$ 3. $\rho_{e_j,e_k} = 0$
TEST PARALELOS:	Dos tests, j y k , son paralelos, si sus varianzas de error $[\sigma_{e_j}^2 = \sigma_{e_k}^2]$ y las puntuaciones verdaderas de un sujeto en ambas formas son equivalentes $[V_j=V_k]$.

Sin embargo con la definición del modelo es imposible estimar el efecto del error de medida sobre las puntuaciones observadas. Por ello a la formulación básica del modelo es necesario añadir ciertos supuestos que permitan llevar a cabo la estimación deseada.

La puntuación obtenida en un test (X) es una variable compuesta por dos componentes, la puntuación verdadera (V) y la puntuación de error (E). El error aleatorio puede incrementar o reducir la puntuación verdadera, y para estimarla es necesario considerar los siguientes supuestos:

Supuesto 1

La puntuación verdadera de un sujeto es la esperanza matemática de su puntuación empírica. Es decir, si se obtuvieran repetidas veces las puntuaciones de un sujeto en un mismo test, la puntuación verdadera se obtendría como la media aritmética de todas las puntuaciones obtenidas en las distintas aplicaciones del test.

Supuesto 2

No existe correlación entre la puntuación verdadera y la puntuación de error. Los errores por definición son errores aleatorios y por tanto no determinan la puntuación verdadera.

Supuesto 3

La correlación entre los errores es 0.

Definición de formas paralelas

Dos test son paralelos si las puntuaciones verdaderas son equivalentes y si ambas formas tiene varianzas de error equivalentes.

De esos supuestos se derivan los siguientes teoremas:

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

1.- El error de medida es la diferencia entre la puntuación empírica y la puntuación verdadera.

$$e = X - V$$

2.- La esperanza matemática del error de medida es 0. Es decir los errores no son sesgados.

$$E(e) = 0$$

3.- La media aritmética de la puntuación empírica es igual a la media aritmética de la puntuación verdadera.

$$\mu_X = \mu_V$$

4.- La covarianza entre la puntuación verdadera y el error de medida es 0.

$$Cov(V, e) = 0$$

5.- La covarianza entre la puntuación empírica y la puntuación verdadera es la varianza de la puntuación verdadera.

$$Cov(X, V) = \sigma_V^2$$

6.- La covarianza entre las puntuaciones empíricas de dos formas paralelas es igual a la covarianza entre las puntuaciones verdaderas de dos formas paralelas.

$$Cov(X_j, X_k) = Cov(V_j, V_k) = \sigma_V^2$$

7.- La varianza de la puntuación empírica es igual a la suma de la varianza de la puntuación verdadera y de la varianza de error.

$$\sigma_X^2 = \sigma_V^2 + \sigma_E^2$$

8.- La correlación entre la puntuación empírica y el error de medida es igual a la razón entre el error estándar de medida y la desviación estándar de la puntuación empírica.

$$\rho_{Xe} = \frac{\sigma_e}{\sigma_X}$$

9.- En K formas paralelas las medias aritméticas, las varianzas y las correlaciones entre ellas son equivalentes.

5.3.2 Coeficiente de fiabilidad

Según el modelo clásico de tests el coeficiente de fiabilidad ($\rho_{XX'}$) se define como la correlación producto momento de Pearson entre las puntuaciones empíricas (X y X') obtenidas en dos formas paralelas. Si la influencia de los errores de medida en las formas paralelas fuese inexistente, las puntuaciones empíricas obtenidas serían equivalentes y por tanto la correlación entre ellas sería 1. A medida que el coeficiente de correlación se aleja de 1 se incrementa el efecto del error de medida y por tanto se pierde estabilidad en la misma.

El desarrollo de esa definición (Elosua, 2005) deriva en el significado formal del coeficiente de fiabilidad:

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2}$$

El coeficiente de fiabilidad es la razón entre dos varianzas, la varianza de la puntuación verdadera y la varianza de la puntuación empírica. Tratándose de varianzas el valor del coeficiente de fiabilidad siempre será positivo.

$$0 \leq \rho_{XX'} \leq 1$$

Del mismo modo puede expresarse el coeficiente de fiabilidad como,

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_e^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{\sigma_X^2}$$

En el caso hipotético de ausencia de error de medida las varianzas empírica y verdadera coincidirían alcanzando el coeficiente de fiabilidad su máximo valor ($\rho_{XX'}=1$).

5.3.2.a Índice de fiabilidad

La correlación entre la puntuación verdadera y la puntuación empírica en dos formas paralelas da lugar al índice de fiabilidad del test (ρ_{XV}), que no es más que la raíz cuadrada del coeficiente de fiabilidad.

$$\rho_{XV} = \sqrt{\rho_{XX'}} = \frac{\sigma_V}{\sigma_X}$$

5.3.2.b Error típico de medida

La desviación típica de los errores de medida se conoce como *error típico de medida*.

$$\sigma_e = \sigma_X \sqrt{1 - \rho_{XX'}}$$

El error típico de medida podría interpretarse como la desviación estándar de las puntuaciones empíricas obtenidas por un sujeto en aplicaciones reiteradas de un test, es decir, su variación. A medida que aumenta la distribución de las puntuaciones empíricas disminuirá el coeficiente de fiabilidad del test. El error típico de medida es conocido como *fiabilidad absoluta* y el coeficiente de fiabilidad como *coeficiente de fiabilidad relativo*.

5.3.3 Procedimientos empíricos para la estimación del coeficiente de fiabilidad

La estimación del coeficiente de fiabilidad tal y como ha sido formulado, razón entre varianzas, no es posible dado que la única información de que se dispone tras la aplicación de un test es la puntuación empírica. Por ello es necesaria la aplicación de diseños que permitan su estimación. Entre los

Conceptos básicos y aplicaciones prácticas con Rcommander

utilizados en el contexto de la teoría clásica de test destacan tres; formas paralelas, test-retest y método de las mitades. Cada uno de ellos tiene ventajas e inconvenientes que es preciso conocer para una correcta interpretación del coeficiente de fiabilidad.

Formas paralelas. Es el diseño directamente derivado del modelo. La estimación del coeficiente de fiabilidad descansa en la construcción de dos formas paralelas de un mismo test. Una vez aplicadas a la misma muestra se estimará la correlación entre las puntuaciones obtenidas. El coeficiente de fiabilidad así estimado indica el grado de equivalencia entre las formas paralelas.

Test-retest. Este diseño no exige la construcción de formas paralelas. Asumiendo que un test es paralelo a sí mismo, la obtención del coeficiente de fiabilidad exige la administración en dos ocasiones diferentes del mismo test a la misma muestra. Una vez obtenidos los datos, la fiabilidad del test se obtendría por medio de una correlación entre puntuaciones empíricas. Este coeficiente de fiabilidad sería interpretado en el sentido de estabilidad temporal de las puntuaciones. La validez de este diseño está amenazada por los efectos del factor tiempo; es decir por el tiempo entre las dos administraciones del test. Si es demasiado reducido, el “recuerdo” podría producir un incremento de las puntuaciones, y si el tiempo es demasiado prolongado las puntuaciones podrían verse afectadas por el cambio en el constructo a medir.

Método de las mitades. Es el diseño más simple. No son necesarias formas paralelas ni aplicaciones dobles del mismo test. El coeficiente de fiabilidad se estima a partir de los datos obtenidos en una única administración del test. Para ello se asume que el test está compuesto por dos tests (medio test) que generan dos grupos de puntuaciones. El coeficiente de fiabilidad se estima por medio de un ajuste aplicado a la correlación entre ambas mitades (Spearman-Brown). El coeficiente de

Conceptos básicos y aplicaciones prácticas con Rcommander

fiabilidad así obtenido sería un indicador de la consistencia interna del test. La amenaza a la validez de este diseño está relacionada con la definición de dos mitades paralelas. El procedimiento habitual para ello consiste en dividir el test en ítems pares e ítems impares con el fin de garantizar que la dificultad de las dos mitades sea equivalente.

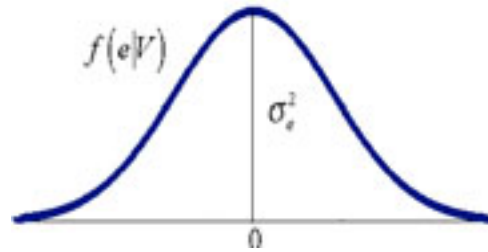
5.3.4 Estimación de la puntuación verdadera

La puntuación verdadera de un sujeto puede estimarse a partir del coeficiente de fiabilidad del test. Básicamente existen dos procedimientos para su estimación; el primero se apoya en el concepto de error típico de medida y el segundo aplica el modelo de regresión lineal simple definiendo la variable dependiente como la puntuación verdadera y la variable predictora como la puntuación observada.

5.3.4.a Distribución de los errores

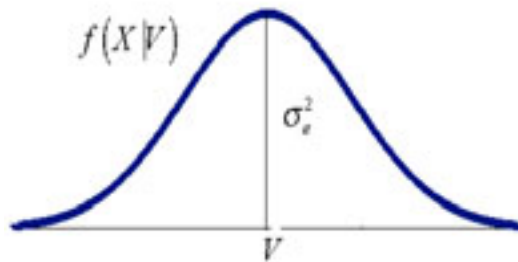
Suponiendo que la administración repetitiva de un cuestionario a un sujeto genera una distribución de valores empíricos X_i en torno a un único valor verdadero (V_i), es posible considerar cada una de las puntuaciones empíricas como una variable aleatoria extraída de esa distribución que se supone normal. En estas condiciones cada X_i es un estimador insesgado de la media aritmética de la distribución (V_i). Por lo tanto, una vez prefijado un nivel de confianza es posible estimar un intervalo de probabilidad para V_i a partir de la desviación de las puntuaciones empíricas, es decir, a partir del error típico de medida.

$$f(e|V) \sim N(0, \sigma_e^2)$$



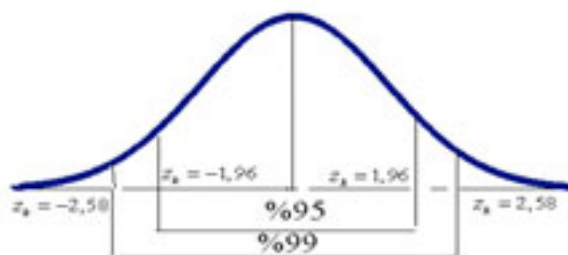
5.1 Figura. Distribución de los errores de medida

$$f(X|V) \sim N(V, \sigma_e^2)$$



5.2 Figura. Distribución de la puntuación empírica

1.- *Determinar el nivel de confianza* $(1-\alpha)$. Es habitual fijarlo en 95% o 99% ($\alpha=0,05$ y $\alpha=0,01$). Dada la simetría de la distribución normal los valores Z_k absolutos correspondientes a $\alpha/2$ y $(1-\alpha/2)$ son los mismos. Las áreas están limitadas por los valores críticos $z_c=\pm 1,96$ y $z_c=\pm 2,58$.



5.3 Figura. Distribución normal. Valores críticos

Conceptos básicos y aplicaciones prácticas con Rcommander

2.- *Estimación del error máximo.* Para ello es suficiente multiplicar el valor crítico ($Z_{\alpha/2}$) por el error típico de medida.

$$E_{Max} = |Z_K| s_e$$

3.- *Límites del intervalo para V.* Los límites del intervalo se obtendrán añadiendo y sustrayendo el error máximo a la puntuación empírica correspondiente.

$$L_{Sup} = X_i + E_{Max}$$

$$L_{inf} = X_i - E_{Max}$$

Donde L_{sup} y L_{inf} son los límites superior e inferior del intervalo para la puntuación verdadera.

En la estimación de la puntuación verdadera se ha utilizado el error típico de medida; es un concepto aplicable al grupo y que asume igualdad de varianzas para todas las puntuaciones empíricas; es decir, homocedasticidad.

La igualdad de la distribución de los errores de medida en todo el continuo de puntuaciones empíricas es tal vez una de las mayores debilidades de la teoría clásica de test; en las situaciones prácticas es difícil asumir ese supuesto. Ante esta violación del supuesto de homocedasticidad es posible buscar alternativas en modelos derivados de la teoría de respuesta al ítem, o en su defecto sería posible establecer intervalos de puntuaciones empíricas y estimar para cada una de ellas sus correspondientes errores típicos de medida.

5.3.4.b Modelo de regresión simple

Las variables incluidas en un modelo de regresión simple X (predictor) e Y (criterio) serían asimilables a la puntuación empírica (X) y a la puntuación verdadera (V); la primera sería la variable predictora y la segunda sería la

Conceptos básicos y aplicaciones prácticas con Rcommander

variable criterio. En este contexto, el problema se reduce a la estimación por medio de la recta de regresión del valor de V correspondiente a una puntuación empírica. La estimación en tanto en cuanto modelo de regresión asume la igualdad de las distribuciones condicionadas de V sobre X (homocedasticidad). La recta de regresión es aquella que une las medias aritméticas de estas distribuciones condicionadas.

El modelo de regresión lineal utiliza la siguiente formulación para la estimación puntual de la variable criterio.

$$Y' = \rho_{XY} (\sigma_Y / \sigma_X) (X_i - \bar{X}) + \bar{Y}$$

Donde X_i puntuación observada,
 ρ_{XY} , correlación entre las variables X e Y,
 σ_X , Desviación estándar de la variable X,
 σ_Y , desviación estándar de la variable Y,
 \bar{X} , media aritmética de la variable X,
 \bar{Y} , media aritmética de la variable Y.

En el contexto de la teoría clásica de test la formulación anterior adquiere esta forma:

$$V' = \rho_{XX'} (X_i - \bar{X}) + \bar{X}$$

Donde $\rho_{XX'}$ es el coeficiente de fiabilidad,
 X_i es la puntuación observada,
 \bar{X} , es la media aritmética de las puntuaciones empíricas.

La fórmula anterior es la ecuación de Kelley (Kelley, 1947). Se trata de una regresión a la media que está condicionada por la fiabilidad del test. Si el coeficiente de fiabilidad es 0 la puntuación verdadera estimable a un sujeto se equipara a la media aritmética de la distribución de puntuaciones empíricas. Existe una versión multivariable de la formula de Kelley conocida como estimación Bayes empírica que mejora la estimación de la fiabilidad de una escala parcial a partir de la información aportada por la relación entre la escala de interés y otras parciales incluidas en el test (Elosua, 2008).

Conceptos básicos y aplicaciones prácticas con Rcommander

En todo proceso de estimación existe una diferencia entre el valor pronosticado (V_i') y el valor verdadero (V_i). Esta diferencia es conocida como error de estimación. Su distribución, el error estándar de estimación ($\sigma_{V.X}$) se utiliza en la estimación de intervalos de confianza en torno a un valor pronosticado.

$$\sigma_{V.X} = \sigma_X \sqrt{1 - \rho_{XX'}} \sqrt{\rho_{XX'}} = \sigma_e \sqrt{\rho_{XX'}}$$

El error estándar de estimación se ajusta a una distribución normal, que es además homocedástica, es decir equivalente para todo el continuo de valores de la variable predictora (V'). Seguiríamos el procedimiento habitual de estimación:

1.- *Determinar el nivel de confianza* ($1-\alpha$). Es habitual fijarlo en 95% o 99% ($\alpha=0,05$ y $\alpha=0,01$). Dada la simetría de la distribución normal los valores Z_k absolutos correspondientes a $\alpha/2$ y $(1-\alpha/2)$ son los mismos. Las áreas están limitadas por los valores críticos $z_c=\pm 1,96$ y $z_c=\pm 2,58$.

2.- *Estimación del error máximo*. Para ello es suficiente multiplicar el valor crítico ($Z_{\alpha/2}$) por el error típico de medida.

$$E_{Max} = |Z_K| s_e$$

3.- *Límites del intervalo para V*. Los límites del intervalo se obtendrán añadiendo y sustrayendo el error máximo a la puntuación estimada correspondiente.

$$L_{Sup} = V_i' + E_{Max}$$

$$L_{inf} = V_i' - E_{Max}$$

Donde L_{sup} y L_{inf} son los límites superior e inferior del intervalo para la puntuación verdadera.

Conceptos básicos y aplicaciones prácticas con Rcommander

5.3.5 Coeficiente alfa

Entre todos los procedimientos disponibles para la estimación del coeficiente de fiabilidad basados en una única aplicación del test se ha impuesto el coeficiente alfa de Cronbach (1951). Es un indicador del grado de covariación de los ítems de una escala, y se interpreta como coeficiente de consistencia interna.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_X^2} \right) = \alpha = \frac{n}{n-1} \left(\frac{\sum_{j=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n Cov(i, j)}{\sigma_X^2} \right)$$

Donde n es el número de ítems del test,
 σ_i^2 , la varianza del ítem i ,
 σ_X^2 , la varianza del test
 $Cov(i, j)$, la covarianza entre i y j .

Un modo de conceptualizar el significado del coeficiente alfa podría ser la consideración de que los n ítems de los que consta un test son paralelos y cada uno de ellos presenta un coeficiente de fiabilidad ρ_{ii} . En esas circunstancias y considerando que el test ha incrementado K veces su longitud, se podría obtener el coeficiente de fiabilidad del nuevo test aplicando la formula “profética” de Spearman-Brown ($n=K$) que relaciona la longitud de un test y su fiabilidad.

$$\rho_{XX'} = \frac{n\rho_{ii'}}{1 - (n-1)\rho_{ii'}}$$

Donde n es el número de ítems del test
 $\rho_{ii'}$, es el coeficiente de fiabilidad de cada ítem.

5.3.5.a Casos particulares del coeficiente alfa

Conceptos básicos y aplicaciones prácticas con Rcommander

Alfa y los procedimientos de dos mitades

La literatura psicométrica tradicional muestra multitud de modos de estimar el coeficiente de fiabilidad de un test basándose en una sola aplicación. Todos estos métodos asimilables a la categoría genérica de métodos de las mitades pueden considerarse casos particulares del coeficiente alfa.

Alfa analiza la consistencia de un test compuesto por n ítems. Si el test estuviera compuesto por dos elementos, la fórmula de alfa sería equivalente a las formulas de Guttman (1945), Flanagan (1937) o Rulon (1939). Estos procedimientos, claramente en desuso, dividen el test en dos mitades y aplican las siguientes fórmulas para la estimación de la fiabilidad.

Rulon

$$\rho_{XX'} = 1 - \frac{\sigma_d^2}{\sigma_X^2}$$

Donde σ_d^2 es la diferencia entre las puntuaciones empíricas obtenidas en las dos mitades
 σ_X^2 , es la varianza del test completo.

La hipótesis básica de este procedimiento es la equiparación entre las diferencias entre las dos mitades equivalentes extraídas de un mismo test y el error de medida. En estas condiciones se asume la equivalencia entre la varianza de las diferencias y la varianza de error.

Guttman-Flanagan

$$\rho_{XX'} = 2 \left[1 - \frac{\sigma_{pa}^2 + \sigma_{im}^2}{\sigma_X^2} \right]$$

Donde σ_{pa}^2 es la varianza de las puntuaciones pares,
 σ_{im}^2 , es la varianza de las puntuaciones impares
 σ_X^2 , es la varianza del test completo

Esta formulación es algebraicamente equivalente a la formulación de Rulon.

Conceptos básicos y aplicaciones prácticas con Rcommander

Coefficiente alfa y Kuder-Richardson

Si los ítems son dicotómicos es posible aplicar la formula de Kuder-Richardson conocida como KR_{20} para la estimación del coeficiente de fiabilidad.

$$KR_{20} = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n p_i q_i}{\sigma_X^2} \right]$$

Donde n es el número de ítems del test,
 p_i , la proporción de sujetos que responden correctamente al ítem i
 q_i , es la proporción de sujetos que responden incorrectamente al ítem i
 σ_X^2 , es la varianza del test.

Esta fórmula es un caso especial de alfa, dado que la varianza de un ítem dicotómico puede expresarse como el producto de p_i y q_i .

Kuder-Richardson propusieron otro índice de consistencia interna aplicable cuando las medias aritméticas de los ítems fueren equivalentes, KR_{21} ,

$$KR_{21} = \frac{n}{n-1} \left[1 - \frac{\bar{X} - (\bar{X}^2/n)}{\sigma_X^2} \right]$$

Donde n es el número de ítems del test,
 \bar{X} , es la media aritmética del test
 σ_X^2 , es la varianza del test.

5.3.6 Índices derivados del modelo factorial

Considerando sólo la matriz de correlaciones de Pearson y el modelo del análisis factorial, McDonald (1985, pág. 217) describió como computar un coeficiente de consistencia interna desde el modelo del factor común. Para una escala compuesta por n ítems el coeficiente alfa podría estimarse como:

$$\alpha = \frac{n}{n-1} \left[\frac{n(\bar{\lambda})^2 - \overline{\lambda^2}}{n(\bar{\lambda})^2 + (\overline{u^2})} \right]$$

Donde n es el número de ítems

$\bar{\lambda}$ es la media aritmética de los pesos factoriales

$\overline{\lambda^2}$ es la media aritmética de los cuadrados de los n pesos factoriales

$\overline{u^2}$ es la media aritmética de las unicidades de las n variables

Armor (1974) basándose en el modelo de *Componentes Principales*, introdujo una estimación de la fiabilidad, *el coeficiente theta*, que fue desarrollado para escalas multidimensionales. *El coeficiente theta* para una solución factorial simple se estimaría de acuerdo con la siguiente expresión:

$$\theta = \frac{n}{n-1} \left[1 - \frac{1}{Eigen_1} \right]$$

Donde, n es el número de ítems de la escala

$Eigen_1$ es el valor propio mayor correspondiente a la matriz de correlaciones

5.3.6.a El modelo clásico de tests y el modelo del factor común

Dentro del modelo clásico de test la puntuación observada de un sujeto en un ítem (X_i) se descompone en puntuación verdadera (V_i) y componente aleatorio o error (E_i):

$$X_i = V_i + E_i$$

Esta descomposición de la puntuación observada en puntuación verdadera y error podría generalizarse al modelo unidimensional del factor común:

$$X_i = \lambda f_i + u_i$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Donde λ es al peso factorial

f_1 es el factor común, que podría asociarse con la puntuación verdadera
 u_i es la unicidad de la variable que podría asociarse al componente de error.

Dentro del modelo factorial la fiabilidad de la puntuación observada podría estimarse por medio de la razón entre la suma de los elementos de la matriz de varianzas/covarianzas entre puntuaciones verdaderas y la matriz de varianzas/covarianzas entre puntuaciones observadas. Novick y Lewis (1967) demostraron que el coeficiente alfa ofrece una estimación insesgada de la fiabilidad cuando los pesos del factor común sobre las variables son iguales, es decir, cuando las medidas son tau-equivalentes. La fórmula de la fiabilidad para una puntuación compuesta, conocida como coeficiente omega (McDonald, 1999) se formalizaría como:

$$\omega = \frac{\left(\sum_{i=1}^n \lambda_i \right)^2}{\left(\sum_{i=1}^n \lambda_i \right)^2 + \sum_{i=1}^n \text{var}(e)_{ii}}$$

Donde $\text{var}(e)_{ii}$ denota la varianza de error en el modelo factorial

Si en la estimación de los valores propios, en lugar de la matrices de correlaciones producto momento de Pearson, se utilizara una matriz de correlaciones policóricas (o tetracóricas) sería posible obtener coeficientes de fiabilidad para escalas de respuesta ordinales que ofrecerían estimaciones no-sesgadas del valor del coeficiente de fiabilidad (Elosua y Zumbo, 2008).

5.3.6.b Interpretación de alfa

Alfa es un indicador de consistencia interna que ha sido sobre-utilizado en la literatura psicométrica aplicada. El debate sobre su uso está abierto en la investigación psicométrica actual (Elosua y Zumbo, 2008). Una correcta

Conceptos básicos y aplicaciones prácticas con Rcommander

interpretación del significado de alfa habría de tener en cuenta los siguientes puntos:

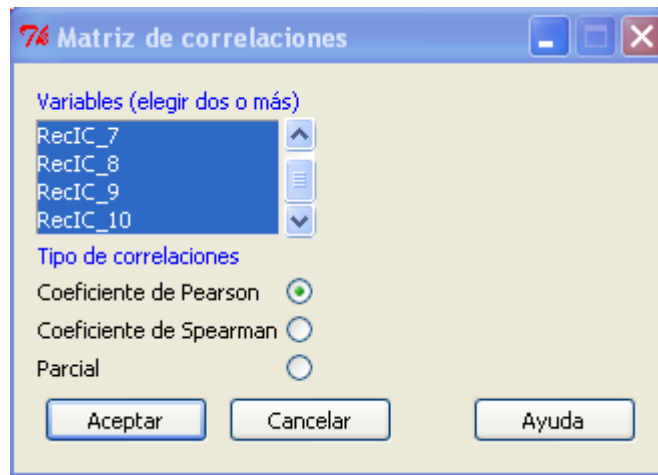
- La estimación de alfa no aporta información alguna sobre la estabilidad de las medidas obtenidas por un test. Alfa es un indicador de consistencia interna, no de estabilidad temporal.
- A partir de un test compuesto por n ítems es posible definir test-mitades utilizando multitud de criterios diferentes y a cada uno de ellos se le asociaría un coeficiente de fiabilidad. Alfa podría interpretarse como la esperanza de los coeficientes estimables (Cronbach, 1951).
- Alfa no es un indicador de unidimensionalidad. Aunque el valor de alfa viene determinado por el grado de covariación entre ítems su interpretación no es equiparable al concepto de unidimensionalidad. La presencia de varios factores correlacionados puede ser causa de covariación (Hattie, 1984).

5.4 Rcommander y la estimación de la fiabilidad

5.4.1 Matriz de correlación

La obtención de la matriz de correlaciones entre ítems por medio de Rcommander se efectúa con la opción: Estadísticos > Resúmenes > Matriz de correlaciones de la barra de menús. La ventana de diálogo que se abre permite seleccionar aquellas variables, ítems, cuyas correlaciones definirán la matriz de correlación. En este caso se seleccionan los ítems recodificados de la escala insatisfacción corporal.

Conceptos básicos y aplicaciones prácticas con Rcommander



5.4 Figura. Matriz de correlaciones

La función que utiliza Rcommander para obtener la matriz de correlaciones es `cor` :

```
cor(Edi.data[,c("RecIC_1", "RecIC_2", "RecIC_3", "RecIC_4", "RecIC_5", "RecIC_6", "RecIC_7", "RecIC_8", "RecIC_9", "RecIC_10")], use="complete.obs")
```

La ventana de salidas ofrecerá el siguiente resultado:

	RecIC_1	RecIC_2	RecIC_3	RecIC_4	RecIC_5	RecIC_6	RecIC_7	RecIC_8	RecIC_9
RecIC_10									
RecIC_1	1.00	0.61	0.73	0.63	0.40	0.50	0.18	0.49	0.59
0.50									
RecIC_2	0.61	1.00	0.52	0.61	0.47	0.62	0.19	0.70	0.70
0.56									
RecIC_3	0.73	0.52	1.00	0.69	0.48	0.47	0.13	0.56	0.48
0.57									
RecIC_4	0.63	0.61	0.69	1.00	0.60	0.58	0.16	0.65	0.61
0.67									
RecIC_5	0.40	0.47	0.48	0.60	1.00	0.48	0.08	0.60	0.53
0.62									
RecIC_6	0.50	0.62	0.47	0.58	0.48	1.00	0.18	0.55	0.67
0.66									
RecIC_7	0.18	0.19	0.13	0.16	0.08	0.18	1.00	0.11	0.19
0.12									
RecIC_8	0.49	0.70	0.56	0.65	0.60	0.55	0.11	1.00	0.58
0.73									
RecIC_9	0.59	0.70	0.48	0.61	0.53	0.67	0.19	0.58	1.00
0.57									

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

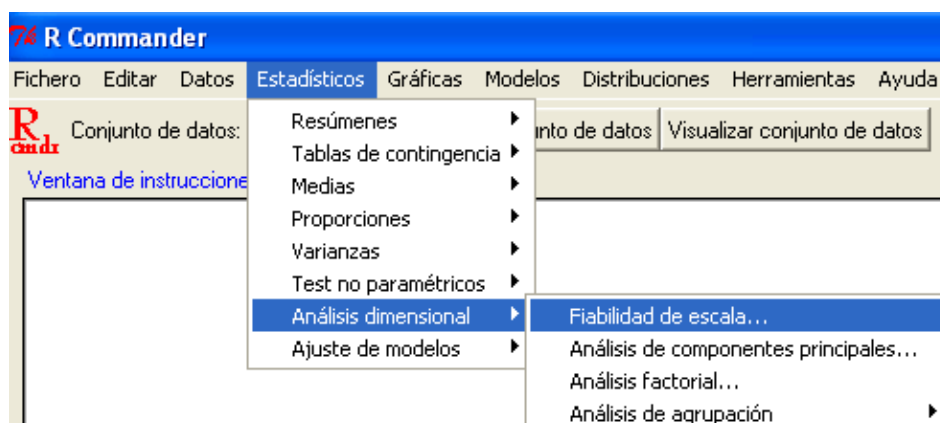
RecIC_10	0.50	0.56	0.57	0.67	0.62	0.66	0.12	0.73	0.57
1.00									

Puede apreciarse en la tabla que los elementos de la diagonal principal de la matriz, son 1 y que la matriz es simétrica. Todos los valores estimados son positivos. La correlación más alta corresponde al par formado por el ítem 3 y el ítem 1 ($r_{13}=0,73$), y el coeficiente de correlación más bajo se ha obtenido en el par compuesto por los ítems 7 y 8 ($r_{78}=0,11$). Una correcta interpretación de esos valores va siempre más allá de la simple interpretación cuantitativa y ha de considerar siempre el contenido de los ítems.

- 1.- *Creo que mi estómago es demasiado grande*
- 3.- *Creo que mi estómago tiene el tamaño adecuado*
- 7.- *Me siento hinchado después de una comida normal*
- 8.- *Creo que el tamaño de mis muslos es adecuado*

5.4.2 Índice de discriminación y coeficiente alfa

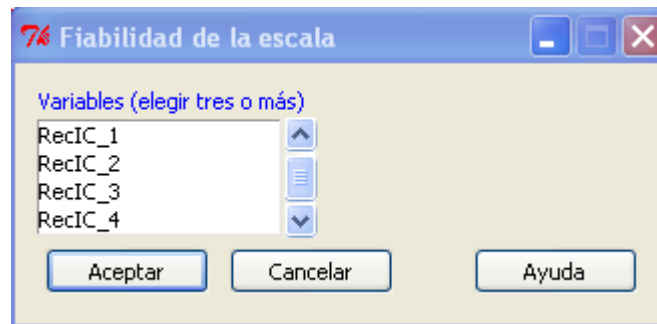
Para obtener los indicadores de consistencia interna, índice de discriminación y alfa, la opción viene dada por Estadísticos > Análisis Dimensional > Fiabilidad de escala....



5.5 Figura. Fiabilidad de la escala.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

Esa opción abre un cuadro de dialogo en el que se seleccionarán los ítems a analizar.



5.6 Figura. Fiabilidad de la escala. Variables

Rcommander incorpora una función específica para la estimación de la fiabilidad según la formulación de Cronbach, *reliability*.

```
reliability(cov(Edi.data[,c("RecIC_1", "RecIC_2", "RecIC_3", "RecIC_4",  
"RecIC_5", "RecIC_6", "RecIC_7", "RecIC_8", "RecIC_9", "RecIC_10")],  
use="complete.obs"))
```

La salida generada por Rcommander se divide en dos partes. En la primera puede leerse el valor del coeficiente alfa de Cronbach (Alfa *reliability*); en nuestro caso su valor es 0,9109. Tras este valor, aparece la estimación del coeficiente estandarizado que se obtendría utilizando las variables estandarizadas (Standardized alfa). Entre estos dos coeficientes es habitual utilizar el primero.

La tabla de valores que genera Rcommander incluye información relevante para el análisis de ítems porque permite analizar el peso que tiene cada uno de los ítems sobre el coeficiente de consistencia interna e incluye los índices de discriminación de cada uno de los mismos. En la primera columna de la tabla aparecen los nombres de los ítems analizados. La segunda columna titulada Alfa, ofrece los valores de los coeficientes alfa estimados para la escala de la que se ha eliminado el ítem de la fila. Por ejemplo, el primer valor

Conceptos básicos y aplicaciones prácticas con Rcommander

0,9007 es el coeficiente alfa de la escala compuesta por 9 ítems, los ítems 2,3,4,5,6,7,8,9,y 10. La tercera columna bajo el título Std. Alfa estima los valores estandarizados de alfa para las escalas mutiladas, es decir, las escalas a las que se ha eliminado el ítem de la fila. La última columna ofrece el valor de la correlación entre el ítem de la fila y la escala mutilada, es decir, ofrece información sobre el índice de discriminación corregido ($r(\text{ítem}, \text{total})$). Todos los valores de la columna excepto uno, superan el valor establecido como punto de corte 0,30. El valor del índice de discriminación para el ítem 7 es 0,1881. Este ítem estaba implicado en la correlación más baja estimada entre todos los pares de ítems.

Puede apreciarse la relación inversa que existe entre la columna que ofrece información sobre el valor alfa de la escala “mutilada” y el índice de discriminación. Cuando el descenso en el valor de alfa tras la eliminación de un ítem es brusco, el índice de discriminación de ese ítem es elevado.

```
Alfa reliability = 0.9109
Standardized alfa = 0.9077

Reliability deleting each item in turn:
      Alfa Std. Alfa r(item, total)
RecIC_1 0.9007    0.8967    0.6969
RecIC_2 0.8970    0.8931    0.7528
RecIC_3 0.9006    0.8967    0.6975
RecIC_4 0.8948    0.8908    0.7930
RecIC_5 0.9043    0.9005    0.6386
RecIC_6 0.9001    0.8959    0.7079
RecIC_7 0.9258    0.9261    0.1881
RecIC_8 0.8968    0.8932    0.7566
RecIC_9 0.8982    0.8938    0.7413
RecIC_10 0.8966    0.8929    0.7600
```


Conceptos básicos y aplicaciones prácticas con Rcommander

5.4.3 Coeficiente Theta

Para estimar el coeficiente theta es necesario teclear en la ventana de edición de Rcommander la siguiente función. Una vez escrita la función es suficiente indicar el número de ítems de la escala (n) y la matriz de correlaciones de los ítems (kor), para que R nos devuelva el valor del coeficiente theta.

```
theta <- function (n,kor) {

  lambda <- eigen(kor, only.values=TRUE)$values[1]
  theta.indice <- (n/(n-1))*(1-(1/lambda))
  print(theta.indice)}
```

La función exige que la matriz de correlaciones entre ítems haya sido asignada a un objeto de R. Dicha matriz se obtiene con la función cor. Por ejemplo:

```
kor <- cor(Edi.data[,c("RecIC_1", "RecIC_2", "RecIC_3", "RecIC_4",
"RecIC_5", "RecIC_6", "RecIC_7", "RecIC_8", "RecIC_9",
"RecIC_10")], use="complete.obs")
```

Tras la regla de asignación es suficiente ejecutar la función definida por nosotros como theta. Para ello basta teclear theta (10, kor). En nuestro caso el valor estimado para el coeficiente theta es 0,9165.

5.4.4 Error estándar de medida y estimación de la puntuación verdadera

Rcommander a través de la función reliability ofrece una estimación del coeficiente alfa; para obtener una estimación del error estándar de medida es necesario escribir en la ventana de instrucciones su fórmula

$$\sigma_e = \sigma_x \sqrt{1 - \rho_{xx'}}$$

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

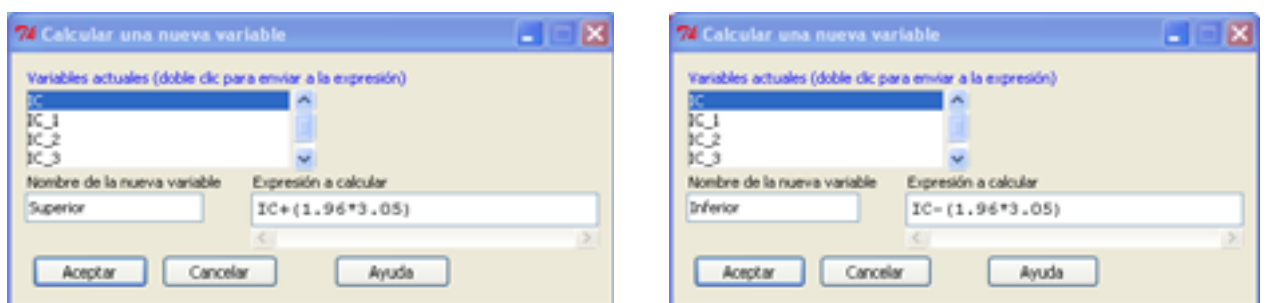
```
des <- sd(Edi.data$IC, na.rm=TRUE)

alfa <- 0.9109
Se <- des * (sqrt(1-alfa))
print (Se)
```

La función `sd` estima la desviación típica de la variable insatisfacción corporal que se ha obtenido previamente por medio de la adición de los ítems que la definen. Una vez escritas esas simples instrucciones en la ventana de comandos de Rcommander es suficiente marcar esas tres líneas con el ratón y presionar la opción `ejecutar` para que en la ventana de resultados aparezca el valor del error estándar de medida para la escala insatisfacción corporal, 3,05244.

Conocido el valor del error estándar de medida es posible estimar con un grado de confiabilidad determinado los intervalos de confianza para la puntuación verdadera a partir de cualquier puntuación observada. Suponiendo que el nivel de confianza es del 95% ($z = \pm 1,96$), podemos estimar los límites inferior y superior de los intervalos generando nuevas variables a través de la opción `Datos` de la barra de menús de Rcommander.

`Datos > Modificar variables del conjunto de datos activo > Calcular nueva variable` es la opción requerida:



5.7 Figura. Estimación de las puntuaciones verdaderas

Si se hubiera utilizado directamente la sintaxis,

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

```
Edi.data$Superior <- with (Edi.data, IC+1.96*Se)
Edi.data$Inferior <- with (Edi.data, IC-1.96*Se)
```

Este procedimiento generará dos nuevas columnas en el conjunto de datos activo, y sus valores representan los límites superior e inferior del intervalo de probabilidad en que con una confianza del 95% se situará la puntuación verdadera de cada uno de los sujetos participantes en la investigación. Por ejemplo sería posible afirmar con una confianza del 95% que la puntuación verdadera de aquellas personas que obtienen una puntuación de 35 en la escala insatisfacción corporal estaría situada en el intervalo 29 - 41 (40.978).

Superior	Inferior
40.978	29.022
7.978	-3.978
19.978	8.022
28.978	17.022
12.978	1.022
23.978	12.022
24.978	13.022
41.978	30.022

5.8 Figura. Intervalos de puntuación verdadera.

6 Validez

La construcción y utilización de tests y cuestionarios se asientan sobre una base que siendo tan primaria como fundamental, a veces en la práctica diaria se obvia; la validez. El uso de un test, independientemente de su ámbito de aplicación u objetivo concreto, ha de estar respaldado por un cúmulo de experiencias que avalen la certeza de que las puntuaciones con él obtenidas permiten la interpretación inicialmente propuesta y por ende, no representan cualquier otro atributo o artefacto originado por el propio instrumento o proceso de medida.

La validez es el primer requisito que ha de cumplir todo instrumento de medida. Su fiabilidad, aún siendo esencial, nunca supondrá una garantía de uso suficiente. El concepto psicométrico de validez es el más importante y fundamental (Angoff, 1988).

Sin embargo, es un concepto huidizo y complejo, que no se apoya como la fiabilidad en un modelo matemático que la delimite, por lo que resulta complicado de aprehender y tratar. La dificultad de representación que entraña es pareja al problema de la validación de las teorías científicas con la que podría equipararse (Cronbach y Meehl, 1955).

Bajo su carácter evolutivo que a lo largo del desarrollo de la psicometría la ha envuelto de mil formas, acepciones o enfoques (convergente, discriminante, factorial, sustantiva, estructural, externa, de población, ecológica, temporal, de tarea; Messick, 1980), la validez se mantiene preeminente respecto a los demás conceptos psicométricos.

Su evolución permite diferenciar tres etapas:

Conceptos básicos y aplicaciones prácticas con Rcommander

- Una primera etapa *operacional* dominada por una visión pragmática en la construcción y uso de instrumentos de medida, en la que prima sobre las demás la validez relacionada con el criterio ("*Un test es válido para aquello con lo que correlaciona*" Guilford, 1946; pág. 429).
- Un segundo estadio *teórico*, en el que se reconoce la importancia de la teoría psicológica sobre el constructo medido, y se adopta una visión integradora en la que se impone el análisis de la validez de constructo, como concepto unificador que abarca aspectos de contenido y de relaciones con otras variables. ("*Toda validación es una*" Cronbach, 1980; pág. 99; "...*toda validación es validación de constructo*" Cronbach, 1984; pág. 126)).
- La fase actual, a la que podríamos denominar, de *adecuación contextual*, en la que se amplía la acepción anterior y se delimita con el concepto de *uso propuesto*. Su objetivo primordial sería el de dotar a los instrumentos de medición de avales científicos y éticos. ("*Una visión integradora de la validez...debe distinguir dos facetas interconectadas del concepto unitario de validez. Una faceta es la fuente de justificación..... La otra faceta es la función o resultado del test...*"Messick, 1989; pág. 20).

Los primeros trabajos sobre la validez, basados en una perspectiva pragmática, casi atórica, la definen como una simple correlación con alguna otra medida de la que el test es instrumento de medición, o aseveran que un test es válido para aquello con lo que correlaciona (Guilford, 1946).

Esta visión, plasmada en las *recomendaciones* de 1954 (APA; AERA; MCME, 1954), diferencia entre 4 tipos de validez: las consabidas, validez de contenido, validez predictiva, validez concurrente, y finalmente validez de constructo. Si bien es cierto que en la edición de los estándares de 1966 y 1974 (APA; AERA;MCME, 1966, 1974), se combinan la validez predictiva y

Conceptos básicos y aplicaciones prácticas con Rcommander

concurrente, en la validez referida al criterio, se mantiene la visión tripartita, trinitaria de la validez, que no se romperá oficialmente hasta la publicación de los estándares de 1985, gracias sobre todo a la contribución de autores como Angoff (1988), Cronbach (1971), Cronbach y Meehl (1955) o Messick (1980, 1981, 1988,1989).

Según los estándares de 1985, "La validez se refiere a la adecuación, significado y utilidad de las inferencias específicas hechas con las puntuaciones de los tests. La validación de un test es el proceso de acumular evidencia para apoyar tales inferencias....la validez es un proceso unitario" (APA; AERA; NCME; 1985, pág.9). En esta revisión, se reconoce el concepto integrador y único de la validez, ("...toda validación es validación de constructo." Cronbach, 1984, pág. 126). A partir de ahí, se recomienda tomar los tipos de validez reconocidos anteriormente como meras etiquetas, o formas de expresar la validez de constructo. Supone en definitiva, el reconocimiento de la validez como proceso único de recogida de evidencias a través de estrategias de investigación diferentes, relacionadas con el constructo, con el contenido o con el criterio.

La evolución del concepto avanza un paso más en la revisión de 1999, donde no se encuentra ya referencia ni explícita ni implícita a estas formas o facetas. Manteniendo la visión integradora, se amplía incorporando a su definición un aspecto hasta entonces olvidado, el uso. Ya no es suficiente la justificación teórica del significado de las puntuaciones obtenidas, es necesario delimitar los fundamentos teóricos en un contexto externo, con relación al *propósito o interpretación propuesta*.

"La validez se refiere al grado en que la evidencia y la teoría soportan las interpretaciones de las puntuaciones del test implicadas por los usos propuestos del test.... El proceso de validación implica acumular evidencia para proveer de una base científica las interpretaciones propuestas de las puntuaciones. Lo que se evalúa, son las interpretaciones de las puntuaciones del test requeridas para los usos propuestos, no el test en sí mismo." (APA; AERA; NCME; 1999, pág.9).

La interpretación propuesta aporta un punto de vista nuevo a la teoría de la validez que diferencia entre *garantía científica* y *garantía ética*. Bajo ella, descansa la aseveración de que interpretar un test es usarlo, y todos los usos incluyen una interpretación del test.

6.1 Sesgo

La importancia concedida a las implicaciones derivadas de una contextualización práctica o uso de un determinado instrumento de medida, ha permitido la irrupción y afianzamiento de un concepto tan complejo como el de la validez: el sesgo. Es un término con connotaciones políticas, sociales, estadísticas y psicométricas, que comienza a cobrar relevancia en la década de los 20 debido a la controversia surgida en Estados Unidos acerca de la parcialidad de los tests respecto a determinados grupos (Cole y Moss, 1989; Cronbach, 1975; Jensen, 1980).

Desde un punto de vista estrictamente psicométrico el sesgo queda definido como error sistemático (Camilli y Shepard, 1994; Osterlind, 1983) originado por deficiencias en el test o en el modo en que este es usado, que produce una distorsión en el significado de las puntuaciones y contamina su interpretación.

Es necesario minimizar las causas que puedan estar en el origen del sesgo, garantizando las propiedades técnicas del test, y explicitando las condiciones que maximicen su validez.

Validez y sesgo se convierten en las dos caras de una misma moneda. El sesgo siempre supondrá falta de validez, y la falta de validez puede ser el origen del sesgo. Para maximizar una y consecuentemente minimizar otra, el test habrá de incorporar una descripción detallada de cada uno de los ámbitos de uso que ofrecerá el marco conceptual básico para la recopilación de evidencias e

Conceptos básicos y aplicaciones prácticas con Rcommander

interpretación de puntuaciones. El objetivo sería el análisis de una posible *infrarrepresentación* del constructo o de la existencia de *varianza irrelevante* para el mismo.

La infrarrepresentación del constructo se refiere al grado en que el test no incluye aspectos relevantes del mismo. Esta carencia, puede corromper el significado de las puntuaciones porque los ítems no son una muestra representativa, no abordan algún proceso psicológico importante, o elicitan un tipo de conducta o componente que no está entre los propósitos del test.

La varianza irrelevante del constructo hace referencia al grado en que las puntuaciones del test se ven afectadas por procesos que son extraños a su propósito. Las puntuaciones pueden estar sistemáticamente influenciadas por componentes que no forman parte del constructo.

Con la inclusión de estos aspectos dentro del proceso de validación, se adopta implícitamente un punto de vista multidimensional sobre el origen del sesgo. Se acepta el hecho de que un instrumento de medida, o mejor su uso concreto puede ser origen de sesgo, si mide más de una dimensión en dos grupos y además las distribuciones multidimensionales de estos grupos difieren (Ackerman, 1992; Mellenbergh, 1989; Camilli y Shepard, 1994).

Las implicaciones más directas de estos principios suponen que el proceso de validación ha de incluir una atención cuidadosa a posibles distorsiones en el significado de las puntuaciones causadas por la influencia de variables ajenas a los objetivos propuestos. Si su origen puede situarse en una inadecuada representación del constructo o aspectos tales como el formato de la prueba, condiciones de administración, o nivel del lenguaje utilizado, el proceso continuo de recogida de evidencias para garantizar la validez debería de considerarlas objeto de análisis.

6.2 Fuentes de evidencia

La recopilación de evidencias para prestar una base científica a la interpretación de las puntuaciones en un uso concreto puede provenir de diversas fuentes. La importancia otorgada a cada una de ellas dependerá siempre de los objetivos del test, que determinarán en cada caso el tipo de evidencia más significativa. Los últimos estándares (AERA, APA y NCME, 1999) diferencian entre fuentes relacionadas con el contenido, el proceso de respuesta, la estructura interna, las relaciones con otras variables y las consecuencias del test, que en ningún modo suponen distintos tipos de validez sino aspectos complementarios.

6.2.1 Contenido

El análisis del contenido de la prueba aglutina dos tipos de estudios suplementarios. Unos encaminados a evaluar las relaciones entre el constructo y el contenido del test, y otros dirigidos a valorar los factores contextuales internos y externos que puedan añadir varianza no deseada.

El objetivo de los primeros es garantizar que la muestra de ítems que componen la prueba es además de relevante, representativa del constructo (Messick, 1981). Su dependencia respecto del uso concreto obliga a evaluar las distintas acepciones o significados que un mismo término puede poseer para diferentes grupos. Este punto es especialmente importante en la adaptación de tests (Elosua, Hambleton y Zenisky, 2006; Elosua y López, 1999, 2007). Una cuidadosa revisión desde el momento mismo de la construcción del instrumento por parte de expertos podrá evitar potenciales fuentes de dificultad irrelevante, o de un uso sesgado del lenguaje, que pueda mostrar un punto de vista estereotipado o incluso irrespetuoso sobre un determinado grupo cultural o étnico (Title, 1982). Se recomienda también, sobre todo en los tests de rendimiento, asegurar o garantizar que la experiencia curricular de los sujetos sea la misma (Elosua, López y Egaña, 2000; Linn y Harnisch, 1981).

Conceptos básicos y aplicaciones prácticas con Rcommander

El estudio de los factores contextuales cubre un abanico amplio de condiciones, que incluyen un análisis del formato de los ítems (Scheuneman, 1984, 1987) o el tipo de tareas exigidas, y la evaluación de la propia situación de test. Dentro de esta última se incluiría un estudio de las instrucciones para la administración y corrección de la prueba, de la interacción entre examinador-examinado, de la familiaridad con la situación, de las diferencias de motivación o ansiedad o del tipo de material utilizado (Hambleton, 1993; van de Vijver y Poortinga, 1991).

La evidencia basada en el contenido, aunque en su mayoría es cualitativa y viene sustentada por análisis lógicos, puede incluir, sobre todo en tests de rendimiento y referidos al criterio, estudios empíricos sobre la adecuación con que el contenido del test representa el dominio de contenido y sobre la relevancia del dominio de contenido respecto a las interpretaciones de las puntuaciones.

El análisis de contenido cubre básicamente tres aspectos: la definición del dominio, su representación y la relevancia del mismo. La primera está relacionada con las especificaciones del test y las dos restantes analizan el test mismo. La definición del dominio equivale a la definición operacional del dominio teórico, es decir con su explicitación. Tras el análisis teórico del dominio el objeto de medida quedaría definido por medio de aspectos, dimensiones o factores. Además del análisis de contenido, en pruebas de rendimiento sobre todo, es interesante valorar aspectos cognitivos relacionados con ellos. Con este fin han sido propuestas varias clasificaciones siendo la más utilizada la taxonomía de Bloom que diferencia seis aspectos cognitivos (Bloom, 1956; Bloom, Hastings y Madaus, 1971):

- *Conocimiento*: Recuerdo de la información aprendida durante la fase de instrucción.
- *Comprensión*. Modificación de un concepto, interpretación y extrapolación. .

Conceptos básicos y aplicaciones prácticas con Rcommander

- *Aplicación.* Resolución de problemas a partir de principios generales.
- *Análisis.* Descomposición de un problema en partes. Para ello se exige el análisis de varios elementos, de sus relaciones y de principios organizativos
- *Síntesis.* Combinación de elementos, utilizando principios generales o su adecuación a nuevos contextos.
- *Evaluación.* Ofrecer valoraciones utilizando criterios externos o internos.

Una vez finalizado el análisis se obtendría tablas similares a la que se muestra a continuación; en ellas se conectan las áreas de contenido y las áreas cognitivas que se quieren evaluar. La clasificación mostrada o cualquier otra derivada de los objetivos del tests permiten racionalizar el proceso de construcción y evaluación de ítems.

	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación
Operaciones						
Lógica						
Geometría						
Resolución de problemas						

6.1 Tabla. Definición del dominio

Representación del dominio: Consiste en analizar el contenido de los ítems en función de las especificaciones del test. Para ello se clasifica cada uno de los ítems en las dimensiones o factores definidos tras el estudio del constructo; además del contenido, es posible analizar las demandas cognitivas o de otro tipo relacionadas con cada uno de los ítems.

Es posible construir índices numéricos para cada uno de los ítems basándose por ejemplo en el grado de acuerdo entre jueces en la asignación de los ítems a las distintas áreas de dominio. Por ejemplo el índice de congruencia para un ítem evaluado por 10 jueces expertos entre los cuales el nivel de acuerdo es de 7 sería 0,7 (7/10). Esta sencilla proporción ofrece información válida que

Conceptos básicos y aplicaciones prácticas con Rcommander

permite ir construyendo un test con las garantías mínimas de validez imprescindibles.

Relevancia del dominio. Una correcta asignación de ítems a un área de dominio es importante, pero es imprescindible, además, analizar la relevancia de ese ítem respecto al dominio. Un procedimiento que permite obtener información al respecto exigiría la colaboración de un grupo de expertos en el dominio de contenido que valoren la relevancia de cada uno de los ítems. Para la recogida de información pueden crearse escalas como la que siguen, que permitirían estimar para cada ítem un índice de relevancia.

<i>Valora la relevancia del ítem en la evaluación de cada uno de las áreas de dominio definidas</i>									
1	2	3	4	5	6	7	8	9	10
<i>Nada relevante</i>								<i>Muy relevante</i>	

Es posible definir distintos índices de relevancia, sin embargo, todos tienen en común la participación de grupos de expertos; elemento indispensable y muchas veces obviado en la fase de construcción de un test en cualquier área de contenido.

Es importante en la evaluación de la validez de contenido, que las áreas de dominio definidas representen correctamente el constructo a medir evitando infrarrepresentaciones o varianzas relevantes.

En general pueden recomendarse estas etapas en la evaluación del contenido de test:

1.- *Selección del grupo de expertos.* Los expertos además de conocer en profundidad el dominio de contenido deben conocer las características de la población destino con el fin de que la adecuación entre el contenido y el grupo sea máxima. Para ello es recomendable que en el grupo de expertos participen personas con características geográficas, demográficas y étnicas representativas

Conceptos básicos y aplicaciones prácticas con Rcommander

de la población destino. No es suficiente una buena representación del dominio, ésta tiene que venir acompañada con un conjunto de especificaciones contextuales que adapten la representación a las circunstancias de administración del test y las características del grupo destino.

Además del análisis del contenido es necesario cuidar especialmente la calidad lingüística del test; No deben de descuidarse aspectos sintácticos, semánticos o incluso fonológicos que puedan generar sesgo en grupos específicos de respondientes (Elosua y López, 1999, 2007; Elosua, López y Egaña, 2000).

2.- *Formación de los expertos.* Es necesario formar al grupo de expertos en dos puntos que garanticen la corrección del proceso de estudio del contenido. Por un lado hay que instruirlos en los objetivos del test, y por otro hay que familiarizarlos con las tareas que se les van a solicitar.

3.- *Recogida de datos.* El diseño para la recogida de datos en la fase de análisis de contenido ha de ser sencillo y simple, de modo que no confunda a los expertos ni se confunda con ningún otro tipo de fuente de error.

4.- *Análisis de datos.* Una vez recogidos los datos se estimarán para cada ítem sus índices de congruencia y relevancia. Estos índices permitirán una evaluación dirigida a mejorar la representación del dominio, modificando o eliminando algunos de los ítems construidos.

En definitiva el estudio de la validez de contenido está estrechamente relacionado con la fase de construcción de los test, que aglutina tanto el estudio teórico del constructo del que partirá su definición operacional, como el contexto de aplicación y la población destino.

6.2.2 Proceso de respuesta.

Conceptos básicos y aplicaciones prácticas con Rcommander

La influencia que está ejerciendo la psicología cognitiva sobre la psicometría tradicional, está obligando a reanalizar la medición por medio de test para que fije su atención más que en la utilidad del constructo, en su representación (Prieto y Delgado, 1999; 2000; Snow y Lohman, 1993). En este contexto, el estudio de los procesos cognitivos implicados en la resolución de los ítems es un elemento de análisis en la búsqueda de instrumentos de medida que se ajusten a un marco que no sea estrictamente estadístico. Se convierten de este modo en un instrumento útil para la constatación de modelos cognitivos concretos (Embretson, 1985) que permiten además indagar posibles discrepancias cognitivas entre grupos.

La metodología descansa sobre el análisis de protocolos de respuesta, interrogatorios, y en general procedimientos que permitan indagar y descomponer la respuesta a un ítem en diferentes procesos.

Este tipo de estudio no recae únicamente sobre los sujetos. Es aconsejable analizar a los expertos encargados de evaluar o recoger los datos para garantizar una interpretación consistente entre todos ellos. De este modo, la validación puede recoger estudios empíricos de como los jueces recogen y evalúan datos, junto con análisis sobre la adecuación de esos procesos con la definición del constructo.

6.2.3 Estructura interna

El estudio de la evidencia centrada en el análisis de la estructura interna evalúa el grado en que las relaciones entre los ítems y los componentes del test conforman el constructo que se quiere medir y sobre el que se basarán las interpretaciones. Según los últimos estándares, se centra en el estudio de la dimensionalidad de la prueba, y del funcionamiento diferencial de los ítems.

Conceptos básicos y aplicaciones prácticas con Rcommander

6.2.3.a Dimensionalidad

El estudio de la dimensionalidad cubre habitualmente dos aspectos: 1) la verificación o refutación de la hipótesis de unidimensionalidad, y 2) la descripción, si procede, de la estructura multidimensional de un conjunto de datos. Estos dos objetivos implican la utilización de una metodología encaminada a determinar qué y cuántos factores o en su caso rasgos explican la matriz de correlaciones, o determinan la respuesta de un sujeto.

Generalmente, su objetivo es especificar y demostrar la existencia de una "*estructura simple*" (Thurstone, 1947), aquella, que resulta de la agrupación de ítems en núcleos dimensionalmente homogéneos entre ellos, y dimensionalmente distintos entre sí. De entre todos los procedimientos utilizables para ello, tal vez sea el modelo de análisis factorial el que ha gozado de mayor popularidad, convirtiéndose en el más utilizado.

El análisis factorial engloba un conjunto de técnicas de análisis multivariadas (McDonald, 1985) cuyo objetivo es resumir la información contenida en un conjunto de variables observadas por medio de un número reducido de variables hipotéticas, conocidas normalmente como factores. Basado en el modelo lineal de Spearman, pretende reproducir las matrices de covarianzas o correlaciones entre variables, a través de la siguiente ecuación

$$Z = Af + e$$

Dónde el vector de variables observadas Z se explica a través de un vector de factores f , ponderado por una matriz de pesos factoriales (A) que representa las correlaciones entre las variables observadas y latentes, más un término de error (e), que habitualmente en psicología se descompone, en varianza única y aleatoria.

Es un procedimiento que puede utilizarse como técnica de análisis exploratoria (Análisis Factorial Exploratorio), o confirmatoria, para contrastar la

Conceptos básicos y aplicaciones prácticas con Rcommander

hipotética estructura subyacente a un conjunto de datos (Análisis Factorial Confirmatorio).

En este último caso el modelo de análisis factorial se integra en el conjunto de modelos de estructuras de covarianza, que han de especificar tanto el modelo de medida, o relación entre las variables observadas e hipotetizadas, como el modelo de ecuaciones estructurales, que define las relaciones entre factores .

El análisis factorial se ha convertido en el procedimiento de análisis de la estructura interna por excelencia en psicología, debido entre otros factores a su implementación en paquetes estadísticos tan conocidos como el SPSS, SAS o SYSTAT. Al respecto, señala Ferrando (1993) " se utiliza excesivamente y por lo general se utiliza mal, sin una adecuada comprensión del método, con finalidades que no le corresponden y con datos para los que no resulta apropiado" (pág. 24).

La inadecuada aplicación de este método se deriva de la violación de los presupuestos del modelo lineal en que se apoya. Entre ellos, a) la métrica de las variables tanto observadas como latentes corresponde a la escala estandarizada, b) los factores comunes y los únicos son estadísticamente independientes, c) los factores únicos son estadísticamente independientes, d) las correlaciones entre las variables observadas y los factores son las saturaciones factoriales de la variable en el factor (Martínez Arias, 1995).

6.2.3.b Funcionamiento diferencial del ítem

La definición más general de funcionamiento diferencial del ítem podría ser la aportada por Mellenbergh (1989), según la cual dada una variable Z , y con respecto a otra variable G , el ítem i presenta funcionamiento diferencial (FDI), si y sólo si, se satisface la siguiente desigualdad para todos los valores g y z de las variables G y Z .

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

$$f(X|g,z) \neq f(X|z)$$

Donde X es el ítem estudiado; Z la variable medida, y G una variable aparentemente irrelevante al objeto de medida, habitualmente de naturaleza nominal (sexo, idioma, raza, edad...) y cuyos valores determinan la diferenciación entre los grupos de referencia y focal.

Esta formulación se traduce en que la probabilidad de respuesta correcta a un ítem no depende únicamente del nivel del sujeto en la variable medida, sino que está condicionada por la pertenencia a un determinado grupo social, cultural, lingüístico, instruccional... Esta dependencia del ítem conduce a una falta de equivalencia métrica que cuestiona su validez.

La importancia de garantizar la equidad en el proceso de medición, implícita en el concepto de validez, es el origen de la multitud de trabajos destinados tanto a la elaboración y estudio de técnicas diseñadas para la detección del FDI (Berk, 1982; Camilli y Shepard, 1994; Clauser y Mazor, 1998; Elosua, 1996; Fidalgo, 1996; Holland y Wainer, 1993; Martínez Arias, 1995; Millsap y Everson, 1993; Osterlind, 1983; Potenza y Dorans, 1995), como a la búsqueda de teorías explicativas que analizan sus causas (Elosua, Hambleton y Zenisky, 2006; Elosua y López, 2007; Mellenbergh, 1989).

Son muchos los métodos diseñados para detectarlo y muchos los estudios que comparan su efectividad bajo distintos supuestos. Básicamente, los procedimientos de detección del FDI pueden clasificarse en función del carácter de la variable condicionante. Según sea ésta observada o latente, nos encontramos con técnicas conocidas como *invarianza condicional observada* o *invarianza condicional latente* (Millsap y Everson, 1993). Todas ellas comparan la ejecución en un ítem por sujetos que perteneciendo a distintos grupos, presentan el mismo nivel de habilidad, de ahí su calificación de técnicas condicionales.

Conceptos básicos y aplicaciones prácticas con Rcommander

Estas técnicas definen Z como la puntuación total observada. En ellas se encuadran los procedimientos *chi-cuadrado* (Scheuneman, 1984), el estadístico *Mantel-Haenszel* (Holland y Thayer, 1988), *la estandarización* (Dorans y Kulick, 1986), *modelos log-lineales* y los derivados de *la regresión logística* (Swaminathan y Rogers, 1990).

En el segundo grupo de técnicas pueden incluirse los procedimientos derivados de los modelos de respuesta al ítem (MRI), que para muchos autores ofrecen el marco teórico más apropiado para la detección del FDI. En este contexto un ítem muestra funcionamiento diferencial cuando su función de respuesta depende de la pertenencia al grupo.

Las técnicas de detección de FDI derivadas de los MRI incluyen tres tipos de procedimientos. Aquellos que comparan en los grupos de estudio los parámetros de las curvas características del ítem (a, b, c) (Lord, 1980), los basados en el cálculo de la superficie que limitan las curvas características producidas por un ítem en dos poblaciones distintas (Linn y Harnisch, 1981), y por último los basados en la comparación de modelos.

El desarrollo de todos los procedimientos anteriores, parte de la consideración de la unidimensionalidad de la variable condicionante, sea esta observada o latente. Sin embargo, desde una perspectiva multidimensional (Ackerman, 1992; Shealy y Stout, 1993) surgen procedimientos como el SIBTEST (Shealy y Stout, 1993) o los modelos de estructuras de medias y covarianzas (MACS; Elosua, 2005) derivados del modelo del factor común.

A pesar de la estrecha relación entre los conceptos de sesgo y FDI, es importante reseñar el hecho de que no existe correspondencia biunívoca entre ambos. Aunque consideremos los índices de FDI arrojados por los distintos procedimientos definiciones operacionales del sesgo, no hay que olvidar que son procedimientos estadísticos que pueden producir falsas detecciones. Es siempre necesario complementar todo estudio empírico de detección de FDI (Elosua,

Conceptos básicos y aplicaciones prácticas con Rcommander

López y Torres, 2000) con procedimientos de juicio que evalúen y contextualicen en cada caso el contenido de cada uno de los ítems (Elosua, Hambleton y Zenisky, 2006; Scheuneman, 1987), para así poder concluir la presencia o no de sesgo, y poder evaluar los resultados en un marco sustantivo de estudio de la validez. En definitiva, es necesario buscar el origen del FDI dentro de teorías explicativas generales, que permitan su control desde el estadio inicial de construcción de tests.

6.2.4 Relaciones con otras variables

El estudio de las relaciones entre la medida obtenida por el test y otras variables externas, conocida como *el aspecto externo de la validez* por Loevinger (1957) y Messick (1989), o como *amplitud nomológica* por Embretson (1983, 1985), tal vez sea la fuente de información más utilizada en el proceso de validación. Su defensa como fuente de validez por al enfoque funcionalista (Prieto y Delgado, 1999) en la construcción de tests ha avalado su uso desde los primeros estándares de la APA.

Se centra en la búsqueda de evidencias que relacionen la puntuación con algún criterio que se espera prediga el test, o con otros tests que hipotéticamente midan el mismo constructo, constructos relacionados o constructos diferentes (AERA, APA y NMCE, 1999). Los resultados de estos análisis podrían evaluar el grado en que las relaciones hipotetizadas son consistentes con la interpretación propuesta.

Es un apartado general en el que se incluyen la evidencia convergente/discriminante, las relaciones test/criterio y los estudios de generalización de la validez.

6.2.4.a Evidencia convergente y discriminante.

Conceptos básicos y aplicaciones prácticas con Rcommander

Las relaciones entre métodos de medida diferentes aplicados a distintas variables se han estudiado tradicionalmente por la matriz multirasgo/multimétodo (Campbell y Fiske, 1959). Evalúa las correlaciones entre distintas variables medidas por distintos métodos, analizando así la influencia del instrumento de medición sobre la medida. La validez convergente (valores monorasgo-heterométrodo) se refiere al grado de relación entre distintos procedimientos que miden el mismo constructo, mientras que la validez discriminante (valores heterorasgo-monométodo) hace referencia a la evaluación de distintas variables medidas con el mismo método. Aunque en primera instancia se trate de un procedimiento heurístico sin validez estadística (Martínez Arias, 1995), se han propuesto distintos modelos para su estudio basados en el análisis factorial confirmatorio (Bagozzi, 1993, Browne, 1984; Byrne, 1989; Marsh y Bailey, 1991; Schmitt y Stults, 1986)

La matriz multirasgo-multimétodo puede complementarse con estudios experimentales, que incluyan diseños pre-post-tratamiento o comparación de grupos que supuestamente presentan valores diferentes en la variable medida, distinta instrucción, o incluso entrenamiento (Cronbach, 1975).

6.2.4.b Relaciones test-criterio

El estudio de las relaciones test-criterio adquiere una gran relevancia en contextos de utilidad donde es fundamental la precisión con que se efectúa una predicción. Su estudio incluye además del grado de precisión, un análisis de los factores que inciden directamente sobre ella, como las características propias del instrumento evaluado, la relevancia, fiabilidad y validez del criterio, y en general todos aquellos que afecten a la fiabilidad de las variables como el tamaño muestral o la restricción del rango.

Los diseños utilizados para la obtención de índices de este tipo dependen del tiempo transcurrido entre la recogida de datos en el test y en el criterio,

Conceptos básicos y aplicaciones prácticas con Rcommander

siendo habitualmente conocidos como el predictivo, concurrente o retrospectivo.

En función tanto del número de variables empleadas como de su carácter sería posible la utilización de diversas técnicas de análisis multivariado. Entre ellas, regresión y correlación simple (un test / un criterio), regresión múltiple, regresión logística, análisis discriminante (varios predictores / un sólo criterio) o la correlación canónica y el análisis de regresión multivariante para el caso de varias variables predictoras y varios criterios.

6.2.5 Generalización de la validez

La posibilidad de que los procesos de validación locales puedan extenderse a nuevas situaciones está constituida por los estudios de meta-análisis, que en este ámbito cumplen dos objetivos complementarios. Por un lado, unificar los resultados de aplicaciones particulares de un mismo test, y por otro, estimar la variabilidad de los resultados locales obtenidos debidos a artefactos estadísticos. En los estudios de validación Hunter y Schmidt (1991) diferencian entre 11 posibles artefactos que sería necesario neutralizar:

Error de medida en la variable independiente	Error de medida en la variable dependiente
Dicotomización de la variable independiente	Dicotomización de la variable dependiente
Variación del rango en la variable independiente	Variación del rango en la variable dependiente
Validez de constructo imperfecta en la variable independiente	Validez de constructo imperfecta en la variable dependiente
Errores de informe y transcripción	Varianza debida a factores extraños
Error muestral	

Conceptos básicos y aplicaciones prácticas con Rcommander

En definitiva se trata de calcular un promedio de validez, corrigiendo los efectos de cada uno de los factores mencionados.

6.2.6 Consecuencias

El estudio de las consecuencias de la utilización de los tests es una fuente de evidencia compleja que puede traspasar los límites de la validez como concepto enmarcado en el proceso de construcción de teorías científicas, para llegar a tocar aspectos relacionados con la política social. Esta distinción e interconexión vuelve a enfatizar la importancia de la adecuación contextual en el proceso de validación de un instrumento de medida, que cobrará una relevancia especial en las situaciones en las que puedan existir consecuencias diferenciales para grupos socioculturales diferentes (selección y promoción, educación especial...)

La integración del test con sus consecuencias, ha sido ilustrada por Messick (1989), que propone un marco de estudio de la validez en el que incluye un componente pragmático con el que intenta enfatizar:

- a.- la importancia de la relación entre la connotación teórica y las connotaciones prácticas atribuidas a las puntuaciones
- b.- la necesidad de valorar la relevancia y la utilidad de las puntuaciones en cada uno de los usos propuestos
- c.- la necesidad de conocer y en su caso controlar las consecuencias sociales del uso propuesto.

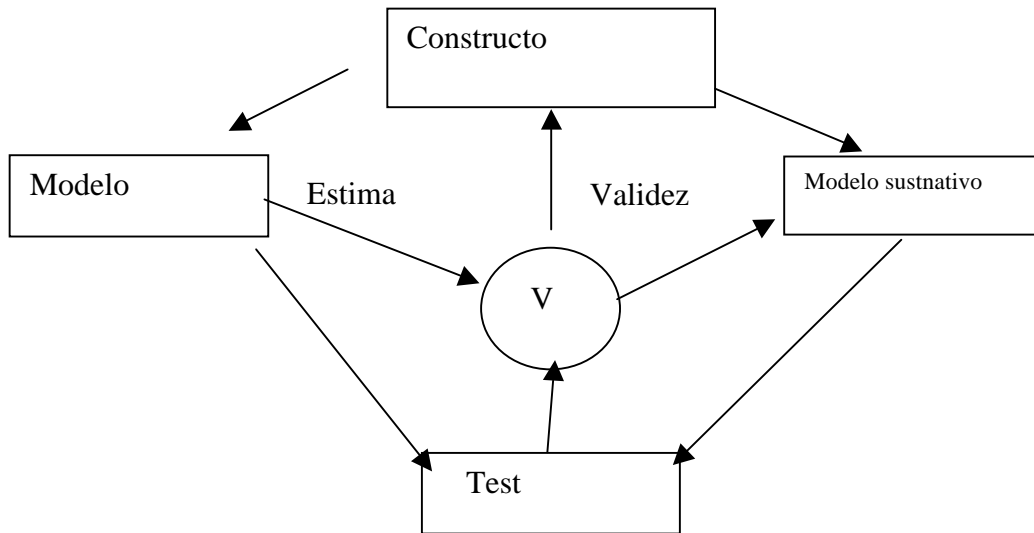
Con ello se busca un equilibrio entre el valor instrumental del test o su finalidad y los efectos derivados de su uso, que sólo se consigue haciendo al usuario cómplice y responsable del valor terminal del test.

Conceptos básicos y aplicaciones prácticas con Rcommander

En definitiva, el proceso de validación de un instrumento aglutina un conjunto de estudios encaminados a proveer a las puntuaciones del test de una interpretación teórica coherente en un contexto de uso bien delimitado. Es un análisis que ha de comenzar en el momento previo a la construcción del instrumento, y que ha de guiar y acompañar su desarrollo, asegurando un test técnicamente fiable, y sostenido por el suficiente cúmulo de evidencias que garanticen equidad tanto en la administración como en la puntuación e interpretación.

La inferencia psicológica derivada de las puntuaciones de los test, exige la conjunción entre los modelos formales utilizados en la estimación de los errores aleatorios que interfieren todo proceso de medida y el modelo sustantivo base, utilizado en la construcción del test. Sólo una correcta conexión entre modelo formal y modelo sustantivo garantizará la validez de las inferencias derivadas de las puntuaciones.

Todo test se construye sobre dos bases, una formal y una sustantiva. Los modelos psicométricos (la teoría clásica de tests, por ejemplo) se encargarán del análisis formal de las puntuaciones; los estudios de validez por su parte, tienen como objetivo el análisis sustantivo de las puntuaciones obtenidas. Todo ello, siempre orientado o dirigido por los objetivos y finalidades del test, lo cual viene delimitado por el contexto de aplicación y las características de la población destino.



6.1 figura. Componentes formal y sustantivo de un test

7 Estructura interna. Dimensionalidad

Los constructos teóricos que son objeto de investigación por las ciencias sociales no son directamente medibles y es necesaria la utilización de indicadores manifiestos para su medición. La relación entre el constructo teórico y sus indicadores manifiestos definirá la validez de la medida obtenida. Es fundamental la correcta definición de los indicadores manifiesto para que representen el constructo sin varianza irrelevante y sin infrarepresentación. Existen modelos formales cuyo cometido se centra en el análisis empírico de las relaciones entre variables observadas y variables latentes; entre ellos, el análisis factorial es el modelo más utilizado en la investigación psicométrica.

El análisis factorial engloba un conjunto de modelos matemático-estadísticos que analizan las relaciones de dependencia entre variables. Su objetivo es explicar la variabilidad contenida en n variables observadas por medio de m variables latentes, es decir, analizar la estructura interna o dimensionalidad de los datos. El análisis factorial se haya estrechamente unido al estudio de la validez interna de un test o cuestionario y en el ámbito psicométrico es la técnica más utilizada, aunque en muchas ocasiones su utilización no sea correcta.

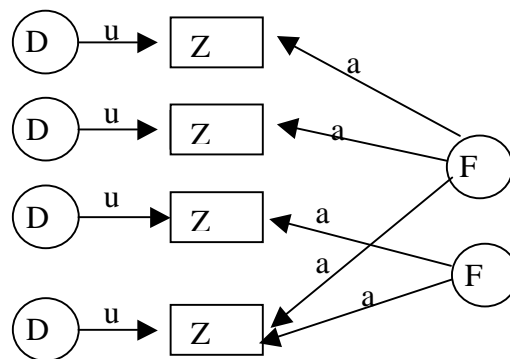
Los datos a analizar están contenidos en la matriz de covarianzas (correlaciones) entre variables; y el objetivo es explicar la variabilidad contenida en ellas asumiendo la existencia de factores o variables latentes comunes a todas ellas. Los objetivos del análisis factorial son la parsimonia y la simplicidad de la solución que garantizan la significatividad e interpretabilidad formal y sustantiva de los factores extraídos.

Conceptos básicos y aplicaciones prácticas con Rcommander

Es una técnica surgida de los trabajos pioneros de autores de la talla de Lawley (1940), Spearman (1904) o Thurstone (1947), que se ha desarrollado de la mano de Gorsuch (1974, 1983) o McDonald (1981, 1985), y que hoy se extiende prácticamente a todos los campos de investigación social. La bibliografía sobre el tema es abundante y puede consultarse entre otros, Gorsuch (1974, 1983), McDonald (1985) o Mulaik (1972). En castellano es especialmente recomendable el libro de Ferrando (1993), además de las obras de Comrey (1985) o Martínez Arias (1995).

7.1 Modelo de análisis factorial

El modelo de análisis factorial asume la presencia de factores latentes que dan cuenta de la variación observada en la matriz de correlaciones y establece relaciones lineales entre los factores y las variables manifiestas. Utilizando los esquemas de Wright, un modelo bifactorial podría representarse gráficamente del siguiente modo:



7.1 Figura.

Modelo bifactorial

Esta representación gráfica se formaliza utilizando el siguiente conjunto de ecuaciones:

$$\begin{aligned}
 Z_1 &= a_{11}F_1 + a_{12}F_2 + u_1D_1 \\
 Z_2 &= a_{21}F_1 + a_{22}F_2 + u_2D_2 \\
 Z_3 &= a_{31}F_1 + a_{32}F_2 + u_3D_3 \\
 Z_4 &= a_{41}F_1 + a_{42}F_2 + u_4D_4
 \end{aligned}$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Donde Z_i son las variables observadas,
 F_j , son los factores comunes,
 D_i , son los factores específicos
 a_{ij} , j son los pesos o saturaciones factoriales.

Los elementos que integran el modelo son:

- *Variables observadas*. Son variables estandarizadas; en el ejemplo son 4 ($n=4$).
- *Variables latentes*, también conocidas como factores o dimensiones. Son de dos tipos:
 - *Factores comunes*. Son estructuras que se relacionan con más de una variable observada; En este ejemplo existen dos factores ($m=2$).
 - *Factores específicos*. Son estructuras latentes asociadas a cada una de las variables observadas. Su número es n .
- *Pesos o saturaciones*. Son las saturaciones de las variables sobre los factores, en función de los factores se distingue entre:
 - a_{ij} son los pesos asociados a los factores comunes.
 - u_i son los pesos asociados a los factores específicos.

Siguiendo las especificaciones gráficas sería posible representar el modelo del siguiente modo:

$$\begin{aligned}Z_1 &= a_{11}F_1 + u_1D_1 \\Z_2 &= a_{21}F_1 + u_2D_2 \\Z_3 &= a_{32}F_2 + u_3D_4 \\Z_4 &= a_{41}F_1 + a_{42}F_2 + u_4D_4\end{aligned}$$

La ecuación general del modelo factorial podría expresarse como:

$$Z_i = a_{i1} F_1 + a_{i2} F_2 + a_{i3} F_3 + \dots + a_{im} F_m + u_i D_i$$

$$Z_i = \sum_{j=1}^m a_{ij} F_j + u_i D_i$$

7.1.1 Supuestos y consecuencias

Supuesto 1. Tanto los factores como las variables observadas están estandarizados.

$$Z_i \sim N(0,1)$$

$$F_i \sim N(0,1)$$

$$D_i \sim N(0,1)$$

Supuesto 2. La correlación entre los factores comunes y los específicos es nula.

$$\rho_{F_i D_j} = 0 \quad (i = j \vee i \neq j)$$

Supuesto 3. Los factores específicos son estadísticamente independientes.

$$\rho_{D_i D_j} = 0$$

Supuesto 4. Los factores comunes son estadísticamente independientes.

$$\rho_{F_i F_k} = 0$$

A partir de estos supuestos se derivan las siguientes consecuencias:

Consecuencia 1. Varianza de las variables observadas. La varianza de las variables observadas se descompone en comunalidad y unicidad.

$$\sigma_i^2 = 1 = h_i^2 + u_i^2$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Donde h_i^2 es la comunalidad de la variable i
 u_i^2 , es la unicidad de la variable i .

La comunalidad de una variable es el sumatorio del cuadrado de las saturaciones de los factores comunes en esa variable e indica la parte de varianza explicada por los factores comunes.

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 = \sum_{j=1}^m a_{ij}^2$$

Por lo tanto,

$$\sigma_i^2 = h_i^2 + u_i^2 = \sum_{j=1}^m a_{ij}^2 + u_i^2$$

Existe una clara relación entre el concepto de fiabilidad y el modelo factorial. En el contexto de la teoría clásica de tests la unicidad de una variable puede descomponerse en dos: parte sistemática o específica y parte de error. La comunalidad y la parte sistemática de la unicidad contribuirían a la puntuación verdadera siendo equiparable la parte de error de la varianza al error de medida.

$$u_i^2 = b_i^2 + e_i^2$$

$$\sigma_{Z_i}^2 = a_i^2 + u_i^2 = a_i^2 + b_i^2 + e_i^2 = 1$$

Consecuencia 2. Correlación entre variables observadas. La correlación entre dos variables observadas se reproduce por medio del sumatorio de los productos de las saturaciones de los factores comunes sobre ellas. El teorema básico del análisis factorial establece:

$$\rho_{ik} = \sum_{j=1}^m a_{ij} a_{kj}$$

Donde ρ_{ik} es la correlación entre las variables observadas i y k
 a_{ij} , a_{kj} , son las saturaciones del factor j en las variables i y k .

Conceptos básicos y aplicaciones prácticas con Rcommander

Consecuencia 3.- Correlación entre los factores comunes y las variables manifiestas. La correlación entre un factor común y una variable manifiesta es igual al peso del factor en la variable.

$$\rho_{Z_i F_j} = a_{ij}$$

Donde $\rho_{Z_i F_j}$ es la correlación entre el factor común j y la variable observada i
 a_{ij} , es el peso del factor j sobre la variable i

7.1.2 Reproducción de la matriz de correlación

A partir de los supuestos del modelo, es posible reproducir la matriz de correlaciones entre variables observadas. La matriz reproducida es una fuente de información básica en el estudio del ajuste del modelo a los datos; en la medida en que las diferencias entre los valores reproducidos y observados sean equivalentes la solución factorial podrá considerarse óptima.

$$\begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdot & \cdot & a_n \end{bmatrix} = \begin{bmatrix} a_1^2 & a_1 a_2 & \cdot & \cdot & a_1 a_n \\ a_2 a_1 & a_2^2 & \cdot & \cdot & a_2 a_n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_n a_1 & a_n a_2 & \cdot & \cdot & a_n^2 \end{bmatrix} = \begin{bmatrix} h_1^2 & \rho_{12}^* & \cdot & \cdot & \rho_{1n}^* \\ \rho_{21}^* & h_2^2 & \cdot & \cdot & \rho_{2n}^* \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{n1}^* & \rho_{n2}^* & \cdot & \cdot & h_n^2 \end{bmatrix}$$

Donde a_i es el peso del factor común en la variable i ,
 h_i^2 , es la comunalidad de la variable i
 ρ_{ij}^* , es la correlación reproducida entre las variables i y j

$$\mathbf{A} \times \mathbf{A}' = \mathbf{R}^*$$

La diagonal principal de la matriz de correlaciones reproducida está ocupada por las comunalidades. El objetivo principal del análisis factorial es maximizar la comunalidad de las variables, es decir la parte de la varianza común debida a los factores comunes. A medida que aumenta la comunalidad de las variables los valores de la diagonal principal de la matriz de correlaciones tenderán a 1.

La diferencia entre la matriz de correlaciones observadas y la matriz de correlaciones reproducida o matriz reducida es la matriz de residuales.

$$res_{ij} = \rho_{ij} - a_i a_j$$

$$res_{ij} = \rho_{ij} - \rho_{iF} \rho_{jF}$$

$$\mathbf{res}_{ij} = \begin{bmatrix} u_1^2 & res_{12} & \cdot & \cdot & res_{1n} \\ res_{21} & u_2^2 & \cdot & \cdot & res_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ res_{n1} & res_{n2} & \cdot & \cdot & u_n^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdot & \cdot & \rho_{1n} \\ \rho_{21} & 1 & \cdot & \cdot & \rho_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{n1} & \rho_{n2} & \cdot & \cdot & 1 \end{bmatrix} - \begin{bmatrix} h_1^2 & \rho_{12}^* & \cdot & \cdot & \rho_{1n}^* \\ \rho_{21}^* & h_2^2 & \cdot & \cdot & \rho_{2n}^* \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{n1}^* & \rho_{n2}^* & \cdot & \cdot & h_n^2 \end{bmatrix}$$

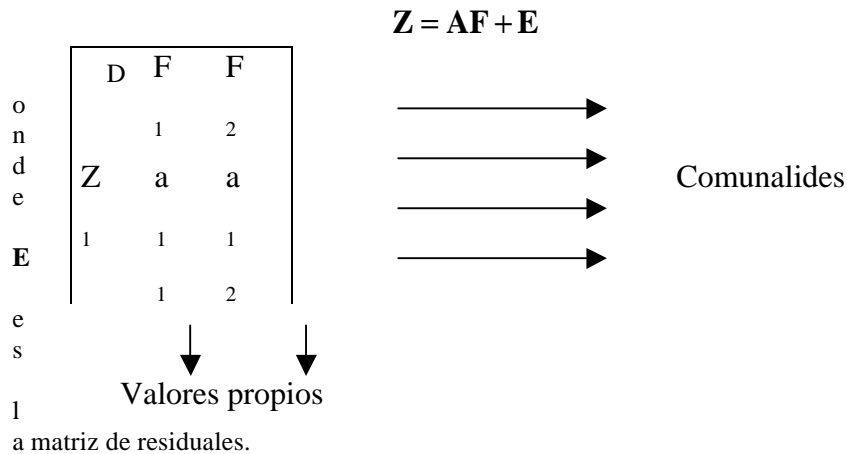
7.1.3 Representación matricial

La ecuación básica del modelo factorial expresada en lenguaje matricial adquiere la siguiente forma:

$$\mathbf{Z} = \mathbf{AF} + \mathbf{UD}$$

- Donde **Z** es la matriz de correlaciones entre variables observadas,
F, es la matriz de factores comunes,
A, es la matriz de pesos,
U, es la matriz de pesos de los factores específicos
D, es la matriz de factores específicos.

En la mayoría de las ocasiones sin embargo, la expresión se presenta como:



$$\begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \cdot \\ \cdot \\ F_m \end{bmatrix} + \begin{bmatrix} u_{11} & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ D_n \end{bmatrix}$$

$$[n \times 1] = [n \times m] [m \times 1] + [n \times n] [n \times 1]$$

7.2 Figura. Matriz factorial.

La matriz A recibe el nombre de matriz factorial, y recoge las saturaciones de los factores comunes en las variables observadas (a_{ij}). Cada uno de los elementos de la matriz (a_{ij}) expresa la correlación entre la variable observada y el factor común. El sumatorio del cuadrado de cada uno de los elementos de cada una de las columnas es el valor propio o *eigenvalue* del factor común correspondiente. Los valores propios son característicos de los factores y se utilizan para interpretar la parte de la varianza de las variables explicada por el factor.

La lectura de la matriz factorial por filas proporciona información sobre las comunalidades de las variables, es decir, la parte de varianza de cada variable explicada por los factores comunes.

Conceptos básicos y aplicaciones prácticas con Rcommander

El producto de la matriz factorial por su transpuesta dará como resultado la matriz de correlaciones reproducida:

$$\mathbf{R}^* = \mathbf{A}\mathbf{A}'$$

La matriz original se obtendrá sumando a la matriz de correlaciones reproducida o matriz reducida (\mathbf{R}^*) la matriz de unicidades.

$$\mathbf{R} = \mathbf{R}^* + \mathbf{U}$$

7.1.4 Extracción factorial

La extracción factorial es el núcleo del análisis factorial. Son muchos los procedimientos algebraicos y estadísticos contruidos con tal fin. Su exposición sobrepasa los límites de este manual y remitimos al lector a la bibliografía especializada sobre el tema, entre la cual destacaríamos Comrey (1985), Gorsuch (1974), Harman (1980), Maxwell (1977), McDonald (1985) ó Mulaik (1972).

Expondremos brevemente las bases del procedimiento de estimación por máxima verosimilitud, por ser el que implementa `factanal`, la función que utiliza `Rcommander` para ejecutar el análisis factorial.

Método de máxima verosimilitud

Se trata de un procedimiento de estimación estadístico que obtiene estimadores con propiedades estadísticas deseables. Es un método simple aplicable a multitud de modelos diferentes. Una vez obtenido un estimador máximo-verosímil la teoría general sobre estimación por máxima verosimilitud proporciona errores estándar, test de significación y demás resultados que son útiles en la inferencia estadística.

Conceptos básicos y aplicaciones prácticas con Rcommander

Ejemplo. Supongamos que se quiere estimar la probabilidad de obtener cara en un experimento de lanzamiento de monedas.

.- Se lanza la moneda 10 veces ($n=10$) y se obtiene el siguiente resultado:

CC+CCC++CC.

.- La probabilidad de obtener esa secuencia es una función de un parámetro desconocido Π :

$$\Pr(\text{datos}|\text{parámetro})=\Pr(\text{CC+CCC++CC}|\Pi)=\Pi\Pi(1-\Pi)\Pi\Pi\Pi(1-\Pi)(1-\Pi)\Pi\Pi=\Pi^7(1-\Pi)^3.$$

Los datos de nuestro experimento son fijos; los conocemos.

El parámetro Π también tiene un valor fijo, pero este valor es desconocido, y podemos estimar cual sería su valor tratando la probabilidad de los datos observados como una función de Π . Esta función es la función de verosimilitud:

$$L(\text{parámetro}|\text{datos}) = L(\Pi|\text{CC+CCC++CC}) = \Pi^7(1-\Pi)^3$$

La función de probabilidad y la función de verosimilitud vienen dada por la misma ecuación, pero la función de probabilidad es una función de los datos con el valor del parámetro fijo, mientras que la función de verosimilitud es una función del parámetro con los datos conocidos.

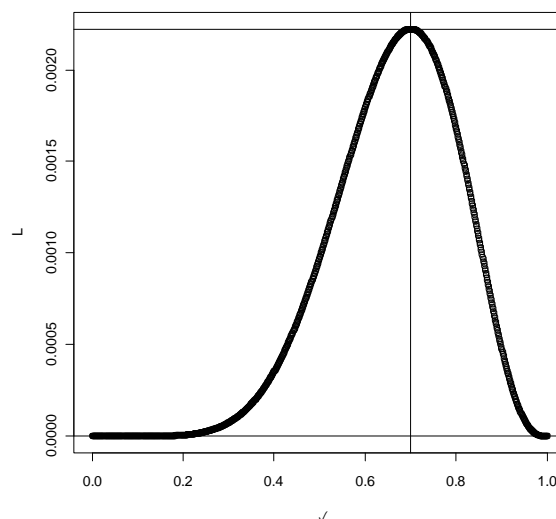
En la tabla siguiente se muestran distintos valores de la función de verosimilitud para distintos valores de Π .

Π	$L(\text{parámetro} \text{datos}) = \Pi^7(1-\Pi)^3$
-------	---

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

0,0	0,0
0,1	0,0000000729
0,2	0,00000655
0,3	0,0000750
0,4	0,000354
0,5	0,000977
0,6	0,00179
0,7	0,00222
0,8	0,00168
0,9	0,000478
1,0	0,0

La función de verosimilitud completa ha sido reproducida en la siguiente ilustración. Aunque cada valor de $L(\Pi|\text{datos})$ tiene una notación probabilística, la función no es una función de probabilidad y el área encerrada no es 1.



7.3 Figura. Función de Verosimilitud.

La probabilidad de obtener la muestra de datos es pequeña en relación al parámetro Π . La función de probabilidad contiene información muy importante respecto al parámetro desconocido Π . Por ejemplo, mencionaremos que Π no puede adoptar los valores 0 o 1 y es poco probable que su valor sea cercano a 0

Conceptos básicos y aplicaciones prácticas con Rcommander

o 1. Invirtiendo el razonamiento se podría afirmar que el valor de Π más probable vistos los datos es aquél que genera una verosimilitud mayor, este valor es el *estimador máximo verosímil*. El *estimador máximo verosímil* es el máximo de la función de verosimilitud, es decir, aquél cuyo valor maximiza la probabilidad de obtener los datos conocidos. En este ejemplo el estimador máximo verosímil del parámetro Π es 0,7, que coincide con la proporción de caras en el experimento de lanzamiento de monedas, 7/10.

En términos de cálculo, es más sencillo maximizar el logaritmo de la función de verosimilitud que la función de verosimilitud, porque la transformación logarítmica convierte los términos del producto en términos aditivos.

$$\text{Log}L(\pi) = x \log \pi + (n - x) \log(1 - \pi)$$

Diferenciando la función respecto al parámetro e igualando a 0 se obtendría el estimador máximo verosímil.

$$\frac{d \log L(\pi)}{d\pi} = 0$$

Los estimadores máximo verosímiles son consistentes, asintóticamente insesgados, asintóticamente eficientes y distribuidos según la ley normal.

Siendo $L(\hat{\pi})$ el valor de la función de verosimilitud en el punto MLE, y $L(\pi)$ la función de verosimilitud en el parámetro verdadero, la razón de verosimilitudes sigue asintóticamente una distribución Chi cuadrado con un grado de libertad bajo la hipótesis nula.

$$G^2 \equiv -2 \log \frac{L(\hat{\pi})}{L(\pi)} = 2 [\log L(\hat{\pi}) - \log L(\pi)]$$

7.1.5 Modelo de componentes principales

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Aunque, desde un punto de vista formal, el modelo de componentes principales no es un modelo de análisis factorial, su uso como tal es común en la investigación aplicada. De hecho, programas con licencias privativas como el SPSS incluyen el modelo de componentes principales como un método de extracción factorial. El objetivo del modelo de componentes principales es reproducir la varianza total mientras que el análisis factorial está interesado en la varianza común. Esa diferencia básica en los objetivos marca la diferencia entre los modelos.

Es cierto, sin embargo, que cuando el modelo de componentes principales se utiliza como un método de extracción factorial los resultados pueden interpretarse de modo análogo. De hecho, a medida que aumenta la varianza común, aumenta la equivalencia entre ambos modelos.

El modelo de componentes principales define a partir de de combinaciones lineales entre n variables observadas, Y componentes que explican el máximo de varianza observada. Una vez estimado un componente, su efecto es eliminado y se vuelve a estimar un segundo componente relacionado con la varianza residual.

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{n1}X_n \\ Y_2 &= a_{12}X_1 + a_{22}X_2 + \dots + a_{n2}X_n \\ &..... \\ Y_n &= a_{1n}X_1 + a_{1n}X_2 + \dots + a_{nn}X_n \end{aligned}$$

Las diferencias entre el modelo factorial y el modelo de componentes principales podría sistematizarse del siguiente modo:

- a.- El análisis de componentes principales define n componentes a partir de n variables observadas, y esos componentes se ordenan en función de la varianza explicada por cada uno de ellos; aunque cuando se utiliza en el marco del análisis factorial no se retienen todos los componentes extraídos. En el modelo factorial se estiman m factores comunes y n

Conceptos básicos y aplicaciones prácticas con Rcommander

factores específicos a partir de un número de variables observadas igual a n .

b.- La parte más significativa de la solución factorial está constituida por las comunalidades; en el modelo de componentes principales el concepto fundamental es la varianza. Las primeras son desconocidas las segundas, sin embargo, no.

7.1.6 Rotación factorial

La solución factorial generada tras someter a un conjunto de datos a un análisis factorial cumple la ecuación del modelo ($\mathbf{AA}'=\mathbf{R}^*$), sin embargo, eso no significa que la solución sea única. Para cada análisis es posible encontrar varias matrices que cumplan la condición anterior.

Entre todas las posibles soluciones habría que optar por aquella que mejor pueda interpretarse desde un punto de vista sustantivo; esta solución habitualmente se aproxima a la estructura simple definida por Thurstone (1947). Una matriz factorial \mathbf{P} tiene una estructura simple cuando:

- Cada fila en la matriz \mathbf{P} tiene al menos un 0 (o valores próximos a 0). Es decir, cuando para cada variable existe al menos un factor que no contribuye a la varianza de la variable.
- Para cada factor en la matriz \mathbf{P} existe un grupo de variables con peso nulo (o valores próximos a 0).
- Para cada par de columnas existirán variables con valores próximos a 0 en una y elevados en la otra.

La estructura de una matriz factorial con estructura simple podría ser similar a:

	Factor 1	Factor 2	Factor 3
--	----------	----------	----------

Conceptos básicos y aplicaciones prácticas con Rcommander

1	XX		
2	XX		
3		XX	
4		XX	
5			XX
6	XX		XX

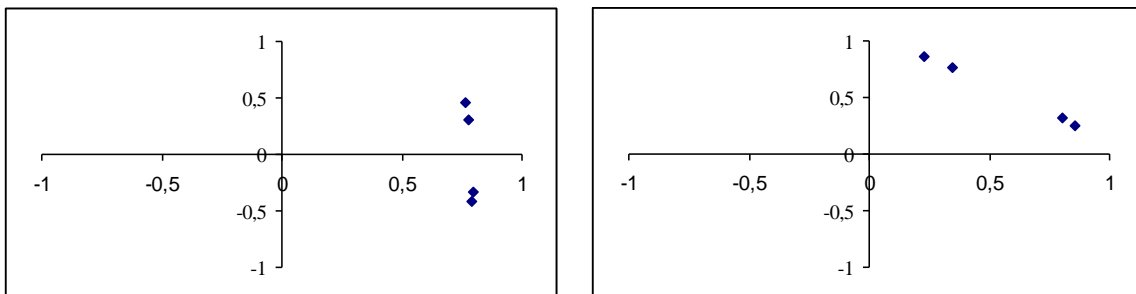
8.1 Tabla. Estructura simple

La rotación factorial consiste en encontrar una nueva matriz factorial que se aproxime a la estructura simple,

$$P = AT$$

Donde **A** es la matriz factorial
T, es la matriz de transformación.

Si partiéramos de una solución factorial como la expresada en la siguiente matriz factorial y la rotáramos, podríamos obtener una nueva solución, la solución rotada. Formalmente equivalente a la solución original, pero más fácil de interpretar; con pesos que se distribuyen en bloques claramente interpretables. La representación gráfica en coordenadas cartesianas, muestra la proximidad de las variables a los ejes que representan los factores en la solución rotada, y la difícil interpretación sustantiva de la solución original.



7.4. Figura. Representación Solución Factorial.

Conceptos básicos y aplicaciones prácticas con Rcommander

La rotación factorial no afecta a la calidad formal de la solución factorial; las comunalidades de las dos soluciones (no-rotada, rotada) son equivalentes; la rotación produce una redistribución de las cargas factoriales que afecta a la parte de varianza común asociada a cada factor. Los procedimientos de rotación factorial se agrupan en dos grandes bloques; los procedimientos ortogonales y los procedimientos oblicuos. Los primeros mantienen la independencia entre factores, y los segundos estiman factores correlacionados.

Rotaciones ortogonales

- *Varimax*: Formulada por Kaiser (1958), la rotación *varimax* maximiza la varianza factorial. Para ello minimiza el número de variables con altas saturaciones. Es el procedimiento más empleado.
- *Qartimax*: El objetivo de la rotación *qartimax* es simplificar la explicación de las variables, para ello minimiza el numero de factores comunes asociados a ellas. Es un procedimiento recomendado cuando el número de variables observadas es elevado.
- *Equamax*: Es un procedimiento que combina los dos anteriores.

	Solución no-rotada			Solución rotada	
	F ₁	F ₂		F ₁	F ₂
Z ₁	0,780	0,304	Z ₁	0,348	0,761
Z ₂	0,799	-0,328	Z ₂	0,802	0,321
Z ₃	0,767	0,462	Z ₃	0,229	0,865
Z ₄	0,789	-0,417	Z ₄	0,857	0,250

Rotaciones oblicuas

Las rotaciones oblicuas son muy utilizadas en la investigación aplicada porque habitualmente ofrecen soluciones más interpretables, sin embargo la generalización de la solución es menor que la atribuible a una rotación ortogonal. Una rotación oblicua de una solución factorial genera dos matrices

Conceptos básicos y aplicaciones prácticas con Rcommander

diferentes, la matriz de estructura (S) y la matriz de configuración o matriz factorial (P). Los elementos de la primera son correlaciones entre las variables y los factores extraídos; los elementos de la matriz de configuración muestran las saturaciones de los factores sobre las variables. La obtención de la matriz de correlaciones reproducida tras una rotación oblicua se ajustaría a la siguiente expresión:

$$\mathbf{R}^* = \mathbf{P}\mathbf{\Phi}\mathbf{P} + \mathbf{U}^2$$

Entre los procedimientos de rotación oblicua,

- *Oblimin* (Jennrich y Sampson, 1966) es la más utilizada en la investigación aplicada. Con el fin de controlar la correlación entre factores, es necesario la especificación del parámetro Delta. Un valor 0 expresaría la máxima correlación.
- *Promax* (Hendrickson y White, 1964), este procedimiento maximiza la razón entre los pesos mayores y menores.

7.2 Etapas en la ejecución de un análisis factorial

1.- *Hipótesis*. Todo análisis factorial ha de ir guiado por un conocimiento de los datos a analizar que permita formular hipótesis sobre su estructura interna. Es cierto, sin embargo que en muchos contextos de investigación no hay hipótesis relativas a la dimensionalidad, convirtiéndose el análisis factorial en una herramienta meramente descriptiva.

En general, el estudio de la dimensionalidad latente a un conjunto de datos puede adoptar dos perspectivas; exploratoria y confirmatoria. En la primera se obtendrá una descripción de la estructura interna de los datos, en la segunda se contrastarán hipótesis referidas a la dimensionalidad. El contexto de

Conceptos básicos y aplicaciones prácticas con Rcommander

investigación delimitará el tipo de acercamiento a la estructura de los datos, siendo el confirmatorio preferido al exploratorio.

2.- *Extracción factorial.* Viene determinada por la elección de un método de extracción que se fijará en función de las características de los datos.

3.- *Adecuación de la solución.* La adecuación de la solución factorial dependerá en parte del grado en que los datos a analizar cumplen los supuestos del modelo factorial. A medida que aumenta la distancia entre los supuestos del modelo y las características de los datos la solución obtenida será más débil.

4.- *Rotación factorial.* En muchas ocasiones la primera solución factorial es difícil de interpretar desde una perspectiva sustantiva, por ello se opta por buscar una solución formalmente equivalente pero cuya interpretación sea más clara desde el punto de vista sustantivo.

5.- *Interpretación.* La interpretación de la solución factorial ha de aunar el aspecto formal y el sustantivo del análisis. Las decisiones que se tomen relativas al número de factores a retener, su interpretación psicológica o la adecuación del modelo no puede guiarse nunca por criterios exclusivamente numéricos; el análisis está siempre supeditado a la teoría.

6.- *Estimación de las puntuaciones factoriales.* Tras estimar el modelo y valorar positivamente su grado de adecuación es posible estimar para cada uno de los sujetos de la muestra las puntuaciones factoriales, o valores asignados en cada uno de los factores extraídos.

7.2.1 Interpretación de la solución factorial

7.2.1.a Número de dimensiones

Conceptos básicos y aplicaciones prácticas con Rcommander

Una vez ejecutado un análisis factorial, es importante interpretar correctamente la solución para que las conclusiones tengan un sustento formal y sustantivo. Es posible extraer tantos factores como variables observadas y es importante decidir sobre la solución más adecuada. Son varios los criterios formales disponibles que puedan ayudar en tal proceso (Elosua y López, 2002). Entre ellos es posible distinguir entre indicadores basados en el porcentaje de varianza explicada, indicadores basados en el análisis de residuales e índices relativos.

Índices relacionados con la varianza

Son los índices más utilizados en la investigación aplicada para la determinación del número de factores a retener e interpretar tras un análisis factorial, entre otras razones porque son muy sencillos de generar y de interpretar. Su justificación para el caso de un modelo unidimensional por ejemplo, es muy simple. A medida que aumenta la diferencia entre la varianza asociada al primer factor extraído, y la varianza que explica el segundo factor extraído, mayor será la evidencia acerca de la unidimensionalidad de los datos. Son varios los índices construidos sobre esa base:

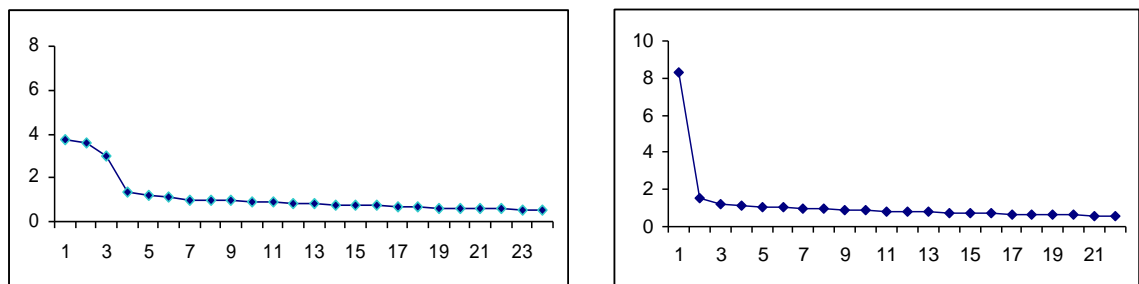
- *Criterio de Carmines y Zeller (1979)* Un conjunto de datos es unidimensional si el primer factor extraído explica el 40% de variabilidad observada.
- *Criterio de Reckase (1979)*. Esta autor propone un punto de corte del 20% en la determinación de la varianza que tiene que explicar un factor para poder considerar la existencia de una dimensión dominante.
- *Índice de Lord (Lord, 1980)*. Para establecer el grado de relevancia del primer factor sobre el resto de factores extraídos, este autor estima la razón entre la diferencia entre los dos primeros factores y la diferencia entre el segundo factor y el tercer factor.

Conceptos básicos y aplicaciones prácticas con Rcommander

En las situaciones de multidimensionalidad, los siguientes criterios pueden ser útiles en la fijación del número de factores a retener:

- *Criterio de Kaiser-Guttman.* (Guttman, 1954; Kaiser, 1960). Estos autores equiparan el número de dimensiones con el número de factores que poseen valores propios mayores a la unidad. A pesar de ser un criterio muy utilizado en muchas ocasiones sobreestima el número de dimensiones.
- *Scree-Plot de Cattell:* (Cattell, 1966, 1978) Este autor propone un procedimiento gráfico en el que se representan los valores propios de cada factor extraído. Dibujando una línea recta en el punto de inflexión de la curva así construida es posible extraer conclusiones sobre el número de factores necesarios para explicar los datos observados. La simplicidad y efectividad del procedimiento hace de él el método de decisión más utilizado en la investigación aplicada.

En la siguiente figura se muestran dos *screenplots* o gráficos de sedimentación que reflejan dos estructuras dimensionales diferentes; una de ellas es unidimensional y la otra representa una solución en tres factores.



7.5 Figura. Gráficos de sedimentación

Índices relacionados con los residuos

Conceptos básicos y aplicaciones prácticas con Rcommander

La adecuación del modelo a los datos queda reflejada en la similitud/diferencia entre la matriz de correlaciones reproducida y la matriz de correlaciones observadas. La diferencia entre ambas se aproximará a 0 a medida que lo haga la adecuación del modelo. El estudio de los residuales o diferencias entre ambas matrices, es un modo de aproximación al problema de la adecuación que está siendo objeto de estudio como fuente de información sobre el origen de la falta de ajuste y sobre la magnitud de esta.

Uno de los procedimientos básicos en el estudio del ajuste es el estudio de la distribución de los errores. El modelo asume que los errores siguen una distribución normal con media 0.

Error cuadrático residual medio (Root mean square error). Es uno de los índices más utilizados; cuando la adecuación del modelo es buena su valor se aproxima a 0

$$ECMR = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (r_{ij} - r_{ij}^*)^2}{n(n-1)}}$$

Donde r_{ij} es la correlación entre las variables i y j ,
 r_{ij}^* , es la correlación reproducida entre i y j
 n , número de variables analizadas.

Índices relativos

Cuando la rotación factorial se ha efectuado por medio del procedimiento de estimación por máxima verosimilitud se puede utilizar la razón de verosimilitud como criterio para la comparación de dos soluciones factoriales diferentes. En la comparación de soluciones con un número de factores diferente, este contraste ofrece un estadístico que asintóticamente se distribuye como una Chi cuadrado con grados de libertad igual a la diferencia entre los grados de libertad de los modelos evaluados (Lawley, 1940).

$$G = -2n \log \frac{P(\text{datos} | \text{modelo reducido})}{P(\text{datos} | \text{modelo extendido})}$$

Dados los problemas relacionados con el estadístico Chi cuadrado referidos a su sensibilidad con respecto al tamaño de la muestra, es habitual valorar el grado de ajuste atendiendo a estadísticos bondad de ajuste. Los estadísticos de bondad de ajuste ofrecen valores que dentro de un contexto de interpretación deberán ser valorados por el investigador; es decir, no son pruebas estadísticas que rechacen o acepten una hipótesis nula. Entre ellos podríamos citar el propuesto por McDonald (1999), cuyo valor es menor a 0,05 cuando el ajuste es adecuado:

$$RMSEA = \sqrt{d / gl}$$

$$d = \frac{G - g.l.}{N}$$

Donde G es la razón de verosimilitud,
gl son los grados de libertad
N, es el tamaño de la muestra.

7.2.1.b Criterios generales para la interpretación

- ¿Qué se ha de interpretar en una solución factorial? En la interpretación de los factores es importante tener en cuenta las saturaciones. Tras una rotación oblicua es necesario especificar la matriz que está siendo interpretada (correlaciones o saturaciones).

- ¿Qué saturaciones han de interpretarse? Aquellas que son significativas; en ausencia de los errores de estimación, Stevens (1992) propuso como punto de

Conceptos básicos y aplicaciones prácticas con Rcommander

corte 0,40 . Sin embargo la contextualización correcta del trabajo determinará el punto de corte en cada caso.

- ¿Cuál es el tamaño mínimo de muestra necesario? Son varias las respuestas que se han dado a esta cuestión; entre ellas algunos autores afirman que como mínimo son necesarios 5 sujetos por variable (Gorsuch, 1983; Stevens, 1992). Sin embargo, a pesar de la recomendación es fundamental analizar la solución factorial.

a.- Comunalidades:

- Cuando las comunalidades son altas, ($>0,60$) y los factores aparecen claramente definidos es suficiente un tamaño de muestra de 100.
- Cuando las comunalidades giran en torno a 0,5 y los factores están claramente definidos el tamaño de muestra suficiente es 100-200.
- Cuando las comunalidades son menores a 0,5 y los factores están bien definidos el tamaño de muestra anterior seguiría siendo aceptable.
- Cuando las comunalidades son bajas, el número de factores es reducido y cada factor tiene 4 o 5 indicadores manifiestos es necesaria una muestra mínima de 300.
- Cuando las comunalidades son bajas, el número de factores elevado y su definición pobre, son necesarios tamaños de muestra superiores a 500 para obtener una estructura factorial adecuada.

b.- Sobre la definición de los factores (Guadagnoli y Velicer, 1988):

- Cuando un factor tiene cuatro o más variables y las saturaciones son superiores a 0,60, ese factor es fiable.
- Siendo las saturaciones bajas, un factor será fiable cuando además de estar definido por 10 indicadores o más, y el tamaño de la muestra es superior a 150.

Conceptos básicos y aplicaciones prácticas con Rcommander

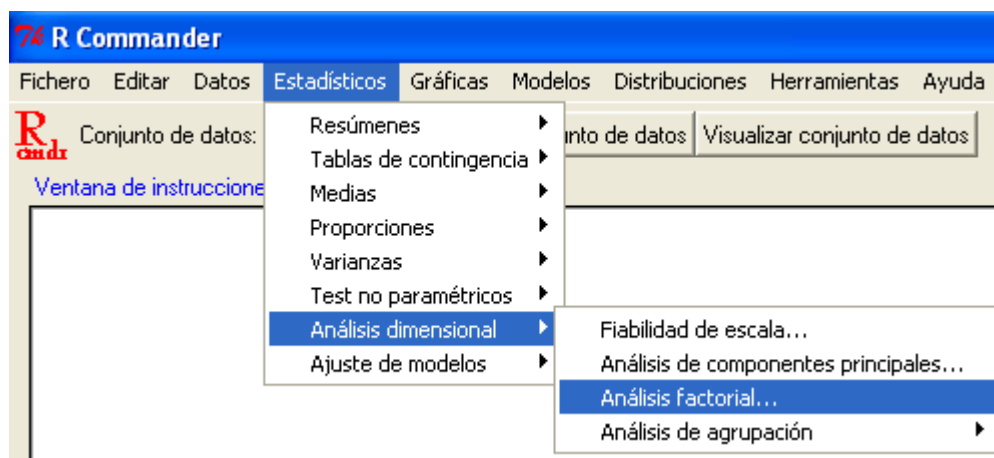
- Cuando el tamaño de la muestra es menor a 300 no se tendrán en cuenta factores con bajas saturaciones.
- En general es recomendable reducir el número de variables y factores e un análisis factorial, para asegurar comunalidades significativas. Sin embargo todas las consideraciones anteriores han de estar guiadas por un buen conocimiento teórico de las variables estudiadas; nunca se ha de anteponer la solución estadística a la solución teórica.

7.3 Análisis factorial y Rcommander

Dentro de los posibles modos de evaluar la dimensionalidad de un conjunto de variables, Rcommander incorpora el análisis factorial exploratorio, el análisis de componentes principales y el análisis de conglomerados. Se trata de funciones básicas en el estudio de la dimensionalidad que pueden extenderse con la utilización de paquetes adicionales (por ejemplo, *sem*).

7.3.1 Modelo unidimensional

La secuencia de selección para la ejecución de un análisis factorial es: Estadísticos>Análisis Dimensional>Análisis factorial.



7.6 Figura. Análisis Factorial con Rcommander.

Conceptos básicos y aplicaciones prácticas con Rcommander

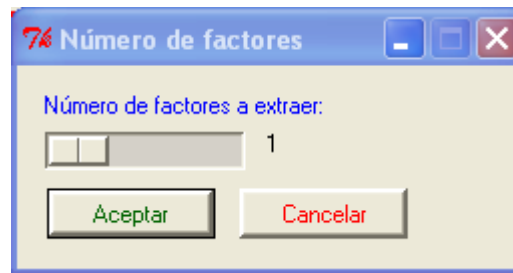
Esa selección abre una ventana donde se especificarán las variables a analizar, el método de rotación a aplicar y el procedimiento de estimación de las puntuaciones factoriales. Rcommander ejecutará el análisis factorial sobre el conjunto de datos activo, o en función de los intereses del investigador, sobre un subconjunto de datos que cumplan una determinada condición.



7.7 Figura. Menú para el análisis factorial.

Dado que estamos analizando la escala Insatisfacción Corporal, se seleccionarán los 10 ítems que la componen; los 10 ítems recodificados. En esta ocasión dejaremos la opción “Expresión de selección” en blanco porque nos interesa evaluar el modelo en el conjunto de la muestra. Como el objetivo es evaluar la unidimensionalidad de los datos la opción de rotación de los factores carece de sentido y por ello se marcará el cuadro “Ninguno”. Una vez especificados los argumentos, Rcommander nos pide el número de factores a extraer; fijaremos el valor 1. El procedimiento de estimación incorporado en Rcommander es máxima verosimilitud.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



7.8 Figura. Factores a extraer.

La función utilizada por Rcommander para ejecutar el análisis factorial es `factanal`:

```
factanal(x = ~RecIC_1 + RecIC_2 + RecIC_3 + RecIC_4 + RecIC_5 +  
RecIC_6 + RecIC_7 + RecIC_8 + RecIC_9 + RecIC_10, factors = 1,  
data = Edi.data, scores = "none", rotation = "none")
```

Las salidas generadas por Rcommander aparecen agrupadas en tres bloques: Unicidad (*Uniquenesses*), pesos factoriales (*Loadings*) e información sobre la adecuación de los datos al modelo.

Unicidad

La unicidad es la parte de varianza específica de cada variable. Como puede comprobarse, a medida que aumente la unicidad de una variable disminuye su comunalidad o varianza común. De hecho, la varianza de una variable se descompone en unicidad y varianza común, que en el caso de variables estandarizadas sumarán 1. Por ejemplo, para el primer ítem de la escala, $1 = 0,479 + 0,722^2$. Los elementos son la varianza total (1), la unicidad (0,479) y la varianza común ($h^2=0,722^2$). En los modelos unifactoriales la varianza común (h^2) es el cuadrado de la saturación factorial ($h^2=a^2$).

En el modelo de un factor, las saturaciones de las variables en los factores son correlaciones entre el factor y la variable. Desde el punto de vista de la regresión lineal se trataría de coeficientes de regresión. En nuestro ejemplo el ítem con una saturación más baja es el ítem 7, que por tanto mostrará una

Conceptos básicos y aplicaciones prácticas con Rcommander

unicidad mayor que el resto de variables. En el análisis de la fiabilidad de la escala, observamos ya un comportamiento algo anómalo de este ítem que obtuvo el índice de discriminación más bajo (0,1881).

Unicidades:

```
Uniquenesses:

RecIC_1  RecIC_2  RecIC_3  RecIC_4  RecIC_5  RecIC_6  RecIC_7
RecIC_8  RecIC_9  RecIC_10
0.479    0.380    0.464    0.316    0.534    0.448    0.962    0.355
0.407    0.356
```

Saturaciones:

```
Loadings:

          Factor1
RecIC_1  0.722
RecIC_2  0.787
RecIC_3  0.732
RecIC_4  0.827
RecIC_5  0.683
RecIC_6  0.743
RecIC_7  0.196
RecIC_8  0.803
RecIC_9  0.770
RecIC_10 0.803
```

Información sobre la adecuación del modelo:

```
          Factor1
SS loadings      5.30
Proportion Var   0.53

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 822.41 on 35 degrees of freedom.
The p-value is 4.29e-150
```

Conceptos básicos y aplicaciones prácticas con Rcommander

El valor propio (SS loadings) del factor extraído y la parte de varianza común asociada (Proportion Var) son ofrecidas por `factanal` en el último bloque informativo. El valor propio es la suma al cuadrado de las saturaciones ($5,30 = 0,722^2 + 0,787^2 + 0,732^2 + 0,827^2 + 0,683^2 + 0,743^2 + 0,196^2 + 0,803^2 + 0,770^2 + 0,803^2$). Dividiendo ese valor entre el número de elementos ($n=10$) se obtiene la proporción de varianza explicada por el factor ($5,30/10 = 0,53$). En nuestro caso, el factor extraído es capaz de dar cuenta del 53% de la variabilidad observada en los datos.

Ese valor es utilizado en la literatura psicométrica para evaluar el grado de unidimensionalidad de la escala. El 53% es superior al punto de corte establecido por Reckase (20%), y es superior incluso al 40% a partir del cual todos los procedimientos de evaluación de la dimensionalidad consideran un conjunto de datos como unidimensional (Elosua y López, 2002). Por consiguiente podemos concluir que a los 10 ítems que conforman la escala analizada subyace un factor principal relacionado con la Insatisfacción Corporal.

Por último, y dado que se ha utilizado el procedimiento de estimación por máxima verosimilitud, `Rcommander` ofrece el estadístico Chi-cuadrado bajo la hipótesis nula de que el modelo se ajusta a los datos. El valor 822,41 que se distribuye con 35 grados de libertad ofrece evidencia en contra de la hipótesis nula. Sin embargo, en el contexto del modelo del factor común, la información ofrecida por este estadístico no es concluyente y ha de ser completada con indicadores e índices desarrollados para tal fin que se obtendría en un análisis factorial confirmatorio.

Análisis de contenido

El análisis formal de un conjunto de datos, en este caso el análisis factorial, ha de complementarse en toda situación con un análisis sustantivo basado en el contenido de las variables analizadas. En el ejemplo que estamos abordando, el análisis del contenido de los ítems, apoyaría las conclusiones numéricas en el sentido de covariación entre los ítems; el contenido de todos los

Conceptos básicos y aplicaciones prácticas con Rcommander

ítems a excepción del ítem séptimo son apreciaciones subjetivas sobre el conjunto del cuerpo o partes de él. El contenido del séptimo ítem, que presenta una saturación algo menor que el resto y tiene un índice de discriminación menor, se “escapa” de este patrón de contenido general haciendo referencia a una situación concreta.

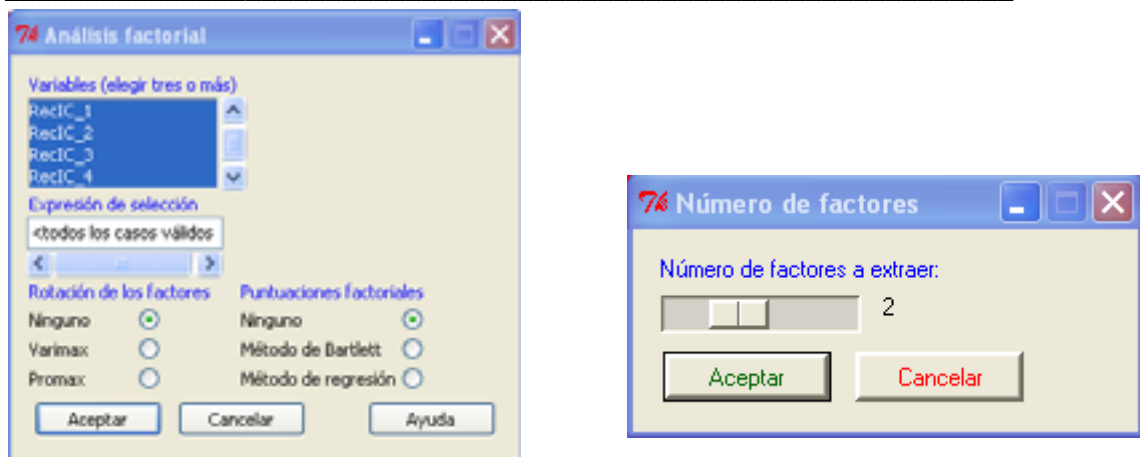
- 1.- Creo que mi estómago es demasiado grande.
- 2.- Pienso que mis muslos son demasiado gruesos.
- 3.- Creo que mi estómago tiene el tamaño adecuado.
- 4.- Me siento satisfecho con mi figura.
- 5.- Me gusta la forma de mi trasero.
- 6.- Creo que mis caderas son demasiado anchas.
- 7.- Me siento hinchado después de una comida normal.
- 8.- Creo que el tamaño de mis muslos es adecuado.
- 9.- Creo que mi trasero es demasiado grande.
- 10.- Creo que mis caderas tienen el tamaño adecuado.

7.3.2 Modelo bidimensional

Para mostrar la aplicación e interpretación de un modelo multidimensional analizaremos dos escalas provenientes del *Eating Disorder Inventory-3*; insatisfacción corporal y obsesión por la delgadez. En aras a obtener una solución más clara se analizarán conjuntamente cuatro ítems provenientes de cada una de ellas.

En la opción de Rcommander Estadísticos>Análisis dimensional>Análisis factorial podrán seleccionarse las variables a analizar, se definirá el tipo de rotación factorial y finalmente se definirá el número de factores a extraer.

Conceptos básicos y aplicaciones prácticas con Rcommander



7.9 Figura. Análisis factorial bidimensional.

Dado que la configuración a evaluar tiene dos factores, expondremos la solución factorial no-rotada y una solución factorial rotada con el fin de comprobar el efecto de la rotación sobre la matriz factorial. Para ello en la opción “Rotación de los factores” se marcará en función de cada caso, la ventana requerida (“Ninguno”, “Varimax” o “Promax”). Rcommander ofrece dos procedimientos de rotación factorial, varimax y promax. La rotación Varimax (Kaiser, 1958) es una rotación ortogonal, y Promax (Hendrickson y White, 1964) es una rotación oblicua. La primera mantiene la independencia entre factores fijando su correlación a 0; la segunda ofrece una solución en la que los factores están relacionados.

La elección entre un tipo de rotación y otra está siempre en manos del investigador. La solución ortogonal ofrece una generalización mayor que la solución oblicua. La interpretación de la solución ortogonal es más sencilla que la interpretación de la solución oblicua. En el primer caso las saturaciones mostradas en la matriz factorial representan las correlaciones entre las variables y los factores extraídos. La solución oblicua genera dos matrices diferentes, la matriz de configuración que muestra las saturaciones de las variables en los factores (es la salida por defecto de la función `factanal`) y la matriz de estructura que recoge las correlaciones entre las variables y los factores extraídos.

Conceptos básicos y aplicaciones prácticas con Rcommander

Analizando las dos soluciones, no-rotada y rotada, se puede comprobar que las unicidades o varianzas específicas no cambian. Los valores son iguales para los 8 ítems analizados (0,340, 0,559, 0,376, 0,474, 0,555, 0,363, 0,353 y 0,386). La diferencia fundamental entre ambas soluciones se centra en la matriz factorial. La rotación reorganiza las saturaciones; los pesos de las variables en los factores se redistribuyen. En caso preliminar, el primer factor presenta saturaciones altas en todas las variables (0,790, 0,638, 0,735, 0,656, 0,659, 0,727, 0,781 y 0,751) y los pesos asociados al segundo factor no son altos y en algunos casos son negativos (0,191, 0,183, 0,289, 0,309, -0,103, -0,330, -0,194 y -0,223). La variabilidad común explicada por ambos factores es 57,4%.

<i>Uniquenesses:</i>				<i>Uniquenesses:</i>			
<i>RecIC_2</i>	<i>RecIC_3</i>	<i>RecIC_9</i>	<i>RecIC_10</i>	<i>RecIC_2</i>	<i>RecIC_3</i>	<i>RecIC_9</i>	<i>RecIC_10</i>
0.340	0.559	0.376	0.474	0.340	0.559	0.376	0.474
<i>RecOD_3</i>	<i>RecOD_4</i>	<i>RecOD_5</i>	<i>RecOD_7</i>	<i>RecOD_3</i>	<i>RecOD_4</i>	<i>RecOD_5</i>	<i>RecOD_7</i>
0.555	0.363	0.353	0.386	0.555	0.363	0.353	0.386
<i>Loadings:</i>				<i>Loadings:</i>			
	<i>Factor1</i>	<i>Factor2</i>			<i>Factor1</i>	<i>Factor2</i>	
<i>RecIC_2</i>	0.790	0.191		<i>RecIC_2</i>	0.687	0.434	
<i>RecIC_3</i>	0.638	0.183		<i>RecIC_3</i>	0.576	0.330	
<i>RecIC_9</i>	0.735	0.289		<i>RecIC_9</i>	0.719	0.327	
<i>RecIC_10</i>	0.656	0.309		<i>RecIC_10</i>	0.679	0.255	
<i>RecOD_3</i>	0.659	-0.103		<i>RecOD_3</i>	0.385	0.545	
<i>RecOD_4</i>	0.727	-0.330		<i>RecOD_4</i>	0.270	0.751	
<i>RecOD_5</i>	0.781	-0.194		<i>RecOD_5</i>	0.404	0.695	
<i>RecOD_7</i>	0.751	-0.223		<i>RecOD_7</i>	0.363	0.694	
	<i>Factor1</i>	<i>Factor2</i>			<i>Factor1</i>	<i>Factor2</i>	
<i>SS loadings</i>	4.138	0.456		<i>SS loadings</i>	2.298	2.296	
<i>Proportion Var</i>	0.517	0.057		<i>Proportion Var</i>	0.287	0.287	
<i>Cumulative Var</i>	0.517	0.574		<i>Cumulative Var</i>	0.287	0.574	
<i>Solución no rotada</i>				<i>Rotación Varimax</i>			

Conceptos básicos y aplicaciones prácticas con Rcommander

La matriz factorial obtenida tras la rotación es diferente. Las variables aparecen distribuidas en dos bloques claramente diferenciados que se asocian con dos factores bien definidos. Las primeras cuatro variables presentan altas correlaciones con el primer factor, y las últimas cuatro variables presenten niveles de correlación elevados con el segundo factor. Las primeras variables provienen de la escala Insatisfacción corporal (IC) y las cuatro últimas variables pertenecen a la escala Obsesión por la delgadez (OD). La varianza asociada al primer factor es 28,7 y el segundo factor explica el mismo porcentaje de variabilidad. Entre los dos factores dan cuenta del 57,4% de la varianza observada. Este valor es equivalente para las dos soluciones: no-rotada y rotada.

```
IC-1. Creo que mi estómago es demasiado grande.  
IC-8. Creo que el tamaño de mis muslos es adecuado.  
IC-9. Creo que mi trasero es demasiado grande.  
IC-10. Creo que el tamaño de mis caderas es adecuado.  
OD-4. Me aterroriza la idea de engordar.  
OD-5. Exagero o doy demasiada importancia al peso  
OD-6. Estoy preocupado, porque querría ser una persona más  
delgada  
OD-7. SI engordo un Kg. me preocupa que pueda seguir ganando  
peso
```

7.3.3 Representación gráfica de la solución factorial

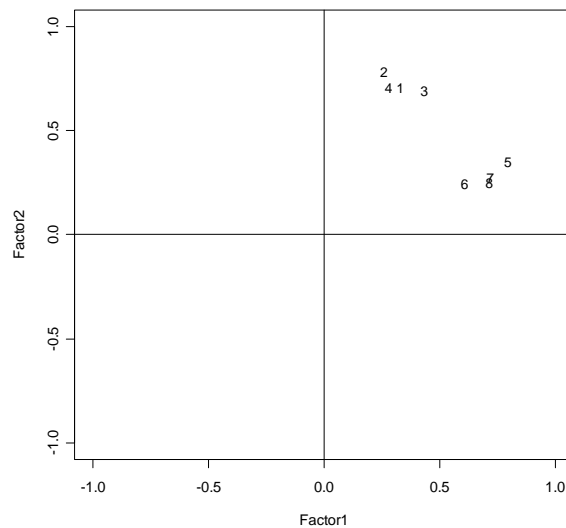
Sería posible obtener una representación gráfica de la distribución de los ítems en el espacio bidimensional generado por ambos factores por medio de la función `plot`; que aplicada a nuestro ejemplo escribiríamos del siguiente modo:

```
plot(.FA$loadings,xlim=range(-1,+1),ylim=range(-1,+1), type="n")  
abline(h=0,v=0);  
text(.FA$loadings,substring(rownames(.FA$loadings),5,6),cex=0.9)
```


Conceptos básicos y aplicaciones prácticas con Rcommander

La función `plot` ordena a R que represente gráficamente los valores de la matriz factorial (`FA$loadings`) generada por el análisis factorial. Los parámetros `xlim` e `ylim` indican el rango de valores para el eje X y para el eje Y, que representan en este caso a los dos factores. “`abline`” es utilizado para añadir las líneas que cortan el gráfico en el punto 0, y finalmente con `text` estamos indicando que queremos visualizar el número del ítem que representan cada punto.

En la solución gráfica se aprecia claramente la existencia de dos grupos de puntos relacionados con el factor 1 y con el factor 2, cuyos contenidos psicológicos vendrían definidos por las escalas “Insatisfacción Corporal”, y “Obsesión por la delgadez”.



7.10 Figura. Solución factorial gráfica

7.3.4 Estimación de las puntuaciones factoriales

En muchos contextos de investigación el análisis factorial exploratorio concluirá con la evaluación del ajuste formal del modelo después de un proceso que incluye la secuencia de estimación de los parámetros del modelo y rotación de los factores extraídos. Sin embargo, son numerosas las situaciones en las que

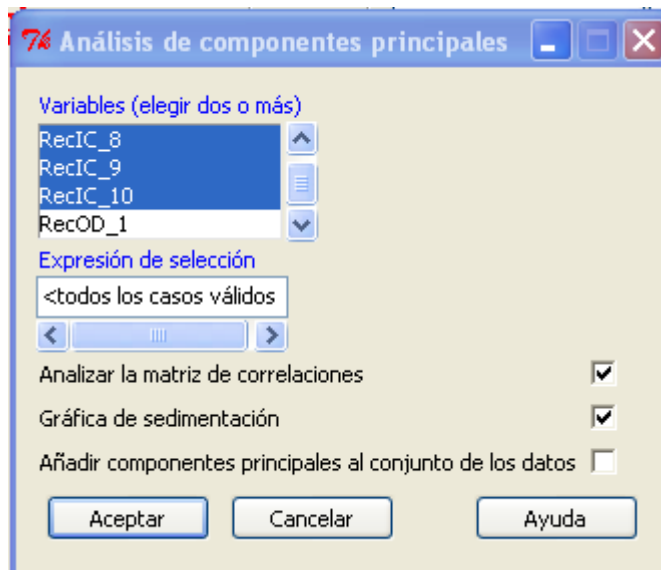
Conceptos básicos y aplicaciones prácticas con Rcommander

el investigador está interesado en la estimación de las puntuaciones obtenidas por los sujetos en los factores extraídos. La estimación de las puntuaciones factoriales no es sencilla. En la exposición del modelo del factor común, las variables observadas eran explicadas por medio de factores, pero en el caso que nos ocupa tendríamos que invertir la relación. Por ello sería incorrecto hablar de la estimación de las puntuaciones factoriales, en tanto en cuanto, éstas no son parámetros sino que son variables aleatorias, y el problema por tanto debería de centrarse en su predicción.

Rcommander incorpora a través de la función `factanal` la posibilidad de pronosticar las puntuaciones factoriales utilizando el método de Bartlett y el método de la regresión. El primero sugirió el uso de mínimos cuadrados ponderados para regresar las observaciones sobre los pesos ya ajustados tratando las unicidades como términos aleatorios y las puntuaciones factoriales como parámetros. En el método de la regresión, los factores son tratados como variables aleatorias.

7.3.5 Análisis de componentes principales

El objetivo del análisis de componentes principales es describir la variación encontrada en un conjunto de datos correlacionados en términos de un nuevo conjunto de variables no correlacionadas cada una de las cuales es una combinación lineal de las primeras. Las nuevas variables son derivadas en un orden decreciente de importancia en el sentido de varianza explicada. Estas nuevas variables se conocen como Componentes Principales.

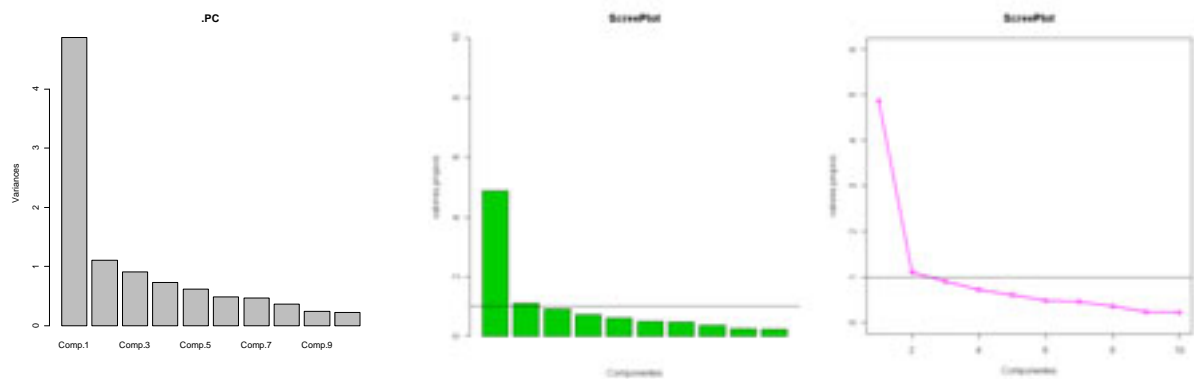


7.11 Figura. Análisis de Componentes Principales con Rcommander.

En un análisis de componentes principales interesa que los primeros componentes den cuenta de un porcentaje de varianza elevado. Con ello se conseguiría reducir la estructura de los datos a un pequeño número de dimensiones. La función que incorpora Rcommander (PC) para llevar a cabo un análisis de componentes principales es accesible por medio de las opciones Estadísticos > Análisis Dimensional > Análisis de componentes principales. La ventana a la que da acceso esta opción permite analizar la matriz de correlaciones, representar la gráfica de sedimentación y añadir los valores de los sujetos en los componentes extraídos como variable/s en el conjunto de datos activo.

El *screep*lot (gráfico de sedimentación) de Cattell es uno de los procedimientos gráficos más utilizados como apoyo en la determinación del número de factores a retener en un análisis de dimensionalidad. El *screep*lot generado por Rcommander presenta un aspecto que es fácilmente modificable variando los parámetros gráficos que utiliza la función *screep*lot.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



7.12 Figura. Screeplot bruto y screeplots modificados

Para generar un *screeplot*, independientemente del análisis de componentes principales que utiliza Rcommander se podría transcribir cualquiera de las siguientes dos funciones en la ventana de instrucciones. La primera función (`barplot`) genera el gráfico que muestra los valores propios de la matriz de correlaciones entre los ítems. Al gráfico generado se le ha añadido una línea que marca el punto de corte establecido por Kaiser-Guttman en la consideración de los factores a retener; aquellos con valores propios superiores a la unidad.

```
barplot(eigen(cor(Edi.data[,16:25]))$values,col=3,ylim=c(0,10),y  
lab="valores propios",xlab="Componentes", main="ScreePlot")  
abline(1,0)
```

Si en lugar de barras se quisieran representar los valores propios de los factores extraídos por medio de puntos (como lo hace el SPSS) el comando a utilizar sería el siguiente,

```
a <- eigen(cor(Edi.data[,16:25]))$values  
plot(a, col=6, pch=9, ylim=c(0,6),ylab="valores  
propios",xlab="Componentes", main="ScreePlot")  
lines(a, lwd=2, col=6)  
abline(1,0)
```

8 Funcionamiento diferencial del ítem

El sesgo es una de las mayores amenazas contra la validez de los tests porque su presencia puede alterar el significado de las puntuaciones de un modo sistemático y puede interferir en las inferencias derivadas de ellas. Su estudio es importante y para ello la psicometría ha diseñado métodos empíricos basados en el concepto de funcionamiento diferencial del ítem. Sesgo y funcionamiento diferencial del ítem no son conceptos sinónimos ni conceptos equivalentes; si bien es cierto que en los primeros estadios de los estudios sobre FDI era habitual referirse a ellos como estudios de sesgo. El funcionamiento diferencial es un característica formal de cuyo resultado se podría inferir (o no) la presencia de sesgo. Un estudio de funcionamiento diferencial que se limite a su detección en un conjunto de ítems no es válido en sí mismo. El funcionamiento diferencial es

Conceptos básicos y aplicaciones prácticas con Rcommander

una herramienta de trabajo que permitirá, en función de los resultados obtenidos, adoptar decisiones que afecten, bien a la prueba que se está construyendo o adaptando (Elosua, 2006; Elosua, López y Torres, 1999; Elosua, López y Egaña, 2000), bien a su utilización en poblaciones determinadas (López y Elosua, 2002).

8.1 Conceptos básicos

Un ítem presenta funcionamiento diferencial cuando la probabilidad de respuesta correcta (o la probabilidad de respuesta sobre una categoría) no depende únicamente del nivel de la persona en el rasgo intencionadamente medido por el test. Es decir, el nivel en la variable medida no explica la probabilidad de respuesta. Existen factores que ejercen su influencia sobre la probabilidad de respuesta que interactúan entre ciertas características del ítem y un grupo demográfico determinado originando una alteración entre la relación original establecida entre el ítem y la variable medida. El funcionamiento diferencial del ítem no es una propiedad del ítem, tampoco lo es de una población concreta de examinados; la utilización de ese ítem en ese grupo de población es el origen de una interacción que causa la presencia de funcionamiento diferencial del ítem. ¡Los ítems no son sesgados en modo absoluto! El sesgo está relacionado con el uso.

Mellenbergh (1989) sintetiza perfectamente la idea en una definición formal según la cual, dada una variable Z (constructo o variable medida), y con respecto a otra variable G (grupo), el ítem i presenta funcionamiento diferencial si y sólo si,

$$f(X|g, z) \neq f(X|z)$$

Conceptos básicos y aplicaciones prácticas con Rcommander

para todos los valores g y z de las variables G y Z . Donde, $f(X|g,z)$, representa la distribución de las respuestas al ítem dada g , y, $f(X|z)$, representa la distribución de respuestas dada z .

Las respuestas dadas al ítem habrían de depender únicamente de los valores z que puede adoptar la variable medida (Z), y no de una combinación de valores z y g de las variables que representan el objeto medido (Z) y la pertenencia a determinado grupo (G).

En todo estudio de FDI se trabaja simultáneamente con tres variables; ítem, variable medida y grupo.

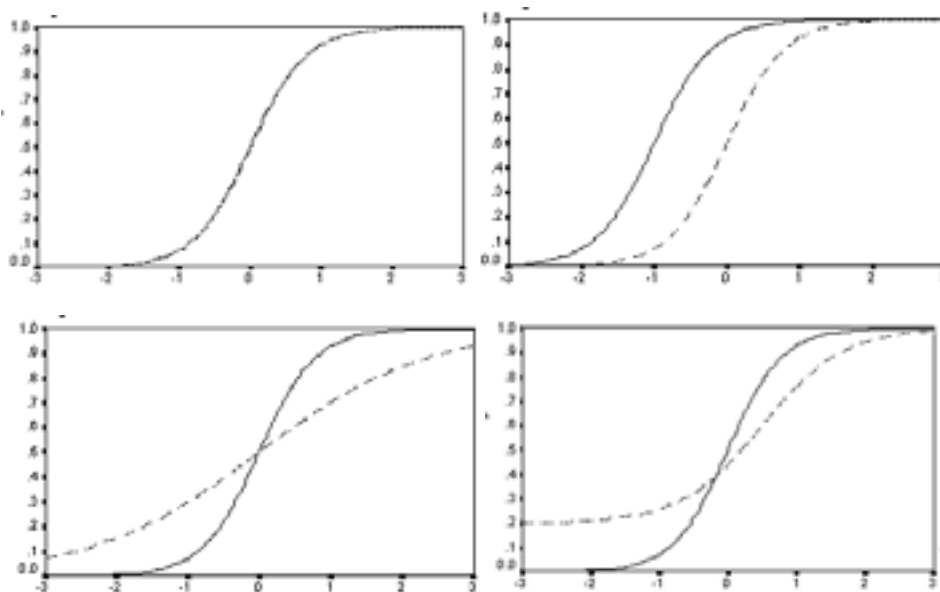
- *Ítem* es la variable observada objeto de estudio; su naturaleza puede ser dicotómica o politómica.
- *Grupo* es una variable de naturaleza nominal que puede definirse en función de distintos criterios (edad, sexo, idioma, cultura, raza, experiencia curricular...). Habitualmente los grupos se identifican como grupo de referencia y grupo focal (Holland y Thayer, 1988). El grupo de referencia, también denominado grupo mayoritario o principal es tomado como grupo estándar sobre el que se realizan las comparaciones. El grupo focal o grupo minoritario es el grupo de interés. Normalmente es el que hipotéticamente está en desventaja.
- *Variable medida* Se refiere al objeto de medida (puntuaciones observadas o variables latentes) que se supone es de naturaleza continua. El estudio del FDI exige comparar las respuestas dadas a un ítem por sujetos que proviene de distintos grupos pero que presentan el mismo nivel en la variable medida.

El funcionamiento diferencial del ítem supone la existencia de diferencias entre grupos en el modo de operar sobre el ítem; en función de estas diferencias se habla de FDI uniforme, FDI no uniforme y FDI mixto.

Conceptos básicos y aplicaciones prácticas con Rcommander

- *FDI uniforme*: Un ítem presenta FD uniforme, cuando el sentido del mismo se mantiene constante a lo largo de todo el continuo de habilidad. Es decir, las diferencias de funcionamiento detectadas entre los dos grupos de sujetos son constantes. Haciendo uso de la terminología aportada por la TRI, diríamos que las curvas características que definen el ítem en las dos poblaciones, son paralelas.
- *FDI no uniforme*: Las diferencias percibidas entre las respuestas del grupo focal y del grupo referencia no se mantienen constantes a lo largo del continuo de respuesta. Las curvas características del ítem se cruzan en algún punto del continuo.
- *FDI mixto*: En esta situación las curvas características de los ítems se cruzan y además las dificultades de los ítems son diferentes a lo largo del continuo de habilidad. En este tipo de FDI el cruce entre curvas se da en alguno de los extremos del continuo de habilidad.

Las figuras siguientes representan gráficamente las curvas características de los ítems estimadas en situaciones de ausencia de funcionamiento diferencial, funcionamiento diferencial uniforme, funcionamiento diferencial no-uniforme y funcionamiento diferencial mixto.



8.1 Figura. Patrones de funcionamiento diferencial

Conceptos básicos y aplicaciones prácticas con Rcommander

La psicometría ha desarrollado varios procedimientos y métodos para la detección del funcionamiento diferencial del ítem; tras más de dos décadas de trabajo sobre este tema se dispone de métodos altamente eficaces, que pueden clasificarse en tres grandes bloques:

- **Modelos formales latentes.** Agrupamos en esta categoría los procedimientos de detección construidos sobre modelos formales que relacionan funcionalmente variables latentes y las respuestas al ítem. Los más importantes son los derivados del modelo del factor común y de los modelos de respuesta al ítem (TRI). Los primeros por medio del estudio de estructuras de medias y covarianzas (MACS) comparan modelos de medida anidados que van restringiendo (o liberando) parámetros en los grupos de interés (Elosua y Muñiz, 2010; Elosua, 2005). Los segundos han diseñado varios procedimientos que básicamente comparan curvas características del ítem, parámetros de los ítems o modelos anidados derivados de la aplicación de algún modelo de respuesta al ítem (Elosua y Wells, 2008). La característica principal de estos procedimientos es las fuertes restricciones sobre las que se asientan.
- **Modelos formales manifiestos.** La detección del FDI se basa en la aplicación de modelos formales menos restrictivos que los anteriores, y por lo tanto su aplicabilidad es mayor. Entre los más utilizados cabe destacar los conocidos como Mantel-Haenszel (y sus extensiones polinomiales y politómicas) y la regresión logística. El primero está basado en el análisis de tablas de contingencia multidimensionales, y el segundo en una comparación de modelos anidados de regresión logística.
- **Procedimientos numéricos.** Los procedimientos incluidos en este apartado no proponen modelos de medida ni latentes ni manifiestos; sin embargo, son muy útiles. Incluiríamos aquí los procedimientos gráficos de análisis de medias ponderadas, y los índices numéricos como la estandarización que a

Conceptos básicos y aplicaciones prácticas con Rcommander

veces son utilizados como medidas del tamaño del efecto de otros procedimientos.

Dados los objetivos de este volumen describiremos someramente la regresión logística y los procedimientos numéricos. Ambos son sencillos de implementar por parte de lectores no expertos, eficaces y fácilmente interpretables.

8.2 Estandarización

La estandarización, es una técnica situada dentro de los límites generales de la TRI pues puede considerarse una versión no paramétrica de la misma y que se asienta a medio camino entre ella y los procedimientos basados en el análisis de tablas de contingencia.

Básicamente, compara las medias en un ítem obtenidas por sujetos provenientes de dos grupos que han sido emparejados en función de los valores alcanzados en la variable medida. Para ello se segmenta la puntuación total en K niveles de puntuación y se estiman las medias aritméticas en cada uno de los niveles y en cada uno de los grupos.

Con ello conseguimos comparar *las curvas características empíricas del ítem* o *regresiones ítem-test no paramétricas* estimadas en dos poblaciones distintas. Para cada uno de los niveles de habilidad la diferencia entre las medias aritméticas del ítem reflejaría las diferencias medias entre el grupo de referencia y el focal:

$$Dif_{iK} = \bar{X}_{iFK} - \bar{X}_{iRK}$$

Donde \bar{X}_{iFK} y \bar{X}_{iRK} , son las medias condicionadas al nivel de aptitud k en los grupos focal y de referencia

Conceptos básicos y aplicaciones prácticas con Rcommander

A partir de este concepto básico y simple (diferencia en un nivel de habilidad entre las medias aritméticas de un ítem obtenido en dos grupos) es necesario definir un índice que tenga en cuenta las diferencias a lo largo del continuo de habilidad, y que además pondere éstas por algún criterio que “corrija” el efecto que grandes diferencias estimadas en grupos muy reducidos de sujetos tengan en el índice final. En la práctica, uno de los factores de ponderación (w) más utilizado es la proporción de sujetos pertenecientes al grupo focal situados en el nivel K . Entre los índices numéricos más conocidos destaca el siguiente (STDDIF):

$$STDDIF_i = \sum_{k=1}^k w_k [\bar{X}_{iFK} - \bar{X}_{iRK}] = \sum_{k=1}^k \frac{K_k [\bar{X}_{iFK} - \bar{X}_{iRK}]}{\sum_{k=1}^k K_k}$$

dónde $W_k = [K_k / \sum K_k]$ es el factor de ponderación .

Los valores de este índice se sitúan en el intervalo (-1, 1). Mientras un valor positivo indica que el ítem favorece el grupo focal, un valor negativo indica cierta “ventaja” del grupo de referencia. Estudios recientes aconsejan “vigilar” un ítem cuando el valor del índice supera el 5% del rango de las puntuaciones del ítem.

8.3 Regresión logística

Este método modela la probabilidad de respuesta a un ítem en función de la puntuación empírica obtenida en el test, de la pertenencia a un grupo y de la interacción entre ambos factores. Evalúa la presencia de FDI a través del estudio de la mejora en el ajuste que produce la incorporación sucesiva de los parámetros mencionados al modelo de regresión logística (Puntuación Total, Puntuación total+ Grupo, Puntuación total+Grupo+Puntuación total×Grupo). Para evaluar el FDI se comparan las razones de verosimilitud de los modelos anidados (total, total+grupo, total+grupo+interacción). El modelo base se construye únicamente respecto al parámetro de la variable que indica el nivel de habilidad (Total; Modelo 1). La existencia de FDI uniforme se concluiría cuando la diferencia entre el modelo base y el modelo que incluye el parámetro

Conceptos básicos y aplicaciones prácticas con Rcommander

de pertenencia al grupo (Total+Grupo; Modelo 2) es significativa. El FDI no uniforme compara este segundo modelo con el modelo que incluye el término de interacción (Total+Grupo+Interacción; Modelo 3). Este método además de un test de significación incluye una medida del efecto del FDI basada en la diferencias en las R^2 entre dos modelos (Gierl y McEwen, 1995). Esta medida, R^2 , representa la proporción de variación de las respuestas al ítem explicada por el modelo de regresión. Un ítem presenta FDI moderado cuando el estadístico Chi-cuadrado es significativo y cuando el incremento en R^2 se sitúa entre los valores 0,035 y 0,070. Un ítem presenta un FDI severo cuando además de la significación del Chi-cuadrado, la diferencia entre dos R^2 es superior o igual a 0,070.

Para las situaciones de respuestas politómicas, el modelo se extiende dando lugar a tres variaciones básicas que dependen de la definición de los logit: el modelo acumulativo, el modelo continuo y el modelo de categorías adyacentes siendo de todos ellos el más utilizado, el modelo acumulativo. En este modelo se compara la probabilidad de que la respuesta al ítem (Y) sea menor o igual que la opción de respuesta j , con la probabilidad de que la respuesta (Y) sea mayor a la opción de respuesta j .

Podríamos representar el modelo de regresión logística del siguiente modo:

$$P(u = 1) = \frac{e^z}{[1 + e^z]}$$

donde,

$$z = \beta_0 + \beta_1 X + \beta_2 g + \beta_3 (Xg)$$

La probabilidad de respuesta depende de tres fuentes de variación: grupo de pertenencia, nivel de aptitud, e interacción entre ambos. Siendo g la variable indicadora del grupo de pertenencia, puede adoptar dos valores; 1 si el sujeto pertenece al grupo 1 (referencia), y 0 si pertenece al grupo 2 (focal). X representa

Conceptos básicos y aplicaciones prácticas con Rcommander

el nivel en la variable medida y se equipara con la puntuación obtenida por los sujetos en la prueba. Por último, el término, Xg se refiere a la interacción entre las dos variables independientes. Un ítem presenta FDI uniforme si, $\beta_3 = 0$ y $\beta_2 \neq 0$, y por el contrario presenta FDI no uniforme si se cumple $\beta_3 \neq 0$, sea cual fuere el valor de β_2 .

En este modelo, la hipótesis nula para la detección del FDI queda definida como:

$$H_0: \beta_2 = 0 \quad \beta_3 = 0$$

Para su comprobación, se compara el ajuste de modelos especificados con distintos parámetros. El procedimiento más sencillo y rápido es evaluar el modelo base, aquel que incorpora únicamente el nivel de habilidad, con el modelo de interacción. En caso de que se optara por el modelo extendido, es posible estudiar el origen del funcionamiento diferencial estimando el modelo para el FDI uniforme. La significación estadística del parámetro incorporado, se evalúa a través de la razón de verosimilitud, que analiza el incremento en el ajuste producido por la incorporación de un nuevo parámetro. Este estadístico sigue una distribución Chi cuadrado con grados de libertad igual a la diferencia entre los grados de libertad entre los modelos (3-1=2).

8.3.1 Consideraciones sobre el funcionamiento diferencial del ítem

Los estudios de funcionamiento diferencial del ítem agrupan estudios sobre la detección del FDI, causas del FDI y consecuencias de la presencia del FDI. La literatura especializada ha dedicado mucho tiempo y esfuerzo en al diseño de procedimientos de detección (Fidalgo, 1996; Holland y Wainer, 1993) fruto de los cuales disponemos de todo un arsenal de métodos que detectan con un alto grado de eficacia la presencia de FDI. Aunque todavía van surgiendo nuevos procedimientos de detección que intentan pulir los anteriores es un área de estudio que en la psicometría puede considerarse prácticamente cerrada.

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Sin embargo los estudios de FDI no terminan con la detección; de poco sirve concluir que una escala presenta FDI sin no se da un paso más en el análisis de sus consecuencias o en el estudio de la significación formal y aplicada de sus consecuencias.

Causas. Los estudios de FDI aportan información sobre factores que influyen en las repuestas a los ítems que pueden incrementar el corpus sustantivo referido a la variable de interés, y mejorar de ese modo el proceso de construcción de tests. Por ejemplo, se ha demostrado que los ítems pueden presentar funcionamiento diferencial relacionado con el momento de administración de la prueba, experiencia curricular del alumno, idioma de aplicación o causas relacionadas con la adaptación lingüística (Elosua y López, 2008; López y Elosua, 2002).

Detección. Contaminación del criterio. Los procedimientos de detección del FDI corren el riesgo de utilizar como criterio de emparejamiento una variable contaminada, pues en su definición se incluyen ítems con funcionamiento diferencial. Para superar este problema de circularidad relacionado con la detección de FDI utilizando una variable contaminada por ítems con FDI, cada procedimiento ha diseñado pautas de *purificación* del criterio (Holland y Thayer, 1988). El objetivo del mismo, es liberar al criterio de los ítems con FD, como paso previo a la detección del sesgo. El proceso, resumido en tres fases seguiría el siguiente esquema:

1. Detección
2. Redefinición de la variable de emparejamiento excluyendo los ítems con FDI detectados en la fase 1.
3. Reestimación del FDI

Consecuencias: ¿Cuáles son las consecuencias de la presencia de FDI en un test? Si alguno de los grupos de interés está afectado negativamente por la

Conceptos básicos y aplicaciones prácticas con Rcommander

presencia de ítems que actúan en su contra, existen procedimientos que permitan equiparar las puntuaciones. En la práctica diaria el evaluador ha de analizar detenidamente cuales son las consecuencias prácticas del FDI, y en caso de que existiera un error sistemático ha de tomar medidas para neutralizarlo. ¿Cómo? Dependiendo del modelo formal que utilice podría optar por la estimación de puntuaciones factoriales que tengan en cuenta la existencia de modelos diferentes para cada grupo, o podría utilizar técnicas de equiparación de puntuaciones basadas en el concepto de Curva Característica del Test (Elosua y López-Jáuregui, 2008).

8.4 Funcionamiento diferencial del ítem y Rcommander

8.4.1 Regresión logística

Son varios los paquetes y funciones que incorporan la regresión logística binaria, ordinal y politómica en R. Entre ellas, utilizaremos la función `lrm` que viene incorporada en el paquete `Design`. El primer paso por tanto, será cargar (o instalar en su defecto) el paquete.

```
library(Design,T)
```

Una vez instalado, son dos los modelos de regresión que habremos de aplicar a los ítems objeto de estudio. El modelo base definido como aquél que explica las respuestas al ítem en función de la puntuación total, y el modelo de interacción que además de la puntuación total incorpora como parámetro la interacción entre el grupo (sexo) y la puntuación total (IC:SEXO). Estas dos sencillas instrucciones harán que el modelo de regresión logística para datos ordinales se ejecute sobre los datos.

```
fit0<-lrm (RecIC_1~IC, data=Edi.data)$stats  
fit2<-lrm (RecIC_1~IC+SEXO+IC:SEXO, data=Edi.data)$stats
```

Conceptos básicos y aplicaciones prácticas con Rcommander

Los resultados de la aplicación de ambos modelos (`$stats`) han sido asignados a dos objetos `fit0` u `fit1`. En ellos se almacena información sobre el ajuste del modelo de regresión; entre los estadísticos que se incluyen nos interesan especialmente el valor de la razón de verosimilitud, guardada en la tercera posición (`[3]`) y el valor del R^2 que ocupa la décima posición en la lista de resultados.

Dado que el modelo de regresión logística se basa en la comparación de modelos anidados solo quedará estimar la diferencia entre esos estadísticos para obtener información sobre el posible funcionamiento diferencial del ítem. Para ello se han creado dos nuevas variables (`LR.DIF` y `R2.DIF`) que recogen la diferencia entre los valores estimados en el modelo base y en el modelo de interacción.

```
LR.DIF <- fit2[3]- fit0[3]
R2.DIF <- fit2[10]-fit0[10]
```

El valor estimado de `LR.DIF` es 15,48555 y el de `R2.DIF` 0.008411846. El valor `LR.DIF` sigue una distribución Chi cuadrado con dos grados de libertad (diferencia entre los grados de libertad del modelo de interacción y los grados de libertad del modelo base). La probabilidad asociada a este valor ($\chi^2_{(15,48; 2)}=0,00045$) es muy baja lo cual nos lleva a rechazar la hipótesis nula de que el incremento en el número de parámetros del modelo (Grupo, Grupo: Puntuación) no mejora el ajuste del modelo base, y deberíamos de aceptar el modelo de la interacción. Sin embargo, no es una buena práctica llegar a conclusiones basadas en criterios de decisión estadística construidos sobre el valor Chi cuadrado. Es necesario incorporar información sobre el tamaño del efecto, que en el contexto de la Regresión Logística es aportada por la diferencia en los coeficientes de determinación. En este caso concreto el incremento en este índice provocado por la incorporación de parámetro en el modelo fue, 0,008. Este tamaño del efecto no supera el punto de corte establecido en la literatura

Conceptos básicos y aplicaciones prácticas con Rcommander

psicométrica para la categorización del funcionamiento diferencial de un ítem como severo, que suele fijarse en el punto de 0,07.

La tabla siguiente muestra el análisis efectuado sobre el total de los ítems de la escala Insatisfacción Corporal. Aunque la significatividad de algunas de las diferencias entre razones de verosimilitud pudiera inducir a concluir la presencia de funcionamiento diferencial del ítem, los valores estimados para la diferencia entre los coeficientes de determinación no apoyan esa conclusión.

	Regresión logística			
Creo que mi estómago es demasiado grande	15,49	<.001	0,008	-0,10
Creo que mis muslos son demasiado gruesos	27,26	<.001	0,012	0,26
Creo que mi estómago tiene el tamaño adecuado	39,25	<.001	0,019	0,05
Estoy satisfecho con mi cuerpo	6,16	<.001	0,002	0,02
Me gusta mi trasero	4,51	0,10	0,002	0,12
Creo que mis caderas son demasiado anchas	2,55	0,28	0,001	0,09
Me siento hinchada después de una comida normal	1,19	0,55	0,001	-0,09
Creo que el tamaño de mis muslos es adecuado	20,51	<.001	0,008	0,25
Creo que mi trasero es demasiado grande	1,26	0,53	0,001	0,00
Creo que mis caderas tienen el tamaño adecuado	4,95	0,08	0,002	0,05

8.1. Tabla. Resultados del funcionamiento diferencial del ítem

8.4.2 Diferencias entre medias ponderadas

Uno de los acercamientos más interesantes y prometedores en el análisis y estudio del funcionamiento diferencial (al igual que en multitud de análisis) descansa sobre la descripción gráfica de los datos. Un gráfico que plasme las diferencias condicionadas al nivel de habilidad de las medias de respuesta en función de los grupos ofrece información fácil de entender, fácil de comunicar y fácil de interpretar. Este tipo de representación ofrece información sobre la magnitud y sentido de las diferencias existentes entre dos grupos a lo largo de todo el continuo de habilidad de la variable medida. En el contexto de los

Conceptos básicos y aplicaciones prácticas con Rcommander

modelos de respuesta al ítem esa información es aportada por las curvas características del ítem, que podrían reemplazarse por curvas empíricas como las mostradas en la siguiente figura para cada uno de los ítems de la escala de Insatisfacción Corporal.

Su obtención es sencilla con R (o Rcommander); basta segmentar en varios niveles la puntuación total y para cada uno de ellos estimar la mediana aritmética de cada uno de los ítems para cada uno de los grupos (la información aportada en el último capítulo será suficiente para poder ejecutar estas acciones).

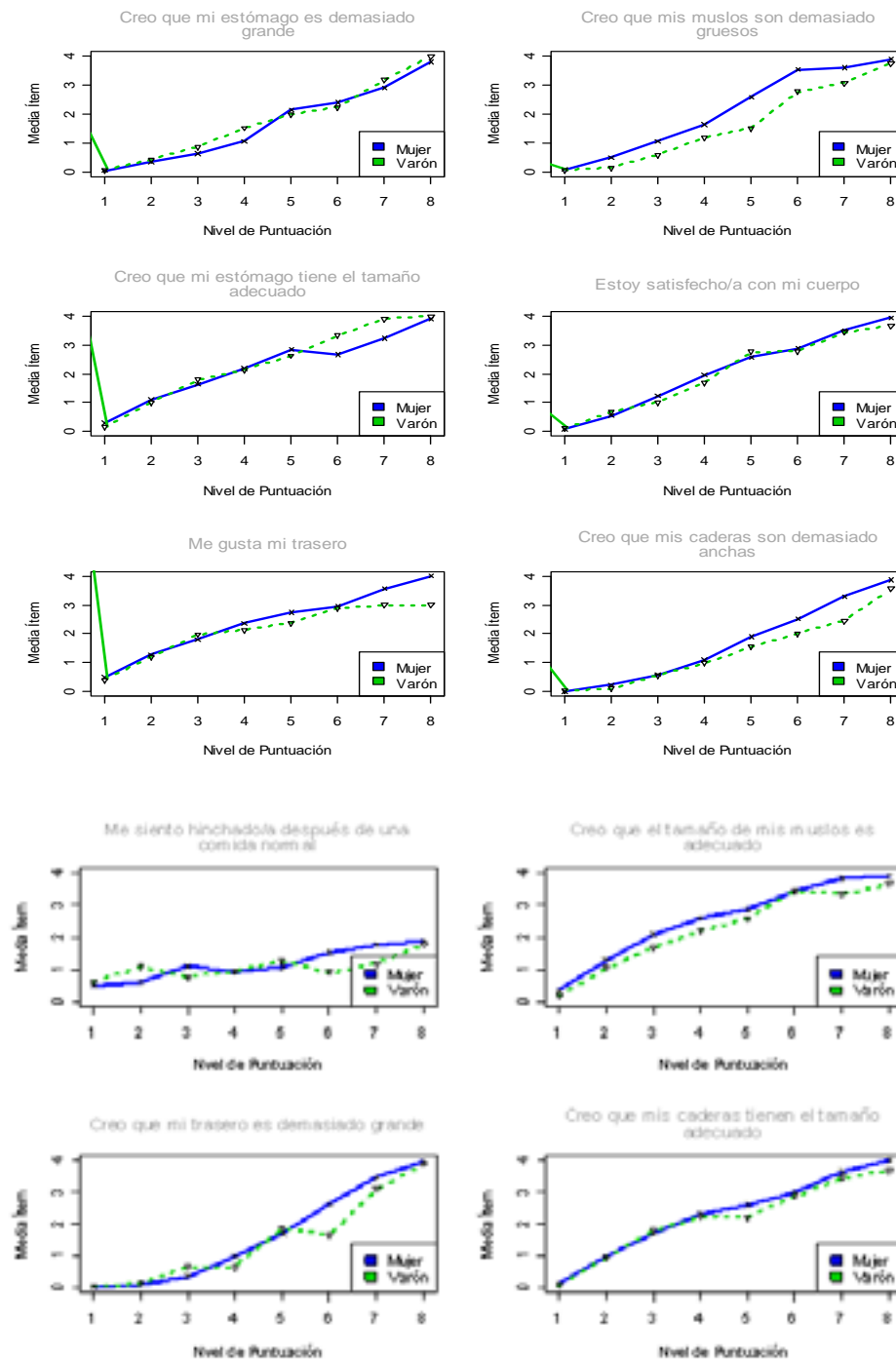
El análisis gráfico de las curvas corrobora la información ofrecida por el modelo de regresión logística ordinal. Las curvas de medias condicionadas prácticamente se superponen en el grupo de varones y en el grupo de mujeres analizados. Las mayores divergencias se producen en el segundo ítem de la escala (“Creo que mis muslos son demasiado gruesos”). La diferencia entre varones y mujeres es constante a lo largo del continuo de habilidad. Es un ítem a cuyo contenido habría que prestar atención; el cuestionario que estamos analizando fue construido para ser usado en poblaciones clínicas de mujeres con trastornos de la conducta alimentaria. La muestra de estudio actual proviene de un grupo no-clínico formado por mujeres y varones. Dadas las diferencias existentes entre la percepción corporal e insatisfacción corporal entre mujeres y hombres este tipo de resultado hace reflexionar sobre el contenido de los ítems en la evaluación de la insatisfacción corporal.

La tercera curva también refleja divergencias, sin embargo las divergencias entre curvas que aparecen en los gráficos se sitúan en los valores extremos del continuo; para la correcta interpretación de estas diferencias es menester considerar que en los niveles de puntuación extremos existen muy pocos sujetos y que por tanto es necesaria la cautela en las conclusiones que derivaremos de ellas.

Conceptos básicos y aplicaciones prácticas con Rcommander

La representación gráfica de las medias condicionadas puede complementarse con un índice numérico que además del tamaño de las diferencias tenga en cuenta la necesidad de ponderación de las mismas en función del número de sujetos situado en cada nivel de habilidad. No es lo mismo una diferencia obtenida en un nivel que agrupa a muchos individuos que una diferencia extrema en cuyo cómputo intervienen muy pocos sujetos. Este planteamiento es el origen de la fórmula de la estandarización cuya obtención en Rcommander sería tan sencilla como la elaboración de los gráficos de medias ponderadas. En nuestro caso y utilizando como elemento ponderador el número de varones en cada uno de los niveles de la puntuación total en Insatisfacción Corporal se obtuvieron los índices que se leen en la tabla de resultados mostrada. Los resultados numéricos respaldan en todo momento los resultados gráficos y existe acuerdo entre los dos procedimientos numéricos empleados.

Conceptos básicos y aplicaciones prácticas con Rcommander



8.2 Figura. Curvas empíricas de funcionamiento diferencial

9 Comparación entre grupos. (Alicia López-Jáuregui)

El estudio de las relaciones entre las puntuaciones y variables externas al test es una fuente de evidencia importante en el estudio de la validez. EL concepto de variables externas incluye medidas de algún criterio que se espera prediga el test, relaciones con otros tests que miden el mismo constructo, constructos relacionados o constructos diferentes. También es posible analizar las relaciones entre las puntuaciones y variables categóricas como la pertenencia a un grupo que aportaría información relevante cuando el marco teórico sobre el constructo medido hipotetiza o sugiere la presencia/ausencia de diferencias en función de variables definitorias de grupo (edad, sexo, experiencia curricular...). Este tipo de estudios aportan evidencia sobre el grado en que las relaciones observadas son consistentes con el constructo objeto de medida. Por ejemplo, y ajustándonos a los datos derivados del Eating Disorder Inventory-3, habría que analizar su capacidad discriminativa en relación a la evaluación de trastornos alimentarios; en este sentido se hipotetizaría que aquellas personas con un diagnóstico relacionado con un trastorno de la conducta alimentaria obtendrían puntuaciones significativamente superiores a aquellas personas pertenecientes a un grupo no-clínico. Este tipo de estudio podría acometerse utilizando modelos de análisis discriminativo o regresión logística, sin embargo, en este manual nos centraremos en instrumentos más sencillos para la comparación de grupos.

El objetivo es responder a cuestiones como, ¿es mayor la proporción de varones en riesgo de desarrollar un trastorno de la conducta alimentaria que la proporción de mujeres?. Puede afirmarse que existe algún tipo de relación entre la baja autoestima y los trastornos de la conducta alimentaria? En el estudio de estas cuestiones de investigación el investigador no está únicamente interesado en describir sus datos (muestra) el objetivo es la generalización de resultados (inferencia) más allá de la muestra de estudio. El investigador busca extender las conclusiones derivadas de su estudio a la población de la que la muestra no es más que una representación. El “salto” de la muestra a la población se apoya

Conceptos básicos y aplicaciones prácticas con Rcommander

en la estadística inferencial que ha desarrollado multitud de técnicas con ese fin. La utilización de una u otra dependerá de la cuestión de investigación y de la naturaleza de los datos a analizar. La selección de la prueba más adecuada a cada caso es una decisión del investigador que `Rcommander` ejecutará.

9.1 Comparación de dos grupos. Pruebas de comparación de medias

Las pruebas de este epígrafe son aplicables a variables métricas y obviamente están destinadas a valorar cuestiones que involucran a las medias. Nos podemos plantear dos tipos de cuestiones distintas; ¿Se puede mantener, una vez observados los datos, que la media de la población en la escala de insatisfacción corporal tiene un valor determinado digamos “X”? ¿Existen diferencias entre dos grupos (por ejemplo chicos y chicas) en insatisfacción corporal?

Por ejemplo, supongamos que hemos aplicado *Eating Disorder Inventory-3* a un grupo de bailarinas de ballet y que se conoce la media en la escala de bulimia en la población general. Una hipótesis a demostrar podría ser que este colectivo presenta una media en la escala de bulimia más alta que el resto de la población. Alguien podría pensar que es suficiente con calcular la media de nuestros datos y compararla con el valor de μ (media poblacional). Ahora bien, la conclusión obtenida con esta comparación sólo sería aplicable a esa muestra de sujetos pero el investigador lógicamente pretende la generalización. La prueba estadística pertinente es la aplicación de la *prueba de comparación de una media observada (el valor de la media de la muestra) a una media teórica (el valor conocido de la media en la población general)*.

El segundo tipo de cuestión tiene que ver con la comparación de medias observadas en dos grupos o muestras distintas. Siguiendo con la aplicación del *Eating Disorder Inventory-3* podríamos preguntarnos por la diferencia en la escala de satisfacción corporal entre chicos y chicas. Podríamos sostener la hipótesis de que estas últimas están más insatisfechas con su cuerpo que

Conceptos básicos y aplicaciones prácticas con Rcommander

aquellos. A la vista de las medias observadas en los dos grupos observados, ¿Podríamos generalizar los resultados a la población y dar por demostrada la hipótesis? La prueba adecuada sería la de comparación de dos medias.

9.1.1 Prueba para una media

Continuando con el ejemplo de las bailarinas de ballet, supongamos que nuestro conjunto de datos procede de ese colectivo (976 chicas) a las que se ha administrado *Eating Disorder Inventory*, y que la media en la escala de bulimia en la población general (μ) es 3,30. Nos podríamos preguntar si el colectivo de bailarinas de ballet puntúa más alto en la escala de bulimia que el conjunto de la población:

La hipótesis nula sería:

$$H_0: \mu=3,30$$

y la hipótesis alterna,

$$H_1: \mu>3,30$$

La media en la escala de bulimia de las bailarinas (\bar{x}) es 3,72; Este valor es más alto que el poblacional. La diferencia entre el grupo de bailarinas (3,72) y la población (3,30) es 0,42. Ahora bien, ¿podemos considerar estadísticamente significativa esta diferencia? Para saberlo necesitamos aplicar la prueba de “t” de una muestra.

La expresión que permite obtener el estadístico t de significación es:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

Donde, \bar{x} media muestral

μ media poblacional

s desviación típica muestral

N tamaño muestral

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Este estadístico se distribuye según la distribución t de Student con $N-1$ grados de libertad. La magnitud de este estadístico tiene que ver con la diferencia entre la media observada o empírica (\bar{x}) y el valor poblacional (μ). Cuánto mayor sea esta diferencia mayor será el valor de t . El denominador de la expresión es lo que se denomina *error estándar de la media muestral* (Standard Error, SE), y conceptualmente podemos concebirlo como una medida del “ruido” o variación aleatoria muestral. A mayor error estándar mayor variación aleatoria entre las diferentes medias muestrales posibles. Si esta variación o ruido es grande, necesitaremos más “señal” (valor del numerador) para que ésta sea destacable.

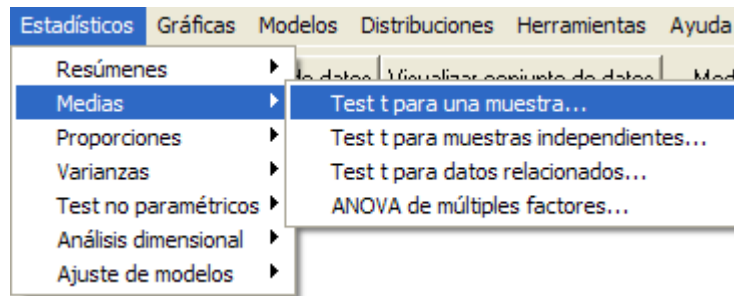
A partir del valor de t se obtendrá el “grado de significación” que no es sino una probabilidad aparejada a t . El grado de significación (p) evalúa la probabilidad de obtener un valor muestral como el observado, si se hubiera extraído efectivamente de una población con un valor 3,30, es decir evalúa en cierto modo el grado de compatibilidad de nuestros datos con la hipótesis de partida (Pardo y San Martín, 1998).

El valor “ p ” es una probabilidad y por lo tanto su valor estará entre 0 y 1. Si el grado de significación es pequeño, (en nuestro ejemplo es 0,003) es pequeña la probabilidad de haber obtenido un valor de 3,72 de una población con una media de 3,30. Por tanto optamos por rechazar este último valor como hipótesis plausible a la luz de nuestra evidencia y concluimos que, en efecto, el grupo de bailarinas es significativamente superior en la escala de bulimia a la población general.

Ahora procedamos a contestar a la cuestión con la ayuda de RCommander .

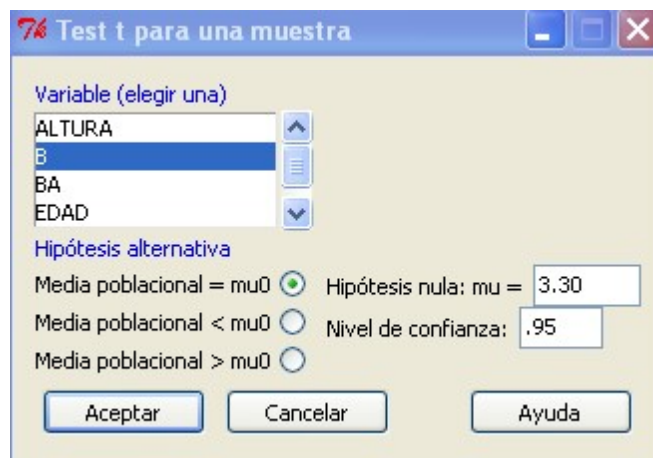
La opción del menú que nos interesa es Estadísticos>Medias>test t para una muestra.

Conceptos básicos y aplicaciones prácticas con Rcommander



9.1 Figura. Prueba de comparación de una media observada a una teórica. Rcommander

En la ventana emergente que se abre con esta opción del menú seleccionaremos la variable de interés (B es el nombre adjudicado a la variable “Bulimia”). Debemos escribir el valor hipotético (3,30) y seleccionaremos el tipo elegido de hipótesis alterna. También podemos especificar el nivel de confianza elegido (1- α). En Rcommander el valor por defecto es 0,95. El nivel de confianza servirá para el cálculo del correspondiente intervalo de confianza para la media. El valor de t obtenido con Rcommander 2,0319 y el número de grados de libertad (degrees of freedom, df) son 901 ($n-1= 902-1 = 901$; N es el número de casos válidos en la variable bulimia), y el grado de significación (p -value) es 0,04245. Por último aparece la conclusión de la prueba de decisión estadística: “true mean is not equal to 3.30”; Esto es, se rechaza la hipótesis nula y en consecuencia se acepta la hipótesis alterna.



Conceptos básicos y aplicaciones prácticas con Rcommander

9.2 Figura. Selección de variable y opciones para la prueba de comparación de una media observada a una teórica.

```
> t.test(Edi.data$B, alternative='two.sided', mu=3.30,
conf.level=.95)

One Sample t-test

data: Edi.data$B
t = 2.0319, df = 901, p-value = 0.04245
alternative hypothesis: true mean is not equal to 3.3
95 percent confidence interval:
 3.310151 3.884971
sample estimates:
mean of x
 3.597561
```

En la salida de Rcommander aparece también el intervalo de confianza. El límite inferior es aquí 3,310151 y el límite superior 3,884971. ¿Qué quiere decir esto? A la luz de los datos, la media de la población de la que proceden los mismos (μ) estará comprendida, con un nivel de confianza del 95%, entre estos dos valores. Nuestra hipótesis para μ era 3,30, y este valor queda fuera de éstos límites por lo que no es un valor asumible para la media poblacional.

9.1.2 Comparación de dos medias

Veamos ahora cómo contestar a la pregunta de si hay relación entre el sexo y la insatisfacción corporal, más específicamente, si esta última es mayor en las mujeres que en los varones. Las medias en la escala de insatisfacción corporal de chicos y chicas (Estadísticos >Resúmenes >Tablas de estadísticas) son $\bar{x}_e = 15,71$ y $\bar{x}_g = 8,09$. La diferencia es por tanto 7,62.

Conceptos básicos y aplicaciones prácticas con Rcommander

Ahora bien, la cuestión que nos interesa no es la diferencia entre los hombres y las mujeres en nuestra *muestra* concreta; queremos saber si existen diferencias entre las *poblaciones* de mujeres y hombres y por ello necesitaremos aplicar una prueba de significación estadística.

El test clásico para la comparación entre dos medias es la prueba *t* de Student. Esta prueba, una de las más clásicas en Estadística, no funciona del todo bien cuando se comparan grupos con varianzas diferentes (sobre todo si los tamaños de los grupos son muy distintos, como de hecho ocurre en el ejemplo). Para estos casos se ha propuesto una modificación de la misma; la prueba de Welch. (Pardo y San Martín, 1998)

En el caso de dos muestras, con varianzas poblacionales distintas, el estadístico *t* se calcula mediante:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s^{*2}}{N_1} + \frac{s^{*2}}{N_2} \right)}}$$

Donde, \bar{x}_1 la media de la primera muestra

\bar{x}_2 media de la segunda muestra y

$$s^{*2} = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

S_1^2 varianza de la primera muestra

S_2^2 varianza de la segunda muestra

N_1 número de sujetos de la primera muestra

N_2 número de sujetos de la segunda muestra

En el numerador tenemos de nuevo una diferencia, en este caso entre dos muestras observadas o muestrales, y el denominador es una estimación del error de muestreo de la diferencia muestral.

Los grados de libertad (gl) del estadístico estarán comprendidos entre N_1-1 y N_2-1 y su valor viene dado por la fórmula de Welch (Welch, 1938):

$$gl = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)}{\frac{1}{N_1 - 1} \left(\frac{s_1^2}{N_1} \right)^2 + \frac{1}{N_2 - 1} \left(\frac{s_2^2}{N_2} \right)^2}$$

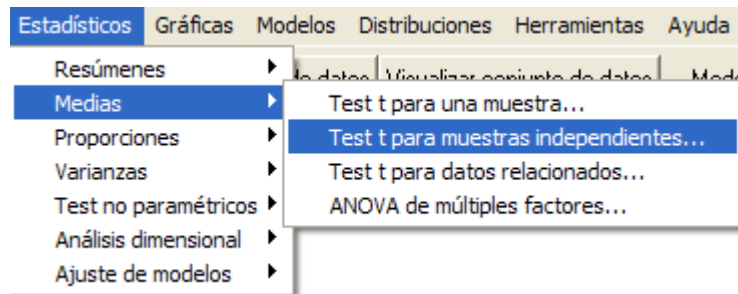
Donde, S_1^2 varianza de la primera muestra

S_2^2 varianza de la segunda muestra

N_1 número de sujetos de la primera muestra y

N_2 número de sujetos de la segunda muestra.

A partir del valor de t se obtendrá el “grado de significación”, es decir la probabilidad asociada con el valor t . El grado de significación (p) evalúa la probabilidad de la diferencia observada en los datos en el caso de que la hipótesis nula fuese cierta, es decir, si efectivamente las dos poblaciones de hombres y mujeres tuvieran medias idénticas en la escala de insatisfacción corporal ($\mu_1 = \mu_2$). Con la ayuda de Rcommander obtendremos los análisis:



9.3 Figura. Prueba de comparación de dos muestras independientes

En la ventana emergente seleccionaremos la variable que define los grupos (sexo) y la variable cuyas medias queremos comparar (insatisfacción corporal; IC). Elegiremos también el tipo de hipótesis alternativa y el nivel de confianza, así como la suposición acerca de la igualdad de varianzas en los grupos. En este ejemplo seleccionamos “Diferencia>0” y dejamos las últimas dos opciones por defecto de Rcommander; esto es, nivel de confianza 0,95 y grupos con varianzas distintas.



9.4 Figura. Selección de variables y opciones para la prueba de comparación de dos muestras independientes

```
> t.test(IC~SEXO, alternative='greater', conf.level=.95,
var.equal=FALSE,
+ data=Edi.data)
Welch Two Sample t-test
data: IC by SEXO
t = 12.2259, df = 905.177, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 6.59255      Inf
sample estimates:
mean in group chica mean in group chico
      15.714286          8.095694
```

En la salida de Rcommander se recogen el valor del estadístico t (12,2259), los grados de libertad (905,177) y el grado de significación (p -value = $2,2e^{-16}$).

¹ Notación esponencial: $2.2e^{-16} = 2.2 * 10^{-16} = 0.000000000000000022$

Conceptos básicos y aplicaciones prácticas con Rcommander

Aparecen también la hipótesis alterna elegida (`true difference in means is not equal to 0`) y el intervalo de confianza. El valor del límite inferior es 6,59255 y el límite superior infinito. Esto quiere decir que la diferencia de medias poblacionales en la variable insatisfacción corporal será mayor que 6,59255 con una probabilidad del 95%. La diferencia entre las medias ha resultado significativa. Hemos encontrado evidencia empírica que apoya la hipótesis del investigador; las mujeres muestran más insatisfacción con su cuerpo que los varones.

Varianzas iguales en los dos grupos

Si las varianzas son iguales en los dos grupos, podremos utilizar la “versión clásica” de la prueba *t* de Student. (Pardo y San Martín, 1998)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s^{*2}}{N_1} + \frac{s^{*2}}{N_2}\right)}}$$

Donde, \bar{x}_1 es la media de la primera muestra

\bar{x}_2 la media de la segunda muestra y

$$s^{*2} = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

S_1^2 es la varianza de la primera muestra y

S_2^2 la varianza de la segunda muestra

N_1 número de sujetos de la primera muestra y

N_2 número de sujetos de la segunda muestra.

Los grados de libertad de este estadístico son $N_1 + N_2 - 2$. El procedimiento de obtención del análisis en Rcommander es idéntico al explicado anteriormente.

9.1.2.a Dos medias. Datos relacionados

Conceptos básicos y aplicaciones prácticas con Rcommander

En algunos casos las medias proceden de dos conjuntos de datos referidos a sujetos emparejados; los sujetos han sido medidos en ocasiones distintas (test/postest), o los sujetos han sido emparejados por alguna variable de interés introducida por el investigador a fin de mejorar su diseño. Estos casos tienen en común desde el punto de vista estadístico que los conjuntos de datos a comparar están correlacionados (un sujeto se parece a sí mismo). En estos casos la estrategia a utilizar es diferente, y la prueba estadística es muy similar a la prueba para una media. En concreto se calcularán las diferencias entre los pares, y la media correspondiente a esas diferencias \bar{x}_d , que se contrastará con la hipótesis nula $\mu_d = 0$

El estadístico t es análogo al utilizado para el contraste de una media;

$$t = \frac{\bar{x}_d - \mu_d}{s_d / \sqrt{N}}$$

Donde \bar{x}_d es la media de las diferencias entre los dos grupos,

μ_d la media de diferencias en la población y

s_d el error estándar de las diferencias entre los grupos, y

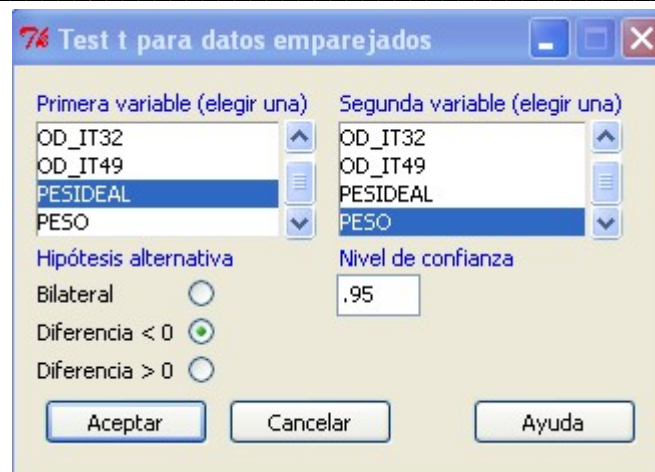
N , el número de diferencias

Los grados de libertad del estadístico son $N-1$.

Por ejemplo supongamos que deseamos indagar en la conformidad de la población con su peso, concretamente si el que se declara como “peso ideal” es inferior al peso real. Los sujetos de nuestra base de datos han contestado a las dos cuestiones por lo que podemos utilizar la prueba descrita.

Escogemos la opción adecuada en Rcommander, Estadísticos>Medias>Test t para datos relacionados, y en la ventana emergente debemos seleccionar las variables a comparar (PESO y PESIDEAL), el tipo de hipótesis alternativa y el nivel de confianza

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



9.5 Figura. Selección de variables y opciones para la prueba de comparación de datos relacionados.

```
Paired t-test

data: Edi.data$PESIDEAL and Edi.data$PESO
t = -8.8598, df = 404, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -2.138298
sample estimates:
mean of the differences
 -2.627160
```

La media observada de las diferencias entre las variables (mean of the differences) es -2,6271, y el valor t correspondiente a esta magnitud es -8,85; Cómo el número de diferencias es 405, los grados de libertad son 404 (df : $405-1=404$) y el grado de significación es extraordinariamente pequeño (p -value= $2,2e-16$). Por lo tanto, tenemos evidencia empírica para apoyar la hipótesis de que el peso ideal está por debajo del peso real. El intervalo de confianza del 95% indica que en la población el peso ideal estará 2,62 kg por debajo del real.

Conceptos básicos y aplicaciones prácticas con Rcommander

9.1.3 Pruebas para las varianzas

9.1.3.a Comparación de dos varianzas

La prueba más común de comparación de varianzas en dos grupos es la prueba F , aunque existen otras con el mismo propósito (por ejemplo la prueba de Lévene). Este test, el test F , se aplica en la mayoría de los casos como paso preliminar a la prueba t de Student de comparación de medias, con el objeto de verificar si las varianzas son iguales. El estadístico F se obtiene mediante la razón entre las varianzas observadas en las muestras.

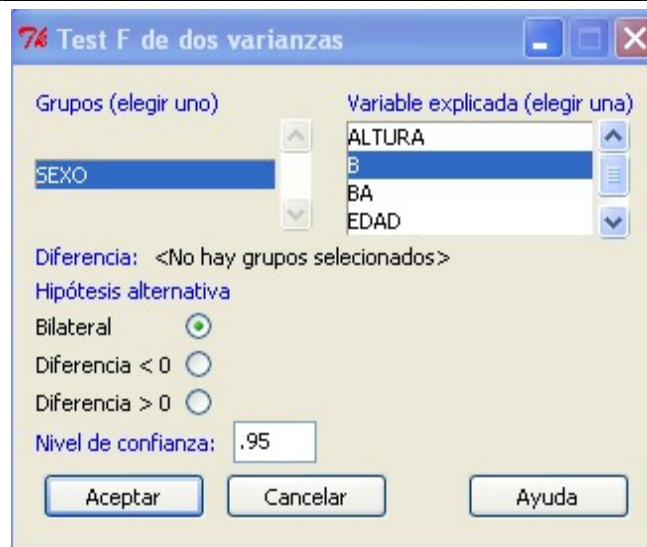
$$F = \frac{s_1^2}{s_2^2}$$

Donde S_1^2 es la varianza de la primera muestra, y
 S_2^2 la varianza de la segunda muestra.

La probabilidad asociada a F (grado de significación) se puede obtener de inmediato mediante la distribución “ F ” de probabilidad; los grados de libertad son N_1-1 y N_2-1 .

Para comprobar si la variable “Bulimia” tiene la misma variabilidad en las poblaciones de mujeres y varones utilizaremos la opción Estadísticos >Varianzas>Test F para dos varianzas. En la ventana emergente se elige la variable que define los grupos a comparar (SEXO) y la variable métrica cuyas varianzas medidas en ambos grupos se van a comparar (B “Bulimia”). Se marca la clase de hipótesis (por defecto “bilateral”) y el nivel de confianza (dejamos el valor 0,95).

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



9.6Figura. Selección de variables y opciones para la prueba de comparación de comparación de varianzas

```
> var.test(B ~ SEXO, alternative='two.sided', conf.level=.95,
data=Edi.data)

F test to compare two variances

data: B by SEXO
F = 0.8926, num df = 480, denom df = 420, p-value = 0.2284
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7411259 1.0738893
sample estimates:
ratio of variances
 0.8926434
```

Rcommander ofrece el valor de F (0,8926), los grados de libertad (480 y 420), el grado de significación (0,2284), el tipo de hipótesis alterna y el intervalo de confianza para la razón de varianzas σ_1^2/σ_2^2 . Por último se muestra la razón de varianzas observada en los datos (s_1^2/s_2^2); en nuestro caso 0,8926. En consecuencia no tenemos evidencia de que las varianzas sean diferentes en los dos grupos.

Conceptos básicos y aplicaciones prácticas con Rcommander

9.1.3.b Comparación de k varianzas

La prueba adecuada para comparar las varianzas de más de dos grupos es la prueba de Bartlett. Para ejecutarla en RCommander, la opción a elegir es Estadísticos>Varianzas> Test de Bartlett, y la salida (al igual que en casos anteriores) incluye el estadístico, los grados de libertad y el grado de significación.

9.2 Pruebas para las proporciones

El recuento de frecuencias o las proporciones son estadísticos adecuados para las variables categóricas (por ejemplo, sexo, trabajo, estado civil...) y de las variables numéricas categorizadas. Por ejemplo, a partir de las puntuaciones en la escala de insatisfacción corporal (IC) se han obtenido dos grupos de sujetos, los que tienen baja insatisfacción corporal y los que tienen alta insatisfacción corporal; esta categorización puede utilizarse como criterio clínico. Para trabajar con las variables categóricas se utilizan las tablas de contingencia.

En las filas figura la insatisfacción corporal, y en las columnas la edad. Tenemos dos categorías de la primera (SI/NO) y tres categorías para la edad (menores de 14 años, entre 14 y 16 años y mayores de 16 años). Para obtener dos grupos de la variable Insatisfacción Corporal se ha utilizado 18 como punto de corte. Este valor separa los dos grupos, 0 y 1. Cada combinación de valores para estas dos variables es una celda. Por ejemplo, la celda que corresponde a los sujetos menores de 14 años con insatisfacción corporal contiene el número 67. Esta tabla es 2 x 3; está compuesta por 2 filas y 3 columnas.

	Edad			
	>14 años	14-16 años	16 años<	
Insatisfacción corporal				Total

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

NO	236	269	146	651
SI	67	115	67	249
Total	303	384	213	900

9.1 Tabla. Tabla de doble entrada

Supongamos que estemos interesados en los sujetos menores de 14 años que están insatisfechos con su cuerpo. Su proporción es simplemente 67 dividido por el total de casos (900) o sea 0,074. Es decir 7,4% de estudiantes son menores de 14 años y están insatisfechos con su cuerpo. Para cada celda se obtiene una proporción, dividiendo la frecuencia por el tamaño total de la muestra. Todas estas proporciones configuran la distribución conjunta de las dos variables categóricas.

Distribuciones marginales

La distribución de una sola variable en una tabla de doble entrada, es la distribución marginal. Hay dos distribuciones marginales, una por cada variable categórica. Según nuestros datos hay 651 sujetos insatisfechos con su cuerpo y 249 insatisfechos. La distribución marginal de la variable insatisfacción corporal es,

	NO	SI
Proporción	0,723	0,277

9.2 Tabla. Distribución marginal de insatisfacción corporal

La segunda distribución marginal para este ejemplo es la correspondiente a la variable edad,

	>14 años	14-16 años	16 años <
Proporción	0,336	0,426	0,238

9.3 Tabla. Distribución marginal de edad

Conceptos básicos y aplicaciones prácticas con Rcommander

Cada distribución marginal en una tabla de contingencia es una distribución de una variable categórica simple.

Tablas de doble entrada. Relaciones

La tabla de doble entrada contiene mayor cantidad de información que las dos distribuciones marginales por separado. Una tabla permite examinar las relaciones mediante el cálculo de ciertos porcentajes a partir de las frecuencias dadas. En el ejemplo, el porcentaje de menores de 14 años que están insatisfechos con su cuerpo es $67/303=0,221=22,1\%$; el porcentaje de sujetos entre 14 y 18 años insatisfechos con su cuerpo es $15/384=0,299=29,9\%$, y $68/221=0,308=30,8\%$, el porcentaje de sujetos mayores de 18 años insatisfechos con su cuerpo.

Distribuciones condicionales

Si nos centramos en el grupo de menores de 14 años y examinamos la distribución de la otra variable categorica, insatisfacción corporal, únicamente en ese grupo, decimos que esa distribución está condicionada al valor de la variable edad.

La distribución condicional ofrece las proporciones o porcentajes para todos los posibles valores de la variable insatisfacción corporal (NO/SI).

	NO	SI
Proporción	77,8%	22,1%

9.4 Tabla. Distribución condicional. Niños e insatisfacción corporal

Realizando cálculos similares para los otros dos grupos de edad,

	NO	SI
--	----	----

Conceptos básicos y aplicaciones prácticas con Rcommander

Proporción	70,05%	29,95%
------------	--------	--------

9.5 Tabla. Distribución condicional. Adolescentes e insatisfacción corporal

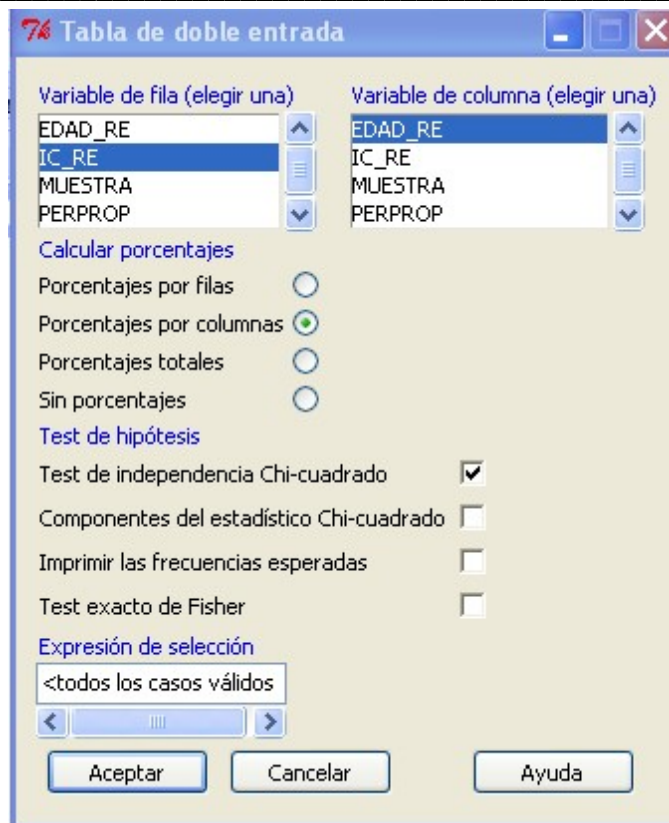
	NO	SI
Proporción	68,5%	31,45%

9.6 Tabla. Distribución condicional. Jóvenes e insatisfacción corporal

Si comparamos las distribuciones condicionales podemos entrever la naturaleza de la asociación entre la edad y la insatisfacción corporal. En este conjunto de datos en el grupo de mayores de 16 años tienen ligeramente más probabilidad de estar insatisfechos con sus cuerpos que los que están entre 14 y 16 años y mucha más probabilidad que los niños. Estos últimos, los niños, muestran el porcentaje más bajo de insatisfacción corporal. Los datos apuntan a que la insatisfacción corporal se incrementa con la edad.

RCommander permite obtener las distribuciones conjuntas, marginales y condicionales; Para ello se elegirá la opción Estadísticos>Tablas de contingencia>Tabla de doble entrada. En la ventana emergente elegiremos las variables fila y columna; insatisfacción corporal y edad (IC_RE y EDAD_RE). Además se pueden especificar los porcentajes a calcular (por filas, por columnas, totales, sin porcentajes), los tests de hipótesis a realizar (Chi-cuadrado, test exacto de Fisher), los componentes del estadístico Chi-cuadrado y las frecuencias esperadas. El análisis puede realizarse sobre todos los casos o bien sobre un subconjunto de ellos utilizando la expresión de selección adecuada.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



9.7Figura. Tabla de doble entrada Rcommander

```
> colPercents(.Table) # Column Percentages
      EDAD_RE
IC_RE  <14 14-16  >16
0      77.9 70.1  68.5
1      22.1 29.9  31.5
Total 100.0 100.0 100.0
Count 303.0 384.0 213.0
```

La opción “Porcentajes totales” permite obtener la distribución conjunta. En la columna y fila rotuladas “Total” aparecen las distribuciones marginales de la insatisfacción corporal y de la edad respectivamente.

```
> totPercents(.Table) # Percentage of Total
      <14 14-16  >16 Total
0      26.2 29.9 16.2  72.3
1       7.4 12.8  7.4  27.7
```

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Total	33.7	42.7	23.7	100.0
-------	------	------	------	-------

9.2.1 Tablas de doble entrada. Inferencia

La adopción de conclusiones estadísticas derivadas del estudio de proporciones hace necesario el uso de pruebas de significación. La hipótesis nula (H_0) de interés en una tabla de doble entrada es que no existe asociación entre la variable fila y la variable columna. En el ejemplo anterior consistiría en afirmar que las variables insatisfacción corporal y edad no están relacionadas. La hipótesis alterna H_1 consistiría en la afirmación de que sí existe una asociación entre las dos variables. La hipótesis alternativa no especifica una dirección concreta para la asociación, no la podemos describir como una hipótesis unilateral o bilateral, ya que incluye todas las clases posibles de asociación. En nuestro ejemplo, la hipótesis H_0 de que no existe asociación es equivalente a la afirmación de la que las distribuciones de insatisfacción corporal en los tres grupos de edad son iguales.

Frecuencias esperadas

Para contrastar la hipótesis nula se comparan las frecuencias observadas con las frecuencias esperadas bajo el supuesto de que la hipótesis nula es verdadera.

¿Cómo se obtienen las frecuencias esperadas? Si observamos los porcentajes en el margen derecho de la tabla veremos que 27,7% (todos los grupos de edad juntos) están insatisfechos con su cuerpo. Si la hipótesis nula de no diferencias por género fuese cierta, esperaríamos que este porcentaje global se aplicara a todos los grupos de edad por igual. Esto es, esperaríamos un 27,7% de insatisfechos en el grupo de menores de 14 años. En la tabla veíamos que el número de sujetos en este grupo es 303. La frecuencia esperada es por tanto

Conceptos básicos y aplicaciones prácticas con Rcommander

27,7 % de 303, es decir 83,9. El resto de frecuencias esperadas se calcularían de modo similar.

Para calcular la frecuencia esperada multiplicamos la proporción de insatisfechos con su cuerpo (249/900) por el número de sujetos menores de 14 años (303). En la tabla vemos que 249 y 303 son los totales por fila y columna de la celda de interés. 900 es el número total de observaciones de la tabla. Por lo tanto la frecuencia esperada de una celda es el producto de los totales de fila y columna divididos por el total de la tabla:

$$\text{Frecuencias esperadas} = \frac{\text{total de la fila} \times \text{total de la columna}}{N}$$

Test de Chi cuadrado

Para verificar la hipótesis nula, esto es, que no existe asociación entre las variables fila y columna, se utiliza un estadístico que compara el conjunto completo de frecuencias observadas con el conjunto de frecuencias esperadas. Consiste en hallar las diferencias entre la frecuencia osbservada y su correspondiente valor esperado y elevar esta diferencia al cuadrado. Una diferencia grande es menos significativa si procede de una celda donde se espere una frecuencia con valor muy alto, por ello se divide cada diferencia al cuadrado por la frecuencia esperada. Finalmente se agregan los valores correspondientes a todas las celdas.

$$X^2 = \sum \frac{(\text{frecuencia observada} - \text{frecuencia esperada})^2}{\text{frecuencia esperada}}$$

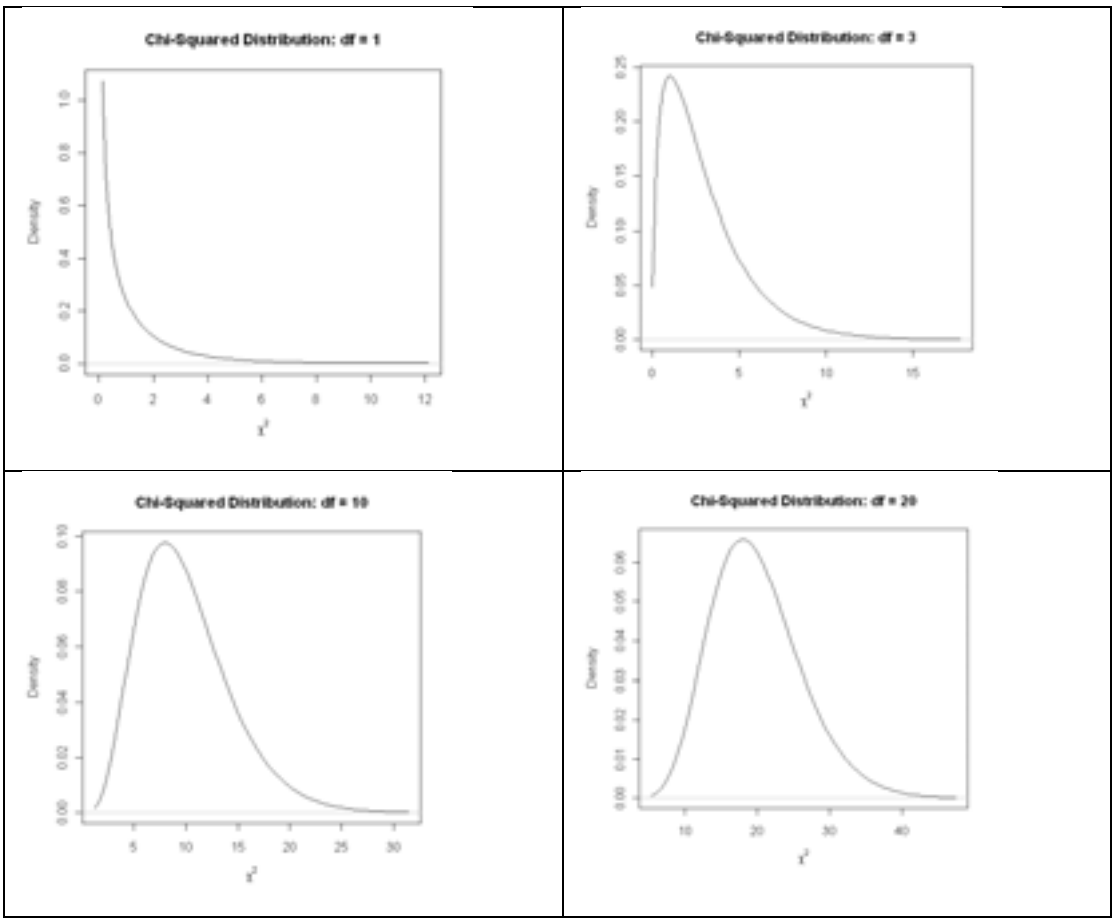
```
> .Test$expected # Expected Counts
      EDAD_RE
IC_RE  <14  14-16  >16
0  219.17  277.76  154.07
```

Conceptos básicos y aplicaciones prácticas con Rcommander

```
1 83.83 106.24 58.93
```

Cuanta más diferencia haya entre las frecuencias observadas y las esperadas mayor será el valor del estadístico Chi cuadrado. Valores altos de Chi cuadrado proporcionan evidencia contra la hipótesis nula. El número de filas y de columnas determinan los grados de libertad de este estadístico $((l-1)(z-1))$.

Las distribuciones chi cuadrado son una familia descrita por un parámetro; los grados de libertad (gl ; df ; degree of freedom). Se escribirá χ^2_{gl} para indicar un miembro particular de esta familia. En la figura siguiente se muestran las curvas de densidad de cuatro distribuciones χ^2_{gl} . Las distribuciones Chi cuadrado toman únicamente valores positivos y son sesgadas a la derecha:



9.8 Figura. Distribuciones Chi cuadrado con diferentes grados de libertad (1,3,10 y 20)

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

El test de chi cuadrado siempre utiliza la cola superior de la distribución χ^2 , puesto que cualquier desviación de la hipótesis nula hace que el estadístico tenga un valor más alto. La aproximación de la distribución del estadístico X^2 a la distribución χ^2 es más adecuada a medida que el número de frecuencias es mayor. Además es más preciso para tablas mayores que 2x2. Para que la aproximación sea adecuada se requiere, en las tablas de 2x2 que las frecuencias esperadas de todas las celdas sean al menos 5. En Para el caso de tablas mayores se requerirá que la media de frecuencias esperadas sea al menos 5. El valor p para el test chi cuadrado es $P(\chi^2_{(f-1)(c-1)} \geq X^2)$. RCommander proporciona el valor de X^2 y el valor de p marcando la opción Test de independencia Chi-cuadrado en la ventana de opciones. En nuestro ejemplo el valor del estadístico chi-cuadrado es 7,1976, el número de grados de libertad de la distribución es $(3-1)(2-1) = 2$, y el valor de p es 0,02736.

```
> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 7.1976, df = 2, p-value = 0.02736
```

El test chi cuadrado confirma que los datos contienen evidencia contra la hipótesis nula. Los sujetos de los tres niveles de edad son distintos en insatisfacción corporal, pero no dice cómo o en qué grado difieren. Hay que completar el test estadístico con una descripción que explique la clase de relación presente en los datos.

Si se obtienen los componentes del estadístico Chi cuadrado, podemos comprobar aquellas celdas que contribuyen en mayor medida a la discrepancia

Conceptos básicos y aplicaciones prácticas con Rcommander

entre lo observado y lo esperado. La tabla de componentes obtenida mediante RCommander se muestra a continuación:

```
> round(.Test$residuals^2, 2) # Chi-square Components
      EDAD_RE
IC_RE  <14 14-16  >16
  0  1.29  0.28  0.42
  1  3.38  0.72  1.11
```

Para calcular el valor correspondiente a la segunda fila y primera columna:

$$\frac{(\text{frecuencia observada} \times \text{frecuencia esperada})^2}{\text{frecuencia esperada}} = \frac{(67 - 83,3)^2}{83,3}$$

9.3 Comparación entre más de dos grupos (ANOVA)

Cuando se precisa la comparación entre más de dos grupos se podría tomar todas las posibles parejas y aplicar a cada una de ellas el test *t*. Sin embargo este procedimiento conduciría a un incremento grave del error Tipo I. El análisis de varianza (ANOVA o AVAR) permitirá probar la significatividad de las diferencias entre varias medias sin que se incremente el error de tipo I. El método ANOVA básicamente consiste en dividir varianza de la variable dependiente en dos o más componentes, cada uno de los cuales puede ser atribuido a una fuente identificable.

9.3.1 ANOVA de un factor

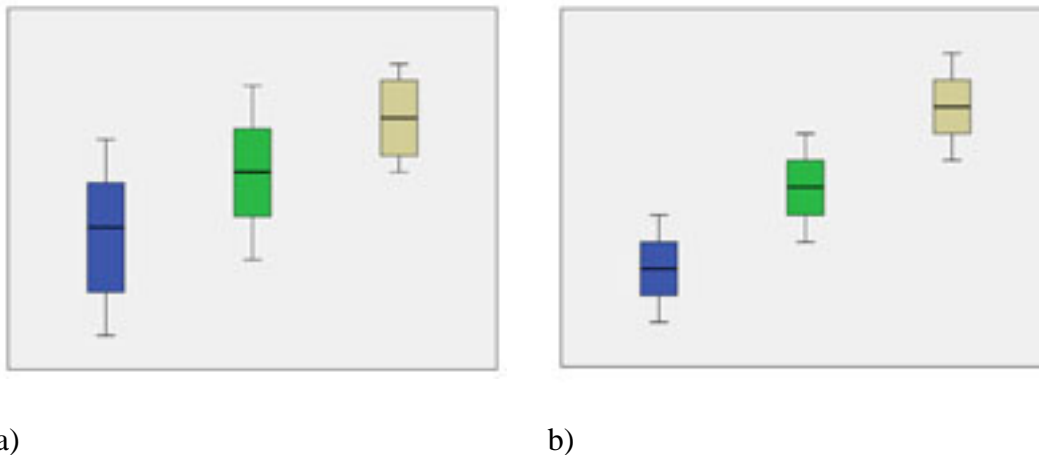
Supongamos que nos interesa averiguar si los tres diferentes grupos de edad (menores de 14 años, 14-16 años y mayores de 16 años) son iguales en insatisfacción corporal, o si por el contrario las medias son diferentes. A la

Conceptos básicos y aplicaciones prácticas con Rcommander

variable independiente (edad) en el ANOVA se denomina *factor* (¡no confundirlo con el “factor” del análisis factorial!) que en este ejemplo tiene tres categorías o *niveles*. El factor puede tratarse de una variable propiamente categórica, o puede haber sido categorizada; como en nuestro caso. La variable dependiente, insatisfacción corporal, (IC) es continua.

Aún cuando la hipótesis nula sea cierta y todas las medias poblacionales sea cual sea la edad fuesen iguales, esperaríamos una pequeña fluctuación de las medias muestrales de insatisfacción corporal. ¿Cuánto se espera que varíen estas medias? Pues bien, a partir de la variación de los datos, podemos tener una idea de la variación entre las medias. Si las medias poblacionales fuesen en realidad diferentes, esperaríamos que esta variación fuese aún mayor. Por ello necesitamos un test de hipótesis con el que se pueda verificar si la variación entre las medias es mayor que la que esperada en el caso de meras fluctuaciones aleatorias. Se precisa un nuevo modelo de distribución muestral, denominada distribución F.

Para aclarar como funciona este test examinemos la siguiente figura que contiene dos conjuntos de diagramas de cajas. Se trata de decidir si las medias muestrales son suficientemente dispares como para rechazar la hipótesis nula. En el segundo gráfico es difícil admitir que las diferencias entre las muestras se deban únicamente a la variabilidad muestral. En cambio en (a) es más fácil deducir que estas observaciones podrían haberse dado a partir de grupos con las mismas medias. Los dos conjuntos (a) y (b) tienen la misma diferencia entre los centros de los grupos. Sin embargo lo que los diferencia es la variación *dentro* de cada grupo. En el segundo es tan pequeña que las diferencias entre los grupos se evidencian. Pues bien, esta es la idea central del test F. Cuando las diferencias entre las medias son grandes comparadas con la variación dentro de los grupos, rechazaremos la hipótesis nula y concluiremos que las medias no son iguales. Sin embargo, en el gráfico (a) las diferencias entre las medias hacen pensar que bien podrían haberse obtenido por azar a partir de grupos con medias iguales, por ello no hay evidencia suficiente como para rechazar H_0 .



9.9Figura. Diferentes variabilidades intragrupo. Diagramas de cajas

Ahora bien, se necesitará una prueba estadística que permita hacer esta comparación de modo más preciso. Los tests vistos hasta aquí se basan en el ratio entre diferencias de alguna clase con alguna medida de la variabilidad muestral; del mismo modo en el ANOVA las diferencias entre las medias figurarán en el numerador, y el término de comparación (en el denominador) se basará en la desviación estándar subyacente, esto es, en la variabilidad *dentro* de los grupos.

9.4 Variación entre grupos y variación intra grupos

Si la hipótesis nula fuese cierta, y todos los grupos procedieran de una sola población, con una sola media (μ), cada media del grupo sería un estimador de esa media y tendríamos estimaciones diferentes e independientes de esa media (en nuestro ejemplo tendríamos tres estimaciones). Trataremos esas medias estimadas como observaciones, y calcularemos su varianza muestral ($s_{\bar{x}}^2$). Esta varianza es la medida que utilizaremos para evaluar la medida en que las medias de los diferentes grupos están alejadas entre sí. Es obvio que cuanto más próximos estén los valores de las medias, menor será la varianza y a la inversa, a medida que estén más alejados la varianza será mayor.

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

Para los datos de nuestro ejemplo, podemos obtener las medias en la variable insatisfacción corporal para los tres grupos de edad, con la opción Estadísticos>Resúmenes>Resúmenes numéricos escogiendo en la ventana emergente la opción “Resumir por grupos” (para mayor claridad, en este ejemplo hemos trabajado con un subconjunto del objeto `Edi.data`):

<i>mean</i>	<i>n</i>	<i>NA</i>
<14	10.82	184 14
14-16	12.89	184 16
>16	13.67	184 8

La varianza entre estas medias es 2,17.

```
Medias <- c(10.82, 12.89, 13.67)
Var(Medias)
[1] 2.17
```

¿Cómo podemos tener idea de si se trata de un valor alto o bajo? Para saberlo necesitamos un modelo, y el modelo se basa en la hipótesis nula de que las medias son iguales. Esto es, que la media de insatisfacción corporal siempre es la misma independientemente del grupo de edad:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_1 : no todas las medias son iguales

La varianza de una media muestral ($s_{\bar{x}}^2$) es σ^2/n . Por lo tanto para estimar σ^2 bastará con multiplicar n por la varianza observada entre las muestras ($ns_{\bar{x}}^2$).

En nuestro caso una estimación de σ^2 sería 400,016 (184 x 2,17). Tenemos así una estimación de σ^2 a partir de la variación *entre* grupos. Esta

Conceptos básicos y aplicaciones prácticas con Rcommander

cantidad es la media cuadrática del factor o media cuadrática *entre* (o inter) grupos (MCG).

Al igual que en el resto de contrastes se necesita un término de comparación, basado en la variabilidad subyacente a las medidas. Necesitamos una estimación independiente de σ^2 , que no dependa de si la hipótesis nula es cierta o falsa. Esta estimación se obtiene a partir de una combinación de las varianzas obtenidas *dentro* de cada grupo. Esta cantidad se denomina media cuadrática del error o media cuadrática *intra* (o dentro) de los grupos (MCE).

Las desviaciones estándar pueden obtenerse desde la misma ventana que hemos utilizado para obtener el listado de las medias, Estadísticos>Resúmenes>Resúmenes numéricos seleccionando la opción "Desviación típica". Rcommander nos proporciona las desviaciones estándar en la variable insatisfacción corporal. Elevando al cuadrado estos valores obtenemos las varianzas; 111,87; 111,99 y 93,412

	<i>sd</i>	<i>n</i>	<i>NA</i>
<14	10.57	184	14
14-16	10.58	184	16
>16	9.66	184	8

Para obtener la varianza conjunta aplicamos,

$$s^{*2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1)}$$

Con nuestros datos,

$$s^{*2} = \frac{(184 - 1)10,577^2 + (184 - 1)10,583^2 + (184 - 1)9,665^2}{(184 - 1) + (184 - 1) + (184 - 1)} = 105,762$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Esta cantidad es la media cuadrática del error o media cuadrática intragrupos (MCE).

9.4.1.a Estadístico F

Contamos con dos estimaciones diferentes de la varianza σ^2 ; La primera se basa en las diferencias entre las medias de los grupos. Si las medias son iguales (H_0 cierta), esta varianza estimará σ^2 . Si H_0 no es cierta, dará sistemáticamente un valor mayor. La otra estimación se basa únicamente en la variación dentro de los grupos en torno a sus propias medias, es independiente de la hipótesis nula. Por lo tanto si H_0 es cierta, estos dos valores, MCE y MCG estimarán σ^2 y su ratio será cercano a 1. Pero cuando H_0 es falsa, la razón entre los dos estimadores tenderá a ser mayor que 1. La estimación MCG será sistemáticamente mayor, (reflejando la diferencia en las medias), mientras que MCE no se verá afectada; el ratio tenderá a ser mayor que 1.

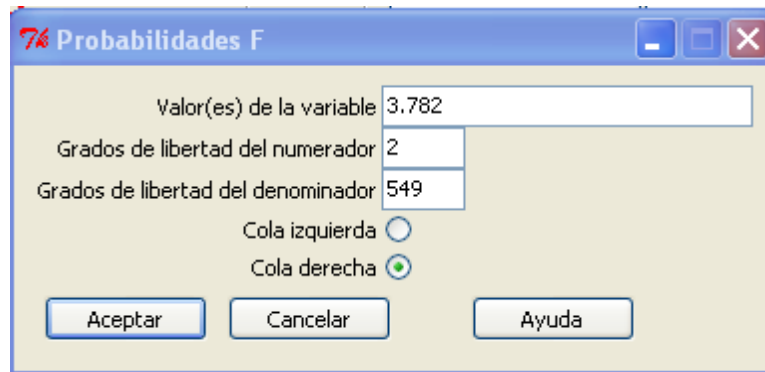
El ratio MCG/MCE es el estadístico F, y cuando H_0 es cierta sigue una distribución F que depende de dos números; los grados de libertad del numerador y los grados de libertad del denominador. En el caso de ANOVA de un factor con K grupos, los grados de libertad para el numerador son $K-1$ y para el denominador $N-K$.

El test F es de una cola puesto que cualquier diferencia en las medias incrementa el estadístico F. Diferencias grandes en los grupos se traducen en un estadístico mayor. El test será significativo si el ratio F es suficientemente grande y su probabilidad asociada (p) suficientemente pequeña. En nuestro ejemplo este ratio es,

$$F = \frac{MCG}{MCE} = \frac{400,016}{105,762} = 3,782$$

Conceptos básicos y aplicaciones prácticas con Rcommander

La distribución de este estadístico es $F_{(K-1)(N-K)} = F_{(3-1),(552-3)} = F_{2,549}$. El valor p del test F es la probabilidad de que una variable aleatoria con distribución $F_{(2,549)}$ sea mayor o igual que el valor calculado. La opción Distribuciones>Distribuciones continuas>Distribución F de Rcommander proporciona ese valor:



9.10 Figura. Distribución de F en RCommander

```
pF(c(3.782), df1=2, df2=149, lower.tail=FALSE)
[1] 0.0233
```

9.4.1.b Modelo

El modelo de ANOVA de un factor intenta describir la puntuación de un sujeto en la variable dependiente a partir de tres componentes: La media general de todas las poblaciones, que representa el efecto de los factores mantenidos constantes (μ), el efecto específico atribuible a la variable independiente (α_k) y el efecto de posibles variaciones aleatorias o residuales no contempladas (ε_{ik}). Si tenemos K grupos y x_{ik} representa a la observación “i” del grupo “k” el modelo establece:

Conceptos básicos y aplicaciones prácticas con Rcommander

$$x_{ik} = \mu + \alpha_k + \varepsilon_{ik} = \mu_k + \varepsilon_{ik}$$

DATOS = AJUSTE + RESIDUAL

Donde α_k es la desviación de la media de la población origen de la muestra “k” respecto de la media general ($\alpha_k = \mu_k - \mu$); esto es lo que se aleja el promedio de ese grupo respecto de la media general. El error o residual representa la desviación de una observación a la media de su grupo, $\varepsilon_{ik} = x_{ik} - \mu_k$; es decir, la fluctuación aleatoria que no se pretende ser explicada en el modelo. En el modelo de ANOVA se asume que estas fluctuaciones o errores cumplen ciertas condiciones; los errores son normales, con media 0 e igual varianza para todos los grupos. Será precisa la comprobación del cumplimiento de estas condiciones para asegurar la adecuación del análisis.

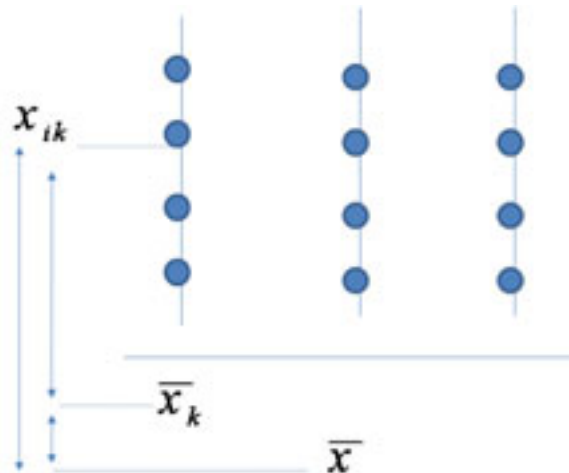
9.4.1.c Descomposición de las sumas de cuadrados

La desviación total de una puntuación x_{ik} a la media general \bar{x} se puede descomponer en dos partes; la desviación de la puntuación a la media de su grupo \bar{x}_k y la desviación de la media del grupo \bar{x}_k a la media general \bar{x} .

$$x_{ik} - \bar{x} = (x_{ik} - \bar{x}_k) + (\bar{x}_k - \bar{x})$$

Esta idea queda reflejada en el gráfico siguiente. Si en lugar de referirnos a una única puntuación lo hacemos al conjunto de las N puntuaciones,

$$\sum_{i=1}^{n_k} \sum_{k=1}^K (x_{ik} - \bar{x})^2 = \sum_{i=1}^{n_k} \sum_{k=1}^K (x_{ik} - \bar{x}_k)^2 + \sum_{i=1}^{n_k} \sum_{k=1}^K (\bar{x}_k - \bar{x})^2$$



9.11 Figura. Descomposición de las sumas de cuadrados (ANOVA de un factor)

Los términos de esta ecuación son sumas de cuadrados (SC); suma de cuadrados total (SCT), suma de cuadrados intergrupo (SCG) y suma de cuadrados del error (SCE).

9.4.1.d Estimación de parámetros

Las medias poblacionales se estiman mediante las correspondientes medias muestrales. μ_k se estima mediante la media del grupo k

$$\bar{x}_k = \frac{1}{N_k} \sum_{i=1}^{N_i} x_{ik}$$

Los errores se estimarán mediante los residuales $e_{ik} = x_{ik} - \bar{x}_k$ o variación en torno a las medias muestrales.

9.4.1.e Tabla de ANOVA

Para llevar a cabo el análisis de varianza toda la información se organiza en una tabla que de modo estándar contiene las columnas fuentes de variación, sumas de cuadrados, grados de libertad, medias cuadráticas, valor F y grado de

Conceptos básicos y aplicaciones prácticas con Rcommander

significación. Las filas son las tres fuentes de variación en el ANOVA de un factor; modelo o variación entregrupos, error o variación intragrupos y total. Los datos de nuestro ejemplo se muestran en la siguiente tabla;

Fuentes de variación	Sumas de cuadrados	Grados de libertad	Medias cuadráticas	F	p
Modelo	799,859	2	399,929	3,781	,023
Error	58065,554	549	105,766		
Total	58865,413	551			

9.7 Tabla. Tabla para el ANOVA de un factor

La fuente de variación *Modelo* ofrece información relacionada con la variabilidad *entre* las medias de los grupos y *Error* da información relacionada con la variabilidad *dentro* de los grupos.

$$\text{TOTAL} = \text{MODELO} + \text{ERROR}$$

Cada suma de cuadrados es una suma de desviaciones al cuadrado. Se utilizan habitualmente SCE, SCG, SCT, para designar las sumas de cuadrados (sums of squares) del error, de los grupos y total respectivamente. Cada suma de cuadrados mide una diferente fuente de variación. SCT mide la desviación entre cada observación y la media global, $x_{ik} - \bar{x}$. SCG recoge la desviación de las medias de los grupos en torno a la media global $\bar{x}_k - \bar{x}$ y por último la variación del error SCE mide la desviación de cada observación a la media de su grupo, $x_{ik} - \bar{x}_k$.

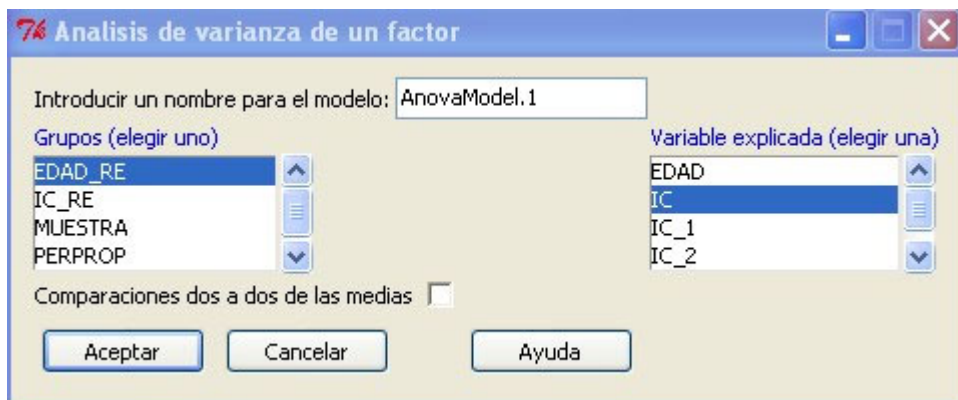
A cada suma de cuadrados se le asocia una cantidad denominada grados de libertad, (degrees of freedom, df). Dado que SCT mide la variación de N observaciones en torno a la media general, los grados de libertad son N-1.

Conceptos básicos y aplicaciones prácticas con Rcommander

Como SCG mide la variación de K medias muestrales en torno a la media global, los grados de libertad correspondientes son $K-1$. Por último SCE es la suma de cuadrados de las desviaciones $x_{ik} - \bar{x}_k$. Son N observaciones que se comparan con K medias muestrales y sus grados de libertad son $N-K$. Para cada fuente de variación se estima la *media cuadrática*; este valor es la suma de cuadrados dividida por los grados de libertad.

9.4.2 ANOVA de un factor mediante Rcommander

En Rcommander el ANOVA de un factor se obtiene a través de la opción Estadísticos>Medias>Anova de un factor. Se seleccionan la variable que define los grupos o variable independiente (EDAD_RE) y la variable explicada o dependiente; esto es, la variable cuyas medias en los diferentes grupos vamos a comparar (en nuestro ejemplo IC). –en este ejemplo trabajamos con un subconjunto;



9.12Figura. Factores en el ANOVA de un factor. Rcommander

La salida obtenida es la siguiente,

```
> AnovaModel.1 <- aov(IC ~ EDAD_RE, data=Edi.data)

> summary(AnovaModel.1)

              Df Sum Sq Mean Sq F value    Pr(>F)
EDAD_RE       2    800    400    3.7813 0.02339 *
```

Conceptos básicos y aplicaciones prácticas con Rcommander

```

Residuals    549  58066    106
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
76 observations deleted due to missingness

> numSummary(Edi.data$IC , groups=Edi.data$EDAD_RE,
statistics=c("mean" ,
+ "sd" ))
      mean      sd   n NA
<14    10.82   10.57 184 14
14-16   12.89   10.58 184 16
>16    13.67    9.66 184  8

```

La primera línea de la tabla que proporciona Rcommander es la correspondiente al modelo, que en este caso incluye la variable independiente (EDAD_RE); La segunda fila es la correspondiente al residual. Se omite la fila correspondiente a la fuente de variación total.

Si la hipótesis nula es cierta, entonces no habrá diferencias entre las medias de los grupos y la razón MCG/MCE se aproximará a 1. En nuestro ejemplo, MCG (*Mean Sq*) es 400 y MCE (*Mean Sq*) es 106; por tanto el valor del estadístico F es (400/106) 3,7813 y su grado de significación 0,02339. Este resultado proporciona evidencia para rechazar la hipótesis nula; los grupos de edad en la población no son iguales en insatisfacción corporal. En definitiva, los datos apuntan a que existe una relación entre las dos variables.

Rcommander proporciona información adicional; la media de cada grupo (10,82; 10,58 y 9,66), la desviación estándar (10,57; 10,58 y 9,66) y el número de sujetos por grupo (184).

Comparaciones múltiples entre medias

Conceptos básicos y aplicaciones prácticas con Rcommander

El test F del ANOVA proporciona una respuesta global a la pregunta ¿Son las diferencias entre las medias de los grupos significativas?; Un valor p pequeño únicamente indica que las medias no son iguales pero no dice específicamente cuáles son las medias que difieren. Aunque una inspección de los gráficos de las medias o de sus valores da alguna pista de donde se producen estas diferencias, es deseable complementarla con otros tests estadísticos suplementarios al ANOVA. Las comparaciones múltiples son las que se llevan a cabo tras un ANOVA y en las que se comparan sistemáticamente todos los pares de medias para encontrar los grupos que han producido esta significación. Las pruebas diseñadas para llevarlas a cabo mantienen el riesgo alfa en cierto nivel para el conjunto de comparaciones que se efectúen.

Rcommander proporciona el intervalo de confianza conjunto para todos los pares de diferencias basado en la prueba HSD (*Honestly Significantly Differences*; Diferencia Honestamente Significativa) de Tukey. En la figura siguiente se muestra la salida. Junto a las diferencias observadas entre los grupos (columna “*Estimate*”), aparecen los límites inferior y superior de los intervalos de probabilidad para las diferencias con un 95% de confianza. El único intervalo que no contiene el valor cero (diferencia significativa) es el correspondiente a la diferencia entre los grupos de mayores de 16 y niños (0,33 - 5,37).

```
Simultaneous Confidence Intervals

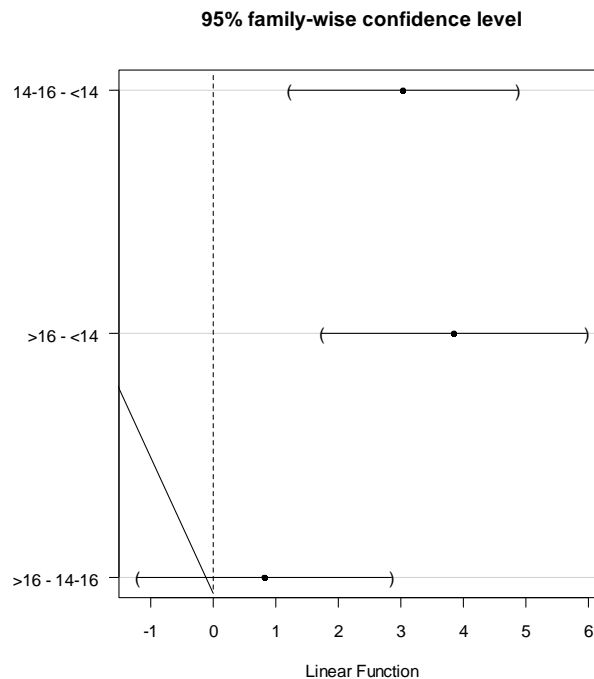
Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = IC ~ EDAD_RE, data = Edi.data)

Estimated Quantile = 2.3503
95% family-wise confidence level

Linear Hypotheses:
              Estimate lwr      upr
14-16 - <14 == 0  2.07   -0.44   4.59
>16 - <14 == 0   2.85    0.33   5.37
>16 - 14-16 == 0  0.7826 -1.73   3.30
```


Rcommander proporciona el gráfico para los intervalos de confianza, que facilita la inspección de datos cuando el número de comparaciones es muy alto.



9.13 Figura. Comparaciones múltiples. Intervalos de confianza.

9.4.3 Análisis de varianza de dos factores

El análisis de varianza de dos factores compara las medias de poblaciones clasificadas con arreglo a dos criterios de clasificación diferentes, o las medias de respuesta en experimentos de dos factores. Al igual que en el modelo de un factor se asume que los datos son aproximadamente normales y que los grupos pueden tener diferentes medias pero tienen la misma desviación estándar. Se utiliza de nuevo un estimador de la varianza común y estadísticos F para los tests de significación. La principal diferencia radicará en el término AJUSTE. El diseño de dos factores ofrece ventajas respecto al anterior,

Conceptos básicos y aplicaciones prácticas con Rcommander

resultando más eficiente estudiar dos factores simultáneamente que por separado. Además, reduce la variación residual en el modelo incluyendo un segundo factor que presumiblemente tendrá influencia en la respuesta y, por último, permite investigar las interacciones entre los factores.

9.4.3.a Modelo

Designaremos con K al número de grupos o niveles del primer factor (Factor “A”) y con L al número de grupos o niveles del segundo factor (Factor “B”). En un diseño de dos factores cada nivel del primer factor (Factor A) se combina con cada nivel o grupo del segundo factor (Factor B) de modo que se compararán $K \times L$ grupos. El tamaño muestral para el nivel k de A y el nivel l de B es N_{kl} . El número total de observaciones es $N = \sum N_{kl}$.

Se asumen muestras independientes aleatorias simples de tamaño n_{kl} procedentes de $K \times L$ poblaciones normales. Aunque las medias poblacionales μ_{kl} pueden ser diferentes, todas las poblaciones tienen la misma desviación estándar. Tanto las medias poblacionales (μ_{kl}) como la varianza común (σ) son parámetros desconocidos. Si representamos por x_{ikl} la i -ésima observación de una población que tiene el nivel k del factor “A” y el nivel l del factor “B” el modelo estadístico es:

$$\begin{aligned}
 x_{ikl} &= \mu + \alpha_k + \beta_l + (\alpha\beta)_{kl} + \varepsilon_{ikl} \\
 &= \mu_{kl} + \varepsilon_{ikl} \\
 \text{DATOS} &= \text{AJUSTE} + \text{RESIDUAL}
 \end{aligned}$$

En el ANOVA de un factor el efecto atribuible a la variable independiente quedaba recogido en el término α_k . Ahora, con dos factores se incluye en el modelo un nuevo término que recoge el efecto del segundo factor (β_l). La interacción o efecto de la combinación de los diferentes niveles de

Conceptos básicos y aplicaciones prácticas con Rcommander

ambos factores se representa con el término $(\alpha\beta)_{kl}$. Estimaremos la media de una población por la media muestral de las observaciones en ese grupo $\bar{x}_{kl} = \frac{1}{N_{kl}} \sum_{k=1}^{n_{kl}} x_{ikl}$. La parte RESIDUAL del modelo contiene la desviación típica desconocida σ . El estimador de σ^2 , s^{*2} , se calcula a partir de las varianzas muestrales;

$$s^{*2} = \frac{\sum (N_{kl} - 1) s_{kl}^2}{\sum (N_{kl} - 1)}$$

Donde N_{kl} número de sujetos en el grupo kl , y
 s_{kl}^2 varianza del grupo kl

9.5 Efectos principales e interacción

Si se tienen muestras independientes de $K \times L$ grupos, podemos concebir el problema inicialmente como un ANOVA de un factor con KL grupos. Cada media poblacional μ_{kl} será estimada por su correspondiente media muestral \bar{x}_{kl} . Las sumas de cuadrados y los grados de libertad pueden calcularse como en ANOVA unifactorial. La suma de cuadrados del modelo se calcula a partir de las desviaciones $x_{kl} - \bar{x}$, donde x_{kl} es la media del grupo kl y \bar{x} es la media de todas las observaciones. Los grados de libertad del modelo (gl) son $KL-1$.

En el ANOVA de dos factores los términos SCM y gl_m se descomponen en: un efecto principal de A (SCA; gl_A), un efecto principal de B (SCB; gl_B), y una interacción entre ambos (SCAB; gl_{AB}).

$$SCM = SCA + SCB + SCAB$$

$$gl_m = gl_A + gl_B + gl_{AB}$$

Conceptos básicos y aplicaciones prácticas con Rcommander

El término SCA representa la variación entre las medias para los diferentes niveles del factor A. Dado que hay K niveles, los grados de libertad (gl_A) son $K-1$. SCB representa la variación entre las medias para los diferentes niveles del factor B, con $L-1$ grados de libertad (gl_B). SCAB es la suma de cuadrados de la interacción y representa la variación del modelo que no es producida por los efectos principales. Sus grados de libertad (gl_{AB}) son: $(KL-1)-(K-1)-(L-1) = (K-1)(L-1)$.

El concepto de interacción se comprenderá mejor con un ejemplo. Se pretende examinar las relaciones entre la edad, la insatisfacción corporal y el sexo. La tabla siguiente muestra las medias en insatisfacción corporal para cada grupo de edad y sexo. Para obtenerla, elegimos en Rcommander la opción Estadísticos > Resúmenes > Tablas estadísticas:

EDAD_RE	chicas	chicos	Media
>14	12,8	7,5	10,0
14-16	17,5	8,2	13,0
<16	16,2	9,0	13,8
Media	15,7	8,0	12,2

9.8Tabla. Insatisfacción corporal. Medias por grupos de edad y sexo

La tabla incluye medias globales en fila y columna. Por ejemplo, la media de insatisfacción corporal para todos los estudiantes entre 14 y 18 años es 13,0. Para calcularlo,

$$\frac{(17,5)(198) + (8,2)(185)}{383} = 13,0$$

La media global del grupo de chicas es 15,7.

$$\frac{(12,8)(142) + (17,5)(198) + (16,2)(141)}{481} = 15,7$$

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Estas medias se denominan medias marginales.

EDAD_RE	chicas	chicos	Total
>14	142	161	303
14-16	198	185	383
<16	141	72	213
Total	481	418	899

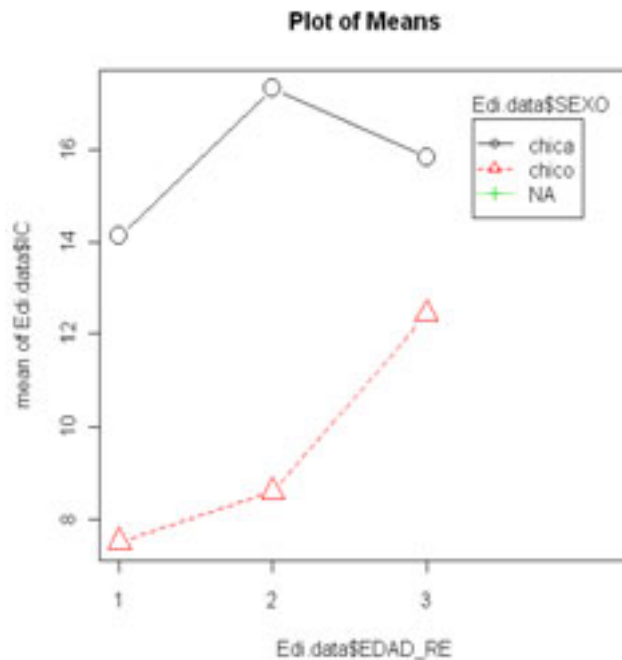
9.9Tabla. Número de sujetos por grupos de edad y sexo

Para obtener el gráfico para representar las medias de los grupos debemos escoger la opción Gráficas> Gráficas de las medias. A partir de este gráfico se ve claramente que las chicas están por encima en insatisfacción corporal que los chicos, esto es, se aprecia un efecto principal para el sexo. También se aprecia que los estudiantes entre 14 y 16 años tienen mayor puntuación en insatisfacción corporal que los niños. En cuanto a los mayores de 16 años, continúan la tendencia ascendente en el caso de los chicos, mientras que las chicas están por debajo del grupo de edad anterior.

Los efectos principales pueden describirse por las diferencias entre las medias marginales. Por ejemplo la media para chicas es 15,7 mientras que para los chicos es 8,0. Este es el efecto principal de la variable sexo. El cuerpo de la tabla muestra que las diferencias en insatisfacción corporal entre chicas y chicos para los tres grupos de edad son 5,3 9,3 y 7,2. Aunque la diferencia es más acusada en el grupo de más edad, la mayor parte de la información relativa a las diferencias por sexo se resume por el efecto principal, la diferencia marginal de 7,7. De modo análogo, el efecto principal de la edad queda resumido en las puntuaciones 10, 13 y 13,8. Ahora bien, el gráfico pone de manifiesto que el efecto de la edad no es igual para chicos que para chicas. En el grupo de 14 a 16 años la insatisfacción es mayor que en el grupo de menores de 14, tanto para los chicos como para las chicas, si bien en el grupo de chicas el incremento es mucho más acusado. Por otro lado, en el grupo de mayores de 16 años, la

Conceptos básicos y aplicaciones prácticas con Rcommander

tendencia se invierte para los dos grupos: en el de las chicas desciende la insatisfacción, mientras que en el de chicos se incrementa. Es decir que la diferencia en insatisfacción corporal entre chicos y chicas depende del grupo de edad que se considere. Esto es lo que se denomina el efecto de *interacción* entre los factores. Cuando existe interacción, las medias marginales no proporcionan un resumen adecuado de la información, aunque esto no quiere decir necesariamente que los efectos principales no sean informativos. En este gráfico y en la tabla es evidente que las chicas tienen mayor insatisfacción corporal que los chicos y también se aprecia la tendencia a que la insatisfacción corporal se incremente con la edad (salvo para el grupo de chicas mayores de 16). Cuando existe interacción, es preciso analizar cuidadosamente las medias para interpretar adecuadamente los datos. Para que no exista interacción, la diferencia en las medias entre niveles de un factor debería ser constante en todos los niveles del otro factor, lo que se traduciría gráficamente en líneas aproximadamente paralelas.



9.14 Figura. Gráfico de medias. Insatisfacción corporal por edad y sexo.

Conceptos básicos y aplicaciones prácticas con Rcommander

9.5.1.a ANOVA de dos factores. Inferencia

La inferencia para el ANOVA de dos factores implica el cálculo de los estadísticos F para cada uno de los efectos principales y para la interacción. Los cálculos se organizan en una tabla basada en la descomposición de la variación total.

$$SCT = SCA + SCB + SCAB + SCE$$

$$gl_T = gl_A + gl_B + gl_{AB} + gl_E$$

A partir de cada suma de cuadrados y sus grados de libertad se calcula la correspondiente media cuadrática (MC=SC/gl.). La significación de cada uno de los efectos principales y la interacción se evalúa mediante un estadístico F que compara la variación debida al efecto de interés con la variación intra-grupo. El diseño general de la tabla es el siguiente:

Fuente	Grados de libertad	Sumas de cuadrados	Media cuadrática	F
A	$I-1$	SCA	SCA / gl_A	MCA/MCE
B	$J-1$	SCB	SCB / gl_B	MCB/MCE
AB	$(I-1)(J-1)$	SCAB	$SCAB / gl_{AB}$	MCAB/MCE
Error	$N-IJ$	SCE	SCE / gl_e	
Total	$N-1$	SST	SCT / gl_T	

9.10. Tabla general ANOVA 2 factores

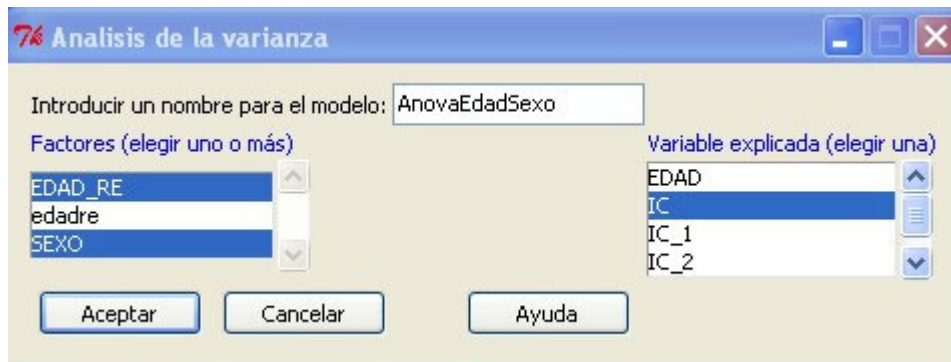
En el ANOVA de dos factores se plantean tres hipótesis nulas, referidas al efecto principal A, al efecto principal B y a la interacción AB y cada una se contrasta con un estadístico F. Primero se examina el efecto de la interacción, puesto que si existe ésta puede condicionar la interpretación del resultado de los efectos principales.

Conceptos básicos y aplicaciones prácticas con Rcommander

Valores altos de F conducen al rechazo de la hipótesis nula. El valor p (grado de significación) es la probabilidad de que una variable aleatoria que siga la distribución F correspondiente sea mayor o igual que el valor calculado.

9.5.2 Análisis de varianza con Rcommander

Para llevar a cabo ANOVA de dos factores elijiremos la opción de Rcommander Estadísticos > Medias > ANOVA de múltiples factores.



9.15 Figura. ANOVA de dos factores. Rcommander.

En primer lugar introducimos el nombre que queremos dar al modelo (en nuestro caso `AnovaEdadSexo`), seleccionamos los factores (`EDAD_RE` y `SEXO`) y por último la variable que se desea explicar; insatisfacción corporal (`IC`).

```
> AnovaEdadSexo <- (lm(IC ~ EDAD_RE*SEXO, data=Edi.data))
```

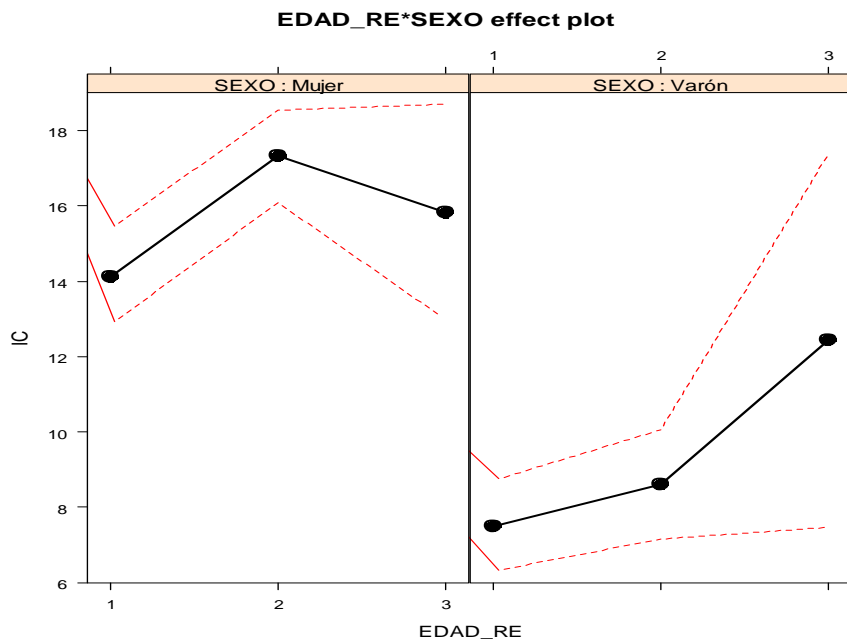
```
> Anova(AnovaEdadSexo)
Anova Table (Type II tests)
Response: IC
      Sum Sq Df F value    Pr(>F)
EDAD_RE    1318  2   7.4349 0.0006275 ***
SEXO       12148  1 137.0088 < 2.2e-16 ***
EDAD_RE:SEXO  652  2   3.6779 0.0256584 *
Residuals  79267 894
```


Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La tabla ANOVA anterior muestra en la primera columna los nombres de los dos factores principales (EDAD_RE y SEXO), la interacción entre ambos (EDAD_RE:SEXO) y los errores o residuales (Residuals). En la segunda columna figuran las sumas de cuadrados (Sum Sq); En la tercera podemos leer los grados de libertad (Df); luego aparece el valor F (F value), y por último el valor de probabilidad que se le asocia o grado de significación. El valor de p para los factores principales es muy bajo; A la derecha de estos valores se imprimen tres asteriscos, que señalan valores de significación inferiores a 0,001. La interacción tiene un grado de significación (p) 0,025.

Una vez generado el modelo, R lo incorpora como un objeto nuevo, con el nombre `AnovaEdadSexo`. Podemos elegir la opción de la barra de menú Modelos y la opción Gráficas > Gráficas de los efectos para obtener el gráfico siguiente:



9.16 Figura. Gráfico de los efectos

Conceptos básicos y aplicaciones prácticas con Rcommander

En el gráfico se representan las medias por sexo y grupo de edad así como los intervalos de confianza. Se puede apreciar que los chicos siguen una tendencia ascendente y uniforme.

10 Test-criterio. Regresión lineal

El fin último de numerosos test o cuestionarios es la predicción de un criterio. Cuando se quiere pronosticar el rendimiento académico, el éxito laboral u otro tipo de conducta se utilizan medidas simples o combinadas que permitan una predicción precisa. Aunque los contextos de predicción son muchos y variados, en la sustitución de un criterio por una variable (compuesto) predictor se persiguen simplicidad, sencillez y ahorro. Si para pronosticar una conducta futura o una conducta presente utilizamos tests o cuestionarios es necesario garantizar la relación estrecha entre las inferencias derivadas de estos y el criterio a predecir para que la sustitución quede justificada.

El modelo de regresión lineal ofrece un marco teórico que posibilita el estudio entre las relaciones test/criterio; su aplicabilidad dependerá en cada caso del cumplimiento de los supuestos del modelo por los datos. En el caso más simple el modelo de regresión lineal explicaría las relaciones entre una variable predictora y un criterio. En el contexto de la teoría de tests, la variable predictora es la puntuación obtenida en el test y la variable criterio es aquella variable que se desea predecir. La relación entre ambas variables se conoce como coeficiente de validez del test.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Donde ρ_{XY} El coeficiente de validez del test X,
 σ_X y σ_Y , desviaciones típicas del test y del criterio
 $Cov(X,Y)$, covarianza entre el test y el criterio.

Conceptos básicos y aplicaciones prácticas con Rcommander

10.1 Regresión simple

El modelo de regresión simple utiliza una sola variable predictora para el pronóstico del criterio. En psicología y educación las predicciones se basan habitualmente en modelos más complejos que utilizan como predictores un conjunto de puntuaciones obtenidas tras la administración de varios tests o de la utilización de métodos basados en la observación, entrevistas...Sin embargo la comprensión de modelos más complejos es más sencilla una vez familiarizados con el modelo simple.

10.1.1 Ecuación de la recta

La utilización y aplicación de un modelo matemático implica la asunción de varios presupuestos que en el caso del modelo de regresión simple se resumen en la relación lineal entre la variable predictora y el criterio. Formalmente, es posible utilizar la ecuación de una recta para explicar los efectos que producen sobre la variable dependiente (Y) los cambios en la variable predictora (X). Por ejemplo, en la siguiente ecuación,

$$Y = a + bX$$

Los parámetros *a* y *b* permiten estimar los valores de Y asociados a los valores de X:

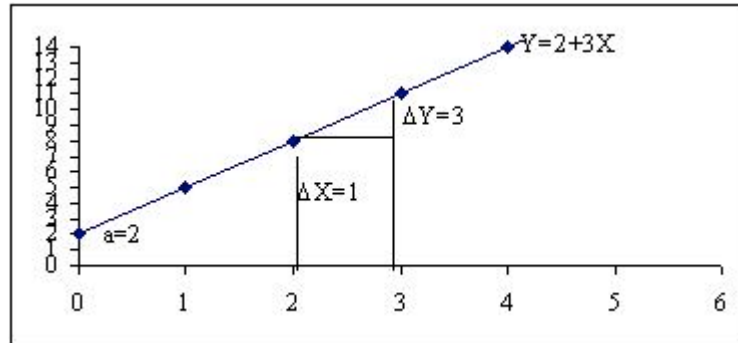
$$Y = 2 + 3X$$

De este modo,

X	0	1	2	3	4	5	6
Y	2	5	8	11	14	17	20

Conceptos básicos y aplicaciones prácticas con Rcommander

Como se aprecia en la tabla el incremento en 1 unidad de X produce un incremento de 3 unidades en Y.



10.1Figura. Ecuación de la recta

La interpretación de los parámetros de la recta es clara a través de la figura anterior:

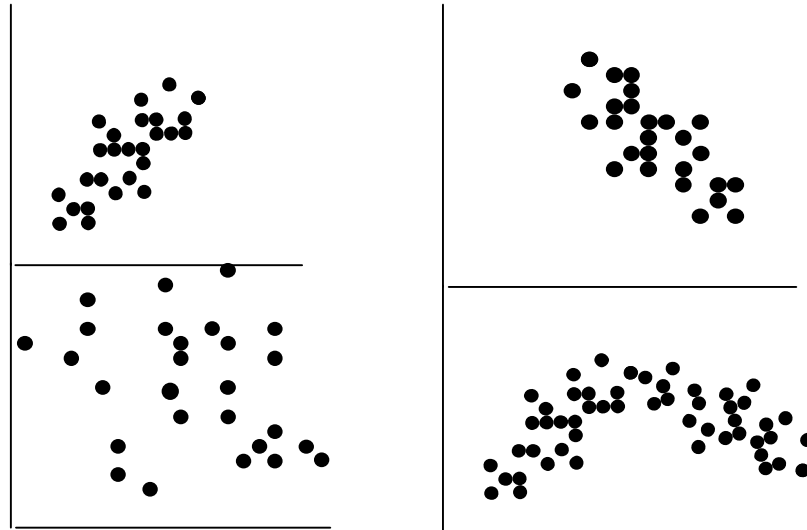
- a. El parámetro a viene definido por el valor en la ordenada cuando la variable predictora es 0. En nuestro caso, $a=2$. Es el punto de corte de la recta en la ordenada.
- b. El parámetro b indica la pendiente de la recta y representa el incremento en la variable dependiente asociado al cambio de una unidad en la variable independiente ($b=3$).

10.1.2 Relaciones entre variables en las ciencias sociales

Cuando el objetivo del investigador es profundizar en las relaciones observadas entre dos variables el primer paso es seleccionar el modelo adecuado para ello. El objetivo es resumir la información en un modelo matemático. Para ello es habitual seguir una pauta que comienza con la recogida de datos (X e Y), y representar por medio de un diagrama de dispersión la distribución de ambas. La información recogida gráficamente es fundamental como fuente de información que guía la elección del modelo adecuado. En un diagrama de

Conceptos básicos y aplicaciones prácticas con Rcommander

dispersión puede apreciarse la relación entre variables, y el sentido e intensidad de la misma.



10.2Figura. Diagramas de dispersión para dos variables

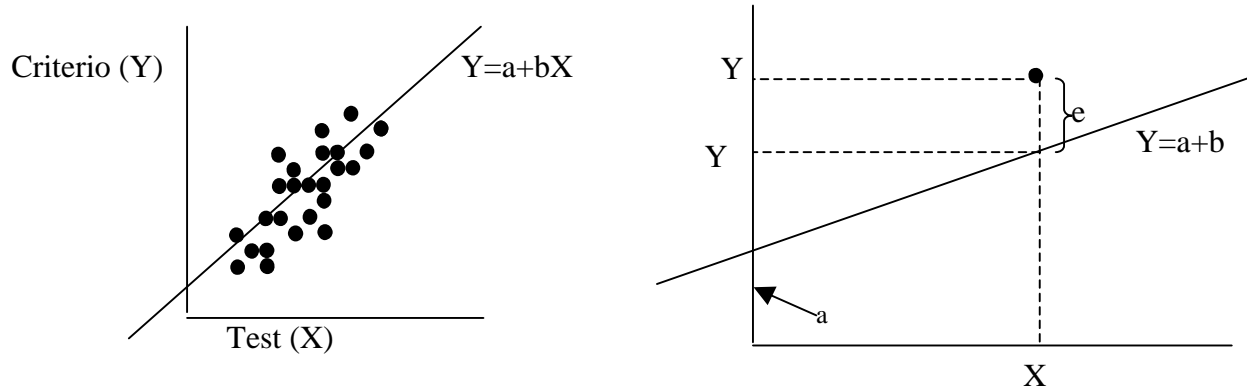
En la figura anterior se muestran cuatro modelos diferentes de relación entre variables. En la primera la relación es lineal y positiva, a medida que se incrementan los valores de X se incrementan los valores de Y. La segunda representación muestra una relación lineal, intensa y negativa. A incrementos de X decrementos de Y. El tercer diagrama muestra la ausencia de relación entre X e Y; los valores no parecen ajustarse a un patrón definido que pueda explicar su relación. El cuarto diagrama muestra una relación intensa y clara entre X e Y; sin embargo esta relación no es lineal.

10.1.3 Estimación de la recta de regresión

La estimación de la recta de regresión tiene como objetivo ajustar una recta a la nube de puntos derivada de la representación gráfica de dos variables de modo que explique el máximo de la variabilidad observada en la variable criterio. Entre todas las posibles rectas que pueden ajustarse a la nube de puntos

Conceptos básicos y aplicaciones prácticas con Rcommander

buscamos aquella que genere el mínimo de errores de pronóstico al predecir los valores de Y a partir de los valores de X.



10.3Figura Regresión simple

Supongamos que un sujeto obtiene una puntuación X_i en un test, y una puntuación Y_i en el criterio. En el plano de la figura anterior puede apreciarse el punto (X_i, Y_i) . Si la relación entre X e Y es lineal es posible ajustar una recta que explique esa relación. Sin embargo, el ajuste de la recta nunca será perfecto y siempre habrá desviaciones entre los valores pronosticados (Y_i') y los obtenidos (Y_i). Esas desviaciones son errores de estimación que el modelo elegido intentará minimizar.

El procedimiento de estimación utilizado es la estimación mínimo cuadrática que minimiza el cuadrado del sumatorio de las desviaciones o errores,

$$SCE = \sum_{i=1}^N (Y_i - Y_i')^2$$

Resolviendo esa función, los parámetros de la recta de regresión quedarían definidos como:

Conceptos básicos y aplicaciones prácticas con Rcommander

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{Cov(X,Y)}{\sigma_X^2} = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$$

Donde \bar{X} es la media aritmética de la variable predictor,
 \bar{Y} , es la media aritmética de la variable criterio,
 ρ_{XY} , es la correlación entre X e Y,
 σ_X y σ_Y , son las desviaciones estándar del predictor y del criterio
 $Cov(X,Y)$, es la covarianza entre el predictor y el criterio.

Conocidos los parámetros de la recta de regresión, puede efectuarse la estimación puntual de Y'_i ; La recta de regresión puede formularse también del siguiente modo:

$$Y'_i = \bar{Y} - \rho_{XY} \frac{\sigma_Y}{\sigma_X} \bar{X} + \rho_{XY} \frac{\sigma_Y}{\sigma_X} X_i = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X_i - \bar{X}) + \bar{Y}$$

Donde \bar{X} es la media aritmética de la variable predictor,
 \bar{Y} , es la media aritmética de la variable criterio,
 ρ_{XY} , es la correlación entre X e Y,
 σ_X y σ_Y , son las desviaciones estándar del predictor y del criterio
 $Cov(X,Y)$, es la covarianza entre el predictor y el criterio

La ecuación de regresión puede obtenerse a partir de puntuaciones diferenciales o típicas:

$$y'_i = bx_i = \rho_{xy} \frac{\sigma_y}{\sigma_x} x_i$$

Donde x_i es la puntuación diferencial en la variable predictor,
 ρ_{XY} , es la correlación entre X e Y,
 σ_X y σ_Y , son las desviaciones estándar del predictor y del criterio

$$z'_y = \rho_{XY} z_x$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Donde x_i es la puntuación típica en la variable predictora,
 ρ_{XY} , es la correlación entre X e Y,

10.1.4 Error típico de estimación

El error de estimación es la diferencia entre el valor pronosticado (Y') y el valor obtenido (Y) y su desviación se conoce como error estándar de estimación ($\sigma_{Y.X}$).

$$\sigma_{Y.X} = \sigma_Y \sqrt{1 - \rho_{XY}^2}$$

Donde $\sigma_{Y.X}$ es el error estándar de estimación,
 ρ_{XY} , es la correlación entre X e Y
 σ_Y , es la desviación estándar de la variable criterio.

Si en lugar de puntuaciones directas se utilizaran puntuaciones típicas el error estándar de estimación se expresaría del siguiente modo:

$$\sigma_{z_{Y'}.z_X} = \sqrt{1 - \rho_{XY}^2}$$

Donde $\sigma_{z_{Y'}.z_X}$ error de estimación en puntuaciones típicas
 ρ_{XY} correlación entre X e Y.

10.1.5 Estimación del criterio

El error estándar de estimación permite estimar un intervalo de probabilidad en torno al valor pronosticado con un nivel de confianza predeterminado. Para ello se asume que el error estándar de estimación se ajusta a una distribución normal, que además es homocedástica, es decir, equivalente para todo el continuo de valores de la variable predictora (X). El proceso seguiría el procedimiento habitual de estimación:

- 1.- *Determinar el nivel de confianza (1- α).* Es habitual fijarlo en 95% o 99% ($\alpha=0,05$ y $\alpha=0,01$). Dada la simetría de la distribución normal los valores Z_k absolutos correspondientes a $\alpha/2$ y $(1-\alpha/2)$

Conceptos básicos y aplicaciones prácticas con Rcommander

son los mismos. Las áreas están limitadas por los valores críticos $z_c = \pm 1,96$ y $z_c = \pm 2,58$.

2.- *Estimación del error máximo.* Para ello es suficiente multiplicar el valor crítico ($Z_{\alpha/2}$) por el error típico de medida.

$$E_{Max} = |Z_K| s_e$$

3.- *Límites del intervalo para Y.* Los límites del intervalo se obtendrán añadiendo y sustrayendo el error máximo a la puntuación empírica correspondiente.

$$L_{Sup} = Y_i' + E_{Max}$$

$$L_{inf} = Y_i' - E_{Max}$$

Donde L_{sup} y L_{inf} son los límites superior e inferior del intervalo para la puntuación verdadera.

10.1.6 Descomposición de la varianza y coeficiente de determinación

La estimación mínimo cuadrática ajusta una recta a la nube de puntos, la recta que genera la mínima desviación de los errores cuadráticos. Sin embargo, el ajuste de la recta no garantiza que el modelo sea correcto para la interpretación de los datos; es necesario analizar el ajuste entre el modelo y los datos por medio de índices construidos con ese fin.

Según el modelo de regresión simple la puntuación criterio (Y) se descompone en dos componentes; la puntuación pronosticada (Y') y el error o desviación (e).

$$e = Y' - Y$$

$$0$$

$$Y = Y' + e$$

Conceptos básicos y aplicaciones prácticas con Rcommander

Dado que la correlación entre ambos elementos es nula ($\rho_{y'e} = 0$), la varianza de la variable compuesta Y es igual a la suma de las varianzas de sus componentes. Es decir, la varianza de la variable criterio (σ_y^2) se descompone en la varianza derivada del modelo de regresión ($\sigma_{y'}^2$) y la varianza debida a las desviaciones o errores de estimación ($\sigma_{y.x}^2$).

$$\sigma_y^2 = \sigma_{y'}^2 + \sigma_{y.x}^2$$

Utilizando el esquema ofrecido por el análisis de varianza podríamos resumir la información acerca de las fuentes de variación en los datos por medio del modelo general según el cual $\text{DATOS} = \text{AJUSTE} + \text{RESIDUAL}$. La variación total en la respuesta Y se expresaría por las desviaciones,

$$(Y_i - \bar{Y}) = (Y_i' - \bar{Y}') + (Y_i - Y_i')$$

Las puntuaciones individuales varían en torno a su media a causa de la variación dentro de la subpoblación de respuestas para un valor determinado de X_i . Esta variación está representada por los residuales $Y_i - Y_i'$ (diferencias entre el valor real y el pronosticado) que registran la dispersión de las observaciones en torno a la recta ajustada y por las diferencias asociadas a que las puntuaciones Y_i corresponden a diferentes valores de la variable predictora X . El valor ajustado Y_i' estima la media de respuesta para un valor determinado X_i específico. Las diferencias de estos valores respecto a la media reflejan la variación en la media de respuesta debida a las diferencias en X_i . La desviación global de cualquier Y_i respecto a la media es la suma de esos dos componentes de desviación.

Si se elevan al cuadrado cada una de las tres desviaciones y se suman para las N observaciones, se obtiene:

Conceptos básicos y aplicaciones prácticas con Rcommander

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i' - \bar{Y})^2 + \sum (Y_i - Y_i')^2$$

(1) (2) (3)

Los términos (1), (2) y (3) son las sumas de cuadrados total, del modelo y del error respectivamente; SCT=SCM+SCE. El término error indica las desviaciones a la recta y se traduce como residuales o fuente de variación no explicada. Esta partición de la variación de los datos en dos fuentes es el núcleo del análisis de varianza. Los grados de libertad asociados con cada suma de cuadrados son $N-1$ para la suma de cuadrados total y I y $N-2$ para la suma de cuadrados del modelo y del error respectivamente. Se relacionan aditivamente:

$$gl_t = gl_m + gl_e.$$

Para cada fuente, la razón entre la suma de cuadrados y los grados de libertad es la media cuadrática (MC). La media cuadrática total MCT, se correspondería con la varianza muestral. La media cuadrática del error MCE sería el estimador de σ^2 , la varianza en torno a la recta de regresión.

La hipótesis nula $H_o : \beta_1 = 0$ puede ponerse a prueba mediante la comparación de MCM con MCE, con el estadístico F .

$$F = \frac{MCM}{MCE}$$

Cuando H_o es verdadera, este estadístico sigue una distribución F con 1 grado de libertad en el numerador y $N-2$ grados de libertad en el denominador, que son los grados de libertad correspondientes a MCM y MCE. Cuando $\beta_1 \neq 0$, MCM tiende a ser alto en relación con MCE. Por tanto valores altos de F proporcionan evidencia contra H_o y en favor de la hipótesis alterna. Este test es equivalente al test t para el parámetro β_0 aunque es preferible el primero porque permite plantear hipótesis alternas bilaterales.

Conceptos básicos y aplicaciones prácticas con Rcommander

La tabla ANOVA para la regresión simple sigue el siguiente formato estándar:

Fuente	Grados de libertad	Sumas de cuadrados	Media cuadrática	F
Modelo	1	$\sum (Y'_i - \bar{Y})^2$	SCM/gl _m	MCM/MCE
Error	N-2	$\sum (Y_i - Y'_i)^2$	SCE/gl _e	
Total	N-1	$\sum (Y_i - \bar{Y})^2$	SCT/gl _t	

Coefficiente de determinación

La descomposición de la varianza es el origen de un índice adimensional, el coeficiente de determinación, que expresa el porcentaje de variabilidad de la variable criterio que es explicada por la variable predictora.

$$\frac{\sigma_y^2}{\sigma_y^2} = \frac{\sigma_{y'}^2}{\sigma_y^2} + \frac{\sigma_{y \cdot x}^2}{\sigma_y^2} = 1$$

El primer componente de la descomposición anterior es el coeficiente de determinación, R². Cuando la relación entre X e Y es perfecta, es decir, en ausencia de errores, el coeficiente de determinación alcanzará el valor 1. Este coeficiente es el índice más utilizado en la determinación del grado de aplicabilidad del modelo.

$$R^2 = \frac{SCE}{SCT} = \frac{\sigma_{y'}^2}{\sigma_y^2}$$

$$R^2 = 1 - \frac{\sigma_{y \cdot x}^2}{\sigma_y^2}$$

Los límites del coeficiente de determinación son 0 y 1. Cuando la relación entre dos variables es lineal, el coeficiente es equivalente al cuadrado de

Conceptos básicos y aplicaciones prácticas con Rcommander

la correlación de Pearson entre las variables (ρ_{XY}^2). A medida que aumenta la dispersión de los valores en torno a la recta de regresión disminuirá el valor del coeficiente de determinación.

$$R^2 = \rho_{XY}^2$$

Además del coeficiente de determinación en el marco de la teoría clásica de tests se han desarrollado varios índices de adecuación que son menos utilizados; entre ellos.

Coeficiente de alienación

$$K = \sqrt{1 - \rho_{XY}^2} = \frac{\sigma_{Y.X}}{\sigma_Y}$$

Coeficiente del valor predictivo

$$E = 1 - \sqrt{1 - \rho_{xy}^2}$$

10.1.7 Modelo lineal

El modelo de regresión simple se construye sobre el modelo lineal que asume además del carácter lineal de la relación entre dos variables los siguientes supuestos:

a.-La esperanza matemática de los errores es 0

$$E(e_{ij}) = 0$$

b.- Homocedasticidad. La variabilidad observada en torno a la recta de regresión es constante para todos los valores de la variable predictora; es decir, $\sigma_{Y \text{ Xi}}$ o σ_e^2 son constantes.

Conceptos básicos y aplicaciones prácticas con Rcommander

c.- Independencia entre los errores y los valores observados.

$$E(e_{ij}, e_{ih}) = 0, \quad j \neq h$$

El conjunto formado por esas tres asunciones son conocidas como las condiciones de Gauss-Markov.

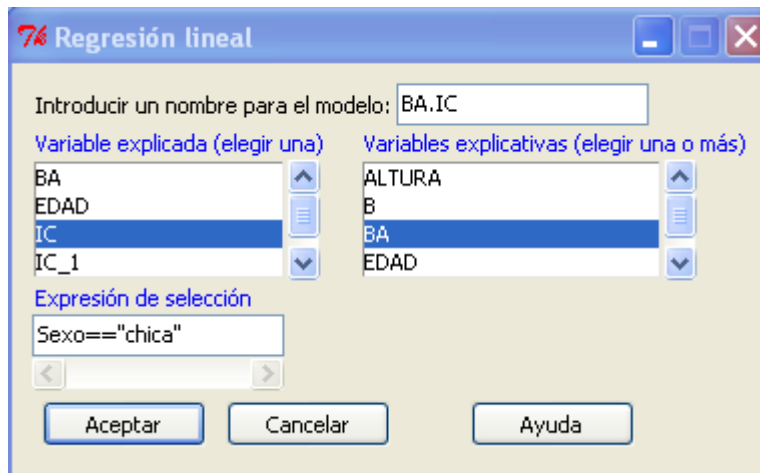
d.- Normalidad de la distribución. Aunque este supuesto no afecta a la estimación mínimo cuadrática es importante evaluarlo cuando el objetivo es la inferencia.

$$Y_{ij} \sim \left(\mu_{Y|X_j}, \sigma_{Y|X_j}^2 \right)$$
$$e_{ij} \sim N(0, \sigma_e^2)$$

10.2 Regresión simple con Rcommander

Como ejemplo de regresión simple se intentará predecir la Insatisfacción corporal a partir de la variable predictora Baja Autoestima. Para ello trabajaremos únicamente con una submuestra formada por mujeres y aplicaremos el modelo de regresión lineal según la siguiente secuencia: Estadísticos > Ajuste de modelos > Regresión Lineal. En la ventana de diálogo que nos abre esta opción se seleccionaran la variable predictora o explicativa (Baja autoestima –BA-) y la variable explicada o variable criterio (Insatisfacción corporal –IC-). La ventana “Expresión de selección” se cumplimentará con la condición deseada; “Expresión de selección: `Sexo=="chica"`. Junto a estas especificaciones Rcommander solicita que se asigne un nombre al modelo “Introducir un nombre para el modelo”. Nosotros lo denominamos BA.IC.

Conceptos básicos y aplicaciones prácticas con Rcommander



10.4Figura. Regresión simple

Esos argumentos generan la siguiente función en la ventana de comandos, `lm` (*linear model*). El símbolo “~” puede interpretarse como “IC explicada por medio de BA”.

```
lm(formula = IC ~ BA, data = Edi.data, subset = SEXO == "chica")
```

En su forma más simple la función `lm` ofrece la siguiente salida.

```
Residuals:
    Min       1Q   Median       3Q      Max
-18.9052  -6.8379  -0.3772   6.1621  25.7014

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.3772     0.6516   12.86 <2e-16 ***
BA             1.4607     0.1036   14.10 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.548 on 446 degrees of freedom
(69 observations deleted due to missingness)
Multiple R-squared:  0.3084,    Adjusted R-squared:  0.3068
F-statistic: 198.9 on 1 and 446 DF, p-value: < 2.2e-16
```


Conceptos básicos y aplicaciones prácticas con Rcommander

En primer lugar se ofrece información sobre la distribución de los residuales (`Residuals`). Por definición la media aritmética de los errores es 0 y según la salida de resultados la mediana de la distribución no se aleja de ese valor (-0,37). Además se ofrece información sobre el primer y el tercer cuartil de la distribución que en valores absolutos y dada la simetría de la distribución han de ser equivalentes. Los cuartiles estimados son -6,8 y 6,16.

A continuación se ofrecen las estimaciones de los parámetros de la recta de regresión (`Coefficients`). El valor interceptal o parámetro *a* es (`Intercept`) 8,37 y la pendiente de la recta es 1,4607. Según esos parámetros la recta de regresión vendría dada por,

$$IC = 8,37 + 1,46 \times BA$$

Junto a los parámetros estimados se ofrecen los errores de estimación para cada parámetro (`Std. Error`), los valores *t* asociados (`t value`) y sus correspondientes probabilidades (`Pr(>|t|)`). Los asteriscos junto a los valores ayudan a la interpretación de los niveles de significación, que vienen explicados al final de la tabla de coeficientes; por ejemplo un asterisco significa que el nivel de significación se sitúa entre los valores 0,01 y 0,05.

Rcommander ofrece también el error estándar de estimación (`Residual Standard Error`), que se utiliza para la determinación de intervalos de probabilidad en torno a la puntuación pronosticada. Para ello es suficiente aplicar la siguiente fórmula,

$$Y' \pm Z \times S_{YX}$$

Por ejemplo, una chica con un nivel de autoestima de 12 puntos se le estimaría una puntuación en la variable Insatisfacción corporal igual a,

$$Y' = 8,37 + 1,46 \times 12 = 25,89$$

Psicometría.

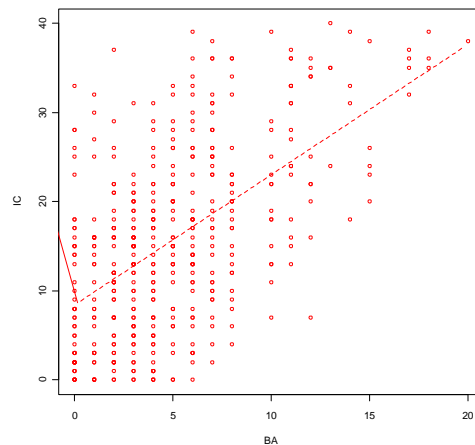
Conceptos básicos y aplicaciones prácticas con Rcommander

Es una estimación puntual. Con un nivel de confianza del 95% se podría afirmar que la puntuación en Insatisfacción Corporal de esa persona se encuentra en el intervalo $25,89 \pm 1,96 \times 8,548$. Si se quisieran generar intervalos de probabilidad para cada uno de los sujetos bastaría con crear dos nuevas variables que marcaran los límites superior e inferior del intervalo tal y como se ha mostrado en la estimación de la puntuación verdadera.

La salida informa que ha habido 69 casos excluidos de los análisis por presentar valores faltantes en alguna de las variables implicadas en el modelo. El coeficiente de determinación del modelo, `Multiple R squared` es 0,3084, lo cual significa que la baja autoestima pronosticaría un 30,84% de la insatisfacción corporal.

Por último `Rcommander` ofrece el estadístico F que evalúa la hipótesis nula de que el coeficiente de regresión sea 0. En el caso de la regresión simple este test no aporta información alguna al test de Student ejecutado para el parámetro de regresión. De hecho, F es el cuadrado de t $198,9 = 14,10^2$. Esta igualdad se cumple en todos los modelos con 1 grado de libertad.

Para obtener una representación gráfica del modelo (dispersión y recta de ajuste) la opción sería `Gráficas > Diagrama de Dispersión`



10.5 Figura. Baja autoestima e Insatisfacción Corporal

Rcommander ofrece además intervalos de probabilidad para los parámetros de cualquier modelo; para ello basta seleccionar de la barra de menús la opción `Modelos>Intervalos de confianza`, y especificar en la ventana de diálogo que se abrirá, el nivel de confianza deseado.

	2.5 %	97.5 %
<i>(Intercept)</i>	6.971232	9.571538
BA	1.278205	1.690892

10.3 Regresión múltiple

10.3.1 Modelo

La predicción de variables en el campo de las ciencias sociales es un proceso complejo en el que intervienen un conjunto de variables predictoras cuyo peso en el pronóstico es necesario evaluar. Ese es el objetivo del modelo de regresión múltiple, pronosticar los valores en una variable (la variable dependiente o variable criterio) a partir de un conjunto de variables predictoras por medio de un modelo entre cuyos parámetros se establece una relación lineal.

En un modelo de regresión múltiple intervienen los siguientes elementos:

Conceptos básicos y aplicaciones prácticas con Rcommander

- Variable criterio o variable a predecir (Y). Es la variable dependiente.
- Variables predictoras ($X_1, X_2, X_3, \dots, X_k$).
- Pesos o ponderaciones ($B_1, B_2, B_3, \dots, B_k$): Cuantifican el peso de las variables predictoras sobre la variable criterio.

Esos elementos definen la ecuación de predicción de la regresión múltiple que podría expresarse del siguiente modo:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_kX_k + \varepsilon$$

Donde Y es el criterio,
 ε , error de predicción,
B, los parámetros del modelo (pesos)
 X_i , variables predictoras.

En función del modelo el valor de un sujeto en la variable criterio dependería de K variables predictoras, K+1 pesos de regresión (B), y un error aleatorio (ε). Utilizando el algebra matricial el modelo podría expresarse del siguiente modo:

$$\mathbf{Y} = \mathbf{XB} + \varepsilon$$

Donde \mathbf{Y} es el vector de puntuaciones de dimensiones (N×1),
 \mathbf{X} , es la matriz de puntuaciones de las variables predictoras, de dimensiones N×(K+1), en la que la primera columna está formada por 1.
 \mathbf{B} , es el vector de parámetros del modelo de dimensiones ((K+1)×1) y
 ε , es el vector de errores aleatorios (N×1).

Es decir,

Conceptos básicos y aplicaciones prácticas con Rcommander

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix} \quad B = \begin{bmatrix} B_0 \\ B_1 \\ \cdot \\ \cdot \\ B_k \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1k} \\ 1 & X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{N1} & X_{N2} & \cdot & \cdot & \cdot & X_{Nk} \end{bmatrix}$$

En el caso de N sujetos y 2 variables predictoras necesitaríamos N ecuaciones,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & X_{N1} & X_{N2} \end{bmatrix} \times \begin{bmatrix} B_0 \\ B_1 \\ B_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix}$$

Una para cada sujeto,

$$\begin{aligned}
 Y_1 &= B_0 + B_1 X_{11} + B_2 X_{12} + \varepsilon_1 \\
 Y_2 &= B_0 + B_1 X_{21} + B_2 X_{22} + \varepsilon_2 \\
 K &K K K K K K K K K K K \\
 K &K K K K K K K K K K K \\
 Y_N &= B_0 + B_1 X_{N1} + B_2 X_{N2} + \varepsilon_N
 \end{aligned}$$

Cuando el modelo general es estimado en una muestra, la ecuación de regresión se representa del siguiente modo,

$$y = Xb + e$$

Donde **y** es el vector criterio,
e, es el vector residual,
b, es la matriz de pesos estimados
X, es la matriz de variables predictoras.

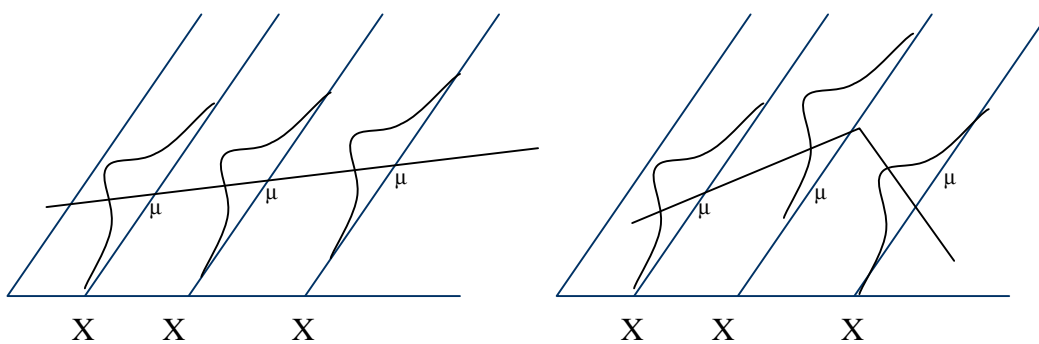
10.3.2 Supuestos

Los supuestos del modelo de regresión múltiple hacen referencia a los residuales, y dada la relación entre estos y la variable criterio pueden ser expresados en función de ésta última:

a.- La linealidad es el supuesto principal del modelo. Las medias aritméticas de Y son una función lineal de los parámetros del modelo. En la representación siguiente la figura de la izquierda cumple el supuesto de linealidad y la figura de la derecha no.

$$\mu_{Y|X_1, X_2, \dots, X_k} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

10.6 Figura. Regresión múltiple. Linealidad



b.- Homocedasticidad: Para todo X las varianzas de Y son iguales,

$$\sigma^2_{Y|X_1, X_2, \dots, X_k} = \sigma^2$$

Conceptos básicos y aplicaciones prácticas con Rcommander

c.- Independencia: Cualquier y_i de Y es independiente de cualquier otro y_j .

d.- Normalidad: Para cada X la distribución de valores de Y es normal.

$$Y \sim N(\mu_{Y|X_1, X_2, \dots, X_k}; \sigma^2)$$

Todos estos supuestos pueden resumirse en la siguiente especificación expresada en términos del error:

$$\varepsilon \sim N_k(\underline{0}, \sigma^2 I_k)$$

Es decir, los errores siguen una distribución normal de orden k, donde la media es un vector nulo ($E(\varepsilon)=0$) y la matriz de varianzas-covarianzas es ($\sigma^2 I_N$), siendo I_N una matriz identidad de orden N. En esa matriz los elementos fuera de la diagonal son 0 (las covarianzas entre errores son 0), ya que los errores son independientes, y los elementos de la diagonal principal (varianzas de error) son iguales (homocedasticidad).

Si en lugar de por medio de los errores exponemos los supuestos del modelo de regresión múltiple en función de la variable criterio, obtendríamos

$$Y \approx N_N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_N)$$

10.3.3 Estimación de los parámetros

El método de estimación de parámetros por mínimos cuadrados minimiza el cuadrado de los errores de estimación. Según este procedimiento de estimación la ecuación matricial de obtención del vector de parámetros sería:

Conceptos básicos y aplicaciones prácticas con Rcommander

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Donde \mathbf{b} es el vector de pesos estimados,
 \mathbf{X} , es la matriz sujetos×variables predictoras con unos en la primera columna
 \mathbf{X}' , es la matriz transpuesta de \mathbf{X}
 $(\mathbf{X}'\mathbf{X})^{-1}$, es la matriz inversa de $(\mathbf{X}'\mathbf{X})$
 \mathbf{y} , es el vector de puntuaciones en la variable criterio.

La ecuación de predicción viene dada por,

$$\mathbf{y}' = \mathbf{X}\mathbf{b}$$

Y el vector de residuales,

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

Si se utilizan puntuaciones diferenciales en lugar de puntuaciones brutas, el parámetro B_0 es 0 y la ecuación de predicción se expresaría como:

$$\mathbf{b} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}$$

Donde \mathbf{b} , es el vector de pesos,
 \mathbf{C}_{xx}^{-1} , es la inversa de la matriz de varianzas-covarianzas de las variables predictoras
 \mathbf{C}_{xy} , es el vector de covarianzas entre las variables predictoras y el criterio.

La utilización de puntuaciones estandarizadas por su parte, situaría a todas las variables en la misma escala facilitando así la interpretación de los pesos, a los que se suele referir como pesos estandarizados β :

$$\beta = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xy}$$

Donde β es el vector de pesos estandarizados
 \mathbf{R}_{xx}^{-1} , es la inversa de la matriz de correlaciones entre variables predictoras,
 \mathbf{R}_{xy} , es el vector de correlaciones entre las variables predictoras y el criterio.

10.3.4 Descomposición de la varianza

Conceptos básicos y aplicaciones prácticas con Rcommander

En el modelo de regresión lineal simple el test F de la tabla ANOVA es equivalente al test bilateral t de hipótesis respecto de β_1 . En el caso de la regresión múltiple también hay un test F de ANOVA pero pone a prueba la hipótesis de que todos los coeficientes de regresión (a excepción del intercepto) son 0. La forma general de la tabla ANOVA para la regresión múltiple es la siguiente:

Fuente	Grados de libertad	Sumas de cuadrados	Media cuadrática	F
Modelo	K	$\sum (Y'_i - \bar{Y})^2$	SCM/gl_m	MCM/MCE
Error	$N-K-1$	$\sum (Y_i - Y'_i)^2$	SCE/gl_e	
Total	$N-1$	$\sum (Y_i - \bar{Y})^2$	SCT/gl_t	

Los grados de libertad del modelo se incrementan de 1 a K reflejando el hecho de que se cuenta con K variables explicativas. En consecuencia los grados de libertad del error disminuyen en la misma cantidad. Las sumas de cuadrados representan fuentes de variación. De nuevo se cumple que tanto las sumas de cuadrados como los grados de libertad son aditivos ($SCT=SCM+SCE$ y $gl_t=gl_m+gl_e$). El estimador de la varianza σ^2 para nuestro modelo de nuevo viene dado por MCE de la tabla ANOVA; $s^2=MCE$.

La razón MCM/MCE es un estadístico F que sirve para verificar la hipótesis nula

$$H_o : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

Contra la hipótesis alterna

$$H_1 : \text{al menos un } \beta_j \text{ es distinto de cero}$$

Conceptos básicos y aplicaciones prácticas con Rcommander

La hipótesis nula mantiene que ninguna de las variables explicativas son predictoras de la variable respuesta. La alternativa establece que al menos una de ellas está linealmente relacionada con la respuesta. Como en la regresión simple, valores altos de F proporcionan evidencia contra H_0 . Cuando H_0 es cierta, F sigue una distribución $F(K, N-K-1)$. Sus grados de libertad son los asociados con el modelo y el error de la tabla ANOVA.

10.3.5 Coeficiente de determinación múltiple

El coeficiente de determinación es el indicador más utilizado para la evaluación del ajuste del modelo y por tanto para la evaluación de su aplicabilidad. Se obtiene a partir de la descomposición de la varianza en sumas de cuadrados, siendo el coeficiente de determinación igual al primer sumando de la descomposición:

$$\frac{SCT}{SCT} = \frac{SCM}{SCT} + \frac{SCE}{SCT} = 1$$

Se interpreta como la proporción de varianza estimada en el criterio por todas las variables implicadas en la predicción; Sus valores están limitados por el máximo teórico de 1 y el mínimo teórico de 0.

$$R^2 = \frac{SCM}{SCT} = \frac{\sum (Y_i' - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

También se puede expresar el coeficiente de determinación en función de la varianza de error,

$$R^2 = 1 - \frac{SCE}{SCT}$$

Conceptos básicos y aplicaciones prácticas con Rcommander

En el modelo de regresión múltiple el valor del coeficiente de determinación es igual al cuadrado del coeficiente de correlación múltiple entre las variables predictoras y el criterio.

La representación matricial del coeficiente de determinación vendría dada por:

$$R^2 = \rho_{y'y}^2 = \frac{\mathbf{b}'\mathbf{C}_{xy}}{S_y^2} = \boldsymbol{\beta}'\mathbf{R}_{xy}$$

Donde \mathbf{b}' es la matriz transpuesta de los pesos de las variables predictoras,
 \mathbf{C}_{xy} , es el vector de covarianzas entre las variables predictoras y el criterio,
 S_y^2 , es la varianza de la variable criterio
 $\boldsymbol{\beta}'$, es el vector transpuesto de pesos beta
 \mathbf{R}_{xy} , es el vector de correlaciones entre las variables predictoras y el criterio.

La correlación múltiple o raíz cuadrada del coeficiente de determinación, siempre tendrá valores positivos ($0 \leq R_{y'y}^2 \leq 1$), y puede interpretarse como el *coeficiente de validez* de un conjunto de variables –tests- predictoras. Su valor nunca será menor que el coeficiente de correlación de cualquiera de las variables predictoras con el criterio.

$$R_{y'y} \geq \max |r_{yj}|$$

El coeficiente de determinación ajustado

El coeficiente de determinación es muy sensible al número de variables predictoras; a medida que aumentan éstas lo hace su valor. Para compensar esa característica se utiliza el coeficiente de determinación corregido o ajustado. El coeficiente de determinación ajustado tiene en cuenta el tamaño de la muestra y el número de variables independientes. El coeficiente de determinación ajustado es especialmente útil para comparar modelos con diferente número de observaciones o variables independientes y su aplicación es especialmente aconsejada en los estudios que incluyen tamaños muestrales reducidos. Su fórmula en función de R^2 es:

$$R_{aj}^2 = 1 - \frac{(N-1)(1-R^2)}{N-K-1}$$

Donde R^2 es el coeficiente de determinación
N es el número de observaciones
K el número de variables independientes.

Cuando el número de observaciones es pequeño o bien el número de variables grande, el valor del ajuste R_a^2 reduce el valor de R^2 , incrementándose para n grande o p pequeño.

Los coeficientes de determinación sin corregir y el ajustado tienen el mismo valor cuando el modelo de regresión utilizado es el modelo simple. A medida que aumenta el número de variables predictoras el coeficiente ajustado va disminuyendo su valor con relación al coeficiente no corregido. El coeficiente de determinación ajustado aumenta a medida que disminuye la varianza residual, lo cual lo convierte en un índice más adecuado que el coeficiente de determinación no corregido.

$$\sigma_e^2 = (1 - R_{aj}^2) \sigma_Y^2$$

10.3.6 Inferencias sobre el modelo

Una vez estimados los parámetros de regresión es importante evaluar su significación estadística. El estudio de la significación se lleva a cabo en dos niveles diferentes; el primero evalúa la significación de todo el modelo, y el segundo se centra en la significación de cada uno de los parámetros de regresión estimados.

La evaluación total del modelo descansa en la evaluación del coeficiente de correlación múltiple. La hipótesis nula a contrastar iguala a 0 el valor del coeficiente de correlación al cuadrado, y se evalúa por medio de un estadístico que sigue una distribución F con k y $N-1$ grados de libertad.

$$F = \left(\frac{N - k - 1}{k} \right) \left(\frac{R_{y'y}^2}{1 - R_{y'y}^2} \right)$$

Obtener intervalos de confianza para R² no es sencillo mediante procedimientos analíticos; Recientemente se han propuesto procedimientos de remuestreo (Simon, 1969) para obtenerlos (López-Jáuregui y Elosua, 2004).

Significación de los predictores

Se pueden obtener intervalos de probabilidad y tests de significación para cada uno de los coeficientes de regresión (β_i) del modelo estimado. Para cada uno de ellos la hipótesis nula a contrastar sería:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

que se verifica por medio del estadístico t que sigue una distribución de Student con $N - K - 1$ grados de libertad.

$$t = \frac{b_i}{ES_{b_i}}$$

Donde b_i es el parámetro de la regresión
 ES_{b_i} es el error estándar del estimador que puede estimarse por medio de,

$$\sigma_{b_i} = \frac{\sigma_y}{\sigma_i} \sqrt{\frac{1 - R_{y'y}^2}{(1 - R_{i,y}^2)(N - k - 1)}}$$

A partir del valor del parámetro y de su error estándar de estimación es posible establecer intervalos de probabilidad para los coeficientes de regresión estimados.

Conceptos básicos y aplicaciones prácticas con Rcommander

1.- *Determinar el valor crítico t_c* en la distribución t de Student en función del número de grados de libertad (N-K-1) y el nivel de confianza prefijado.

2.- *Estimar el error máximo.* Para ello es suficiente multiplicar el valor crítico (t_c) por el error de estimación.

$$E_{Max} = |t_c| ES_{bi}$$

3.- *Límites del intervalo para b.* Los límites del intervalo se obtendrán añadiendo y sustrayendo el error máximo al valor correspondiente.

$$L_{Sup} = b'_i + E_{Max}$$

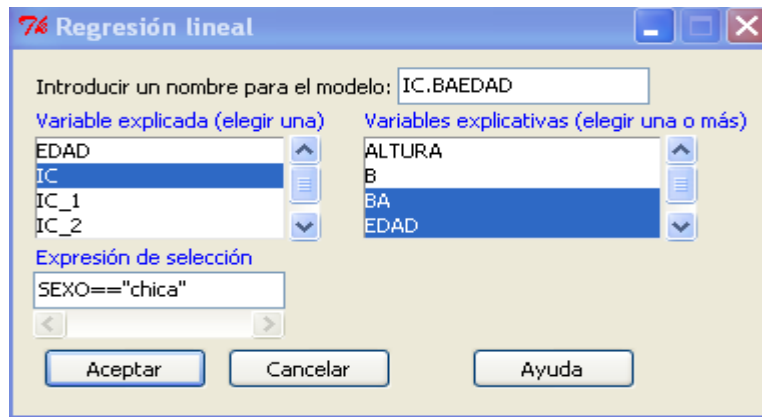
$$L_{inf} = b'_i - E_{Max}$$

Donde L_{sup} y L_{inf} son los límites superior e inferior del intervalo para el coeficiente regresor b_i .

10.4 Correlación múltiple con Rcommander

La utilización de Rcommander en la estimación de un modelo de regresión múltiple se ilustrará evaluando un modelo que predice la Insatisfacción Corporal a partir de las variables Baja Autoestima y Edad. Las opciones a seleccionar de la barra de menús son: Estadísticos>Ajuste de modelos>Regresión Lineal. En este ejemplo la variable pronosticada (criterio) es la insatisfacción corporal (IC) y las variables predictoras son la edad y la baja autoestima (EDAD, BA). Utilizaremos para el análisis una submuestra del conjunto de datos formada íntegramente por mujeres.

Conceptos básicos y aplicaciones prácticas con Rcommander



10.7Figura. Regresión múltiple con Rcommander

Esas especificaciones llaman a la función genérica `lm`

```
IC.BAEDAD <- lm(IC~BA+EDAD, data=Edi.data, subset=SEXO=="chica")
```

En ella se especifica que se desede explicarse la variabilidad en Insatisfacción Corporal (IC) por medio de 2 variables, la baja autoestima (BA) y la edad (EDAD). El modelo de regresión generará un objeto que ha sido asignado a IC.BAEDAD.

```
Call:
lm(formula = IC ~ BA + EDAD, data = Edi.data, subset = SEXO ==
    "chica")

Residuals:
    Min       1Q   Median       3Q      Max
-19.8991  -6.9437  -0.4725   5.7241  25.0760

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.0660     2.4562   1.248  0.2126
BA           1.4626     0.1040  14.066 <2e-16 ***
EDAD         0.3490     0.1563   2.233  0.0261 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conceptos básicos y aplicaciones prácticas con Rcommander

```
Residual standard error: 8.522 on 438 degrees of freedom
(76 observations deleted due to missingness)
Multiple R-squared: 0.3177, Adjusted R-squared: 0.3146
F-statistic: 102 on 2 and 438 DF, p-value: < 2.2e-16
```

En primer lugar, la salida muestra información sobre la distribución de los residuales para este modelo. Los valores mostrados permiten obtener una primera impresión sobre la simetría de los errores en torno a su valor medio teórico. La mediana de la distribución es -0,4725, y el primer y tercer cuartil de la distribución tienen como valores -6,9437 y 5,7241.

A continuación se muestran los valores estimados de los coeficientes de regresión (Estimate), su error estándar (Std. Error) y las pruebas de significación asociadas (t value; $pr(>|t|)$). Las dos variables incluidas en el modelo presenten coeficientes significativos al 5%. Los niveles de significación para las variables baja autoestima ($p<0,001$) y Edad ($p=0,0278$) muestran que su inclusión en el modelo aporta información relevante sobre la variable a pronosticar. Es importante interpretar correctamente el significado de las pruebas para cada coeficiente. El valor t ofrece información sobre el cambio sufrido por el modelo si se elimina el efecto de la variable que se está evaluando y se deja el resto de variables tal y como están definidas en el modelo. La prueba t no permite evaluar la significación de las variables en un supuesto modelo reducido.

El contraste de la hipótesis nula se efectúa utilizando el estadístico t.

$$t = \frac{b_i}{ES_{b_i}} = \frac{1,46}{0,10} = 14,6$$

Para estimar un intervalo de probabilidad para los parámetros estimados se utiliza $(b_j \pm tcES_{b_j})$. Donde ES_{b_j} es el error de estimación del parámetro y tc es el valor crítico de la distribución de Student con $(N-K-1)$ grados de libertad

Conceptos básicos y aplicaciones prácticas con Rcommander

(Número de sujetos- variables predictoras-1). El intervalo de probabilidad para el estimador de la Baja Autoestima sería:

$$1,462+1,64\times 0,1040=1,46\pm 0,17$$

Para obtener el valor crítico de la distribución de Student (1,64) en Rcommander la opción es Distribuciones>Distribuciones continuas>Distribución t>Cuantiles t, en ella se seleccionará al nivel de confianza (0,95) y se especificaran los grados de libertad (517chicas-76 valores faltantes-2 variables predictoras - 1=497).

Es posible obtener los intervalos de probabilidad directamente en Rcommander fijando el nivel de confianza deseado en la ventana que se abrirá utilizando las opciones Modelos>Intervalos de confianza.

```
> Confint(IC.BAEDAD, level=.95)
              2.5 %      97.5 %
(Intercept) -1.76143597  7.8935052
BA           1.25823896  1.6669811
EDAD        0.04179919  0.6561753
```

Por otro lado la evaluación conjunta del modelo es significativa (F(2,4355) = 103,1; p<0,001)) es decir la información aportada por las variables incluidas permite predecir la variabilidad mostrada en la Insatisfacción corporal.

Es importante combinar la información ofrecida por las pruebas de significación t para cada coeficiente con la tabla ANOVA para la regresión. De hecho, es fundamental estudiar la tabla ANOVA del modelo de regresión múltiple. Para ello las opciones de Rcommander serían Modelos > Test de hipótesis > Tabla ANOVA.

```
Anova Table (Type II tests)
```

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

```
Response: IC
          Sum Sq  Df  F value  Pr(>F)
BA          14369   1 197.8417 < 2e-16 ***
EDAD         362   1   4.9855 0.02606 *
Residuals  31811 438
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La información aportada reitera en este caso el resultado anterior. La incorporación de las dos variables predictoras es relevante en el pronóstico de la variable dependiente Insatisfacción Corporal.

Con las estimaciones efectuadas se puede reflejar la ecuación de predicción del siguiente modo:

$$IC=3,06+1,46\times BA+0,34\times EDAD$$

De acuerdo al modelo, el valor pronosticado en Insatisfacción Corporal a una chica de 13 años que tiene un nivel de autoestima igual a 13, sería

$$IC=3,06+1,46\times 5+0,34\times 13=14,78$$

El valor del coeficiente de determinación (Multiple R squared) del modelo estimado es 0,3177 y el coeficiente de determinación ajustado (Adjusted R-squared) tiene un valor de 0,3146.

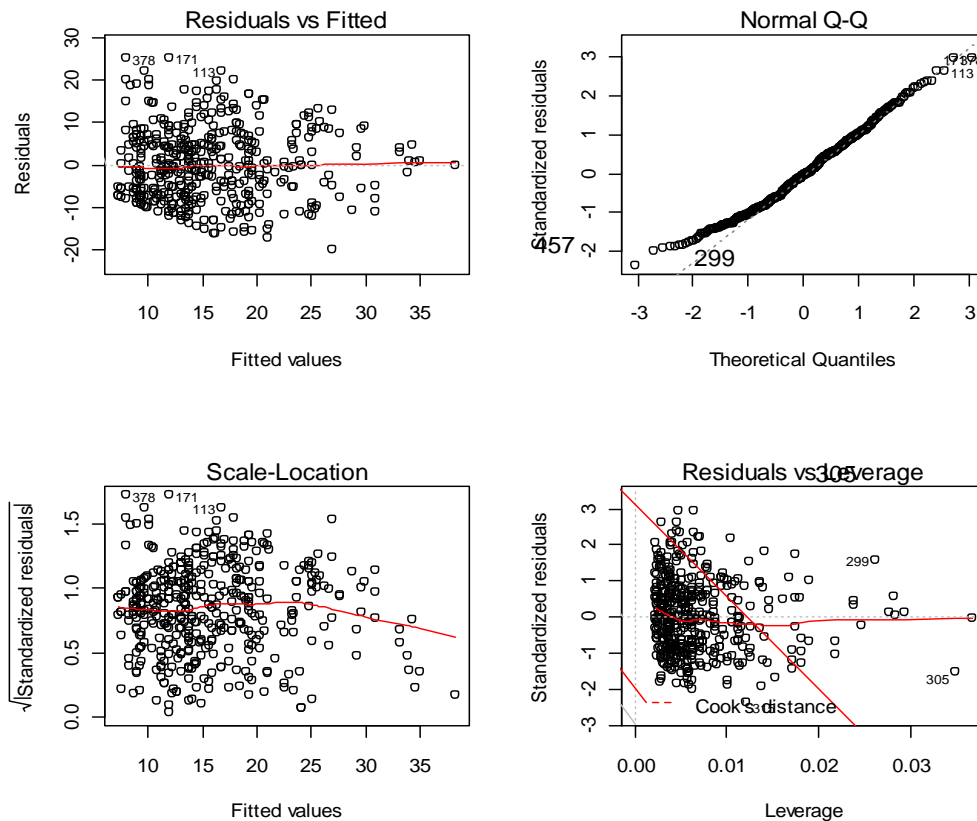
10.4.1 Diagnósticos del modelo de regresión

Rcommander incorpora varias funciones gráficas de gran utilidad en el estudio del cumplimiento de los supuestos en que se basa la regresión múltiple.

Conceptos básicos y aplicaciones prácticas con Rcommander

Para acceder a ellos es suficiente seleccionar la opción Modelos > Gráficas>Gráficas básicas de diagnóstico.

lm(IC ~ BA + EDAD)



10.8 Figura. Diagnóstico gráfico del modelo de regresión

- **Residuos y valores ajustados** (Residuals vs Fitted) Muestra el diagrama de dispersión referido a los valores pronosticados por el modelo y a los valores residuales. La existencia de patrones no aleatorios podría indicar correlaciones significativas entre el modelo y los residuales lo cual incumpliría una de los supuestos del modelo de regresión.
- **Normalidad de los residuos** (Normal QQ) La distribución normal de los errores puede evaluarse por medio del plot de cuantiles de los

Conceptos básicos y aplicaciones prácticas con Rcommander

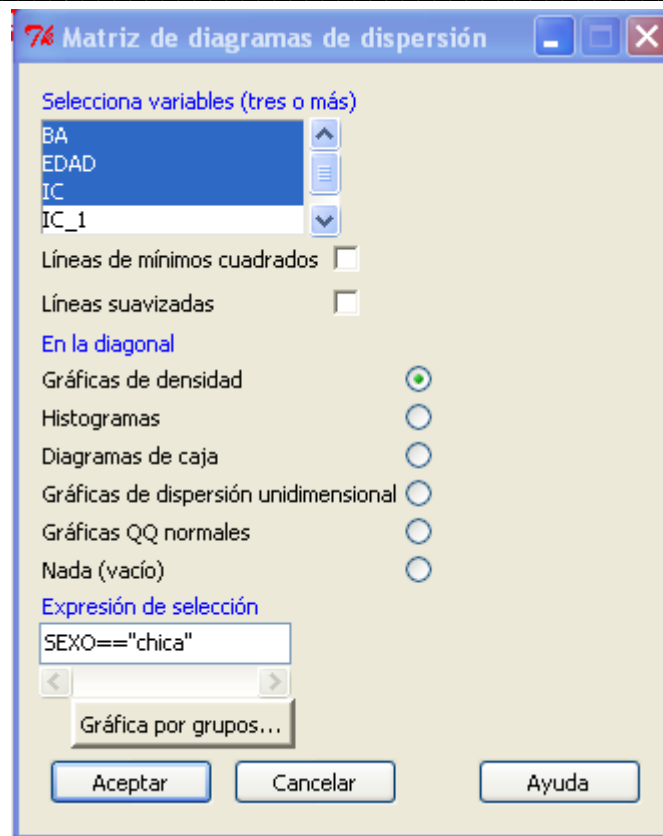
residuales. Los puntos en la gráfica deberían de caer sobre la línea QQ si los residuales se ajustaran a una distribución normal. En este caso parece apropiada la asunción de normalidad. Los puntos que se desvían de la hipótesis son representados con etiquetas que los identifican.

- **Escala-posición** (*Scale-Location*). Representa los valores ajustados frente a la raíz del valor absoluto de los residuos. A diferencia de la primera gráfica en esta se analizan los valores absolutos de los residuos para analizar su magnitud independientemente de su sentido, y se estudian sus raíces con el objetivo de disminuir la asimetría. Facilita el estudio de la homocedasticidad.
- **Distancias de Cook** (*Residuals vs Leverage*). Útil para analizar la influencia de los datos en la estimación de los parámetros del modelo. La distancia de Cook refleja la influencia de cada uno de los valores de la muestra y ofrece la distancia por la que el vector de parámetros estimados se movería si ese valor se quitara de los datos.

10.4.2 Representación gráfica de distribuciones multivariadas

El estudio de la distribución conjunta de variables es una herramienta de análisis exploratorio muy útil que ofrece un apoyo ineludible en la determinación de cualquier modelo matemático. Para ello Rcommander ofrece la opción *Matriz de diagramas de dispersión* dentro del menú *Gráficas*. La ventana a la que se accede por medio de esa selección permite seleccionar las variables de interés,

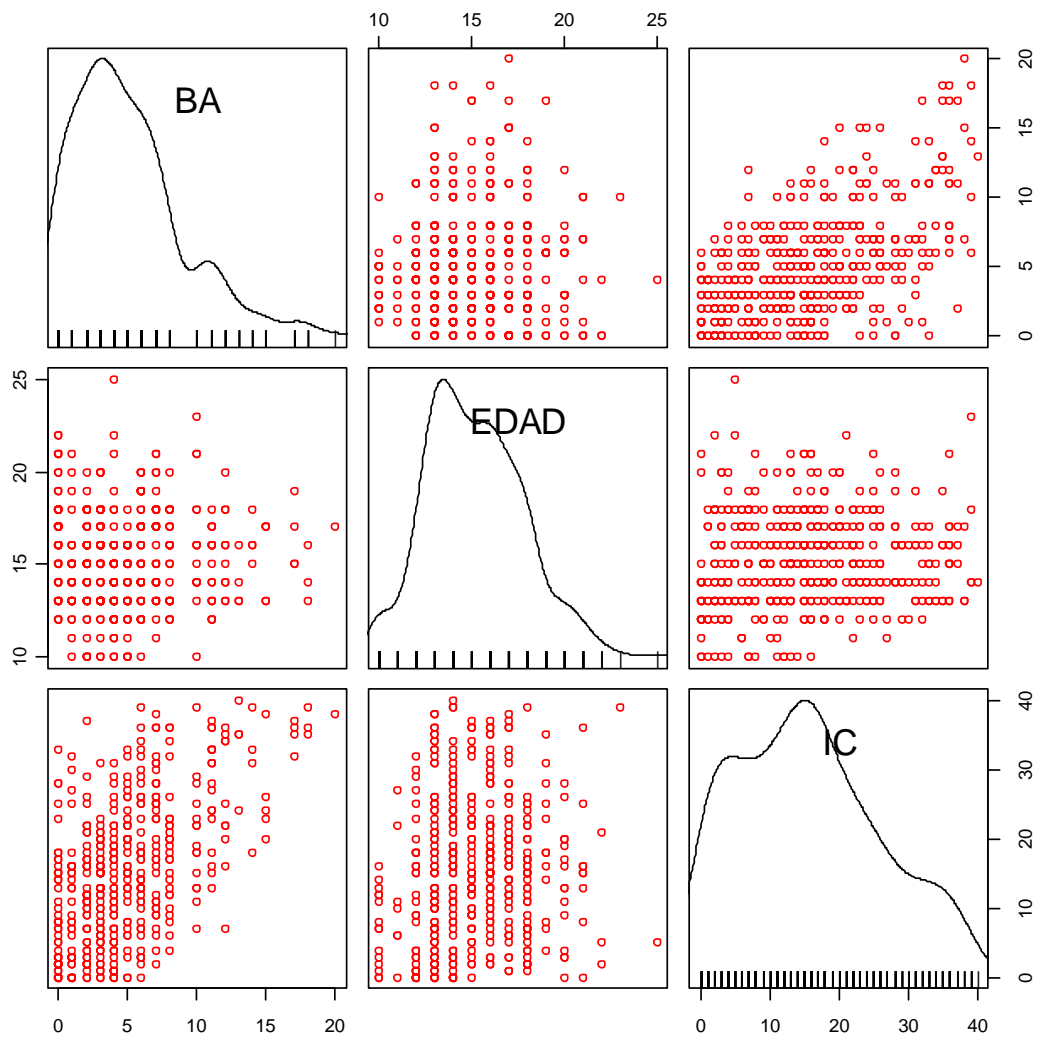
Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander



10.9Figura. Matriz de distribuciones

Esas especificaciones producen que Rcommander llame a la función gráfica `scatterplot.matrix`:

```
scatterplot.matrix(~BA+EDAD+IC,      reg.line=FALSE,      smooth=FALSE,
span=0.5,          diagonal      =      'density',      data=Edi.data,
subset=SEXO=="chica" )
```



10.10 Figura. Matriz de diagramas de dispersión

La diagonal principal de la matriz de gráficas de dispersión está ocupada por los gráficos de densidad de cada una de las variables de interés (Baja Autoestima, Edad e Insatisfacción Corporal). Se aprecia el carácter sesgado de las distribuciones de las variables baja autoestima e Insatisfacción Corporal. El resto de gráficos muestran las distribuciones bivariadas entre todos los pares posibles definidos con las 3 variables.

11 Interpretación normativa de puntuaciones.

Baremos

La puntuación empírica obtenida en un test no puede ser interpretada en términos absolutos en tanto en cuanto las escalas a partir de las cuales se obtienen las puntuaciones no tienen un 0 absoluto; ¿Que se puede inferir de la información contenida en el siguiente enunciado? “Javier obtuvo 16 puntos en una prueba de razonamiento lógico”. Nada; no podemos inferir nada sin tener más información sobre la escala utilizada. La puntuación 16 sólo tiene sentido, referida al test a partir del cual se obtuvo, y con referencia a la muestra (grupo normativo) en la que se estandarizó el test. Sólo con ambas referencias (test y grupo) podremos concluir si la puntuación 16 se sitúa por encima o por debajo de la media aritmética del grupo, o podremos conocer a cuantas desviaciones estándar por encima o por debajo de la media se sitúa el valor 16; de hecho, la media aritmética y la desviación estándar se consideran el punto 0 y la unidad de medida de la escala.

La transformación de puntuaciones engloba el conjunto de procedimientos o medidas encaminadas a situar la puntuación individual con referencia al grupo, y así definir un contexto para la interpretación de las puntuaciones. Para ello se transforman las escalas primarias o escalas brutas (Petersen, Kolen y Hoover, 1989) para que los valores obtenidos puedan ser interpretados sin hacer referencia a parámetros de escala. Básicamente, los tipos de transformación son dos; transformaciones lineales y transformaciones no-lineales.

11.1 Transformaciones lineales

Conceptos básicos y aplicaciones prácticas con Rcommander

La transformación lineal aplicable a cualquier valor o escala está sujeta a la siguiente ecuación. En ella se definen los parámetros de transformación, a y b , que se corresponderían correlativamente con la media aritmética y la desviación estándar de las nuevas escalas o escalas transformadas.

$$Y_i = a + bX_i$$

donde a y b son los parámetros de la transformación lineal

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{S_Y}{S_X}$$

Como consecuencia de una transformación lineal la forma de la distribución original no sufre ningún tipo de alteración; si fuera normal en origen seguiría siendo normal tras la transformación, y si estuviera sesgada la distribución transformada también lo estaría.

Escalas típicas

Las transformaciones lineales típicas, convierten las puntuaciones brutas en puntuaciones estándar o puntuaciones z o puntuaciones típicas (la literatura las reconoce de cualquiera de las tres formas). Las escalas estándar, tienen una media aritmética de 0 y una desviación estándar de 1 y poseen una relevancia especial en las ciencias sociales. A partir de los parámetros de transformación (0,1) es fácil obtener el valor z correspondiente a una puntuación bruta.

$$z_i = \frac{X_i - \bar{X}}{S_X}$$

Donde z_i es la puntuación estándar derivada de la puntuación empírica i ,
 X_i , puntuación empírica,
 \bar{X} , media aritmética de la escala primaria
 S_x , desviación estándar de la escala primaria

Escala típica derivada

Los valores negativos de las escalas típicas (dado que el punto medio de la escala es 0, la mitad de los valores son negativos) son a veces difíciles de interpretar por personas no especializadas ni habituadas a trabajar con este tipo de escala. Con el fin de solventar este inconveniente es habitual utilizar escalas derivadas con medias aritméticas y desviaciones estándar prefijadas. Partiendo de las puntuaciones estándar, el proceso de transformación consistiría en aplicar la siguiente ecuación:

$$Y_i = \bar{Y} + z_i S$$

Donde Y_i es el valor de escala transformado,
 \bar{Y} , la media aritmética de la escala destino
 S , la desviación estándar de la escala destino.

Aunque los parámetros de escala pueden ser definidos por el investigador, existen varias escalas que gozan de tradición en la medición psicológica y educativa; entre ellas, las escalas T, WAISS, Standford-Binet, MMPI o Delta:

Escala	Media aritmética	Desviación estándar
T	50	10
WAIS	100	15
Standford-Binet	100	16
MMPI	50	10
Delta	4	13

10.1 Tabla. Escalas típicas derivadas

11.2 Transformaciones no lineales

Conceptos básicos y aplicaciones prácticas con Rcommander

Además de las transformaciones lineales, existen otras que siendo monótonas crecientes sí alteran la distribución original de las puntuaciones; entre ellas, las más utilizadas son las puntuaciones centiles y las puntuaciones estándar normalizadas.

Centiles o percentiles

Es la transformación más común. El valor de una puntuación centil expresa el porcentaje de la muestra que obtuvo una puntuación igual o menor a ella. Por ejemplo si a una puntuación bruta (p.ej. 13) le corresponde el centil 60, podremos inferir que un 60% de la muestra obtuvo una puntuación menor o igual a 13.

$$P = \frac{f_{ai} + 0,5f_i}{N} \times 100$$

Donde f_{ai} , La frecuencia acumulada correspondiente a la puntuación a transformar,
 f_i , la frecuencia correspondiente a la puntuación a transformar
N, número de sujetos en la muestra.

Para una correcta interpretación de los valores percentiles es necesario considerar que sus valores no definen escalas de intervalo, sino de orden, y las diferencias observadas entre las puntuaciones originales no son mantenidas por las escalas percentiles.

Puntuaciones típicas normalizadas

A partir de los valores centiles es posible obtener puntuaciones típicas normalizadas. Basta con relacionar las puntuaciones centiles con los valores z correspondiente al área de la curva normal que limitan aquellas. Esta transformación normaliza la distribución original de las puntuaciones, por ello antes de llevarla a cabo es conveniente analizar la condición de normalidad de los datos. Si existen diferencias significativas entre ambas (original-normal) esta transformación no sería aconsejable (Angoff, 1984).

Psicometría.

Conceptos básicos y aplicaciones prácticas con Rcommander

Puntuaciones normalizadas derivadas

Una vez normalizadas las puntuaciones pueden aplicárseles transformaciones lineales que las conviertan en cualquier otra escala. En la tradición psicométrica la escala de estaninos o eneatis (Flanagan, 1951) es una de las más utilizadas. Esta escala segmenta el rango de puntuaciones en 9 intervalos de igual amplitud (no frecuencia). Los límites de los intervalos se fijan de acuerdo a los valores de la distribución normal

<i>Eneatis</i>	1	2	3	4	5	6	7	8	9	
Límites	$-\infty$	-1,75	-1,25	-0,75	-0,25	0,25	0,75	1,25	1,75	∞
Percentiles	4	5-11	12-22	23-39	40-59	60-76	77-88	89-95	>95	

10.2 Tabla. Eneatis

La transformación lineal que generaría una escala eneatis es aquella en que los parámetros de transformación, es decir, media aritmética y desviación estándar, son respectivamente, 5 y 2.

Además de los eneatis, la escala de decatipos es también utilizada como referencia para la interpretación de puntuaciones en muchos manuales de tests. Técnicamente es una escala normalizada derivada que divide el intervalo de puntuaciones en 10 categorías, y cuya media aritmética es 5,5, y su desviación estándar es 2.

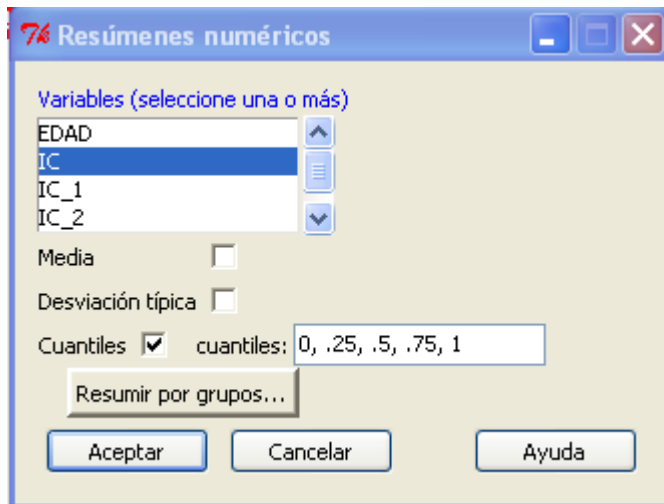
11.3 Baremos con Rcommander

Percentiles

La obtención de valores percentiles con Rcommander es sencilla. La opción `Resúmenes Numéricos` dentro del menú desplegable asociado a

Conceptos básicos y aplicaciones prácticas con Rcommander

Estadísticos > Resúmenes, permite seleccionar los cuantiles asociados a las distribuciones de las variables continuas activas en el conjunto de datos. Por defecto, Rcommander proporciona los cuantiles de la variable continua, es decir, aquellos valores que dividen a la muestra en cuatro intervalos 0-25-50-75-100.



11.1 Figura. Cuantiles de la distribución

Esas opciones generarán la siguiente línea en la ventana de comandos:

```
numSummary(Edi.data[, "IC"], statistics=c("quantiles"),
quantiles=c(0, .25, .5, .75, 1))
```

En la ventana de resultados se recogerán los cuantiles de la distribución de la variable Insatisfacción Corporal. De acuerdo a los valores mostrados se comprueba que la puntuación en la variable Insatisfacción Corporal que deja debajo de ella al 50% de la muestra es 18, y la puntuación que deja sobre ella al 25% de la muestra es 11; la puntuación centil correspondiente a la puntuación bruta 40 es 75.

```
> numSummary(Edi.data[, "IC"], statistics=c("quantiles"),
quantiles=c(0, .25, .5, .75, 1))
      0%  25%  50%  75% 100%   n  NA
      0   3  11  18  40  908  68
```

Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander

La construcción de baremos sin embargo, exige un número de puntuaciones centiles mayores a la opción por defecto de Rcommander. Aunque no existen reglas a seguir en lo que respecta al número de puntuaciones centiles que han de obtenerse, siempre hay que utilizar la norma de que las puntuaciones centiles han de describir a la muestra de la que se extraigan. Es común utilizar incrementos de 5 o incrementos de 10 en función del rango de la puntuación de la variable de la que se van a obtener las puntuaciones centiles. Para obtener estos valores será necesario modificar manualmente el comando anterior en la ventana de comandos especificando los cuantiles que se deseen extraer.

```
numSummary(Edi.data[,"IC"], statistics=c("quantiles"),  
quantiles=c(0,0.10,0.20,0.25,0.30,0.40,0.50,0.60,0.70,0.75,0.80,  
0.85,0.90,0.95,0.99))
```

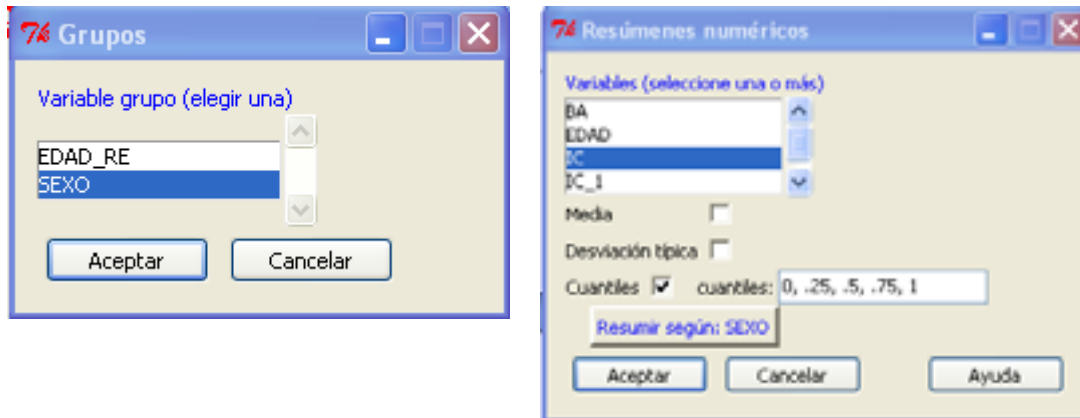
La salida adoptará el siguiente aspecto: La puntuación correspondiente al centil 80 es 21, la asociada al centil 10 es 1, etc.

```
> numSummary(Datos[,"INSA_COR"], statistics=c("quantiles"),  
quantiles=c(0,0.10,0.20,0.25,0.30,0.40,0.50,0.60,0.70,0.75,0.80,  
0.85,0.90,0.95,0.99))  
  
0% 10% 20% 25% 30% 40% 50% 60% 70% 75% 80% 85% 90% 95% 99%  n  NA  
0   1   2   3   4   7  11  14  17  18  21  24  28  33  37  908 68
```

Las puntuaciones centiles obtenidas de este modo, son utilizadas para describir al total de la muestra. Ahora bien, si existen factores a los que la variable de interés es sensible, es decir, si la variable que queremos describir es significativamente diferente en función del sexo, la edad, el estatus socioeconómico...un baremo debería de recoger tal diferenciación. La correcta interpretación normativa de una puntuación empírica debería de ser lo más próxima al grupo normativo del que proviene la persona que se va a evaluar. En

Conceptos básicos y aplicaciones prácticas con Rcommander

el caso que nos ocupa, por ejemplo, siendo los valores obtenidos en la variable Insatisfacción Corporal significativamente diferentes en función del sexo bastaría con seleccionar la opción que permite obtener salidas diferentes por grupo en la ventana de especificación de los descriptivos a obtener:



11.2Figura. Centiles por grupo

Esas especificaciones generarán puntuaciones percentiles diferenciadas por sexo,

	0%	25%	50%	75%	100%	n	NA
chica	0	7	15	23.00	40	481	27
chico	0	2	5	12.75	37	418	41

La orden que produce esa salida es:

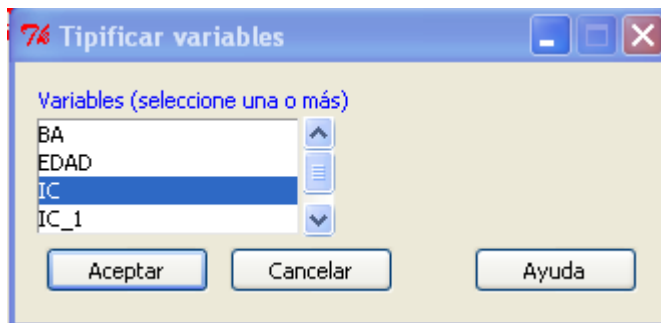
```
numSummary(Edi.data[, "IC" ], groups=Edi.data$SEXO,
statistics=c("quantiles"), quantiles=c(0,.25,.5,.75,1))
```

Para la obtención de un número de valores percentiles mayor sería suficiente añadir en el paréntesis de especificación de cuantiles (`quantiles = c (0 , . 2 5 , . 5 , . 7 5 , 1)`) los valores deseados separados por comas.

Puntuaciones típicas

Conceptos básicos y aplicaciones prácticas con Rcommander

Las puntuaciones típicas correspondientes a las puntuaciones empíricas se obtienen seleccionando la opción **Tipificar variables** perteneciente a las opciones del menú **Datos>Modificar variables** del conjunto de datos activo.



11.3Figura. Tipificar variables

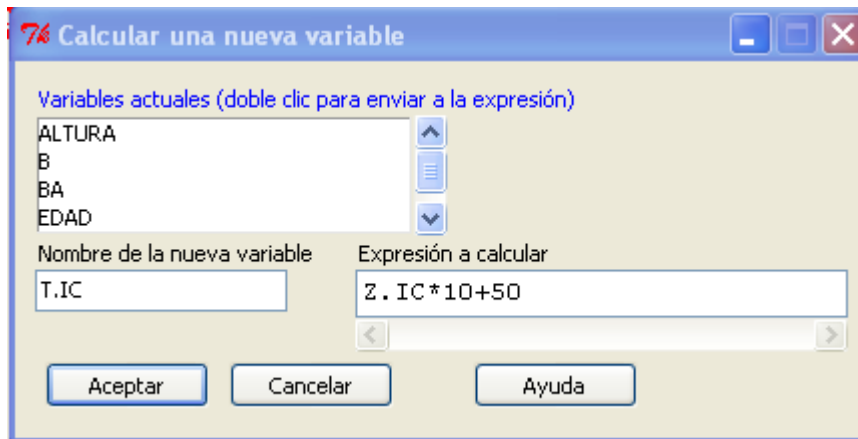
Esta opción permitirá seleccionar la variable del conjunto de datos activo que se desee tipificar. La variable creada de este modo es añadida al conjunto de datos activo con un nombre igual a la variable original precedido por la letra Z. Las opciones para la visualización o edición del conjunto de datos permiten comprobar sus valores. Una vez creada la variable, que evidentemente tendrá una media aritmética de 0 y una desviación estándar de 1, podrá modificarse a cualquier escala con media aritmética y varianza preespecificada con sólo escribir en la ventana de comandos la siguiente instrucción. En este caso, la escala de conversión es una escala T (media 50 y desviación estándar 10).

```
Edi.data$T.escala <- 50 + (Edi.data$Z.IC*10)
```

Este comando crea una variable de nombre `T.escala` que pertenecerá al conjunto de datos actual (`Edi.data$`), a partir de la variable `Z.IC` que está incluida en el conjunto de datos activo (`Edi.data$`). También hubiera sido posible crear la variable `T.escala` a partir de los menús desplegables asociados a **Datos>Modificar Variable** del **Conjunto de Datos>Calcular** una nueva variable, que ha sido explicada en una

Conceptos básicos y aplicaciones prácticas con Rcommander

sección anterior. Esta opción abre una ventana en la que es posible definir el nombre de la variable que se desea crear y la expresión algebraica que la compondrá.



11.4Figura. Escala valores T.

Z . IC	T . IC
2.228908	72.28908
-0.9981409	40.01859
0.1753316	51.75332
1.055436	60.55436
-0.509194	44.90806
0.5664891	55.66489
0.6642785	56.64278
2.326698	73.26698
-0.9003515	40.99648

11.5Figura. Valores típicos y valores T

Si se quisieran estimar los valores típicos o transformados de modo diferencial según el sexo habría que filtrar el conjunto de datos activo (Datos>Conjunto de datos activo > Filtrar el conjunto de datos) para poder ejecutar la orden de tipificación sólo sobre uno de los grupos.

12 Referencias Bibliográficas

- Ackerman, T. A. (1992). Didactic Explanation of Item Bias, Item Impact and Item Validity from a Multidimensional Perspective. *Journal of Educational Measurement*, 29(1), 67-91.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1954). *Standards for Educational and Psychological Tests and Manuals*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1966). *Standards for Educational and Psychological Tests and Manuals*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Angoff, W. H. (1984). *Scales, Norms and Equivalent Scores*. Princeton, New Jersey: Educational Testing Service.
- Angoff, W. H. (1988). Validity: An Evolving Concept. En H. Wainer y H. I. Braun (Eds.), *Test Validity* (19-32 pág.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Armor, D.J. (1974). Theta Reliability and Factor Scaling. En H. Costner (Ed.): *Sociological Methodology* (pp. 17-50). San Francisco: Jossey-Bass.
- Bagozzi, R. P. (1993). Assessing Construct Validity in Personality Research: Applications to Measures of Self-Esteem. *Journal of Research in Personality*, 27, 49-87.
- Becker, R. A., Chambers, J. M. y Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics*. Pacific Grove, CA: Wadsworth.
- Berk, R. A. (Eds.). (1982). *Handbook of Methods for Detecting Item Bias*. Baltimore: John Hopkins University Press.
- Bloom, B. S. (Eds.). (1956). *Taxonomy of Educational Objectives. Handbook I: The Cognitive Domain*. New York: McGraw Hill.
- Bloom, B. S., Hastings, J. T., y Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw Hill.

Conceptos básicos y aplicaciones prácticas con Rcommander

- Braun, W. J. y Murdoch, D. J. (2007) *A First Course in Statistical Programming with R*. Cambridge: Cambridge University Press.
- Browne, M. W. (1984). The Decomposition of Multitrait-Multimethod Matrices. *British Journal of Mathematical and Statistical Psychology*, 37, 1-21.
- Byrne, B. M. (1989). *A Primer of LISREL*. New York: Springer-Verlag.
- Camilli, G., y Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. London: Sage.
- Campbell, D. T., y Fiske, A. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines, E. G., y Zeller, R. A. (1979). *Reliability and Validity Assessment*. London: Sage.
- Cattell, R. B. (Eds.). (1966). *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.
- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis*. New York.
- Chambers, J. M. (2007). *Software for Data Analysis: Programming with R*. New York: Springer.
- Chambers, J. y Hastie, T. (1992). *Statistical Models in S*. Pacific Grove, CA: Wadsworth y Brooks/Cole.
- Clauser, B. E., y Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice*, 18, 31-44.
- Cole, N. S., y Moss, P. A. (1989). Bias in Test Use. En R.L.Linn (Ed.), *Educational Measurement. Third Edition* (pág. 201-219 pág.). New York: American Council on Education and Macmillan Publishing Company.
- Comrey, A. L. (1985). *Manual de análisis factorial*. Madrid: Cátedra.
- Crawley, M. J. (2008). *The R book*. Chichester: John Wiley y Sons.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J. (1971). Test Validation. En R. L. Thorndike (Ed.), *Educational Measurement* (443-507 pág.). Washington, DC: American Council of Education.
- Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist*, 30, 1-14.
- Cronbach, L. J. (1980). *Validity on pEdole: How can we go straight? New Directions for Testing and Measurement: Measuring Achievement over a Decade*. Paper Presented at the ETS Invitational Conference, San Francisco.
- Cronbach, L. J. (1984). *Essentials of Psychological Testing* (4ª Ed.). New York: Harper.
- Cronbach, L. J. (1988). Five Perspectives on the Validity Argument. En H.

- Wainer y H. I. Braun (Eds.), *Test Validity* (3-18 pág.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Cronbach, L. J., y Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52, 281-302.
- Dalgaard, P. (2002). *Introductory Statistics with R, Statistics and Computing*. New York: Springer.
- Dorans, N.J. y Kulick, E.(1986). Demonstrating the utility of the Standardization Approach to assessing unexpected Differential Item Performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23(4), 355-368.
- Ebel, R. L. (1965). *Measuring Educational Achievement*. Englewood, N.J.: Prentice-Hall.
- Elosua, P. (1996). *Froga psikologikoen euskaratzean sorturiko alborapenaren araky y baieztapena*. Euskal Herria, Donostia.
- Elosua, P. (2003). Testen euskaratzea. Balizko alborapenaren iturriak. *Tantak*, 30, 17-38
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321.
- Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema* 17(2), 356-362.
- Elosua, P. (2005). *Psikometria. Testen eraky y erabilpena*. Bilbao, U.P.V./E.H.U.
- Elosua, P. (2006). Funcionamiento diferencial del ítem en la evaluación internacional PISA. Detección y Comprensión. *RELIEVE*, 12(2).
- Elosua, P. (2009). ¿Existe vida más allá del SPSS? Descubre R. *Psicothema*, 21(4), 652-655.
- Elosua, P., Hambleton, R. K., y Zenisky, A. (2006). Improving the methodology for detecting biased test items. Paper presented at the 5th ITC Conference on Adapting Tests, Brussels.
- Elosua, P., López, A., y Torres, A. (1999). Adaptación al euskera de una prueba de inteligencia verbal. *Psicothema*, 11(1), 151-161.
- Elosua, P., López, A., y Torres, E. (2000). Desarrollos didácticos y funcionamiento diferencial de los ítems. Problemas inherentes a toda investigación empírica sobre sesgo. *Psicothema*, 12(2), 198-202.
- Elosua, P., López, A., y Egaña, J. (2000). Fuentes potenciales de sesgo en una prueba de aptitud numérica. *Psicothema*, 12(3), 376-382.
- Elosua, P., López-Jáuregui, A. y Sánchez-Sánchez, F. (2010). *Manual de la adaptación al español del Eating Disorder Inventory-3*. Madrid: TEA
- Elosua, P., y López, A. (2002). Indicadores de dimensionalidad para ítems binarios. *Metodología de las Ciencias del Comportamiento*, 4(1), 121-137.
- Elosua, P. y López, A. (2007). Aplicación de cuatro procedimientos de detección

- del funcionamiento diferencial sobre ítems politómicos. *Psicothema*, 19(2), 329-336.
- Elosua, P. y López-Jáuregui, A., (2008). Equation Between Linguistically Different Tests: Consequences for Assessment. *The Journal of Experimental Education*, 76(4), 387-402.
- Elosua, P. y López-Jáuregui, A., (2010). *Manual de la adaptación al euskera del Eating Disorder Inventory-3*. Madrid: TEA.
- Elosua, P. y Muñiz, J. (2010). Exploring the Factorial Structure of the Self-Concept: A Sequential Approach Using CFA, MIMIC and MACS Models, Across Gender and Two Languages. *European Psychologist*.
- Elosua, P. y Wells, C. (2008). A Comparison of Ordinal Logistic Regression and the IRT Likelihood Ratio Test for Identifying DIF. 6th International Test Commission Conference, Liverpool.
- Elosua, P., y Wells, C. (2008). A Comparison of MACS and the IRT Likelihood Ratio Test for Identifying DIF. III European Congress of Methodology, Oviedo.
- Elosua, P. y Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema* 20(4), 896-901.
- Embretson, S. E. (1983). Construct validity: Construct Representation versus Nomothetic Span. *Psychological Bulletin*, 93(1), 179-197.
- Embretson, S. E. (1985). *Test design: Developments in Psychology and Psychometrics*. New York: Academic Press, Inc.
- Ferrando, P. J. (1993). *Introducción al análisis factorial*. Barcelona: PPU.
- Fidalgo, A. M. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Eds.), *Psicometría* (371-455 pág.). Madrid: Universitas.
- Flanagan, J. C. (1937). A Note of Calculating the Standard Error of Measurement and Reliability Coefficients with the Test Scoring Machine. *Journal of Applied Psychology*, 23, 529.
- Flanagan, J. C. (1951). Units, Scores and Norms. En E. F. Lindquist (Eds.), *Educational Measurement* (1ª Ed.). Washington DC: American Council of Education.
- Fox, J. (2002). *An R and S-plus Companion to Applied Regression*. Thousand Oaks, CA: Sage.
- Fox, J. (2005). The Rcommander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14 (9), 1-42.
- Fox, J. (2007). *Getting Started With the R Commander*. Dirección URL: <http://socserv.mcmaster.ca/jfox/Courses/soc3h6/Getting-Started-with-the-Rcmdr.pdf>
- Garner, D. M. (2004). *Eating Disorder Inventory-3*. Lutz, FL: Psychological Assessment Resources, Inc.
- Gierl, M.J., y McEwen, N. (1995). Differential Item Functioning on the Alberta Educationn Social Stuies 30 Diploma Exams. *Paper Presented at the*

Conceptos básicos y aplicaciones prácticas con Rcommander

Annual Meeting of the Canadian Society for Studies in Education, Ottawa, Ontario, Canada.

- Gorsuch, R. L. (1974). *Factor Analysis*. Philadelphia: Saunders.
- Gorsuch, R. L. (1983). *Factor Analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Guadagnoli, E., y Velicer, W. (1988). Relation of Sample Size to the Stability of Component Patterns. *Psychological Bulletin*, 103, 265-275.
- Guilford, J. P. (1946). New Standards for Test Evaluation. *Educational and Psychological Measurement*, 6, 427-439.
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York: John Wiley and Sons.
- Guttman, L. (1954). Some Necessary Conditions for Common Factor Analysis. *Psychometrika*, 19, 149-162.
- Guttman, L. L. (1945). A Basis for Analyzing Test-retest Reliability. *Psychometrika*, 10, 255-282.
- Hambleton, R. K. (1993). Translating Achievement Tests for Use in Cross-national Studies. *European Journal of Psychological Assessment*, 9(1), 57-68.
- Harman, H. (1980). *Análisis factorial moderno*. Madrid: Saltés.
- Hattie, J. (1984). An Empirical Study of Various Indices for Determining Unidimensionality. *Multivariate Behavioral Research*, 19, 49-78.
- Hendrickson, A. E., y White, P. O. (1964). Promax: A Quick Method for Rotation to Oblique Simple Structure. *British Journal of Statistical Psychology*, 17, 65-70.
- Holland, W. P. y Thayer, D. T. (1988). Differential Item Performance and the Mantel-Haenszel Procedure. En H. Wainer y H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: LEA
- Holland, W. P., y Wainer, H. (1993). *Differential Item Functioning*. Hillsdale: Lawrence Erlbaum Associates.
- Husson, F., Lê, S. y Mazet, J. (2007), FactoMineR: Factor Analysis and Data Mining with R. R package version 1.05.
- Hunter, J. E., y Schmidt, F. L. (1991). My-analysis. En R. K. Hambleton y J. N. Zaal (Eds.), *Advances in Educational and Psychological Testing: Theory and Applications* (157-184 pág.). Boston: Kluwer Academic Publishers.
- Ihaka, R. y Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Jensen, A. R. (1980). *Bias in Mental Testing*. New York: Free Press.
- Jennrich, R. I., y Sampson, P. F. (1966). Rotation for Simple Loading. *Psychometrika*, 31, 313-323.
- Kaiser, H. F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, 23, 187-200.
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor

Conceptos básicos y aplicaciones prácticas con Rcommander

- Analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kelley, T.L. (1947). *Fundamentals of Statistics*. Cambridge, MA: Harvard University Press.
- Kerns, G. J. y Chang, G.A. (2010) *Introduction to Probability and Statistics Using R (IP SUR Taylor & Francis Publishing*.
- Lawley, D. N. (1940). The Estimation of Factor Loadings by the Method of Maximum Likelihood. *Proceedings of the Royal Society of Edinburgh*, 60, 64-82.
- Le, S., Josse, J., y Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1). [<http://www.jstatsoft.org/v25/i01/paper>]
- Linn, R. L., y Harnisch, D. L. (1981). Interactions between Item Content and Group Membership on Achievement Test Items. *Journal of Educational Measurement*, 18(2), 109-118.
- Loevinger, J. (1957). Objective Tests as Instruments of Psychological Theory. *Psychological Reports (Monograph Supp. 9)*, 3, 635-694.
- López-Jáuregui, A. y Elosua, P. (2004). *Revista Electrónica de Metodología Aplicada* 9(2) 1-14.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: LEA.
- Lord, F. M., y Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Magnuson, D. (1967) *Test Theory*. Boston: Addison Wesley
- Marsh, H. W., y Bailey, M. (1991). Confirmatory Factor Analysis of Multitrait-multimethod Data: A Comparison of Alternative Models. *Applied Psychological Measurement*, 15, 47-70.
- Martínez Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis, S.A.
- Maxwell, A. E. (1977). *Multivariate Analysis in Behavioral Research*. London.
- McDonald, R. P. (1981). The Dimensionality of Tests and Items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- McDonald, R. P. (1985). *Factor Analysis and Related Methods*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- McDonald, R. P. (1999). *Test theory. A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mellenbergh, G. J. (1989). Item Bias and Item Response Theory. *International journal of Educational Research*, 13, 127-143.
- Messick, S. (1980). Test Validity and the Ethics of Assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1981). Evidence and Ethics in the Evaluation of Tests. *Educational Research*, 10, 9-20.
- Messick, S. (1988). The Once and Future Issues of Validity: Assessing the
-

Conceptos básicos y aplicaciones prácticas con Rcommander

- Meaning and Consequences of Measurement. En H. Wainer y H. I. Braun (Eds.), *Test Validity* (33-45 pág.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. En R. L. Linn (Eds.), *Educational Measurement* (Third Edition, 13-104 pág.). New York: American Council on Education Macmillan Publishing Company.
- Millsap, R. E., y Everson, H. T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, 17(4), 297-334.
- Muenchen, R. A. (2009). *R for SAS and SPSS Users*. New York: Springer.
- Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York: McGraw Hill.
- Novick y Lewis (1967) Coefficient alpha and reliability of composite measurements. *Psychometrika*, 32, 1-18
- Osterlind, S. J. (1983). *Test Item Bias*. Beverly Hills, California: Sage.
- Pardo, A. y San Martín, R. (1998). *Análisis de Datos en Psicología II*. Madrid. Pirámide.
- Paradis, E. (2005). *R for Beginners*. Institut des Sciences de l'Evolution: Montpellier, France.
- Petersen, N. S., Kolen, M. J., y Hoover, H. D. (1989). Scaling, Norming and Equating. En R. L. Linn (Eds.), *Educational Measurement* (Third edition ed., 221-262 pág.). New York: American Council on Education and Macmillan Publishing Company.
- Potenza, M. T., y Dorans, N. J. (1995). DIF Assessment for Polytomously Scored Items: A Framework for Classification and Evaluation. *Applied Psychological Measurement*, 19(1), 23-37.
- Prieto, G., y Delgado, A. (1999). Medición cognitiva de las aptitudes. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados. Fundamentos y aplicaciones* (207-226 pág.). Madrid: Pirámide.
- Prieto, G., y Delgado, A. R. (2000). Utilidad y representación en la psicometría actual. *Revista de Metodología de las Ciencias del Comportamiento*, 2(2), 111-127.
- Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4(3), 207-230.
- Rulon, P. J. (1939). A Simplified Procedure for Determining the Reliability of a Test by Split-Halves. *Harvard Educational Review*, 9, 99-103.
- Scheuneman, J. D. (1984). A Theoretical Framework for the Exploration of Causes and Effects of Bias in Testing. *Educational Psychologist*, 19(4), 219-225.
- Scheuneman, J. D. (1987). An Experimental, Exploratory Study of Causes of Bias in Test Items. *Journal of Educational Measurement*, 24(2), 97-118.

Conceptos básicos y aplicaciones prácticas con Rcommander

- Schmitt, N., y Stults, D. M. (1986). Methodological Review: Analysis of Multitrait-Multimethod matrices. *Applied Psychological Measurement*, 10, 1-22.
- Shealy, R., y Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- Simon, J. L. (Ed.). (1969). *Basic Research Methods in Social Science*. New York: Random House.
- Snow, R. E., y Lohman, D. F. (1993). Cognitive Psychology, New Test Design and New Test Theory: An Introduction. En N. Frederiksen, R. J. Misley y I. I. Bejar (Eds.), *Test Theory for a New Generation of Tests* (1-18 pág.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1907). Demonstration of Formulae for True Measurement of Correlation. *American Journal of Psychology*, 18, 161-169.
- Spearman, C. (1913). Correlations of Sums and Differences. *British Journal of Psychology*, 5, 417-426.
- Stevens, J. (1992). *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Swaminathan, H. y Rogers, H.J. (1990): "Detecting differential item functioning using logistic regression procedures". *Journal of Educational Measurement*, 27, 4, 361-370.
- Thurstone, L. L. (1947). *Multiple Factor Analysis*. Chicago: University of Chicago Press.
- Title, C. K. (1982). Use of Judgment Methods in Item Bias Studies. En R. A. Berk (Eds.), *Handbook of Methods for Detecting Test Bias*. Baltimore: John Hopkins University Press.
- van de Vijver, F. J. R., y Poortinga, Y. P. (1991). Testing Across Cultures. En R. K. Hambleton y J.N. Zaal (Eds.), *Advances in Educational and Psychological Testing: Theory and Applications* (277-307 pág.). Boston: Kluwer Academic Publishers.
- Venables, W. N. y Ripley, B. D. (2000). *S Programming*. Springer: New York.
- Venables, W.N. y Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth Edition. New York: Springer.
- Venables, W. N., Smith, D. M. y the R Development Core Team (2007). *An Introduction to R*. Vienna, Austria: R foundation for Statistical Computing.
- Welch, B.L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350-362.