

TEMARIO:

PRIMERA PARTE

[Tema 1. La estadística: un método para conocer la realidad](#)

SEGUNDA PARTE: EL ANÁLISIS DE VARIABILIDAD ES UNO DE LOS OBJETIVOS FUNDAMENTALES DE LA ESTADÍSTICA

[Tema 2: En qué consiste la variabilidad](#)

[Tema 3. Cómo descubrir la variabilidad de un conjunto de datos](#)

[Tema 4: El tipo de variabilidad condiciona el modo de analizarla](#)

TERCERA PARTE: CÓMO ANALIZAR Y DESCRIBIR LA VARIABILIDAD CUANDO LOS DATOS DIBUJAN UNA FORMA DE CAMPANA

[Tema 5: Análisis del sector central de la distribución: media y desviación estándar](#)

[Tema 6: Análisis de los extremos de la distribución. La simetría](#)

[Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis](#)

Tema 1. La estadística: un método para conocer la realidad

El interés y la utilidad de la estadística	2
¿De qué hablamos cuando mencionamos “la Estadística”?	5
La estadística nos ayuda a crear conocimiento	7
Razonamiento cotidiano, razonamiento científico y estadística.....	11
▣ Un ejemplo de razonamiento reflexivo : estudio hipotético sobre el comercio urbano.....	11
▣ Razonamiento reflexivo (cotidiano) versus razonamiento científico	16

Tema 1. La estadística: un método para conocer la realidad

EL INTERÉS Y LA UTILIDAD DE LA ESTADÍSTICA

Una de las maneras de luchar contra el rechazo que algunos estudiantes sienten hacia determinadas materias de estudio consiste en ofrecerles argumentos y evidencias de la utilidad “práctica” del conocimiento que proporcionan dichas materias. La hipótesis es que si los estudiantes se hacen conscientes del valor y del interés de determinado tipo de conocimiento, tenderán a suavizar, en menor o mayor medida, su actitud negativa hacia la materia de estudio. Este es precisamente el objetivo del primer tema del curso: proporcionar pruebas y argumentos para intentar modificar la (negativa) concepción que los estudiantes tienen de la Estadística.

Para poder llegar a comprender el valor y el interés de la Estadística es imprescindible entender primero que lo esencial de la Estadística no está en la manipulación de números y fórmulas. Lo esencial en el aprendizaje y en el uso de la estadística es el conocimiento sobre el tipo de problemas y cuestiones que nos puede ayudar a resolver, (esto es así) porque la Estadística es una herramienta y un método para resolver problemas.

¿En qué consiste esa ayuda que, según decimos, podemos encontrar en la Estadística? Podemos considerar que la estadística nos ayuda a resolver problemas en la medida en que nos proporciona los medios para encontrar respuestas a cuestiones que nos preocupan, que nos resultan problemáticas. Muchos de los problemas que afrontamos exigen, además, que tomemos decisiones sobre las acciones posibles para su solución. El método estadístico nos ayuda a conseguir el conocimiento necesario para fundamentar nuestras decisiones. En resumen, la Estadística nos proporciona un método y numerosas herramientas para obtener el conocimiento necesario en la resolución de problemas.

Existen bastantes publicaciones que revelan la utilidad de la estadística en situaciones concretas, en el análisis o resolución de problemas concretos. Como

primera aproximación al tema, recogeremos algunos de las ideas y argumentos que distintos autores exponen para convencernos del valor de la Estadística.

En un artículo titulado *El valor de la estadística para la salud pública*, sus autores nos ofrecen ejemplos concretos sobre la utilidad de la Estadística en el ámbito de la medicina:

(La estadística) Permite comprender los fundamentos racionales en que se basan las decisiones en materia de diagnóstico, pronóstico y terapéutica.

☒ Interpreta las pruebas de laboratorio y las observaciones y mediciones clínicas con un conocimiento de las variaciones fisiológicas y de las correspondientes al observador y a los instrumentos.

☒ Proporciona el conocimiento y comprensión de la información acerca de la etiología y el pronóstico de las enfermedades, a fin de asesorar a los pacientes sobre la manera de evitar las enfermedades o limitar sus efectos.

☒ Otorga un discernimiento de los problemas sanitarios para que eficientemente se apliquen los recursos disponibles para resolverlos.
(Cantú y Guzmán, 2003)

La cita que recogemos a continuación pertenece a un trabajo en el que sus autores exponen y defienden la aportación de la Estadística en el ámbito de la producción industrial:

La Calidad Industrial es un tema de gran actualidad, que cada día cobra un mayor interés tanto industrial como social. Son cada vez más las empresas que implantan sistemas de calidad, para mejorar los productos o los servicios que esta ofrece, siendo de uso cotidiano y coloquial en entornos técnicos el uso de la serie de normas ISO 9000. [...]

El control clásico de la calidad se realizaba sobre el producto acabado de forma que las piezas defectuosas aprovechables eran recicladas mientras que las inservibles pasaban a la chatarra. La idea hoy consiste en llevar el control de calidad a cada fase del proceso de producción y así poder obtener una información que permita controlar la calidad y mejorar el propio proceso productivo. [...]

La estadística es una disciplina que trata de los métodos de obtención de datos, su análisis y resumen y la deducción de consecuencias a partir de la muestra de datos mediante el estudio de una teoría de probabilidades. Esta disciplina, aplicada a la calidad, nos proporciona técnicas rápidas, sencillas y económicas que nos permiten obtener resultados con los que

mejorar los procesos productivos y deducir en que puntos de la organización empresarial puede aumentar la calidad. (Gutierrez Ariza et al., 1998)

En esta ocasión también podemos concluir que la función de la Estadística se sitúa en el ámbito de la resolución de problemas: proporciona información que permite controlar el proceso de producción para lograr un menor número de piezas defectuosas.

En uno de los artículos didácticamente más interesantes que hemos encontrado, el biólogo Marcos Méndez Iglesias argumenta, a través de un ejemplo¹ y de forma especialmente clara, la función de la Estadística en el proceso de creación del conocimiento:

Muchos aficionados a la Ornitología desconocen totalmente los métodos estadísticos. Resulta lamentable ver cómo, ante la creciente presencia en revistas especializadas de trabajos con componente estadístico, la reacción más frecuente es la de rechazo. El argumento que se esgrime es su enorme complejidad, pero la razón subyacente es el total desconocimiento de la utilidad de la Estadística.

[...]

El experimento imaginario presentado más arriba ha pretendido mostrar no ya sólo la utilidad, sino la necesidad de utilizar bases estadísticas para decidir entre opciones. Sin ellas, todas nuestras conclusiones son castillos en el aire y su contribución a la resolución de un problema se transforma en un lastre, al aportar evidencias dudosas.

En cuanto a la complejidad, más aparente que real, no debería constituir un obstáculo para cualquiera que se plantee la Ornitología como algo más que un simple coleccionismo:

-Si has sentido en algún momento la tentación de realizar estudios a partir de tus propias observaciones, probablemente te hayas dado por vencido al ignorar cómo analizar los resultados, o simplemente habrás extraído conclusiones subjetivas y, casi con seguridad, erróneas.

¹ Para no alargar excesivamente la cita del autor se han omitido las explicaciones sobre el ejemplo que desarrolla. Sin embargo, dada la claridad y el interés de la propuesta completa de este autor, se ha incluido el artículo como material de lectura para este primer tema. Se aconseja, por ello, la lectura del artículo completo.

-Si simplemente deseas aumentar tu cultura ornitológica leyendo trabajos de otras personas, habrás encontrado que tu desconocimiento de los métodos estadísticos te impide comprender prácticamente nada. Renunciar en este caso a un conocimiento, por mínimo que sea, de los principios estadísticos conlleva cerrarse muchas puertas. Seguramente supone el limitarse a una observación sin sentido o, en el mejor de los casos, a la eterna confección de estatus y fenologías, actividades que no agotan ni de lejos las posibilidades de la Ornitología.

Hasta aquí, la crítica. Mi esperanza es que la actitud hacia la Estadística cambie. Nadie abominaría del uso de los prismáticos ¿por qué hacerlo de otra herramienta igualmente útil? (Méndez Iglesias, 1996, p. 104-105)

¿DE QUÉ HABLAMOS CUANDO MENCIONAMOS "LA ESTADÍSTICA"?

En el apartado anterior hemos hablado de la utilidad de "la Estadística"; hemos mostrado algunos ejemplos del valor que determinados profesionales le conceden. Ahora bien, antes de continuar hablando de los *posibles* beneficios de la Estadística, es preciso hacer algunas aclaraciones en torno al propio concepto de Estadística. Se trata de un término que se utiliza con distintos significados y no hay nada más desconcertante que leer textos con conceptos cuyos significados, como poco, nos confunden.

El término Estadística se utiliza con distintos significados; la misma palabra se emplea para referirse:

- ▣ A la ciencia o campo de conocimiento,
- ▣ a los listados de datos estadísticos,
- ▣ al conjunto de técnicas que se utilizan para analizar los datos
- ▣ a la forma particular de razonamiento que se utiliza para analizar problemas mediante la aplicación del método estadístico.

De las cuatro acepciones, quizás, la más desconocida para los estudiantes sea la última, es decir, la que considera la estadística como una forma particular de razonamiento. No es de extrañar que sea un significado desconocido, porque, hoy por hoy, la mayoría de cursos de Estadística se presentan mediante un programa que incluye, exclusivamente, un listado, mayor o menor, de técnicas estadísticas. Si la experiencia de los estudiantes se limita a este tipo de cursos, será difícil convencerles de que la Estadística es algo más que la suma de todas las técnicas.

Cada vez son más los autores que en lugar de hablar de la Estadística como de un método prefieren utilizar el concepto de *pensamiento estadístico*, que definen como sigue:

El pensamiento estadístico es la forma en que la información se ve, se procesa y se convierte en pasos de acción. Es una filosofía de pensamiento, no una forma de realizar cálculos matemáticos. El pensamiento estadístico utiliza el concepto de que toda actividad consiste en un conjunto de pasos interconectados que deben complementarse y completarse para lograr una meta planteada, donde se debe investigar cada paso para identificar áreas de oportunidad y mejora a fin de lograr el éxito personal o profesional.

[...]

(El pensamiento estadístico) es el compendio de principios, lenguaje y estrategias para pensar, razonar y plantear una solución a un problema que requiere cuantificación. Cuando se aprende el pensamiento estadístico, se aplica y se entienden mejor los procesos, el trabajo y la vida en general, lo que conlleva a la delimitación, planteamiento y solución apropiada de problemas personales y/o profesionales.

[...]

En la ASQC² (1996b) se plantea primero que el pensamiento estadístico es una filosofía, o sea es una forma de pensar, o un proceso de pensamiento, en lugar de cálculos que a algunos confunden. Ésta es la diferencia clave entre el pensamiento estadístico y las técnicas estadísticas

[...]

En este siglo debe promoverse el pensamiento estadístico en el entrenamiento estadístico porque es necesario insistir que antes de querer aplicar métodos estadísticos debe hacerse uso del pensamiento estadístico, ya que es el compendio de principios, lenguaje y estrategias para pensar, razonar y plantear una solución a un problema que requiere cuantificación. Cuando se aprende el pensamiento estadístico, se aplica y se entienden mejor los procesos, el trabajo y la vida en general, lo que conlleva a la delimitación, planteamiento y solución apropiada de problemas personales y/o profesionales. (López Lozada, 2004, p.4)

² American Society for Quality Control

En resumen, llámesele método estadístico o pensamiento estadístico, lo fundamental es entender que la Estadística es mucho más que un ramillete de técnicas matemáticas que se aplican en un estudio cuando ya se dispone de una serie de datos. Muy al contrario, la Estadística entra a formar parte de un estudio desde su inicio y, sobre todo, mucho antes de recoger los datos.

La definición que hacen Behar y Ojeda (2006) es un fiel reflejo de esta forma de entender la Estadística:

La estadística es una disciplina que proporciona principios y herramientas para hacer juicios sobre colectivos, con base en datos que se han obtenido para un propósito específico. Con sus técnicas y principios brinda la metodología para saber qué datos obtener, cómo obtenerlos y, una vez obtenidos, proporciona métodos y procedimientos para organizarlos y transformarlos con diferentes propósitos, a fin de extraer de ellos la máxima información según nuestros intereses y objetivos. Del análisis de los datos se obtiene la base para la construcción de juicios concluyentes sobre el colectivo bajo estudio; por tal motivo, es muy importante saber de qué colectivo se está hablando.

El curso que aquí se presenta es precisamente un conjunto de materiales destinados a que el estudiante comprenda en qué consiste el pensamiento, o el método si se quiere, estadístico. Se trata, igualmente, de que el estudiante adquiera la capacidad de aplicar este modo concreto de razonar en situaciones adaptadas al nivel de complejidad que las técnicas que desarrollaremos permitan.

LA ESTADÍSTICA NOS AYUDA A CREAR CONOCIMIENTO

Entendida como método o como forma de razonamiento, lo que está claro es que la Estadística nos proporciona recursos para crear conocimiento, para descubrir aspectos del mundo y de la realidad que no se conocían o se conocían insuficientemente. Aunque es mucho el conocimiento que la humanidad ha acumulado durante siglos, sigue siendo infinita la lista de cosas que quedan aun por descubrir y conocer.

De hecho, cada vez está siendo más habitual que los medios de comunicación nos informen de nuevos descubrimientos. Generalmente se trata de cuestiones o temas de gran trascendencia para la sociedad o, incluso, de hallazgos que son escasamente relevantes pero que, sin embargo, resultan curiosos, llamativos. Nos enteramos así de la relación que existe entre el consumo de jamón de pata negra y la tasa de colesterol en sangre.

Pata negra contra el colesterol. Un estudio demuestra que el jamón de bellota es beneficioso para la Salud

MADRID.- Han sido necesarios ciento diecinueve jamones, ocho investigadores y el buen apetito de un grupo de habitantes de un pueblo de Badajoz para dar la razón al paladar. El jamón ibérico de bellota es bueno para la salud.

El lugar elegido para demostrarlo fue la localidad pacense de Valdebotoa. Allí, diecinueve mujeres con una media de 66 años aceptaron participar en un estudio de la Universidad de Extremadura, el Servicio de Medicina Interna del Hospital provincial de Badajoz y la Consejería de Agricultura de la Junta extremeña. (1996, 24 de abril). Diario El Mundo.

En 2004 tuvimos noticia de los resultados de un estudio de dimensiones humanas muy superiores:

La nueva vacuna de la malaria evita el 65% de las infecciones en bebés, donde es más letal. El ensayo del español Pedro Alonso en Mozambique allana el camino a una inmunización a gran escala en África. (2007, 18 de octubre). Diario ABC.

El día 5 de julio de 2007 supimos que habían hecho falta ocho años de estudios de investigación para saber que los hombres hablan tanto como las mujeres.

Se trata del primer estudio que registra conversaciones naturales de cientos de personas durante varios días y revela que las mujeres utilizan alrededor de 16.215 palabras al día y que los hombres emplean 15.669, una diferencia que en términos estadísticos no es significativa.

Investigadores de la Universidad de Arizona en Tucson (Estados Unidos) han estudiado de forma científica el estereotipo de que las mujeres hablan más que los hombres, y han descubierto que tanto mujeres como hombres utilizan una media de 16.000 palabras al día, echando por tierra el tópico. (2007, 5 de julio) Diario el País.

Noticias como las que acabamos de recoger constituyen la cara visible, el resultado resumido y reducido de estudios de investigación que han utilizado el método estadístico.

Cada día se saben, se descubren cosas nuevas aunque, como hemos visto, no todas tienen la misma importancia o trascendencia. En cualquier caso, la mayoría de los descubrimientos se han logrado mediante procesos de investigación cuyas características esenciales son similares:

- ▣ Se inician partiendo de preguntas que queremos responder, de cosas que queremos saber, de aspectos desconocidos de la realidad que queremos averiguar.
- ▣ Se llevan a cabo mediante un proceso planificado de trabajo, siguiendo las pautas del método científico.
- ▣ Las investigaciones han utilizado el método estadístico.

Lo que queremos resaltar aquí es que la investigación y la Estadística están estrechamente relacionadas. La investigación es un proceso ordenado de trabajo mediante el que se intenta descubrir aspectos de la realidad desconocidos o poco conocidos. La investigación es, en definitiva, una actividad conducente a la búsqueda de conocimiento. Para ello el investigador realiza un plan detallado de trabajo, que incluye la búsqueda de información sobre el tema que desea conocer.

La necesidad o el deseo de obtener algún tipo de conocimiento suele ser el inicio de muchas de las actividades que los seres humanos realizamos movidos por nuestra curiosidad o por las exigencias de nuestro trabajo. La búsqueda de conocimiento se produce en contextos en los que las personas, -investigadores, profesionales o particulares-, sienten un problema relacionado con el desconocimiento de algo y deciden hacer una investigación para descubrir lo que necesitan saber.

Llegados a este punto podemos decir que existe un paralelismo -cierta semejanza- entre la actividad científica y determinadas actividades intelectuales propias del ejercicio profesional. En un artículo sobre la educación y las tecnologías digitales, su autor, Ferrán Ruiz Tarragó resume de forma muy clara algunos de los aspectos de este paralelismo:

Todo el mundo tiene que planificar y controlar su trabajo; todas las personas reciben, guardan y buscan información en diversas formas; todos adquirimos información escuchando, leyendo y observando, y también la adquirimos mediante los otros sentidos; en todas las actividades profesionales se produce información escrita u oral, o bien se crean otros tipos de información; todo el mundo interacciona y se comunica con otras personas y con su entorno.

El trabajo mental con la información es el substrato común de todas estas actividades, con independencia de como esté codificada la información, del formato en que se presente y del soporte material que la contenga. (Ruiz Tarragó, 1999, p. 32)

Como señala Ruiz Tarragó, la base común de la actividad científica y la profesional es la gestión de información en un contexto planificado de trabajo. El científico y el que no lo es trabajan con información, la buscan, la procesan, la ordenan...

Una frase de Einstein lleva el paralelismo todavía más lejos:

La totalidad de la ciencia no es más que un refinamiento del pensamiento cotidiano. (Einstein, 1950, citado por Toalongo, 2002, p. 53)

Toalongo nos ofrece la razón última que explica el fondo de la semejanza de toda actividad intelectual: la necesidad, el deseo de averiguar está en la naturaleza humana.

Para precisar, investigar, etimológicamente proviene del latín "in" que significa "en", y, "vestigare" está relacionado con averiguar, inquirir, buscar vestigios. En este sentido, la investigación es una actividad inherente a la naturaleza humana; es decir, todo ser humano trata de comprender, explicar e indagar el sentido de las cosas, sin que esto signifique que todos los seres humanos sean investigadores. (Toalongo, 2002, p. 53)

La envergadura del trabajo con la información y, sobre todo, el método que se utiliza para gestionarla diferencian o particularizan el trabajo científico. Evidentemente, no es lo mismo el proceso de indagación que utiliza una persona para buscar la solución a un problema de vida cotidiana, que el procedimiento que lleva a cabo un investigador profesional en el desarrollo de su trabajo. Pero, en cualquier caso, sea cual fuere el contexto o marco en el que se realice, el proceso de generación de conocimiento requiere de un método: si desconocemos algo y deseamos realizar la actividad conducente a conocerlo necesitamos utilizar un método, una secuencia de procedimientos que nos guíe en el proceso³. Si, además, el objetivo es generar un conocimiento cuya validez y fiabilidad no sea discutida se debe utilizar el método científico.

³ "Método significa camino, vía. Viene del griego métodos, en la acepción común de la palabra significa el medio a través del cual se alcanza un objetivo. En el plano de la filosofía, el término ha estado asociado a la cuestión de descubrir la vía más racional para conquistar la verdad". (Núñez Tenorio, 1974, p 31)

RAZONAMIENTO COTIDIANO, RAZONAMIENTO CIENTÍFICO Y ESTADÍSTICA

A nadie se le escapa que el trabajo de investigación científica requiere una formación especializada. El investigador debe tener un conocimiento profundo no sólo de su materia de estudio sino también de los métodos y técnicas de investigación que permiten seguir avanzando en el conocimiento de su campo de estudio. Es preciso, por tanto, entender que la investigación científica es una actividad compleja que requiere un largo proceso de aprendizaje y de práctica.

Sin embargo, una cosa es estar capacitado para realizar una investigación y otra, muy distinta, es comprender los aspectos esenciales del proceso. Cuando hablamos de aspectos esenciales nos referimos, por ejemplo, al hecho de que toda investigación se inicia como un intento de resolver algún problema o, dicho de otro modo, como una serie de preguntas que se quiere responder. Desde el planteamiento del problema hasta su solución los investigadores aplican un método, es decir, realizan una secuencia de actividades cuya lógica también es fácilmente comprensible⁴. De hecho, como decíamos anteriormente, existe una similitud entre el proceso de investigación y las actividades profesionales o, incluso, los modos de razonamiento que usamos cotidianamente en la resolución de problemas.

Con el fin de ilustrar las similitudes y diferencias entre los distintos modos de razonamiento, hemos creado un ejemplo que muestra el proceso de razonamiento reflexivo que se inicia con un problema al que una persona (imaginaria) trata de encontrar solución. El objetivo es recrear una situación que permita entender mejor la lógica de los procesos de investigación, su semejanza con los procesos de razonamiento reflexivo para la resolución de problemas y, sobre todo, la función que puede cumplir la Estadística en el proceso.

Un ejemplo de razonamiento reflexivo : estudio hipotético sobre el comercio urbano

Alguien de nuestra familia dispone de un pequeño local comercial y ha decidido instalar un negocio que le promocióne un empleo, un modo de ganarse la vida. Ha pensado en varias posibilidades pero no termina de decidirse porque no sabe cuál de ellas podría ser la más rentable. Aunque sabe que el éxito de un negocio depende de muchos factores que no puede controlar

⁴ Para mencionar esta secuencia se habla generalmente de las fases que componen el método científico.

está decidido a indagar, a buscar información que le ayude a fundamentar su decisión.

Por propia experiencia sabe que en los barrios periféricos de la ciudad existen numerosos establecimientos que ofertan bienes y servicios de uso cotidiano. Estos establecimientos cumplen una función importante: evitan que los habitantes de la zona tengan que desplazarse al centro urbano para efectuar cierto tipo de compras.

Con el fin de encontrar información e ideas que puedan orientar su decisión, se le ocurre que puede ser interesante conocer con mayor precisión qué tipo de establecimientos son, cuántos de cada tipo hay en cada barrio y con qué dotación cuenta su propio barrio al respecto. Intuye que, comparando la dotación de servicios de su barrio y la de los otros barrios, podrá saber si en el suyo existe carencia o infradotación de algún tipo de servicios.

Para poner en marcha su plan decide ir a buscar asesoramiento a la Cámara de Comercio de la ciudad. Descubre que este organismo dispone de muchos datos sobre los negocios existentes en la ciudad y que por un precio razonable puede adquirir esta información. Con los datos que le proporciona la entidad de comercio decide elaborar un mapa de localizaciones y además recorrer los distintos barrios de la ciudad para recoger información complementaria.

Un mes más tarde nuestro familiar:

- ▣ dispone de amplias series de datos sobre la red de establecimientos que ofertan bienes y servicios en los distintos sectores urbanos de la ciudad,
- ▣ ha hecho también una cartografía de la distribución de los establecimientos en los distintos barrios.
- ▣ Tiene mucha información pero no sabe cómo gestionarla, cómo analizarla para obtener el conocimiento que necesita.

Estando paralizado por no saber cómo seguir adelante consigue contactar con un amigo que le proporcionará ayuda. Se trata de alguien que está especializado en la realización de estudios estadísticos y que se compromete a hacerle un análisis de los datos que ha recogido.

Es cierto que nuestro familiar no sabe hacer un análisis de datos, pero sí sabe lo que desea conocer y prepara una lista de preguntas para su amigo:

- ▣ Con el fin de comprobar si el número y la tipología de los servicios en otros

barrios de la ciudad es una guía adecuada para fundamentar su decisión, formula la siguiente pregunta: ¿Los barrios con una población similar disponen de un número similar de establecimientos de servicios?

- ▣ ¿Hay en los barrios algún tipo de establecimiento que sea más abundante que otro?
- ▣ ¿Existe en los otros barrios algún tipo de establecimiento que no exista en su propio barrio o cuya presencia sea significativamente inferior?
- ▣ Teniendo en cuenta la diferente distancia que hay de los distintos barrios de la ciudad al centro urbano, existe algún tipo de relación entre la distancia del barrio al centro urbano por un lado y el tipo o a la abundancia de establecimientos por otro?

Si se cuenta con los datos adecuados, el análisis estadístico permite obtener conclusiones sobre el grado habitual de variabilidad que, en este caso, se referiría al número y tipo de establecimientos comerciales. Se podrían detectar de este modo sectores urbanos infradotados o sectores con un desarrollo comercial especialmente elevado. La Estadística dispone también de las herramientas necesarias para analizar si existe una relación entre distintas características de un grupo de objetos. Aplicadas a esta situación, si los datos recogidos son idóneos, podrán dar respuesta a las preguntas sobre la relación entre la distancia de los barrios al centro urbano y la densidad comercial.

El supuesto que hemos planteado contiene una secuencia de acciones que alguien realiza cuando decide buscar o generar el conocimiento que necesita para resolver un problema o, más concretamente en este caso, para fundamentar una decisión. La secuencia incluye una parte de los componentes habituales de una investigación científica, pero carece de algunos otros que son esenciales y que, al no haberse realizado, limitan mucho la validez de los resultados y conclusiones que puedan obtenerse del trabajo.

En el ejemplo planteado las fases de la investigación no están formalizadas como tales pero son fácilmente identificables. En el cuadro que sigue podemos ver los detalles del estudio hipotético estructurados siguiendo el esquema habitual de una investigación científica.

<p style="text-align: center;">FASES DEL PROCESO DE INVESTIGACIÓN CIENTÍFICA</p>	<p style="text-align: center;">FASES DEL ESTUDIO HIPOTÉTICO SOBRE COMERCIO URBANO</p>
<p>Existencia de un problema</p>	<p>De qué forma fundamentar la decisión sobre el tipo de actividad más idónea para un local comercial disponible en un barrio urbano periférico</p>
<p>Búsqueda y análisis del conocimiento ya existente en torno al tema:</p> <p>El investigador busca las teorías existentes sobre las pautas de localización y distribución del comercio urbano, sobre los factores que intervienen en la configuración y en las características de la red establecimientos. El objetivo es conocer lo que ya se sabe al respecto como punto de partida para avanzar en el conocimiento o para corroborar las teorías existentes.</p> <p>En esta fase se obtiene información sobre los aspectos que deben o pueden ser analizados.</p>	<p>En este proceso no se ha realizado esta fase. No se dispone por tanto de información sobre:</p> <ul style="list-style-type: none"> ■ Cuestiones que la ciencia ya conoce en torno al comercio urbano. ■ Características de los establecimientos que se deberían o se podrían analizar. ■ Planes y política municipales sobre el comercio de la ciudad. ■ (...)
<p>El investigador define las cuestiones que quiere conocer en torno al tema. Elaboración de preguntas concretas.</p> <p>Si el conocimiento existente sobre el tema es insuficiente para encontrar las respuestas o decide corroborarlo, pasa a la siguiente fase.</p> <p><u>Si el tema de estudio incluye procesos de cuantificación de datos, la propia formulación de las preguntas está guiada por las posibilidades que ofrece el método estadístico y que el investigador ya conoce.</u></p>	<p>¿Existe una tipología de negocios propia o habitual de las zonas urbanas alejadas de los centros urbanos?</p> <p>¿Existe en los barrios periféricos de tamaños similares una cantidad similar de establecimientos que ofertan bienes y servicios básicos?</p> <p>¿Existen localizaciones preferenciales para alguno de los tipos de actividad? ¿La ubicación del local disponible debería ser un factor de decisión?</p>
<p>Elaboración de hipótesis. Las hipótesis son conjeturas, intentos de respuesta a las preguntas que formulamos en la investigación.</p>	<p>Teniendo en cuenta el tipo de preguntas que ha formulado y el hecho de que ha intentado recoger datos sobre las características del comercio en</p>

<p>Las hipótesis guían el trabajo posterior: definen lo que el investigador debe estudiar para comprobar si sus respuestas hipotéticas son verídicas.</p> <p><u>El investigador ya cuenta, a la hora de formular las hipótesis, con las herramientas que método ofrece para buscar las respuestas.</u></p>	<p>otros barrios, las hipótesis que intuitivamente ha realizado se podrían formular como sigue:</p> <p>Si la tipología plantada existe, un análisis de la oferta de bienes y servicios en las distintas zonas de una ciudad mostrará pautas similares en cada una de ellas.</p> <p>Si el número y tipo de establecimientos en los barrios guarda algún tipo de relación con el número de habitantes o con la distancia al centro, barrios de similares características mostrarán equipamientos similares.</p>
<p>Diseño y planificación de la investigación para comprobar la hipótesis.</p> <p><u>El diseño incluye todos los detalles sobre el tipo de datos que se necesita, sobre la forma de obtenerlos y sobre su tratamiento posterior.</u></p>	<p>No se ha realizado ningún diseño. El autor ha reflexionado sobre las características de los establecimientos que merece la pena analizar. Se ha basado en su experiencia y conocimiento sobre el tema.</p> <p>El autor desconoce las posibilidades que le ofrece la Estadística para el análisis y, por tanto, no la utiliza para tomar decisiones sobre el tipo y la cantidad de datos necesarios para obtener conclusiones fiables.</p>
<p>Recogida de datos, análisis y conclusiones</p>	

Analizado el proceso que ha seguido nuestro supuesto investigador podemos concluir que éste se ajusta a la secuencia que se sigue en cualquier investigación científica, es decir, la que utiliza cualquier investigador cuando se propone estudiar algo en el contexto de su campo de estudio. En el ejemplo hemos planteado la situación de una persona reflexiva y decidida que se aventura a buscar respuestas para un problema personal. En este caso no se trata de un investigador interesado en elaborar un modelo sobre el comercio urbano; no tiene el conocimiento ni la formación y por tanto no ha tenido en cuenta todos los detalles y requisitos relacionados con el método científico, no ha respetado, por tanto, las condiciones que tienen que cumplir la recogida y análisis de datos para que los resultados puedan ser considerados científicamente válidos.

En cualquier caso el ejemplo nos sirve para:

-  Ejemplificar la secuencia básica de las actividades intelectuales

dirigidas a la obtención de conocimiento sobre un tema.

- ✚ Matizar las diferencias entre el modo de razonamiento reflexivo, basado en nuestras observaciones y vivencias, y el modo de razonamiento científico.
- ✚ Mostrar la relación que existe entre la Estadística y la investigación científica o, lo que es lo mismo, el proceso de generación del conocimiento.
- ✚ Presentar la estadística como un método estrechamente relacionado con los procedimientos de generación del conocimiento.
- ✚ Mostrar que el método estadístico no entra a formar parte de la investigación en el momento de analizar cuantitativamente los datos ya recogidos. Desde que el investigador se plantea las cuestiones que desea conocer o corroborar incorpora en la investigación el razonamiento estadístico.

Razonamiento reflexivo (cotidiano) versus razonamiento científico

Puesto que ya hemos visto el paralelismo entre la secuencia de trabajo planteada por el protagonista de nuestro ejemplo y las fases de la investigación científica, es interesante ahondar un poco más en las diferencias entre los dos tipos de razonamiento. La cuestión se podría formular mediante dos preguntas:

- ▣ ¿Que le ha faltado al protagonista del ejemplo para ajustarse a una investigación científica?
- ▣ ¿Por qué los resultados de su estudio no tendrían más validez que servirle de ayuda, más o menos acertada, para fundamentar su decisión?

En el caso del ejemplo que hemos propuesto, el investigador dispondría, para empezar, de un conocimiento previo sobre las pautas de distribución del comercio en los entornos urbanos. Partiendo de dicho conocimiento, y antes de aventurarse a hacer una exploración de los barrios, habría incrementado y actualizado su conocimiento sobre el tema mediante la lectura y el análisis de los textos y materiales más relevantes. El investigador dispondría, además, del conocimiento necesario sobre las fuentes de información existentes para recabar datos sobre los establecimientos comerciales.

Otra de las diferencias fundamentales entre un profano y un investigador sería que este último elaboraría una planificación rigurosa de la investigación, desde el planteamiento del problema hasta el proceso de recogida y análisis de la información. Y es precisamente en esta fase de planificación en la que el investigador utilizaría su conocimiento sobre el razonamiento estadístico, de donde obtendría:

- ▣ El conocimiento sobre las condiciones deben cumplir los datos para que resulten útiles y qué hay que tener en cuenta a la hora de recogerlos.
- ▣ Los procedimientos y herramientas para aceptar o desechar las hipótesis de partida.
- ▣ Las técnicas y procedimientos para analizar los datos y extraer de ellos conclusiones científicamente válidas.

Si nos damos cuenta, las diferencias entre los modos de hacer cotidiano y científico a la hora de generar conocimiento se relacionan más con el rigor y la precisión en el ejercicio de la actividad que con la estructura formal de la misma. La diferencia no es banal y, de hecho, el incumplimiento de ciertas normas durante el proceso de investigación puede invalidar por completo los resultados. Pese a todo, lo que interesa destacar aquí es la idea de que podemos empezar a comprender en qué consiste el procedimiento de investigación científica partiendo de esquemas de razonamiento que nos resultan, en mayor o menor medida, conocidos. Nos interesa también insistir en la idea de método como la secuencia de pasos que estructuran una actividad. Si entendemos de este modo la Estadística, como un método, habremos superado la concepción errónea que la considera como una simple caja de herramientas.

Terminaremos con una cita de Castro Silva (2003) que sintetiza de forma impecable las similitudes y diferencias entre el razonamiento cotidiano y el científico.



En general los hombres de ciencia coinciden en reconocer la fuerte similitud que existe entre los procesos propios del pensamiento con que el hombre resuelve los problemas de su vida cotidiana y los procesos propios del razonamiento científico. [...] Bajo este prisma, Einstein señala que “El modo científico de formar conceptos no difiere básicamente del que usamos en nuestra vida ordinaria sino sólo por una mayor precisión en la definición de conceptos y conclusiones, por una más concienzuda y sistemática selección de material experimental; por una mayor economía lógica”.

Por otra parte, asumimos con J. Dewey que un proceso reflexivo completo se origina sólo cuando, enfrentado a una situación problemática sentida, el hombre cobra conciencia de que los medios disponibles no son suficientes para superar la situación conflictiva en que se encuentra. Pues bien, el curso formal de un proceso reflexivo del carácter señalado, cubre unas etapas equivalentes a las que se dan en un proceso de razonamiento científico completo.

El modo científico de razonar, equivalente a lo que denominamos “método científico”, es - en el decir de Claude Bernard (1865) - “el razonamiento mediante el cual sometemos nuestras ideas a la experiencia de los hechos”. Lo mismo dirá Guy Claxton 125 años más tarde en una frase breve: “El pensamiento científico implica un trío indisoluble de aspectos: tener ideas, meditarlas a fondo y comprobarlas”. Comprendido el proceso del razonamiento científico de este modo, sus diferencias con los procesos del pensamiento reflexivo que gatillan las necesidades sentidas y sin resolver del hombre corriente, son sólo una cuestión de grado, de refinamiento y de mayor rigor en los procesos. Tanto en una de sus obras primeras (*How we think?*. 1905), como en otra de culminación de su pensamiento (*Lógica de la Investigación*. 1958) John Dewey abunda en consideraciones sobre las semejanzas y equivalencias entre “los pasos” del pensar reflexivo y “los pasos” del método (o razonamiento) científico.

SEGUNDA PARTE:
EL ANÁLISIS DE VARIABILIDAD ES UNO DE LOS
OBJETIVOS FUNDAMENTALES DE LA ESTADÍSTICA

Tema 2:
En qué consiste la variabilidad

Segunda PARTE: El análisis de variabilidad es uno de los objetivos fundamentales de la estadística	1
Los conceptos de variabilidad y variación	3
Las características variables de elementos y colectivos	9
Existe un interés creciente por el estudio de la variabilidad.....	18
 Procesos de globalización y cambio	19
 El deterioro medioambiental.....	23
Estadística, variabilidad y futuro.....	25

SEGUNDA PARTE: EL ANÁLISIS DE VARIABILIDAD ES UNO DE LOS OBJETIVOS FUNDAMENTALES DE LA ESTADÍSTICA

En el primer tema decíamos que la Estadística nos ayuda a crear conocimiento. Esta es una afirmación demasiado general que hemos tratado de aclarar ofreciendo ejemplos concretos de algunas investigaciones en las que se ha utilizado el método estadístico.

En esta segunda parte, el título matiza la idea sobre el tipo de conocimiento que se puede obtener mediante la utilización del método estadístico. La Estadística, según una opinión cada vez más extendida, constituye una herramienta esencial para el estudio de la variabilidad, llegando a afirmar muchos autores que el análisis de variabilidad es uno de los objetivos fundamentales de la estadística.

El concepto de variabilidad juega un papel clave dentro de la Estadística. Si los hechos no se repitieran o se repitieran sin variación, la Estadística no tendría razón de ser; pero la realidad es que la mayoría de los fenómenos se repiten y lo hacen mostrando variaciones de mayor o menor intensidad; de ahí la importancia que tiene la Estadística en el mundo moderno, al suministrarle al hombre procedimientos válidos y confiables para analizar esos hechos que se repiten y hacer inferencias acerca de ellos no obstante la variabilidad que presentan. (Gómez Barrantes, 1997, p. 317)

Sintetizando las ideas que proponen los títulos de los dos primeros apartados, podemos concluir que la Estadística resulta útil para generar conocimiento sobre aspectos de la realidad caracterizados por su variabilidad. La Estadística nos ayuda precisamente a conocer dicha variabilidad.

Lo que necesitamos ahora es comprender el concepto de variabilidad, en qué

consiste y qué queremos decir cuando afirmamos que determinado aspecto de la realidad se caracteriza por su variabilidad. Es preciso conocer el significado y el alcance del concepto de variabilidad antes de plantearnos su análisis mediante el método y las herramientas estadísticas.

Tema 2: En qué consiste la variabilidad

LOS CONCEPTOS DE VARIABILIDAD Y VARIACIÓN

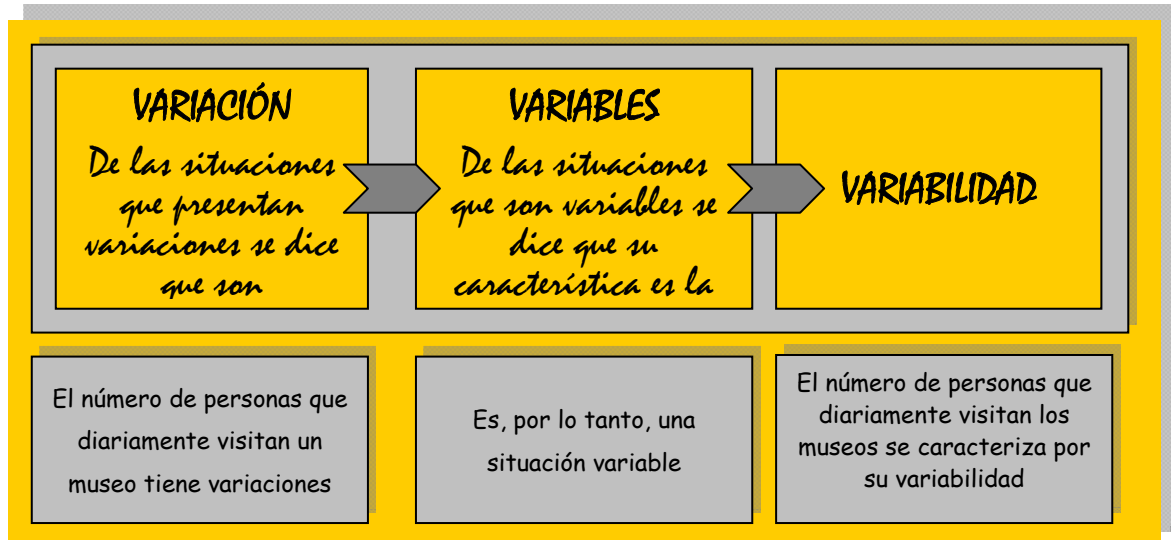
Como primer intento para acercarnos al significado del concepto variabilidad acudiremos al Diccionario de la Real Academia Española¹ de la lengua. El diccionario nos dice que variabilidad es una cualidad de variable. Variable es para la Academia aquello que varía o puede variar.

Sabemos, de momento, que el concepto de variabilidad alude a una cualidad que es propia de las cosas que son variables. Ahora necesitamos saber qué sentido toma esta definición en el contexto que nos ocupa.

Empezaremos diciendo que la variación es una característica de hechos, de personas, de procesos...; la variación está en todo lo que hacemos, en todo lo que nos rodea. El número de personas que visitan un museo varía de un día a otro, la cantidad de fruta que produce un árbol es diferente cada temporada; el número de piezas de fruta que producen dos árboles plantados en la misma parcela suele ser también distinto, al igual que el tamaño de las piezas de fruta de cada uno de ellos....

Puesto que la afluencia de visitantes de un museo presenta variaciones en el tiempo, decimos que es una situación caracterizada por su variabilidad. Hablando de museos y de visitantes, también podríamos comparar el número de personas que visitan diariamente los distintos museos de una ciudad. Si lo hiciéramos, lo más seguro es que obtendríamos cifras distintas para cada uno de los establecimientos. En este caso también concluiríamos que, dado que hay variación en el número de personas que visita cada uno de los distintos museos, la afluencia de visitantes a los museos de la ciudad estudiada es una situación caracterizada por su variabilidad.

¹ <http://www.rae.es/rae.html>



Aunque en los dos ejemplos hablábamos de situaciones de variabilidad en torno a los visitas que reciben determinados museos,

- ▣ En el primer caso analizamos la variaciones de un único elemento a lo largo del tiempo: las diferencias en el número de personas que visitan un mismo museo en distintos períodos de tiempo.
- ▣ En el segundo caso analizamos las variaciones de un elemento a otro, pero en el mismo momento temporal: las diferencias en el número de visitantes que han recibido distintos museos en el mismo período de tiempo.

Todavía podríamos pensar en una tercera situación: las diferencias a nivel territorial, es decir, las variaciones en el número de personas que mensual o anualmente visitan los museos de distintos territorios. Seguiríamos hablando de una variabilidad pero, en este caso, a nivel territorial. Veremos a continuación una serie de datos que permitirán comprender mejor las diferencias entre las tres situaciones planteadas:

Número de visitas que recibieron los museos de la región de Murcia durante 2007 ²	
Museo de la Huerta	20.292
Arqueológico.Encomienda	4.071
Fundación P. Piñero	1.042
Museo Música Étnica	16.796
Arqueológico	22.277
Centro de Artesanía	4.743
Arqueológico	13.662
Museo Siyâsa	14.474
Arqueológico	14.707
Centro de Artesanía	15.130
Museo Monográfico el Cigarralejo	14.707
Museo de la ciudad	19.884
Centro de Artesanía	6.172
Arqueológico de Murcia	46.609
Bellas Artes	59.085
Museo de la Ciencia y el Agua	26.430
Museo Iglesia San Juan de Dios	40.981
Hidráulico Los Molinos del Río Segura	23.808
Salzillo	15.954
Museo Santa Clara	23.017
TOTAL	261.940

VARIACIONES A NIVEL INDIVIDUAL: diferencias en el total de visitantes que ha recibido cada uno de los museos de la región de Murcia en 2007

² Centro Regional de Estadística de Murcia. Anuario estadístico de la ciudad de Murcia. 2007

Número de visitas que recibió el Museo de la Ciudad de Murcia en 2007	
Enero	1.589
Febrero	1.919
Marzo	2.479
Abril	2.453
Mayo	1.751
Junio	1.434
Julio	1.119
Agosto	1.137
Septiembre	1.171
Octubre	1.454
Noviembre	1.724
Diciembre	1.654
Total	19.884

VARIACIONES A NIVEL TEMPORAL: (el mismo museo distintas fechas) diferencias en el número de visitantes que ha recibido el Museo de la ciudad de Murcia en los distintos meses de 2007

Número de visitas que recibieron los museos de la región de Murcia durante 2007	261.940
Número de visitas que recibieron los museos del Principado de Asturias durante 2007 ³	657.77

VARIACIONES A NIVEL TERRITORIAL: (diferencias en el número de visitantes que han recibido en 2007 los museos de la región de Murcia y los del Principado de Asturias)

³ Instituto Asturiano de Estadística. (2007) Anuario Estadístico de Asturias 2007. Gobierno del Principado de Asturias

Podemos concluir que hablar de variabilidad es hablar de una propiedad de un conjunto de elementos que, teniendo algo en común para constituir un colectivo, tienen a su vez alguna característica distinta. En los ejemplos que hemos puesto los colectivos y la característica variable que se analiza serían los siguientes:

- ▣ Los museos de una región constituyen un colectivo cuyo rasgo común es que se trata del mismo tipo de establecimiento cultural y cuya diferencia es, entre otras, el número de visitantes que reciben diariamente.
- ▣ Cuando hablamos de la variabilidad temporal del número de visitas diarias de un museo, nos referimos a un colectivo formado por unidades de tiempo idénticas, pertenecientes al mismo año. Su diferencia -variación- en relación al museo, es la cifra diaria de visitantes.
- ▣ Cuando hablamos de la variabilidad territorial estamos considerando un colectivo cuyos elementos son los distintos grupos de museos de cada uno de los territorios que consideremos.

Escalas de VARIABILIDAD	COLECTIVO	CARACTERÍSTICA
TEMPORAL	Los meses del año 2007	El nº de visitantes del Museo de la Ciudad de Murcia
INDIVIDUAL	Los museos de la región de Murcia	El nº de visitantes que acudió a cada museo en 2007
TERRITORIAL	Las CCAA del estado español	El nº de visitantes que acudió a los museos de cada región en 2007.

Es interesante retener esta tipificación de las escalas de variabilidad (temporal,

individual y espacial o territorial) porque nos resultará muy útil a la hora de entender los estudios estadísticos realizados por otras personas, o de plantear nuestros propios estudios.

Es importante comprender que cuando hablamos de variabilidad nos referimos a una propiedad inherente a la mayoría de las cosas, tangibles e intangibles, que componen nuestro mundo. Varían los caudales de los ríos, de un río a otro y de hoy a mañana, pero varía también el significado de los conceptos, el de las palabras o los gustos y las modas de individuos y sociedades. Decimos que algo varía no necesariamente porque “ese algo” cambie de aspecto o de estado (ayer era azul y hoy es verde) sino porque miramos una característica de un grupo, mayor o menor, de objetos (entidades) y dicha característica es diferente en los distintos objetos.

La variabilidad, omnipresente en nuestras vidas, es objeto de una atención especial en algunos ámbitos. Este es el caso, por ejemplo, del ámbito empresarial, de cuyo amplio espectro de variabilidad el siguiente texto nos da una idea:

Todos percibimos, ya sea en forma tácita o explícita, que en las empresas todo es variable, tanto a través del tiempo como de un departamento a otro. De hecho, no sucede solamente en la vida empresarial; Deming lo resumió diciendo que “la variación es vida; o la vida es variación”. No hay dos clientes idénticos, los efectos de la cultura corporativa sobre el desempeño son diferentes para todos los empleados, los tiempos de llegada y entrega de pedidos o documentación nunca son los mismos; los entornos económico, político y social cambian constantemente, todos aprendemos de forma diferente, etc. (Inda Cunningham, 2000, p. 43)

El interés que en este caso suscita el tema se debe al hecho de que para la empresa la variabilidad es una especie de enemigo que hay que erradicar. Mejorar la calidad implica disminuir la variación en los procesos

En muchos casos la variabilidad no es ningún enemigo, o es sencillamente inevitable, y lo que se necesita es conocer el rango de variación que se produce en determinadas situaciones para poder gestionarlas adecuadamente. Sabemos, por ejemplo, que la cantidad de clientes que acuden a un restaurante varía a lo largo de la semana o del mes; el responsable de la gestión deberá hacer sus previsiones contando con el margen de variabilidad. Poco se puede hacer para aumentar o disminuir la cantidad de agua de lluvia que cae en un territorio a lo largo de un año y, de hecho, esa cantidad puede llegar a ser muy variable de

un año a otro. En este caso también, los encargados de los recursos hídricos de un territorio tendrán que realizar sus previsiones contando con la posible variabilidad de los aportes de agua.

LAS CARACTERÍSTICAS VARIABLES DE ELEMENTOS Y COLECTIVOS

Empezaremos este apartado planteando una serie de ejemplos sobre situaciones de ámbito cotidiano, caracterizadas por su variabilidad. Mediante estos ejemplos presentaremos los conceptos que estructuran un estudio estadístico de variabilidad.

- ▣ Todos conocemos los movimientos migratorios que, a nivel mundial, se están produciendo en los últimos años. Sabemos, también, que los estados más desarrollados de Europa constituyen el destino de muchos emigrantes y estamos cada vez más acostumbrados a la presencia de ciudadanos extranjeros en nuestras ciudades. Por razones distintas, el tema de las migraciones tiene una gran transcendencia y genera numerosas preguntas entre los especialistas e interesados en la cuestión: ¿cuales son, por ejemplo, las ciudades españolas que cuentan con un mayor número de residentes extranjeros? ¿Existen ciudades que resultan más atractivas que otras y, si esto es cierto, por qué razón? ¿Cómo se distribuyen los residentes extranjeros dentro de una ciudad?
- ▣ Muy relacionado con la cuestión de la emigración está el tema de las ayudas sociales que ofrecen los ayuntamientos a los colectivos de ciudadanos más desfavorecidos. La cantidad de dinero que los distintos ayuntamientos destinan a las ayudas sociales suele ser diferente, de modo que no sería extraño preguntarse si la mayor o menor generosidad de las entidades municipales puede tener alguna influencia en la decisión de los emigrantes a la hora de elegir un destino. La pregunta parece clara: ¿a la hora de elegir un lugar para vivir, tienen en cuenta los emigrantes la política de ayudas sociales de los ayuntamientos?
- ▣ Veamos ahora una cuestión que también ha suscitado grandes preocupaciones. Como todos sabemos, el 1 de enero de 2002 el euro se convirtió en la moneda oficial en 12 países de la Unión Europea. Además de las complicaciones propias de un cambio de tal envergadura, en España, la sustitución de la peseta por el euro generó gran inquietud en torno a la posible subida de precios que el cambio acarrearía.

Desde la entrada del euro es frecuente encontrar en los medios de comunicación noticias que dan cuenta del excesivo encarecimiento que han sufrido algunos productos o servicios⁴. Tanto es así, que la entrada del euro marca un antes y un después y ha permitido, en opinión de muchos, camuflar subidas de precio innecesarias. Ahora bien, cabe preguntarse si esos encarecimientos que todos parecemos haber sentido han afectado por igual a todos los productos y servicios o si algunos de estos ha registrado mayores subidas de precios. Nos preguntamos, en definitiva si el encarecimiento ha sido uniforme o variable.

- ▣ Para intentar implicar a la población en las medidas contra el cambio climático y el deterioro ambiental en general las instituciones públicas realizan frecuentes campañas de concienciación. Teniendo en cuenta que los gastos de este tipo de campañas suelen ser elevados resultaría conveniente conocer el nivel de éxito que han logrado. Convendría saber si los mensajes difundidos en las campañas han modificado positivamente la conciencia medioambiental de los habitantes y, en caso afirmativo, si esto ha tenido alguna repercusión en sus prácticas. Sería interesante conocer, por ejemplo, si, como efecto de las campañas de concienciación, se ha producido una mejora en los hábitos de reciclaje.

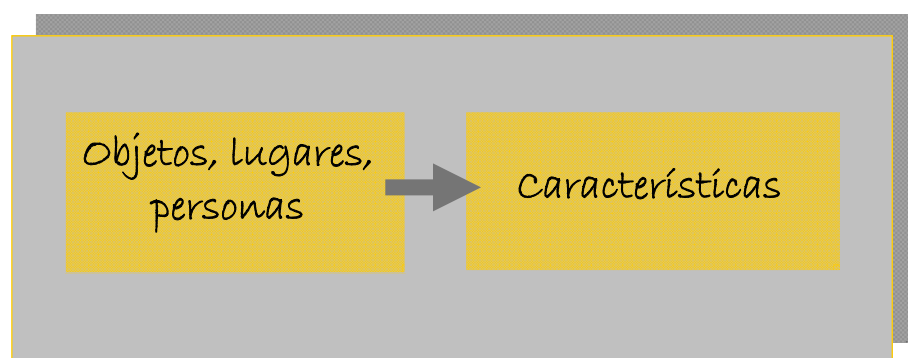
Nadie ignora hoy en día que para responder a preguntas como las que hemos formulado se realizan habitualmente estudios estadísticos. Además, dado que los ejemplos que hemos utilizado se relacionan con temas de gran actualidad, es fácil encontrar en revistas y periódicos los resultados de investigaciones que

⁴ Por otro lado, el efecto de la puesta en circulación del euro sobre los precios ha continuado siendo valorado como inflacionista, aumentando esta sensación con el tiempo. Así, mientras que en noviembre de 2002 el 81% de los encuestados en España afirmaba que el redondeo de precios se había realizado al alza, un año después dicho porcentaje se ha situado en el 87%. Este fenómeno se observa igualmente en el conjunto de la zona euro: el 84,4% consideraba un efecto alcista a finales de 2002, cifra que se elevó a un 89% en los últimos meses de 2003. Este aumento de la percepción inflacionista está relacionado con la comparación de los precios actuales denominados en euros con los precios de los mismos artículos denominados en la moneda nacional a finales del año 2001, sin tener en cuenta la inflación que se habría producido aunque no se hubiera implantado el euro. Además, esta valoración inflacionista de la introducción de billetes y monedas de euro parece haber afectado a la percepción de los agentes sobre algunos de los beneficios potenciales del euro. (Álvarez González, et al., p. 29)

han tratado temas similares y que, por supuesto, han utilizado procedimientos estadísticos similares. Si esto es así, ¿cómo se entiende que investigaciones de temas tan diferentes se lleven a cabo mediante métodos similares? Para encontrar la respuesta debemos fijarnos con atención en las preguntas; si lo hacemos encontraremos rápidamente el rasgo que tienen en común las preguntas y los posibles estudios para encontrar las respuestas:

▣ En todos los casos se mencionan características, tanto materiales como inmateriales, de grupos de personas, de objetos o de lugares:

- ▶ En el ejemplo de los movimientos migratorios, el lugar de residencia elegido sería la característica de la que hablamos. El número de emigrantes recibidos en cada una de las distintas ciudades de una nación europea sería, en este caso, la característica diferencial de un colectivo integrado por dichas ciudades.
- ▶ En el ejemplo sobre el incremento de precios vinculado al cambio de moneda, el precio sería la característica de un colectivo integrado por distintos productos y servicios.
- ▶ En el ejemplo sobre las campañas de concienciación medioambiental la característica podría ser el incremento en peso o volumen de los materiales recogidos, para un período de tiempo dado, en cada uno de los contenedores de recogida de un núcleo urbano.



Aunque no conocemos las respuestas a las preguntas que hemos formulado, es fácil suponer que todas las respuestas serán variables, es decir, reflejarán la variación que existe entre los elementos del grupo en relación a la característica que se analiza:

- En el ejemplo del número de emigrantes asentados en las distintas ciudades, las respuestas serán, seguramente, variables en función no sólo de las ciudades consideradas sino del momento en que se haga el estudio.

El barrio de Cala Major ya tiene un 36% de población inmigrante, Gomila un 28% y la zona centro un 17%

13/6/2005 - Un informe del Observatorio de la Inmigración desvela que el 13% de los residentes de Palma es extracomunitario - El 60% tiene menos de 40 años.

Los barrios palmesanos de Cala Major, El Terreno, Ponent, Estacions y la Plaza de España tiene entre un 20 y un 36% de población inmigrante extracomunitaria. Son datos del Observatori Municipal de la Inmigración, que ha hecho un estudio estadístico a fondo sobre la población inmigrante en Palma en base a los datos del padrón municipal a fecha de 1 de enero de 2004 Concretamente, Cala Major es el barrio que más inmigrantes extracomunitarios concentra, ya que el 36% de sus residentes es extranjero....

<http://www.icare.to/icare-espagnol/news.html> (ICARE. Centro Internet Anti Racismo Europa)

- Las subidas de precio que se produjeron como consecuencia del cambio de moneda de la peseta al euro serán, probablemente, bastante diferentes. También en este caso encontraríamos variaciones no sólo en función de los productos y servicios, sino del lugar e incluso del establecimiento.

En la página Web de la OCU (Organización de Consumidores y Usuarios) se difundieron los resultados de numerosos estudios realizados por dicha organización en las fechas posteriores al cambio de moneda. En uno de estos estudios podíamos encontrar la información siguiente:

Este es el sexto y último control de precios. Nuestro objetivo al plantear estos estudios era controlar el cambio de moneda y algunos de los aspectos relacionados con él: la indicación de precios, los posibles errores en la conversión, la presencia del logotipo de buenas prácticas, la fijación de precios psicológicos en euros y las subidas de precios reales de precios. [...]

Nos centraremos en la evolución de los precios tanto desde implantación efectiva del euro, en enero de 2002, como desde el comienzo de nuestros controles, en septiembre de 2001. [...]

Públicos versus privados:

- Entre enero de 2002 y enero de 2003, los precios públicos se han incrementado mucho menos que los privados. Esto quiere decir que en el sector público las subidas de precios producidas a principios de este año y durante 2002, han sido, globalmente, más moderadas que en el privado. [...]

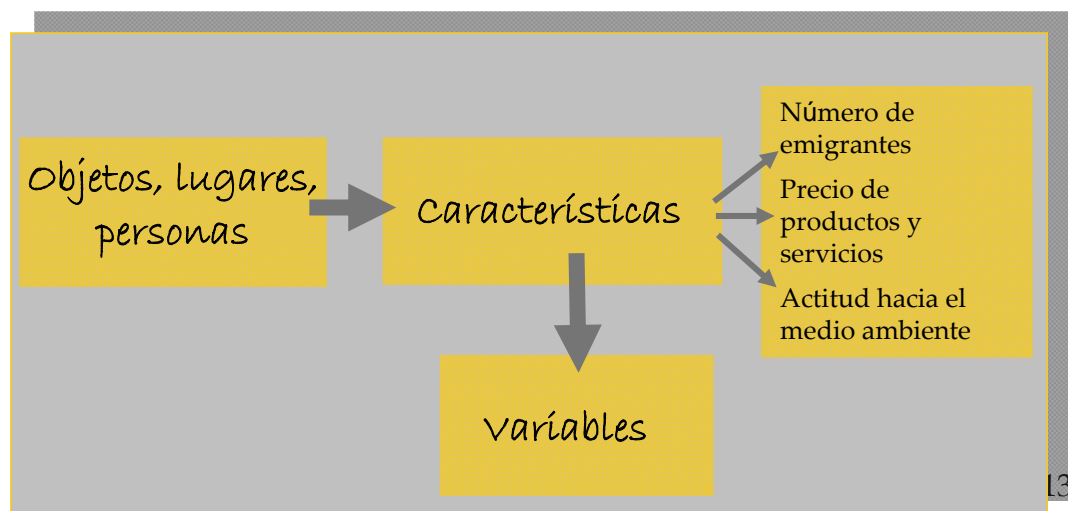
Por sectores:

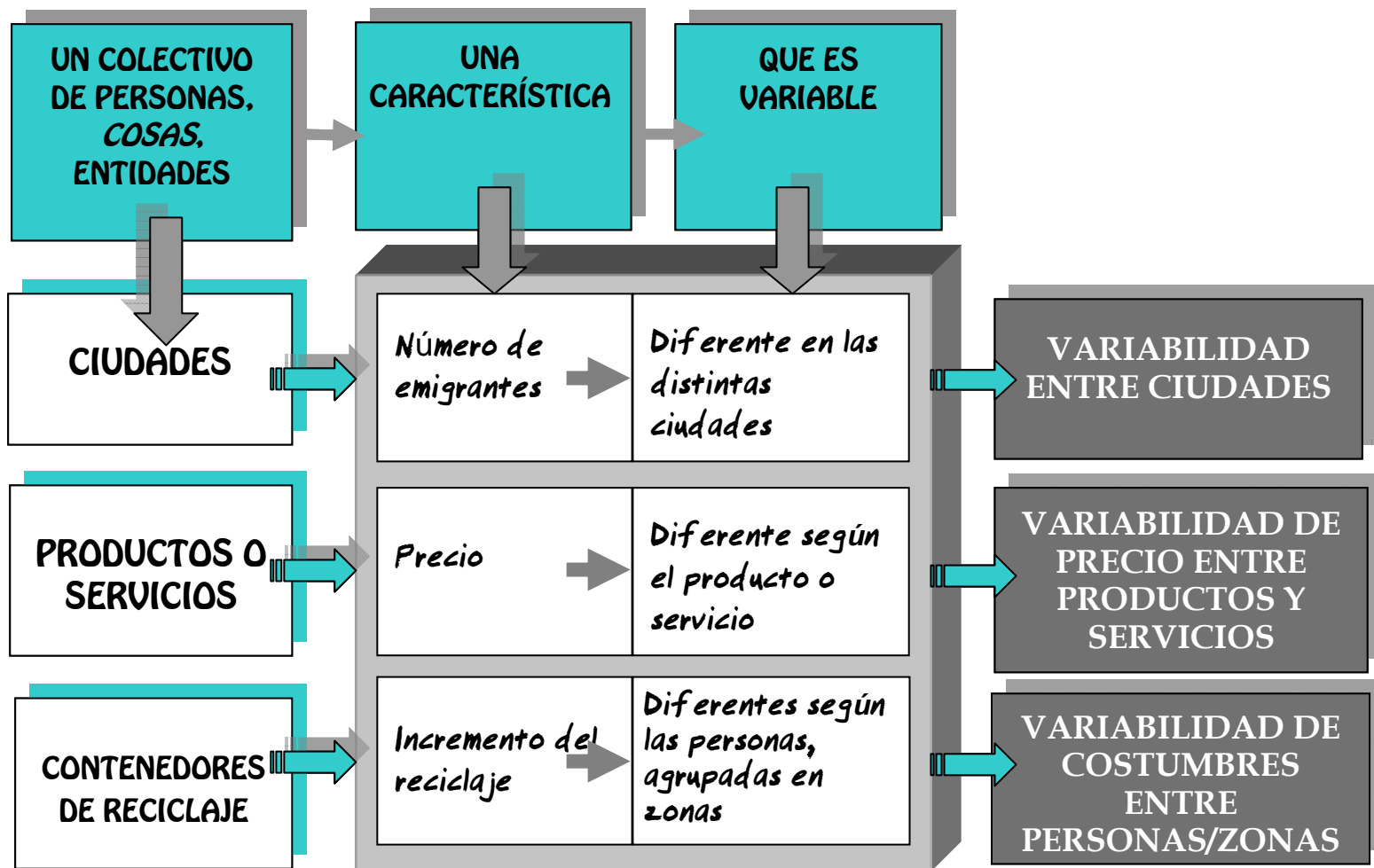
- Tomando como referencia el último año, la subida más espectacular ha sido la de las piscinas cubiertas (un 19,2 %), seguida de peluquerías (12,1 %), papelerías (11,4 %), otras actividades de ocio (11,2%) y el sector de los carburantes (10,1 %).

<http://www.ocu.org/map/src/52121.htm> Enero 2003 - Informes

- La conciencia medioambiental de los ciudadanos y las acciones que estos realizan para frenar o disminuir el deterioro suelen ser también variables. No sería extraño pensar que el alcance de las campañas de sensibilización sea variable en función de distintos factores y que los resultados que puedan notarse en la mejora del reciclaje sean un reflejo de ello.

Independientemente de los resultados posibles de los estudios que proponemos y de las especulaciones que podamos hacer al respecto, lo que nos interesa destacar de los ejemplos que hemos puesto es que, en todos los casos, el objeto de estudio es el comportamiento diferente (variable) de los componentes de un colectivo frente a una característica determinada. En todos los casos, por tanto, lo que se ha propuesto es un estudio de variabilidad.





Cuando los componentes de un colectivo, con respecto a una misma característica, tienen comportamientos distintos, decimos que la situación es variable; es decir, hablamos de variabilidad.

Los ejemplos que hemos puesto son una pequeña muestra de la infinidad de características variables que podemos encontrar. Una vez que comprendamos que el concepto de variabilidad se utiliza para mencionar las diferencias que existen entre los elementos de un colectivo ante una característica dada, nos daremos cuenta de que la lista de colectivos y características resulta inacabable. Bastará pensar en un grupo de entidades para darnos cuenta de la cantidad de elementos diferenciales que existen entre sus elementos:

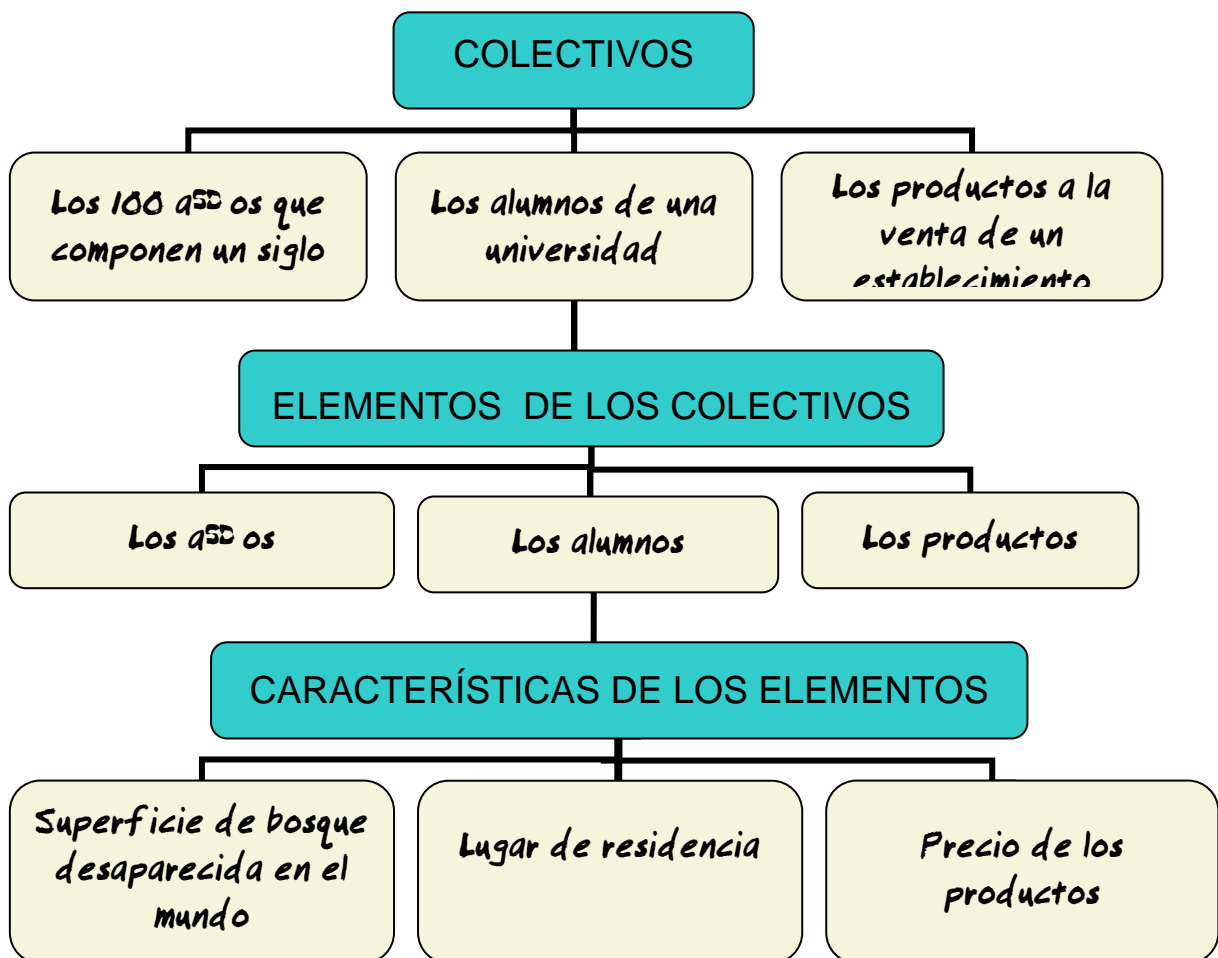
▣ El colectivo puede estar constituido por el conjunto de años que forman un siglo y la características que analicemos pueden ser, entre otras muchas:

↳ El número de muertos que se han producido, en todo el mundo, como consecuencia de las guerras. Dicho número será diferente de un año a otro, es decir, de un elemento a otro del colectivo.

↳ La superficie de bosques que se ha perdido en el planeta también, será diferente de un año a otro.

↳ Cada año será también distinto el número de personas nacidas, o el de catástrofes naturales.

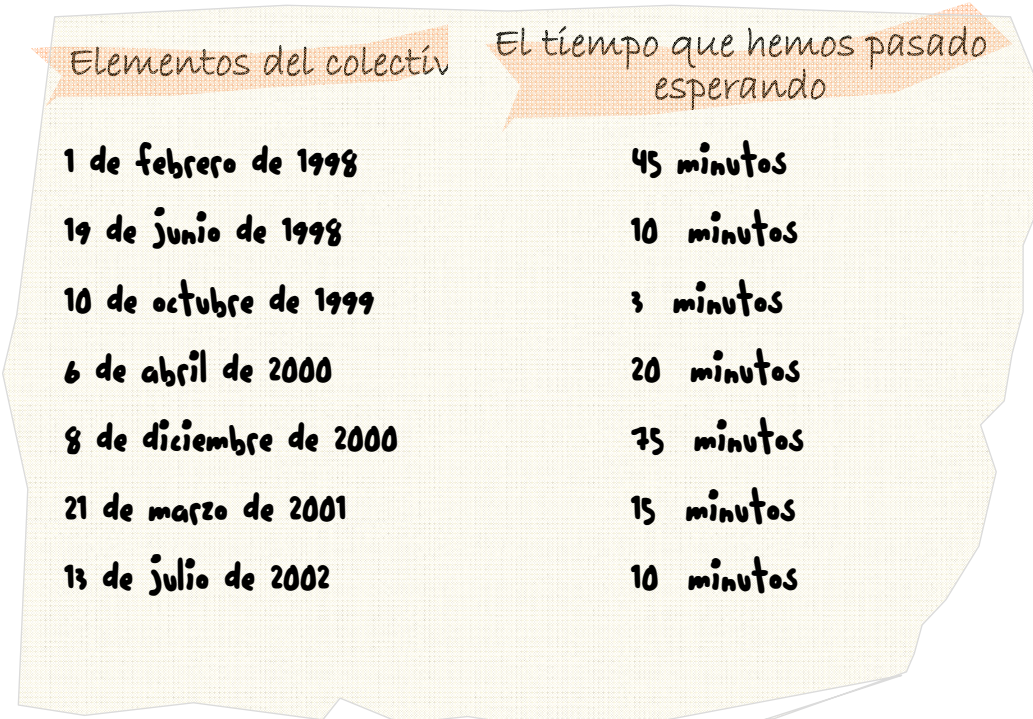
▣ Los alumnos de una universidad constituirían otro colectivo, así como las comunidades o provincias de un país; los que han nacido en la misma fecha; los ríos que pertenecen a un misma cuenca fluvial. En fin, la lista es inagotable.



Lo cierto es que vivimos rodeados de variabilidad y, seamos o no conscientes, estamos acostumbrados a amoldarnos a ella.

Cuando tenemos una consulta médica, no solemos saber de antemano el tiempo que nos llevará porque, además de la entrevista con el médico, hay que contar con el tiempo de espera previo a la consulta. Partiendo de esta idea, podríamos formar un colectivo integrado por todas las visitas al médico que hemos realizado en un período de tiempo determinado. La característica a estudiar, claro está, el tiempo de espera que tuvimos en cada una de ellas.

Colectivo: las visitas realizadas al médico durante un tiempo determinado



Elementos del colectivo	El tiempo que hemos pasado esperando
1 de febrero de 1998	45 minutos
19 de junio de 1998	10 minutos
10 de octubre de 1999	3 minutos
6 de abril de 2000	20 minutos
8 de diciembre de 2000	75 minutos
21 de marzo de 2001	15 minutos
13 de julio de 2002	10 minutos

Cuando decidimos comprar algo tenemos claro, a veces, que por distintas razones, el precio del objeto que buscamos puede llegar a ser muy diferente de un establecimiento a otro. Qué decir, por ejemplo, si hablamos del precio de la vivienda y de las diferencias que pueden existir entre lugares de residencia y entre las propias viviendas.

Los ejemplos sobre el tiempo que pasamos en las salas de espera, o sobre la diferencia de precio de un producto según el lugar o el establecimiento, incluido el de la vivienda, plantean situaciones que forman parte de la experiencia de cualquier persona adulta. A nosotros nos sirven en este momento como ejemplo para mostrar que, incluso al nivel más cotidiano, todos

estamos acostumbrados a convivir con la variabilidad y, a veces, a buscar estrategias para hacerle frente.

EXISTE UN INTERÉS CRECIENTE POR EL ESTUDIO DE LA VARIABILIDAD

No podemos decir que el estudio de la variabilidad sea algo nuevo en los ámbitos técnico y científico. Existen innumerables estudios que se han ocupado de analizar la variabilidad de alguna situación mediante herramientas estadísticas. De hecho, como decíamos al principio, no son pocos los que consideran el estudio de la variabilidad como el corazón de la Estadística.

No es nuevo, por ejemplo, que el clima es un hecho cambiante, que las temperaturas medias varían de un año a otro o que hay años más húmedos que otros. Este tipo de variabilidad ha sido y es un tema habitual de estudio. Lo mismo podríamos decir en relación a las cosechas, cuyo volumen varía de un año a otro o con respecto al número de alumnos que se matriculan en la universidad cada año.

Pero lo que sí parece haber ocurrido es un incremento de la importancia de las situaciones que son variables y de los estudios que se ocupan de analizarlas. Ha aumentado en general el interés de la sociedad por la variabilidad y no son pocos los autores que así lo reconocen. Habría que decir, quizás, que el incremento de la variabilidad en determinados ámbitos se ha convertido en algo más preocupante. Al menos esta es la información que a diario nos transmiten los medios de comunicación a través de inquietantes noticias. En un informe de 2004 sobre seguridad económica, elaborado por la Organización Internacional del Trabajo, se habla de la variabilidad del sistema económico y de sus tremendas repercusiones:

...un incremento de la frecuencia y la gravedad de las crisis económicas en el marco de la globalización. Ha aumentado la variabilidad de las tasas de crecimiento económico y problemas individuales, como el desempleo o la enfermedad, son desplazados por conmociones o crisis de mayor amplitud que afectan a comunidades y regiones enteras. (DCOMM, 2004)

Las páginas que siguen tienen la intención de ser un espacio para recoger los planteamientos de algunos autores y algunas de las ideas más generalizadas en torno a la importancia de la variación en el mundo actual. Para ello, hemos agrupado la información en dos epígrafes que relacionan la intensidad de las variaciones con los procesos de globalización, en primer lugar, y con el deterioro medioambiental en segundo lugar. Vaya por delante entonces, que aunque el discurso nos pueda traicionar a veces, estamos de acuerdo con quienes consideran que *Aún no se ha establecido un consenso metodológico sobre qué*

asuntos son causa y cuáles efecto de los cambios, pero toda caracterización resalta la acumulación de transformaciones. (Alcaraz Ramos, 2009, p. 256). El objetivo es más bien el de ofrecer unas pinceladas para provocar la reflexión que el de intentar un análisis sistemático sobre el tema. Se explica así que nos dejemos fuera muchos de los factores que habitualmente se relacionan con el incremento de la atención a la variabilidad, como son, entre otros, las facilidades de estudio y análisis que ofrecen las nuevas tecnologías de información y los ordenadores.

Procesos de globalización y cambio

Para intentar explicar a qué se debe el incremento del interés por la variabilidad es inevitable dirigir la mirada hacia los procesos de globalización. En las últimas décadas hemos asistido a un conjunto de cambios que han modificado sustancialmente nuestras formas de vida, de relación, de trabajo. Vivimos una situación completamente nueva para la sociedad, una situación sin referentes históricos próximos o lejanos en los que buscar semejanzas, elementos de comprensión o reconocer paralelismos; es la situación que acompaña los procesos de globalización. En la página Web de La Comisión Mundial sobre la Dimensión Social de la Globalización podemos leer algunos comentarios alusivos a la intensidad de los cambios que estamos viviendo:

La Comisión Mundial sobre la Dimensión Social de la Globalización fue instituida por la Organización Internacional del Trabajo (OIT) en febrero de 2002. La Comisión fue un organismo independiente creado para dar respuesta a las necesidades de las personas **que resultan de los cambios sin precedentes** que la globalización provoca en sus vidas, sus familias y en la sociedad donde viven.⁵

Dentro de esta dinámica cambiante generada, o al menos acentuada, por la globalización es posible identificar situaciones en las que la variabilidad y sus consecuencias prácticas resultan especialmente notorias. Nos referimos, por ejemplo, a todas aquellas situaciones que se han visto afectadas por la aparición o el incremento de la competitividad.

El incremento de la competitividad y los efectos que esto produce resultan particularmente evidentes en el ámbito empresarial. La creación de un mercado de ámbito mundial supone para muchas empresas y sectores de actividad un aumento extraordinario del número de competidores con los que tienen que rivalizar para vender su producto. Aumenta el número de competidores y aumenta también el número de posibles consumidores generando con ello un escenario económico extremadamente complejo y cambiante.

⁵ Comisión Mundial sobre la Dimensión Social de la Globalización.

<http://www.ilo.org/public/spanish/wcsdg/>

Las empresas están expuestas a un entorno cada vez más cambiante, incluso en ocasiones acompañado de fenómenos de volatilidad importantes. La variabilidad en los precios (ej. volatilidad de mercados de materias primas), los cambios en las regulaciones nacionales, la variabilidad de los patrones de consumo (ej. mercados de gran consumo), la introducida por los movimientos de los competidores, el dinamismo de la cadena de suministro (distintos proveedores, distintos operadores, reorganizaciones de redes de centros logísticos...), y otros muchos elementos dibujan un panorama de incertidumbres en el que la logística se convierte en un instrumento poderoso de contingencia. (Ciprés et al., 2007)

Pero la competitividad no afecta sólo a las empresas que luchan por producir y vender más, por incrementar sus beneficios. La competitividad se produce también entre personas, instituciones, ciudades, museos, etc. En un texto sobre la importancia de una buena gestión en el futuro de las ciudades encontramos un párrafo revelador:

Existe ya una competencia entre ciudades en el mundo, que se extiende a una lucha por recursos, infraestructura, visibilidad, atracción (de talento, empresas, etc.) y, claro está, también entre ciudades de un mismo territorio. Y sorprende observar, sin ni siquiera tener que acceder a los datos concretos de sus presupuestos, cómo hay ciudades que apuestan básicamente por el bienestar de sus ciudadanos, y hacen de ello bandera (el caso de Barcelona, «la ciudad de la gente»), mientras que otras se presentan como capitales del futuro, lugares donde la economía se va a reinventar (el caso de Singapur, «where great things happen»). La elección de cuál es la identidad que se transmite va a tener, en mi opinión, un papel fundamental en los próximos años. Y va a depender, obviamente, de un modelo ideológico, pero también de la visión que tengan los dirigentes (¿o deberíamos empezar a llamarlos directivos?) de la ciudad. (Cornellá, 2008)

Competir supone medir y comparar, medirse con uno mismo y con los demás. Para poder competir hay que conocer las diferencias (las variaciones) que hay entre los que participan en el juego, en el mercado global. Hay, además, que estar alerta ante las posibles variaciones que se producen en las “puntuaciones” de los rivales, rivales cuyo número ha aumentado por la inserción en el mercado de un buen número de países y de empresas.

Expresiones como *medir para conocer*, *conocer para comparar* o *medir para conocer, conocer para mejorar* resumen bien el papel de la medición en el mundo competitivo

actual. Se ha impuesto la cultura de la medición o, si se prefiere, de la evaluación⁶. Mientras exista la sospecha de que es posible mejorar la calidad de algún producto, proceso o servicio, se analizan una vez tras otras sus características. Como resultado de esta necesidad de conocer-medir se generan y se analizan cantidades impresionantes de datos⁷. Cada día son más los observatorios que se crean para recoger datos sobre todo tipo de cuestiones, para analizar las diferencias (la variabilidad) en el tiempo o en el espacio de numerosos fenómenos. Cada día también aumenta el número de publicaciones que recogen los resultados de estudios que comparan aspectos económicos, sociales etc. entre distintas entidades territoriales.

Los procesos vinculados a la globalización parecen haber generado numerosos e importantes cambios que han acentuado los niveles de variabilidad en muchos ámbitos. El incremento de la competitividad y el fuerte empuje de la cultura de la medición serían algunas de las estrategias para adaptarse a esta nueva situación, caracterizada por la incertidumbre, y explicarían, al menos en cierta medida, el incremento de la preocupación y de los estudios relacionados con la variabilidad.

Es fácil encontrar en Internet referencias que dan cuenta del alcance que tiene la medición en el mundo actual. En un artículo que detalla las medidas del Plan de Modernización de los Servicios de la Administración del Gobierno de Cantabria para el período 2005 - 2007 podemos leer que “la impartición de cursos, seminarios y sesiones informativas, dedicados a fomentar la cultura de la medición” constituye una de de las medidas de dicho plan⁸. (Bedia Miguel,

⁶ Se trata de una cuestión controvertida: mientras para algunos sectores el seguimiento y la medición continuas son positivas e imprescindibles en el desempeño de cualquier actividad, para otros se trata de una obsesión de consecuencias negativas: Esta obsesión por medir resultados buscando mecanismos sistemáticos, baratos, uniformes y homogéneos disparó el frenesí por los sistemas de evaluación que hoy ocupan a la burocracia de los sistemas educativos, olvidados de promover las pedagogías, las didácticas, la actualización y perfeccionamiento docente, dejando librada a la competencia post evaluación la calidad que dicen defender. (Maffei, 2003)

⁷ El Informe del Departamento de Comunicación (DCOMM) de la OIT nos da una idea del volumen de datos que generan algunas de las investigaciones: “Además de servirse de un banco de datos de escala mundial sobre políticas nacionales, el informe utiliza estadísticas de un conjunto de encuestas sobre seguridad de la población llevado a cabo en 15 países, en las que se consultó a 48.000 trabajadores acerca de su labor, las inseguridades que experimentan y su actitud frente a la desigualdad y los aspectos afines de la política social y económica.” (DCOMM, 2004)

⁸ De hecho, el autor encabeza su artículo con una cita muy significativa: “Lo que no se define no se puede medir, lo que no se mide no se puede mejorar y lo que no se mejora se degrada siempre”. Lord

2008, p. 43)

La literatura en torno a la calidad está plagada de menciones sobre la medición como una de las herramientas para el control y la mejora de productos y procesos o de la satisfacción de los clientes. En el ámbito económico de la producción la variabilidad es uno de los principales enemigos de la calidad; los estudios y las mediciones que se hacen buscan las causas de la variación con el fin de eliminarla o, al menos, atenuarla. Behar y Ojeda (2006) afirman que “la variabilidad es inevitable”, que se convive con ella “de la misma manera en que se convive con algunas condiciones adversas del medio ambiente, siempre y cuando éstas no trasciendan ciertos límites.”



Kelvin

La última búsqueda que hemos realizado en Internet mediante Google, utilizando la Expresión “cultura de la medición”, proporcionó el 1 de marzo de 2010 **1.140.000** resultados.

El deterioro medioambiental

Hay un segundo grupo de factores que nos ayuda también a explicar el auge experimentado por los estudios y análisis de variabilidad. Se trata de las consecuencias medioambientales derivadas de la intervención descontrolada del ser humano en el planeta.

Hablamos, también en este caso, de un tema controvertido ya que no todos los especialistas se muestran de acuerdo sobre las causas y consecuencias de tales cambios. En cualquier caso, las sospechas a veces, y las evidencias otras veces, de la existencia de cambios medioambientales no deseables han motivado la recogida de inmensas series de datos. Según Abaurrea et al. (2002) *la creciente preocupación por la incidencia del cambio climático ha contribuido a que los estudios sobre la evolución de la lluvia, a muy diferentes escalas, sean abundantes.* (Abaurrea et al., 2002)

Las alteraciones medioambientales incluyen en muchos casos un incremento de la variación que experimentan algunos procesos. En el caso del clima, por ejemplo, el aumento de la variación se traduce en un incremento de las situaciones climáticas extremas, es decir, de los períodos inusualmente fríos o cálidos, húmedos o secos. En un artículo sobre el aumento de la variabilidad climática en España, el geógrafo José Jaime Capel Molina realiza afirmaciones tajantes al respecto:

En las postrimerías del siglo XX, uno de los principales motivos de incertidumbre medioambiental a nivel global es el del incremento de gases de efecto de Invernadero y ligado al mismo el caldeamiento global inducido por la actividad del hombre. Entre la comunidad científica cada vez es mayor el interés por las numerosas anomalías del clima que están apareciendo en sus diferentes escenarios del planeta. España como tantos otros países de la zona Templada está soportando una anomalía muy reseñable en el clima con fluctuaciones ostensibles en las condiciones de promedio y eventos meteorológicos extraordinarios.

El reciente incremento tan relevante de la variabilidad climática en España, referido a elementos del clima como la temperatura, la precipitación y, en cierta medida, la presión atmosférica, tanto en promedio como en el régimen ha permitido calificar a la última década 1981-1990, por numerosos investigadores y especialistas sobre el clima, como la más anómala del presente siglo (UNES ESCARDÓ, 1989) (FONT TULLOT, 1989) (QUEREDA SALA, 1994) (OLCINA CANTOS, 1994). (Capel Molina, 1995, p. 42)

Antonio Ruiz de Elvira realizó afirmaciones similares en una conferencia sobre el cambio climático:

Así comenzó su intervención Antonio Ruiz de Elvira, catedrático de la

Universidad de Alcalá de Henares y coordinador del ciclo de conferencias y debate *Una discusión sobre el Cambio Climático. ¿Está determinado por la acción del hombre?*, celebrado en CosmoCaixa.

Las mayores temperaturas máximas y menores mínimas son un hecho real. Este incremento de la variabilidad no puede ser explicado sencillamente como un cambio natural. Siempre ha habido inundaciones, pero no, al menos no en los últimos cien años, con la variabilidad que detectamos y que vemos también en los registros de las temperaturas marinas. (Clavería, 2003)

En relación con las modificaciones de origen climático y también con las alteraciones en el tipo y en la densidad de la cubierta vegetal se producen severas alteraciones en el caudal y en la dinámica hidrológica de las cuencas fluviales.

Los cambios ambientales que se están observando en muchas regiones del Planeta, así como las transformaciones que están experimentando los usos del suelo en ámbitos geográficos muy diversos, están suscitando un gran número de investigaciones que tienen por objeto analizar las consecuencias de dichos cambios en medios tan sensibles como los forestales. (Martínez Fernández et al., 2005, p. 7)

Entre las alteraciones que se han podido constatar está el incremento de la variabilidad temporal del caudal de los ríos. Este es, por ejemplo, el resultado que han obtenido Machín et al. (2005) en su estudio sobre la dinámica fluvial del sector de cabecera del río Arlanza. Conclusiones similares, bastante alarmantes, son las que detallan López Bermúdez y Romero Díaz (1998) en un trabajo de balance de las investigaciones realizadas en torno a los procesos de erosión y desertificación en ámbitos mediterráneos. Dentro de la larga lista de consecuencias ambientales de la erosión hídrica y de la desertificación los autores incluyen *la degradación de los recursos hídricos e incremento de la variabilidad en el régimen de los cursos de agua*

ESTADÍSTICA, VARIABILIDAD Y FUTURO

En las páginas anteriores hemos presentado ejemplos suficientes como para adquirir una primera impresión sobre la omnipresencia de la variabilidad en nuestras vidas. Los ejemplos nos han servido también para comprobar que, en muchos ámbitos, la existencia de variaciones es un tema preocupante.

Sin embargo, la cuestión central de este curso no es el estudio de la variabilidad en ámbitos concretos ni el análisis de sus causas. Todas las reflexiones y citas que hemos recogido tienen por objeto contextualizar las afirmaciones en torno al nivel de interés y preocupación que suscitan hoy en día la variabilidad y la incertidumbre a ella asociada y la función que cumple el método estadístico a la hora de afrontar esta situación. Se trata, en definitiva, de profundizar en los objetivos que fundamentan el análisis estadístico. La cuestión es entonces presentar las ideas que nos lleven a preguntarnos qué aporta o qué puede aportar la Estadística a la hora de tratar con la variabilidad. Esta sería la pregunta esencial en el contexto del curso que presentamos.

En el intento de definir de la mejor manera posible la relación entre la Estadística y la variación se han dado multitud de respuestas a la pregunta que hemos formulado. Maxine Pfannkuch, reconocida especialista en didáctica de la Estadística, resume los términos de esta relación de forma breve pero muy significativa:

Los primeros tres mensajes de la variación son: la variación es omnipresente; la variación puede tener consecuencias prácticas serias; y la Estadística proporciona un medio para entender un mundo acosado por la variación. (Pfannkuch, 1999, 209)

De todos modos, también hay que decir que lo que la Estadística pueda aportarnos depende de nuestros deseos o expectativas en relación a la variabilidad de alguna situación. Maxine Pfannkuch (1999), aborda esta cuestión mediante el concepto de *respuestas prácticas a la variación*. Esta autora considera que las reacciones habituales ante la variación se pueden agrupar en tres tipos de situaciones:

- ▣ las situaciones en las que se ignora, en las que se actúa como si no existiera o bien se elaboran estrategias para adaptar las actuaciones al rango de variación;
- ▣ las situaciones en las que se permite la variación porque se desea conseguir productos que se adapten a un amplio rango de situaciones;
- ▣ por último, están las situaciones en las que se intenta modificar el patrón de variación porque lo deseable es el comportamiento homogéneo de determinadas poblaciones.

Podemos resumir los planteamientos de Pfannkuch diciendo que cuando la variación no es algo deseable, podemos ignorarla, amoldarnos a ella o intentar reducirla. Habría que añadir que, a menudo, no se trata de una decisión, es decir, que no podemos optar ya que la variación es inevitable y la única opción posible es adaptarse a ella. Este sería el caso de muchos fenómenos naturales cuyas variaciones son importantes, tienen fuertes repercusiones y, sin embargo, en muchos casos llegan a ser inevitables. Pensemos, por ejemplo, en las lluvias que, debido a su variabilidad en algunas zonas, generan situaciones de escasez o falta de recursos hídricos, disminución de la productividad en los cultivos o, incluso, situaciones catastróficas con pérdidas materiales y humanas⁹.

Adaptarse supone generar estrategias para disminuir o eliminar las consecuencias negativas que pueden derivarse de la variabilidad de una situación. Una de las estrategias esenciales para afrontar la variación es la información. Se trata de aceptar que no podemos evitar la variación pero que si disponemos de suficiente información en torno a ella, podemos realizar actuaciones que nos permitan reducir sus efectos.

La información y el conocimiento del mayor número de detalles posibles sobre el modo en que varían las situaciones que nos interesan es, de hecho, una de las mejores alternativas para hacer frente a la variación, tanto si se trata de variaciones que se pueden modificar como si hablamos de las que son inevitables. El conocimiento sobre los extremos que puede alcanzar la variación, sobre los patrones de cambio habituales, sobre los factores responsables, etc. nos permite hacer la valoración necesaria para fundamentar las posibles decisiones a la hora de afrontar la variación. Nos permite, de algún modo, acotar la incertidumbre.

Si la información y el conocimiento son los antídotos contra la incertidumbre, la Estadística constituye el método para afrontarla, ya que nos proporciona formas de razonamiento y herramientas para conocer el tipo y el grado de variabilidad de situaciones y procesos así como para investigar sobre las causas y factores que la generan.

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y

⁹ Habría que matizar que, en ocasiones, el límite entre lo evitable y lo inevitable no está tan claro y llega a ser objeto de controversia. Nos referimos a situaciones en las que la variación causante del problema se relaciona con la alteración de determinadas condiciones naturales de una zona o un ecosistema. En este tipo de casos, quienes consideran que es posible restablecer las condiciones originales del entorno suelen defender el carácter evitable de la situación, es decir, de las variaciones problemáticas derivadas de la alteración.

cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos. (Ríus, Barón, Sánchez, Parras, 1999, p. 14)

Ahora bien, cuando obtenemos datos sobre las variaciones que se producen en determinadas situaciones dichos datos nos pueden informar sobre el presente o el pasado. La incertidumbre, sin embargo, se siente en relación a acontecimientos futuros. Y entonces, ¿qué podemos hacer con respecto a las variaciones futuras? Y más concretamente, ¿qué puede aportar la Estadística en este caso?

Es evidente que nadie puede conocer lo que ocurrirá en el futuro con ninguna situación que sea variable y de la que no se conoce con la necesaria precisión y seguridad cómo actuarán los factores responsables de la variación. No podemos saber de antemano, por ejemplo, cuántos accidentes de tráfico se producirán en la próxima operación salida de vacaciones o el número de personas que acudirá a votar en las próximas elecciones. Como demuestra la experiencia, el número de accidentes de tráfico o de votantes en unas elecciones varían, mucho o poco, de una ocasión a otra. En el primero de los ejemplos, la atención que los conductores ponen en la carretera, su estado físico, la densidad de tráfico, el estado del firme, las condiciones climatológicas son, entre otros muchos, el tipo de factores que intervienen en el número de accidentes y cuya influencia no es posible conocer con antelación.

Existen, sin embargo, otro tipo de fenómenos cuya situación futura se puede conocer de antemano porque se puede saber con precisión cuál será la influencia de los factores que controlan su funcionamiento. Hablamos de fenómenos que se rigen por leyes conocidas que permiten prever su futuro con total seguridad. Los astrónomos, por ejemplo, son capaces de predecir los eclipses de Sol o de Luna con muchos años de antelación. La física clásica permite predecir con precisión el comportamiento futuro de una partícula o un cuerpo si se conoce su posición y la cantidad de movimiento o momento (masa por velocidad).

Resumiendo, podemos decir que existen dos tipos de fenómenos:

- ▣ Aquellos cuyo funcionamiento se ajusta a leyes físicas, de tal modo que, una vez fijadas las condiciones iniciales, es posible prever su estado futuro.
- ▣ Aquellos cuyo comportamiento futuro no se puede prever porque no es posible conocer las condiciones iniciales o cómo actuarán los factores que influyen en su desarrollo.

En el primer caso, se habla de fenómenos deterministas, y de fenómenos

aleatorios en el segundo. Mauricio Contreras ¹⁰los define como sigue:

Fenómenos o experimentos deterministas son aquellos que, realizados en las mismas circunstancias, sólo tienen un resultado posible.

Fenómenos o experimentos aleatorios son los que pueden dar lugar a varios resultados, sin que pueda ser previsible enunciar con certeza cuál de estos va a ser observado en la realización del experimento.

En el caso de los fenómenos denominados aleatorios, ante la imposibilidad de conocer con antelación su comportamiento futuro, la alternativa que nos queda es intentar hacer predicciones de lo que puede ocurrir y tomar las decisiones necesarias conforme a ellas. Según el diccionario de la Real Academia Española de la Lengua predecir consiste en *anunciar por revelación, ciencia o conjetura algo que ha de suceder*. La aportación de la Estadística al conocimiento sobre las posibles variaciones futuras consiste precisamente en que nos proporciona formas de razonamiento y técnicas para predecir, de forma científica, hechos sobre una realidad que todavía no se ha producido.

Existe una amplia variedad de técnicas que posibilitan la realización de predicciones futuras. Una parte de estas técnicas proceden del ámbito de la Estadística y suponen el manejo de información cuantitativa. Existen también otro tipo de técnicas que utilizan exclusivamente información de tipo cualitativo. Este sería el caso, entre otros, de las técnicas que intentan predecir acontecimientos futuros mediante la consulta a expertos quienes, basándose en una amplia experiencia en el tema, emiten su opinión sobre el futuro de la cuestión que se estudia. Otra de las posibilidades consiste en buscar alguna situación suficientemente similar a la que queremos predecir pero que ya haya ocurrido en otro momento o lugar. En este caso, se trata de predecir un acontecimiento futuro basándose en los resultados que se obtuvieron en una situación análoga a la que se quiere anticipar, es decir, se trata de predecir en base a una supuesta analogía entre dos situaciones.

Sea cual sea el tipo de técnicas que se utilicen para realizar predicciones, lo esencial es comprender que, dado que no es posible determinar con total certeza lo que va a ocurrir en un futuro, toda predicción contiene un margen de riesgo, una probabilidad de resultar errónea.

Es preciso comprender también que, pese al posible riesgo de error, las predicciones resultan imprescindibles en aquellos ámbitos o situaciones en las que es inevitable tomar decisiones de futuro antes de que las cosas ocurran. La realización de predicciones que permitan a su vez hacer previsiones de futuro

¹⁰ Contreras, M. Taller de Estadística. <http://www.mauriciocontreras.es/estadistica4.pdf>

resulta particularmente importante, por ejemplo, en los ámbitos relacionados con el suministro de recursos básicos y con la seguridad de las poblaciones.

La puesta en práctica de planes de emergencia ante riesgo de inundaciones (S.E.I.-M.J.I.; 1995), que eviten en lo posible pérdidas de vidas humanas, impone la necesidad de la previsión hidrológica. Por otra parte, la explotación racional de presas y embalses también obliga a operar con la previsión hidrológica como una de las principales fuentes de información. (Aldana et al., 2004)

La necesidad de prever para decidir está presente en todo tipo de actividades y situaciones. Pensemos, por ejemplo, en algo aparentemente tan simple como las decisiones que tiene que tomar el gestor de compras de un restaurante con respecto a la cantidad o el tipo de alimentos que debe adquirir sin saber con exactitud el número de comensales que acudirán al restaurante durante los próximos días. En este ejemplo, planteamos la incertidumbre con respecto al resultado de un acontecimiento que ocurrirá en un momento concreto: en una fecha cercana un restaurante abrirá sus puertas, ofrecerá su carta habitual de platos y debe contar por tanto con los suministros para elaborarlos, pero no se sabe qué cantidades tendrán que servir de cada uno de los platos ofertados.

Un tipo de incertidumbre diferente se produce cuando sabemos que algo concreto puede ocurrir pero no sabemos cuándo. Las personas que viven en zonas que sufren numerosos movimientos sísmicos conocen bien este tipo de incertidumbre. Saben que, por las características tectónicas de la zona, se pueden producir seísmos de una intensidad tal que provoquen elevados daños humanos y materiales. A diferencia del ejemplo anterior en el que la incertidumbre se refiere a fechas concretas, en el caso de los seísmos, lo que no se sabe es cuándo se producirá uno de estos acontecimientos.

Hemos planteado dos ejemplos muy diferentes en cuanto al tipo de incertidumbre y a la temática. Pese a estas diferencias, en ambos casos se trata de fenómenos de tipo aleatorio sobre los cuales es posible realizar predicciones mediante el uso combinado de herramientas estadísticas y de la Teoría de la Probabilidad.

¿Por qué mencionamos ahora la Teoría de la Probabilidad? Porque, como decíamos anteriormente, toda predicción de fenómenos aleatorios contiene un margen de error y es precisamente la Teoría de la Probabilidad la que nos proporciona los fundamentos para determinar la probabilidad de que se produzcan en la realidad las predicciones realizadas.

La posibilidad de hacer predicciones de fenómenos aleatorios y de calcular la probabilidad de equivocarnos es un hecho de gran importancia, en tanto que permite reducir la incertidumbre con respecto a situaciones futuras. En la

página Web del grupo editorial McMillan podemos encontrar un documento titulado *La idea y el empresario*. En dicho documento encontramos un planteamiento que aplica al mundo empresarial la idea que acabamos de exponer:

La actividad del empresario forma una interminable cadena de toma de decisiones sobre situaciones que se proyectan hacia el futuro. Por tanto, el empresario no toma sus decisiones sobre situaciones de certeza; es totalmente imposible saber con exactitud lo que ocurrirá, por lo que asume la posibilidad de equivocarse. Las consecuencias que tiene una equivocación en la toma de decisiones empresariales pueden ser muy graves y su gravedad será proporcional a la importancia de las decisiones tomadas. [...]

Por tanto, el empresario actúa en situaciones de riesgo, o lo que es lo mismo, la actividad empresarial va permanentemente unida al riesgo empresarial. [...]

Situación de incertidumbre. No es posible determinar lo que va a ocurrir en un futuro. El nivel de la incertidumbre tiene que ver con la probabilidad de que ocurra un suceso u otro. No hay riesgo sin incertidumbre, pero sí existe incertidumbre con muy poco riesgo en los casos en que la incertidumbre esté controlada y la probabilidad de que ocurra algo diferente a lo previsto sea muy pequeña¹¹.

Pese a la diversidad de técnicas existentes¹² para la realización de predicciones, muchas de esas técnicas se fundamentan en el estudio de la variabilidad del fenómeno en el pasado como base para la predicción futura. Se utilizan para la predicción las series históricas de datos porque se considera que la evolución futura del hecho que se estudia será similar a la evolución previa ya conocida. En un portal educativo sobre temas de naturaleza al explicar los estudios meteorológicos que se llevan a cabo en la Antártida, utilizan una expresión que refleja con claridad la idea que queremos transmitir:

En previsión meteorológica, para tener futuro se necesita pasado, y las previsiones que se realicen en el futuro se alimentarán de los datos

¹¹ http://www.macmillanprofesional.es/fileadmin/files/online_files/profesional/LOE/EIE02p022-041.PDF

¹² Resulta ilustrativo al respecto el artículo *El tratamiento de la variabilidad en las predicciones sobre el crecimiento económico* (Moreno y López, 2002), porque muestra que ante la diversidad de procedimientos para la realización de previsiones que proporcionan resultados diferentes, se realizan estudios que analizan la propia variabilidad no ya de un fenómeno, sino de sus predicciones.

precedentes. Cada vez, se irá disponiendo de una base de datos reales creciente que reducirá los márgenes de error.¹³

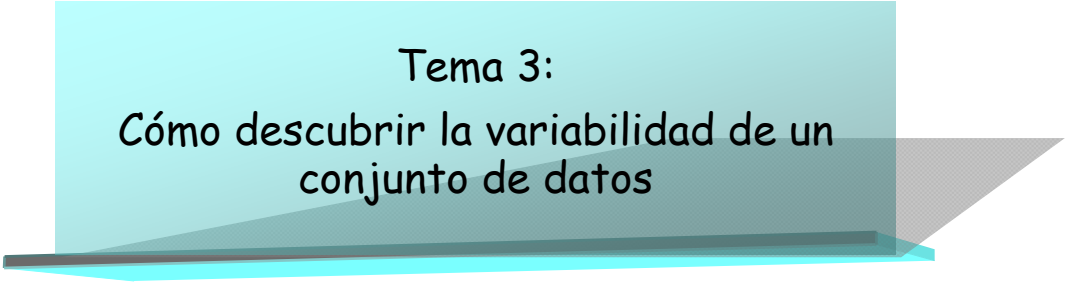
Los fundamentos teóricos que subyacen a la mayoría de los métodos de predicción existentes y la complejidad de los procedimientos para llevarlos a cabo exceden en mucho los objetivos y las posibilidades del presente manual. Pese a ello, hemos considerado necesario hacer esta breve introducción al tema por razones que sí se relacionan con los materiales e ideas que se van a presentar. Los capítulos que siguen están dedicados a explicar los fundamentos del análisis de variabilidad mediante el método estadístico. El objetivo es que quienes decidan trabajar con este material aprendan no sólo a elaborar un análisis básico de la variabilidad de algún aspecto de la realidad. Se trata igualmente de que comprendan la importancia que tienen la variabilidad y su estudio en el mundo actual y que el análisis de las variaciones pasadas o presentes constituyen, en muchas ocasiones, el punto de partida necesario para efectuar predicciones de futuro.

¹³ Antártida. CIENCIAS - METEOROLOGÍA. Tendencia, predicción, estadística.
http://www.natureduca.com/ant_cienc_meteo_tend.php

Tema 3.

Cómo descubrir la variabilidad de un conjunto de datos

Colectivos o poblaciones y características o variables.....	1
Descubriendo la variabilidad: El análisis preliminar de los datos	8
Primera aproximación a la variabilidad: el rango de la variable	9
Detección de posibles errores y anomalías.....	10
La variabilidad interna de los valores. La distribución.....	15
El concepto de distribución de los valores de la variable.....	15
Análisis de la distribución	20
Representación gráfica de la distribución: el histograma.....	27



Tema 3: Cómo descubrir la variabilidad de un conjunto de datos

COLECTIVOS O POBLACIONES Y CARACTERÍSTICAS O VARIABLES

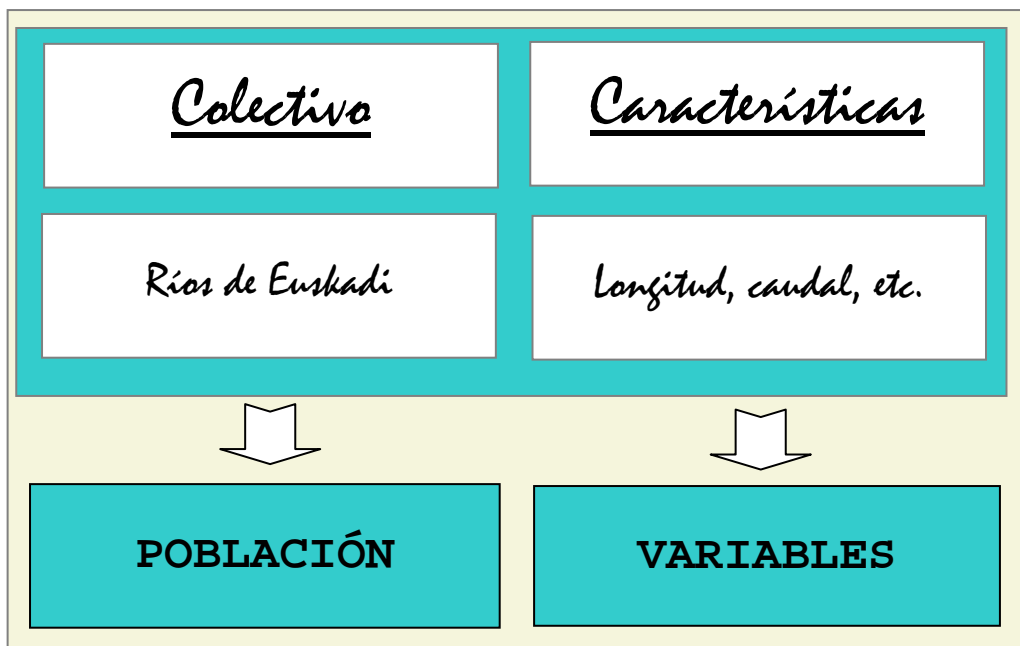
En el tema anterior hemos podido ver que es muy frecuente estudiar la característica de algún colectivo cuyo comportamiento nos interesa conocer. Hablábamos, entre otros ejemplos, de un posible colectivo de ciudades sobre el cual se estudiaba el número de emigrantes que habían fijado su residencia en ellas. Otro colectivo que mencionábamos era el de productos y servicios, del cual nos interesaba estudiar el encarecimiento que habían sufrido tras el cambio de moneda al euro.

Lo que nos interesa ahora es presentar dos conceptos cuya comprensión es esencial en el razonamiento estadístico. No hay que olvidar que en ocasiones resulta difícil entender los procedimientos estadísticos simplemente porque antes no se han entendido bien algunos de los conceptos fundamentales del lenguaje estadístico. Son conceptos sencillos pero, si no se ha asimilado su significado, pueden bloquear la comprensión y el aprendizaje posterior.



Supongamos que nos han encargado un estudio sobre los ríos de Euskadi. El estudio se enmarca en un proyecto de investigación sobre la calidad ambiental de distintos elementos y características del territorio perteneciente a la comunidad autónoma. Se nos ha pedido que hagamos una caracterización de los rasgos que incluya la longitud de los ríos, el caudal medio e indicadores de la calidad ambiental del agua de los ríos y de los bosques de ribera.

El colectivo cuyas características debemos estudiar está integrado en este caso por los ríos que tienen su recorrido dentro del territorio de la comunidad autónoma. Para referirnos a este colectivo utilizaremos en adelante el término correcto que es POBLACIÓN. Una población en Estadística es el conjunto de elementos sobre los que se quiere estudiar y conocer algo. A los integrantes de la población (colectivo de estudio) se les denomina ELEMENTOS DE LA POBLACIÓN.

De cada uno de los ríos estudiaremos algunas de sus características. Estudiaremos su longitud, el caudal máximo, la calidad de sus aguas y el régimen de alimentación. Sabemos que la longitud varía de unos ríos a otros, lo mismo que la cantidad de agua que llevan o el nivel de calidad de sus aguas. A estas características que se estudian para cada uno de los elementos de la población se les conoce, en el lenguaje estadístico, con el nombre de VARIABLES.



Un conjunto de elementos = UNA POBLACIÓN	Características de los elementos = VARIABLES
Los ríos de Euskadi	Longitud
	Caudal máximo
	Calidad del agua en la desembocadura
	Régimen de alimentación

POBLACIÓN Ríos de Euskadi		Característica = VARIABLE LONGITUD de los ríos (km)	
	Nerbioi	43,8	
	Cadagua	48,5	
	Ibaizabal	43,5	
	Deba	57	
	Urola	51	
	Oria	66	
	Bidasoa	70	
	Zadorra	79	

En el ejemplo que vemos en la tabla, en la columna izquierda tenemos los datos correspondientes a la población, que estaría compuesta por todos y cada uno de los ríos de Euskadi. En la columna derecha tenemos los datos correspondientes a los valores de una de las variables que, en este caso, sería la longitud de los ríos.

Veamos ahora algunos detalles de las cuatro características o variables que vamos a estudiar para cada uno de los ríos:

- ▣ La longitud del río, que se expresa generalmente en km.
- ▣ El caudal máximo de un río se refiere al mayor caudal registrado en un punto dado del río, dentro de un período de tiempo determinado. El caudal se mide en litros o en metros cúbicos por segundo.
- ▣ La calidad del agua se suele definir por medio de calificativos que indican niveles diferentes dentro de una escala de valoración. Los calificativos surgen generalmente de una valoración global elaborada a partir de varias medidas de distintos aspectos relacionados con la calidad del agua.¹ En el

¹ Gobierno Vasco (2004-2008): *Red de seguimiento del estado ecológico de los ríos de la CAPV*. Se trata de un estudio exhaustivo desarrollado durante varios años por la empresa Ondotek II para la Agencia Vasca del Agua. Los resultados, divididos en 25 tomos, pueden descargarse en la web de la Agencia Vasca del

estudio sobre el estado ecológico de los ríos de la CAPV utilizaron para la calificación una escala de cinco niveles de calidad: muy bueno, bueno, aceptable, deficiente y malo.

- ▣ El régimen de alimentación hace referencia al origen de las aguas que alimentan al río. El origen del agua puede ser pluvial o nival; puede ser también mixto, con preferencia de uno u otro tipo (agua de lluvia o agua procedente de la fusión de nieve). El régimen se define entonces mediante términos como pluvial nival, pluvio-nival o nivo-pluvial.

El tipo de medida o de datos que corresponde a cada una de las variables es bien diferente. Supongamos que hemos obtenido los datos de las cuatro variables correspondientes a uno de los ríos y que los resultados son los siguientes:

Río	Longitud	Caudal máximo	Calidad del agua en la desembocadura	Régimen de alimentación
Adibide	72 kms	220 m ³ /s	Buena	pluvial

Los datos ficticios de la tabla que hemos elaborado nos sirven para explicar en qué difieren el tipo de datos asociados a cada variable:

- ▣ Las variables longitud y caudal son del mismo tipo en tanto que se definen mediante un valor, un número que expresa la cantidad de kilómetros o la de metros cúbicos por segundo.
- ▣ La variable sobre la calidad del agua no se define mediante un valor sino mediante un calificativo. Ahora bien, aunque no se trata de un valor sí expresa una cuantía, que puede ser mayor o menor. A diferencia de lo que ocurre con el caudal o la longitud, no es una medida precisa. Si la calidad del agua de un río tiene una calificación de muy buena y la de otro río solamente es buena, no podemos saber de forma precisa cuál es la medida de la diferencia entre ambos niveles de calidad. Diremos, en definitiva, que

Agua (Ver bibliografía). Aunque son informes muy voluminosos y también muy técnicos, contienen tablas y gráficos de resumen que permiten ver qué aspectos se han tenido en cuenta y se han medido para otorgar una calificación de calidad a las aguas del río.

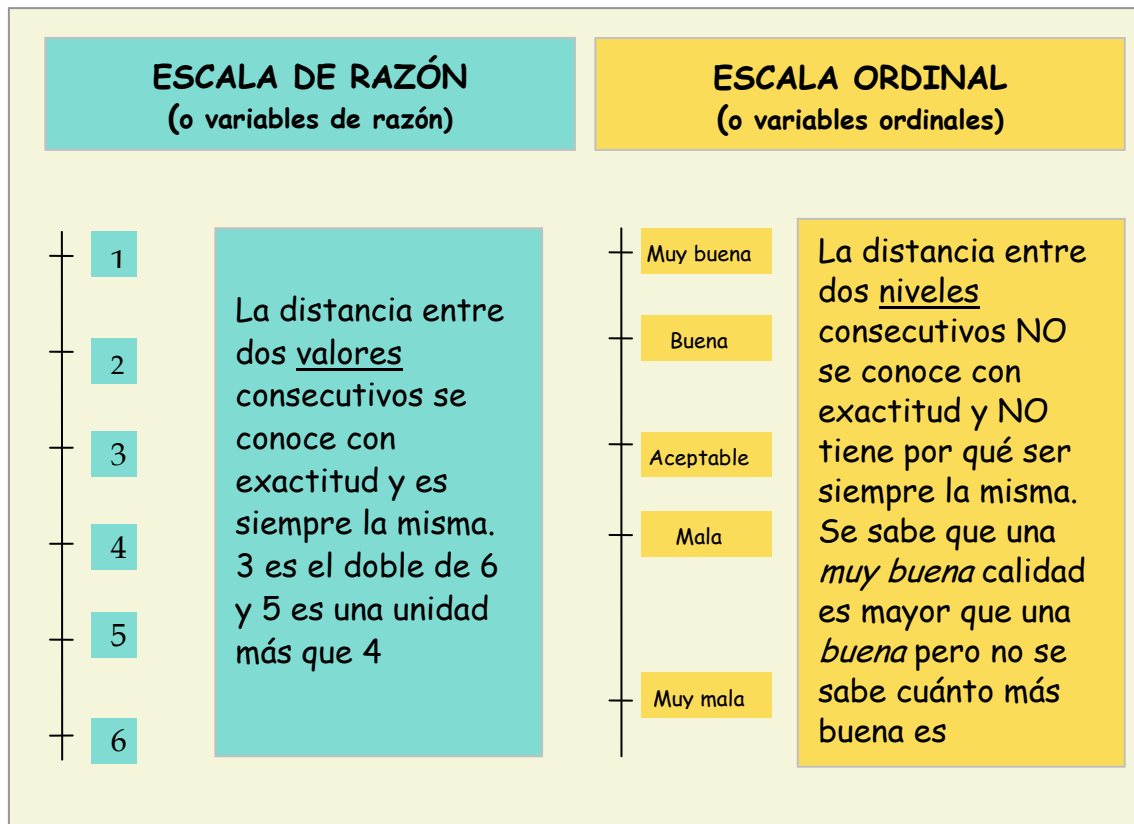
no se puede operar matemáticamente con las mediciones de este tipo.

- Por último, la variable sobre el régimen de alimentación es una característica más de los ríos, que se define mediante un término o una expresión que no alude a cantidad ninguna. El hecho de que la alimentación de un río sea pluvial y la de otro sea nival significa únicamente que son diferentes. El régimen de alimentación de los ríos es una característica variable, es decir, hay variaciones de un río a otro, pero no son de cantidad.

El hecho de que la variable o característica de los elementos de una población se defina de una forma u otra, es decir, mediante un número alusivo a la cantidad o una palabra o expresión, repercute en el tipo de procedimientos que se pueden utilizar para estudiarla. Por este motivo, antes de abordar cualquier tipo de análisis, es importante conocer los distintos tipos de variables y aprender a diferenciarlas entre sí:

Variable	Río Adibide	Río Ibai	TIPO DE VARIABLE O ESCALAS DE MEDICIÓN
Longitud	72 km	48 km	DE RAZÓN: el valor de la variable se establece mediante medidas de precisión estándar, al igual que la diferencia entre valores.
Caudal	220 m ³ /s	110 m ³ /s	
Calidad del agua	Buena	Mala	ORDINAL: las medidas de la variable son niveles dentro de una escala en la que no se conoce la posición exacta de cada nivel
Régimen de alimentación	Pluvial	Nival	NOMINAL: las medidas de la variable no implican cantidad. En lugar de hablar de valores de la variable se habla de modalidades de la variable.

Se verá todavía más claro si comparamos los tipos de variable o escalas de medición ordinal y de razón mediante una ilustración de la escala:



Como decíamos más arriba, el hecho de que las variables sean de un tipo u otro implica que el tratamiento que se puede hacer con ellas es también distinto. Cuando la escala de medición de la variable es *de razón*, se puede hacer todo tipo de operaciones matemáticas con los valores. Se pueden sumar, restar, multiplicar, dividir, etc. Si la escala de medición es *ordinal* los distintos niveles sólo se pueden ordenar de mayor a menor o al revés. Por último, cuando la variable es de tipo *nominal*, es decir, se define mediante una característica o atributo, ni siquiera se pueden ordenar; se habla entonces, no de valores ni de niveles, sino de modalidades de la variable.

Aunque es cierto que el tipo de manipulación y de estudio que se puede hacer con variables es distinto según el tipo o la escala de medición, también lo es el hecho de que la Estadística ofrece posibilidades para el análisis de todas ellas. Lo importante en este caso es saber los procedimientos de análisis son diferentes en función del tipo de variables. A este respecto, hay que aclarar, por último, que **en los materiales que aquí presentamos nos ocuparemos exclusivamente de los procedimientos para tratar con variables que pueden ser medidas en escala de razón.**

Volviendo al tema de la población objeto de estudio en nuestro ejemplo, todavía queda un interrogante:

- ▣ Cuáles son realmente los ríos de Euskadi que debemos analizar. ¿Tenemos que tener en cuenta todos los arroyos y riachuelos o sólo los ríos principales? Y si sólo debemos encargarnos de estos últimos, ¿cómo sabremos cuáles son los ríos principales?. La duda puede surgir, por ejemplo, porque existen algunos riachuelos de pequeña entidad que nacen cerca de la costa y tienen recorridos muy cortos. Nos preguntamos si estos también debemos incluirlos en el estudio.

Sea cual sea la respuesta a nuestra pregunta, lo que queremos destacar aquí es que lo importante en cualquier estudio es disponer de una buena definición de la población que se va a analizar. Definir la población significa fijar o concretar con claridad y precisión las características que debe tener un elemento para ser considerado parte de la población o, en caso contrario, para excluirlo.

La definición nos debe permitir resolver cualquier duda a la hora de decidir si un elemento formará o no parte de la población. En el caso de los ríos, en el supuesto de que no hubiera que estudiarlos todos, se podría utilizar, por ejemplo, un límite mínimo de longitud por debajo del cual el río no formara parte de la población.

DESCUBRIENDO LA VARIABILIDAD: EL ANÁLISIS PRELIMINAR DE LOS DATOS

Una vez que hemos definido de forma correcta y sin ambigüedad la población que estudiaremos, de modo que hemos podido identificar, sin ninguna duda, todos los integrantes de dicha población y una vez, también, que hemos recopilado los datos pertenecientes a la variable la siguiente tarea consiste en **descubrir** la variabilidad de los datos. Para ello se realiza lo que se denomina un análisis preliminar de los datos.

El análisis preliminar de los datos consiste en ordenarlos y elaborar una representación gráfica de los mismos. El objetivo es detectar y corregir posibles errores y anomalías en los datos y obtener el máximo posible de información sobre su variabilidad, es decir, sobre las diferencias que hay entre los elementos de la población en relación a la característica o variable que queremos estudiar. Si la población fueran los ríos de Euskadi y la variable su longitud, analizaríamos en qué medida son diferentes las longitudes de dichos ríos. En esta primera fase de trabajo no se utilizan fórmulas, no se manipulan matemáticamente los datos; simplemente se observan siguiendo una serie de criterios.

El primer paso en cualquier trabajo de análisis es ordenar los datos en función de los valores de la variable. Podemos ordenarlos en orden creciente, del valor más pequeño al más grande, o decreciente. A continuación localizamos el valor más alto y el más bajo y calculamos la diferencia. A esta diferencia entre los valores extremos de la variable se le denomina **rango**. El cálculo del rango es una tarea sencilla (sobre todo si la realizamos mediante un programa de ordenador) pero muy provechoso porque nos proporciona una primera idea sobre la variabilidad de los datos que estamos analizando.

A continuación debemos hacer una revisión cuidadosa de los datos con el fin de detectar y corregir posibles errores y anomalías.

No debemos olvidar que todas las conclusiones que podamos obtener sobre la población que estamos estudiando tienen su punto de origen en los datos que hemos recogido. Entenderemos así que es preciso tener un cuidado exquisito a la hora de comprobar los posibles errores en los datos que puedan comprometer la validez de los resultados.

Primera aproximación a la variabilidad: el rango de la variable

Tal como hemos indicado, la primera operación que realizamos para tener una idea del grado de variabilidad es calcular la diferencia entre el valor más alto y el más bajo de nuestra serie de datos, es decir, el rango. Para calcular el rango utilizaremos, en este caso, los datos sobre la renta personal disponible en los municipios de Euskadi, correspondientes al año 2001²

Explicado de forma resumida, podemos decir que la renta personal disponible es la cantidad de dinero anual de la que una persona dispone para gastar o ahorrar. Esta cantidad disponible surge de restar a la cantidad de dinero que una persona obtiene mediante todo tipo de ingresos la cantidad que destina al pago de impuestos (directos) y de las cuotas obligatorias de la Seguridad Social³.

En nuestro ejemplo lo que nos interesa es conocer en qué medida varía la renta de un municipio a otro. Nos interesa, por ejemplo, conocer si la renta de los habitantes de los distintos municipios es similar o, si por el contrario, existen municipios con una renta sensiblemente inferior o superior a la de otros.

Lanestosa	7.192
Valle de Carranza	7.275
Lapuebla de Labarca	7.800
Trucios	7.813
Kripan	8.131
.....	
Zigoitia	14.294
Gordexola	14.335
Sukarrieta	15.016
Leintz-Gatzaga	15.346
Laukiz	20.627

La tabla muestra los municipios que encabezan la lista con los valores de renta media personal disponible y los municipios que están al final de la lista, con los valores más bajos.

² Ver la serie completa de datos en la tabla nº 1 del Anexo.

³ Una visión sencilla sobre la importancia de los estudios sobre la renta personal en Larrañaga Sarriegi, M. (2006) Distribución de la renta..

Observamos que son los municipios de Lanestosa con 7.192 euros anuales de renta per cápita y Laukiz, con 20.627, los que presentan los valores máximo y mínimo. Hacemos la resta y vemos que el rango de la variable renta personal per capita de los municipios de Euskadi en 2001 fue de 13.435 euros

Tenemos ya una primera idea sobre la variabilidad que estamos tratando de analizar. En este caso, sólo con el dato del rango, podemos afirmar que en Euskadi, las diferencias de renta entre los municipios son muy grandes, o lo que es lo mismo, que la variabilidad entre municipios es muy elevada: Laukiz, el municipio con mayor renta de Euskadi, posee una renta anual casi tres veces superior a la de Lanestosa.

Detección de posibles errores y anomalías

Cuando las diferencias entre los valores máximo y mínimo de la variable son elevadas, como en el caso de la renta personal disponible de los municipios de Euskadi en 2001, lo primero que hay que preguntarse, y comprobar, es si la diferencia observada es real o si podría tratarse de un error de los datos. Nos preguntamos en este caso si es real que los habitantes de Laukiz hayan podido disponer, de media, en 2001 de una cantidad de dinero tan superior a la del resto de habitantes y municipios.

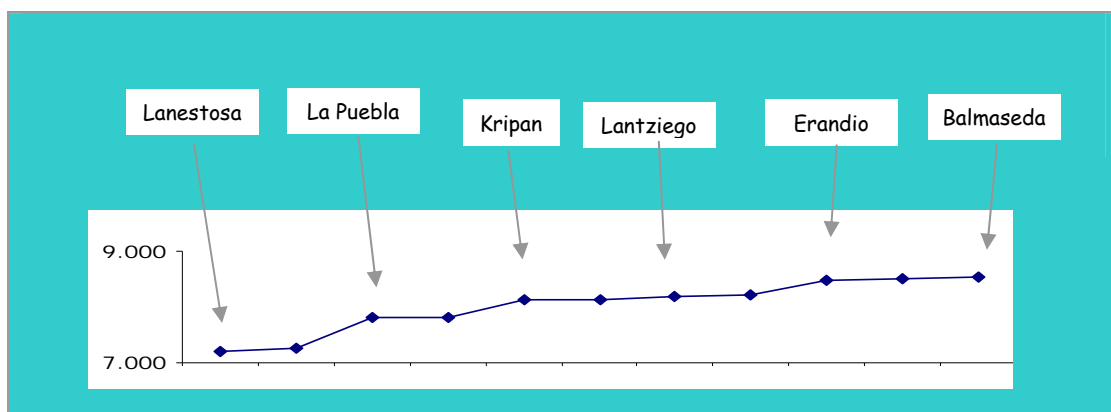
Para una persona recién iniciada en el análisis de datos la duda será, quizás, qué criterio utilizar para decidir que una serie de datos es sospechosa de contener errores. La tarea requiere algo de práctica pero es más sencilla de lo que puede parecer. Veremos cómo proceder con los datos de renta.

Si uno de los dos valores extremos de la serie fuera erróneo, es decir, si fuera anormalmente alto o bajo a consecuencia de un error producido en la fases previas de recogida de datos o de informatización, sería muy diferente del resto de los datos. Sería un valor discordante.

Para comprobar si se trata de valores raros, es decir, si los valores de Lanestosa y Laukiz son una excepción o no, lo que hacemos es mirar en el listado de datos si hay otros municipios que tienen valores de renta similares a los de dichos municipios:

Municipios con los valores de renta más bajos	
Lanestosa	7.192
Valle de Carranza	7.275
Lapuebla de Labarca	7.800
Trucios	7.813
Kripan	8.131
Ekora	8.138
Lantziego	8.185
Valle de Arana	8.218
Erandio	8.492
Elantxobe	8.507
Balmaseda	8.547

En la tabla vemos los 10 municipios que siguen en la lista a Lanestosa y que tienen, por tanto, rentas progresivamente superiores a la suya. Hay, por ejemplo, otros tres municipios más cuya renta no llega a los 8.000 euros. Vemos también que los aumentos de renta de cada municipio al siguiente forman una progresión que asciende suavemente. El gráfico siguiente proporciona una imagen visual de la progresión:



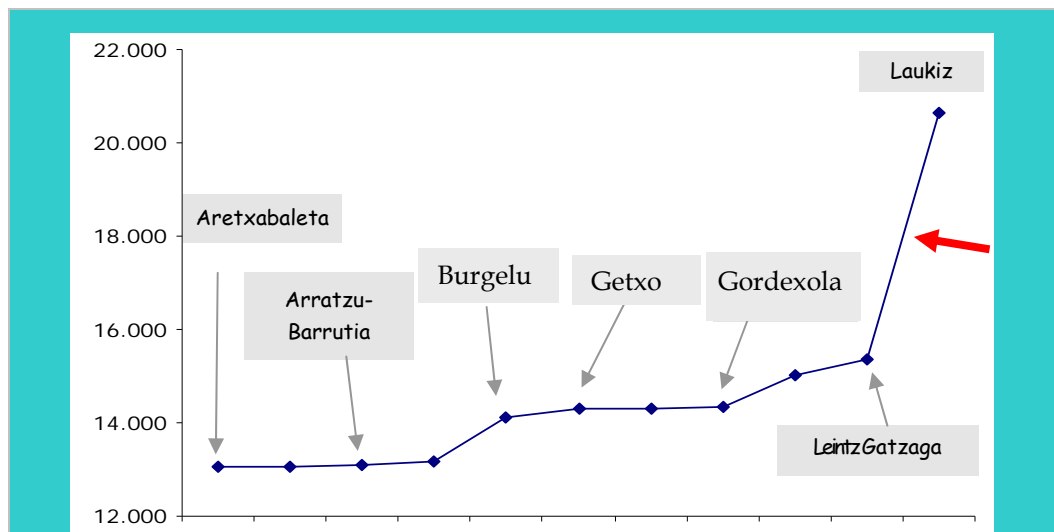
La pendiente suave de la gráfica muestra que las diferencias de valores de un municipio a otro son pequeñas, y, sobre todo, que conforman una progresión muy regular (las diferencias entre municipios son bastante similares)

A continuación haremos lo mismo con los once municipios cuyas rentas son las mayores de la comunidad autónoma:

Municipios con los valores de renta más altos	
Aretxabaleta	13.044
Legutio	13.069
Arratzu-Ubarrundia	13.100
Lezama	13.157
Burgelu	14.105
Getxo	14.291
Zigoitia	14.294
Gordexola	14.335
Sukarrieta	15.016
Leintz Gatzaga	15.346
Laukiz	20.627

La diferencia entre la renta de Leintz Gatzaga y Laukiz es de 5.281 €

En este caso, el aumento en el valor de la renta de un municipio a otro es también progresivo hasta que llegamos a los dos últimos municipios, entre los que existe una diferencia muy importante en sus valores de renta: 5.281 euros separan las rentas de Leintz-Gatzaga y Laukiz.



En esta ocasión el gráfico refleja claramente el salto que se produce entre los valores de renta de Leintz-Gatzaga y Laukiz.

Ya estamos en situación de afirmar que el valor de renta de Laukiz es un elemento raro, ya que no existe otro municipio que tenga un valor similar. El valor de Lanestosa, sin embargo, no es un elemento raro ya que existen otros municipios con valores cercanos.

El hecho de que los valores raros que encontramos a veces en los datos resulten sospechosos no significa que necesariamente tengan que ser errores. A veces ocurre, sin más, que determinados elementos de la población que analizamos tienen valores mucho más altos o bajos que el resto de los elementos. Sea como fuere, lo que debemos hacer es intentar comprobar la veracidad de los datos que nos resulten dudosos. En el ejemplo que estamos analizando se trata de intentar averiguar si el dato de Laukiz se corresponde con una situación real, es decir, si verdaderamente este municipio disfrutó en 2001 de unas rentas muy superiores a las del resto o si, por el contrario, se trata de un dato erróneo. Para intentar resolver la duda no nos queda otro remedio que revisar todo el proceso de recogida y manipulación de los datos, hasta llegar a la fuente original de los mismos.

Los datos de renta que estamos utilizando los hemos obtenido directamente de la página Web del Instituto Vasco de Estadística (EUSTAT). Una segunda inspección de los listados de datos que ofrece la entidad en Internet nos ha permitido comprobar que no se ha producido ningún error y que, por tanto, el dato que tenemos de Laukiz, hasta donde nosotros podemos comprobar, es correcto.

Una vez comprobado que no se ha producido ningún error en el proceso de recogida y manipulación de los datos todavía puede ocurrir que, por el conocimiento que tenemos del tema que estamos estudiando o por alguna otra razón, sigamos pensando que se trata de un valor erróneo que se ha “colado” en los listados originales de la fuente de datos. Ante una situación similar tenemos dos opciones:

- ▣ Eliminar el dato que seguimos considerando sospechoso y realizar todos los análisis posteriores sin dicho dato, o
- ▣ Realizar un doble análisis de los datos incluyendo el supuesto dato erróneo en uno de los análisis y excluyéndolo en el otro.

En cualquier caso, sea cual fuere la decisión que tomemos, todo lo ocurrido debe quedar reflejado en el informe final del estudio.

En ocasiones, todavía es posible encontrar recursos alternativos para seguir indagando. Cuando el tema que analizamos es de los que suscitan algún tipo de interés en la sociedad, puede ocurrir que, de haberse producido alguna situación particularmente llamativa, haya sido recogida por la prensa. Esto es lo que ocurrió con la renta de Laukiz, de cuya situación particular se hizo eco la prensa de Álava en 2007:

PUEBLO RICO, PUEBLO POBRE

Vacas gordas, vacas flacas

A Laukiz y al Valle de Carranza les separan 15.000 euros de renta per cápita. A su pesar, son los municipios más rico y más pobre de Vizcaya. «Ni tanto ni tan poco», matizan sus lugareños. (...)

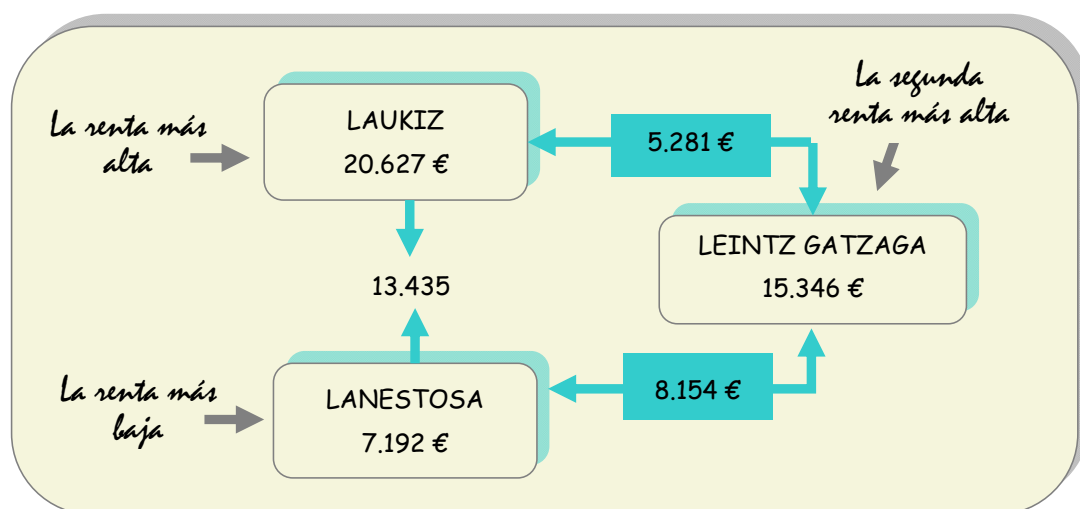
Los vecinos de Laukiz no es la primera vez que escuchan la misma canción. Jóvenes o mayores, la primera reacción se reduce a una carcajada. «Ya estamos. Pregúntale a los de allí arriba, que son los que elevan nuestro nivel», sostiene Íñigo Barrena. No existe una distinción terminológica, pero Laukiz, salvando las distancias, es como uno de esos municipios de anuncio televisivo que compiten en fiestas populares: los de arriba y los de abajo. Pues en las alturas, residen los propietarios de una de las urbanizaciones más lujosas de Vizcaya, Unbe-Mendi, integrada por 127 chalés. Por eso, no es extraño toparse con turismos de marcas prohibitivas kilómetros antes de llegar a sus inmediaciones. «

El correo, 12 de mayo de 2007

Es cierto que la noticia de prensa es de 2007 y nuestros datos son de 2001. No se trata de que esta noticia avale el dato de 2001; sirve solamente para confirmar que, en algunas ocasiones, la renta de Laukiz destaca por su diferencia con la del resto de municipios y para orientarnos sobre las razones de la misma.

LA VARIABILIDAD INTERNA DE LOS VALORES. LA DISTRIBUCIÓN

Para ahora ya tenemos claro que las diferencias de renta entre los municipios de la C.A. no son tan exageradas como parecía en principio. Es cierto que la diferencia de 13.435 euros entre los municipios con mayor y menor renta (Laukiz y Lanestosa) es real, pero también es cierto que hay un único municipio con un valor de renta tan elevado. Si comparamos, por ejemplo, la renta de Lanestosa con la del segundo municipio "más rico", la diferencia se reduce a 8.154 euros.



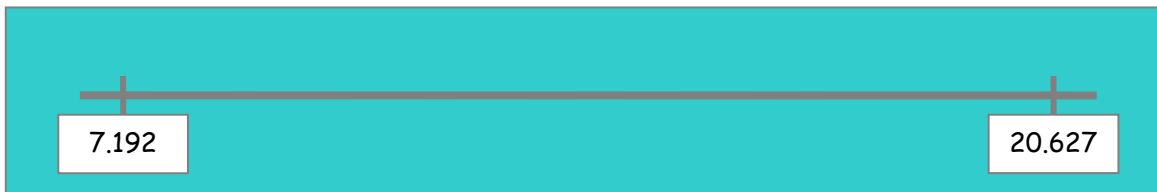
Tal como decíamos anteriormente, el rango nos proporciona una primera idea de las diferencias que presentan los valores máximo y mínimo de la variable, pero no nos dice nada sobre el resto de los valores. En el caso de la renta nos preguntamos, por ejemplo, si son muchos los municipios que tienen rentas próximas a las más bajas o si son mayoría los que tienen valores próximos a los de Leintz-Gatzaga. Podría ocurrir también que una mayoría de los municipios tuviera valores de renta a mitad de camino entre los de Laukiz y Leintz-Gatzaga...

El concepto de distribución de los valores de la variable

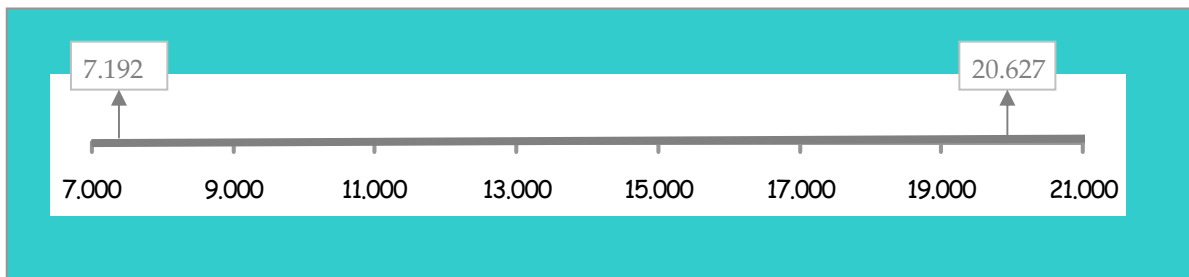
Para visualizar la similitud o diferencia de los datos entre sí, pero de toda la serie, utilizaremos un gráfico que, tras un vistazo, nos permitirá responder a preguntas como las que hemos formulado más arriba.

Iniciaremos el gráfico trazando un segmento de recta mediante el cual representaremos el rango de la variable, es decir, la distancia entre el valor más

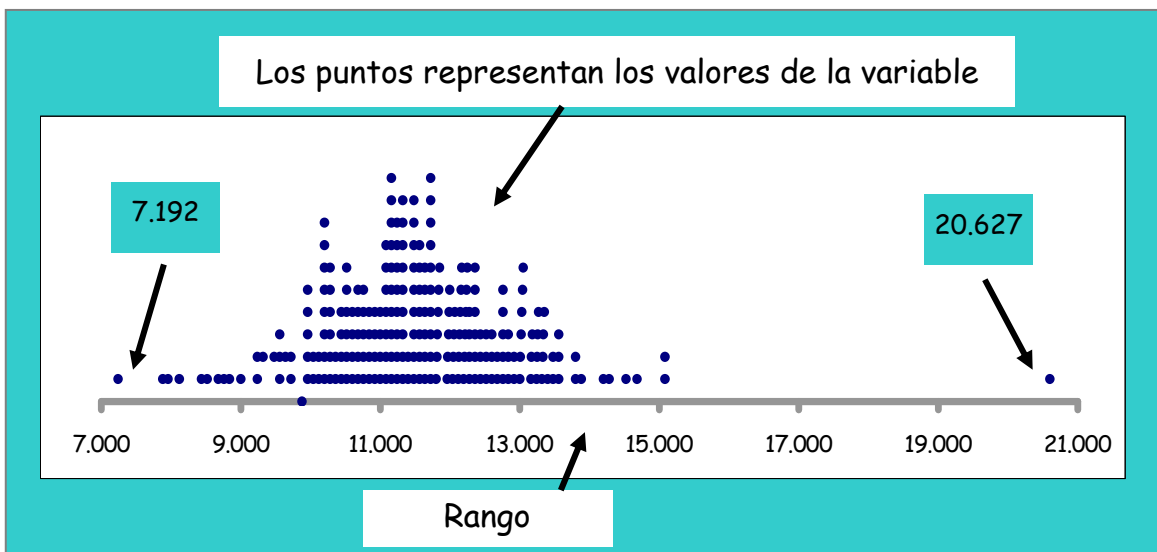
alto y el más bajo. En puntos próximos a los extremos del segmento anotamos los valores extremos del rango.



A continuación dividimos el segmento en partes iguales que representan tramos de incremento del valor de renta. Tenemos una escala graduada:



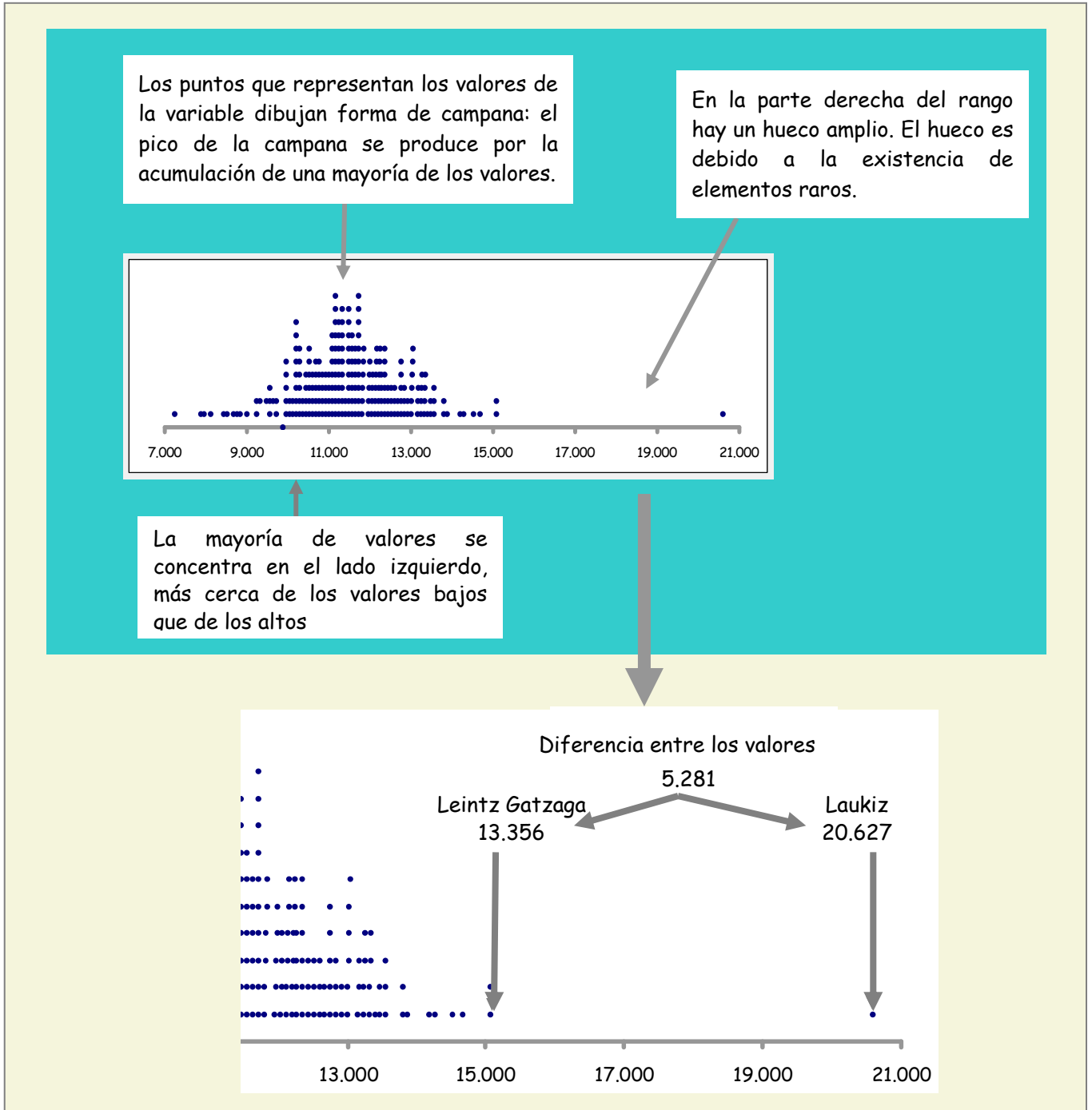
Sobre la recta podemos "colocar" ahora los distintos valores de la variable. Cada dato de la serie lo representamos mediante un pequeño símbolo situado en la posición que ocuparía por su valor. Cuando el espacio en el que debemos situar un punto está ya ocupado, colocamos el nuevo punto encima del anterior.



El resultado final es un gráfico que muestra la distribución de los valores de la variable en el interior del rango. Por ello, para referirnos a la información que nos proporciona el gráfico, se utiliza habitualmente la expresión ***distribución de los valores de la variable***.

El gráfico es una simplificación, ya que hemos sustituido los valores por puntos situados en una escala. La variabilidad la percibimos ahora de forma visual, fijándonos en la posición de los valores dentro del rango y en la distancia que hay entre ellos. Aunque sólo nos proporciona una idea aproximada, su ventaja, con respecto a la serie original de datos, consiste en la facilidad con la que podemos visualizar la distribución y las distancias que hay entre los valores, sin necesidad de hacer ningún cálculo. Un vistazo rápido es suficiente para percibir la variabilidad, ya que la mayor o menor proximidad de los valores en la recta es el indicativo de la misma.

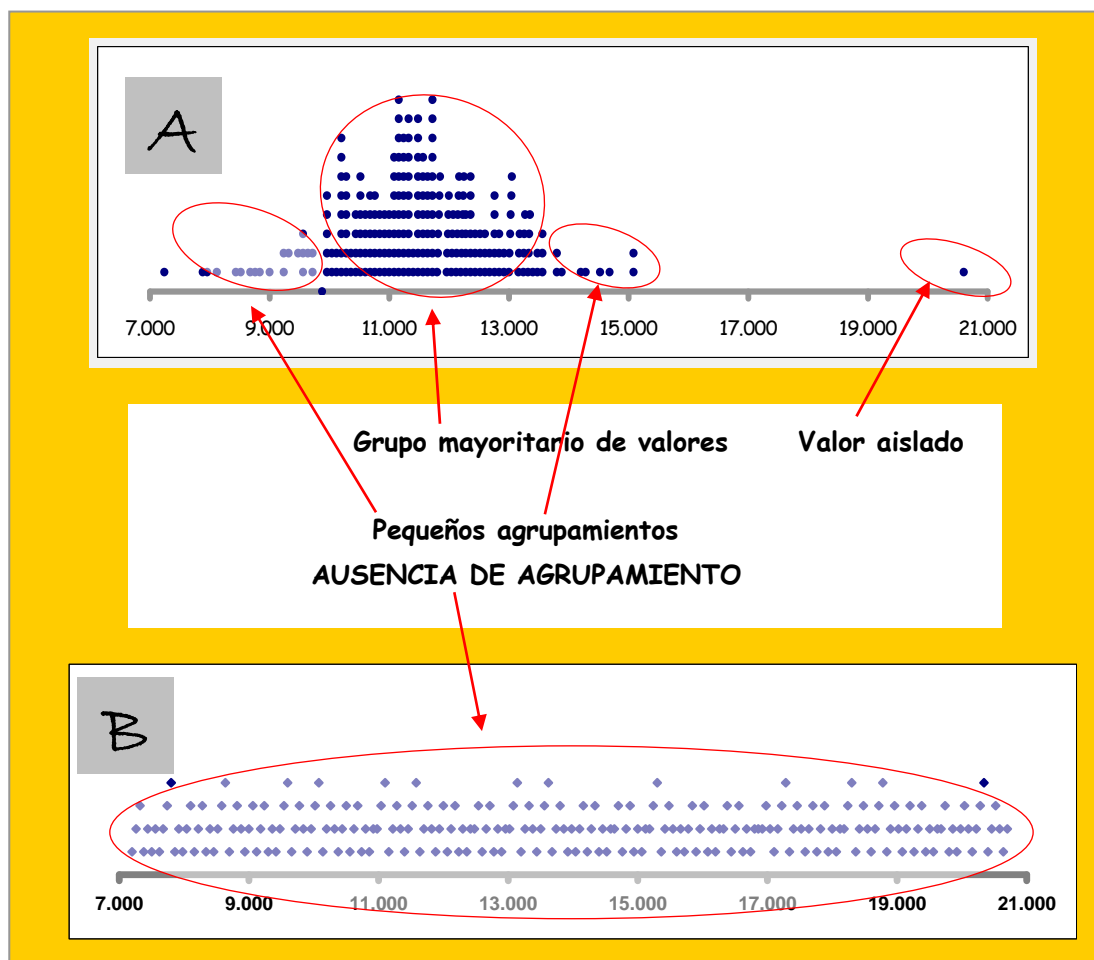
En este punto, el gráfico resulta una herramienta adecuada para preguntarnos sobre la información que se puede obtener a través de la observación de la distribución de valores.



Cuando analizamos una serie de datos, una de las características de ésta que más nos interesa es el modo en que se distribuyen los valores dentro del rango, ya que dicha distribución es un indicador directo de la variabilidad. **Cuanto mayor es el agrupamiento de los valores, menor es la variabilidad y viceversa.**

En la ilustración de la página anterior, elaborada con los datos de renta disponible de 2001, el gráfico nos muestra que existe un agrupamiento notable de los valores. Aunque hay algún valor raro, muy alejado del resto, la mayoría de los valores se encuentran agrupados en un sector del rango

Ahora veremos la comparación entre el gráfico de renta disponible **(A)** y otro que representa una situación completamente diferente **(B)**:



En este último gráfico **(B)**, los valores no se agrupan en un sector del intervalo, sino que se distribuyen de forma regular a lo largo del mismo. Si los puntos del gráfico se correspondieran con valores de renta, podríamos decir que la

variabilidad entre los municipios es total; que no existe grupo alguno de municipio cuyos valores de renta sean más parecidos entre sí que los del resto.

Análisis de la distribución

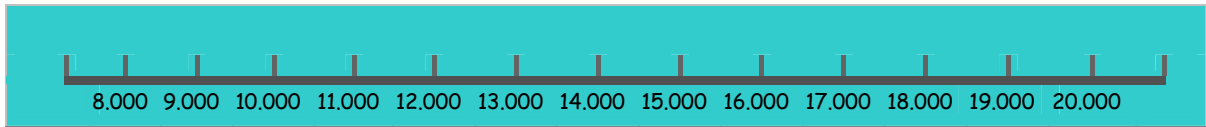
El gráfico que hemos utilizado en el apartado anterior para representar el rango y la distribución de valores resulta muy útil para comprender los conceptos implicados, pero no es el tipo de gráficos que se utilizan habitualmente para analizar o describir la variabilidad. Se trata únicamente de una herramienta didáctica que no resulta válida para el análisis de la distribución porque:

- ▣ No resultaría ni factible ni práctico intentar representar mediante puntitos todos los valores de la variable sobre la recta del rango.
- ▣ Aunque nos proporciona una imagen de la distribución, le falta la precisión necesaria para sustentar una descripción. Por ejemplo, aunque en el gráfico veíamos que la densidad de puntos aumentaba de forma progresiva hacia la zona central del rango no podríamos precisar en qué punto del rango empieza a aumentar la densidad de valores.
- ▣ No cumple con un requisito que resulta esencial para avanzar en el análisis de los datos: reducir el número inicial de datos de forma que sea más manejable.

Puesto que el gráfico que hemos utilizado hasta ahora sólo tiene un valor didáctico, presentaremos a continuación el tipo de gráficos y procedimientos habituales para el análisis de la distribución de valores.

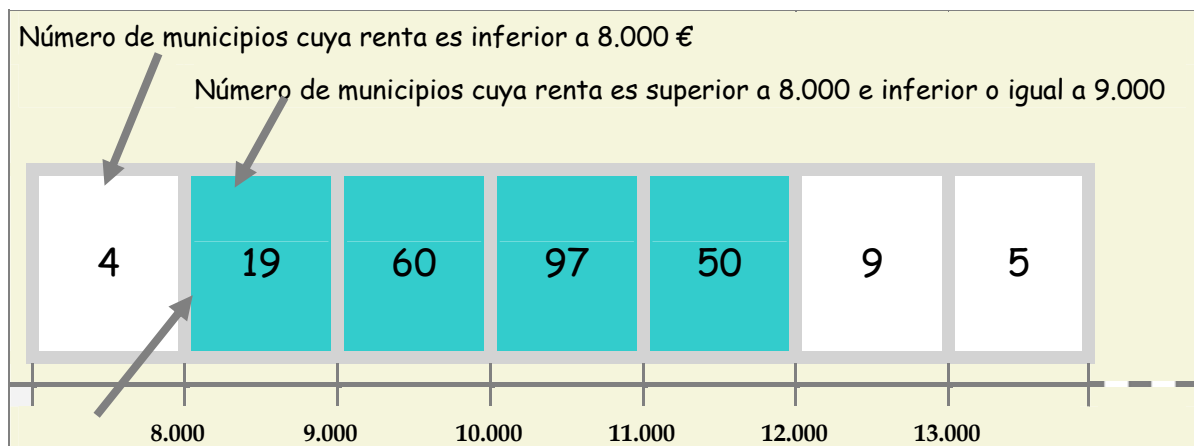
Cuando analizamos la distribución interna de los valores de una variable nuestro objetivo es conocer y caracterizar el comportamiento de los datos como colectivo. Queremos saber si existen agrupamientos de datos y, si los hay, en que sectores del rango se encuentran. Para obtener este tipo de información lo mejor es analizar el rango por sectores y ver cuántos valores de la variable hay en cada sector. Para comprender el procedimiento haremos este análisis con los datos de la renta personal disponible de los municipios de Euskadi en 2001.

Empezaremos dividiendo el rango de la variable en sectores:



A los sectores en que dividimos el rango se les denomina intervalos o clases

A continuación contaremos el número de valores que *caen* dentro de cada uno de los sectores (clases):



Ahora podemos precisar que en el punto del rango correspondiente a 8.000 € se produce un cambio importante en el agrupamiento (frecuencia) de valores: sólo son 5 los municipios con rentas inferiores a 8.000 y sube hasta 19 el número de municipios con rentas entre 9.000 y 10.000 €. El agrupamiento cambia también bruscamente a partir de los 12.000 €

Cuando agrupamos los valores de la variable en segmentos del rango -en intervalos- obtenemos los datos para elaborar una tabla de frecuencias de la variable. Es una tabla de dos columnas: en la columna de la izquierda indicamos los segmentos del rango o intervalos y en la de la derecha el número de valores de la variable que caen dentro de cada intervalo (frecuencia).

Los intervalos o clases están formados por dos valores, que son sus límites. Al valor más pequeño se le denomina límite inferior y al valor mayor se le denomina límite superior.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS	
Intervalo	Frecuencia
≤ 8.000	5
$> 8.000 \leq 9.000$	19
$> 9.000 \leq 10.000$	60
$> 10.000 \leq 11.000$	97
$> 11.000 \leq 12.000$	50
$> 12.000 \leq 13.000$	9
$> 13.000 \leq 14.000$	5
$> 14.000 \leq 15.000$	4
$> 15.000 \leq 16.000$	1
$> 16.000 \leq 17.000$	0
$> 17.000 \leq 18.000$	0
$> 18.000 \leq 19.000$	0
$> 19.000 \leq 20.000$	0
> 20.000	1

INTERVALO ABIERTO
Solo se establece uno de los dos límites, en este caso el superior. Pertenecen al intervalo todo los elementos con un valor inferior a 8.000.

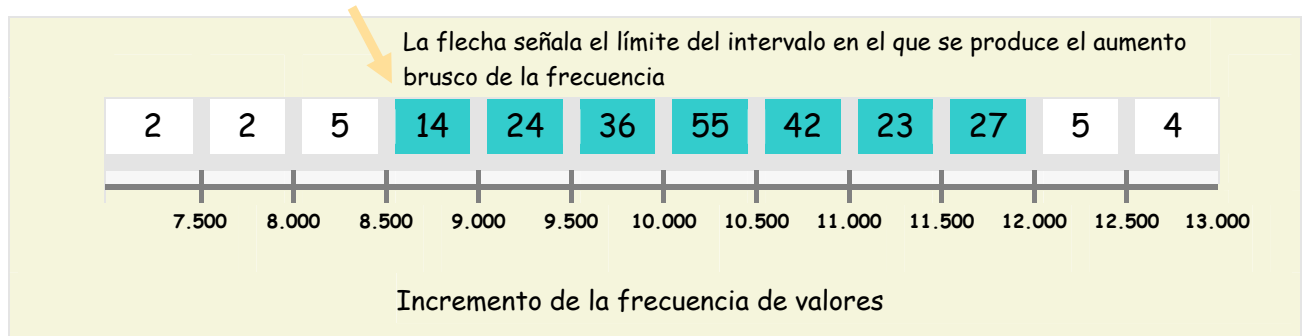
INTERVALO CERRADO
Límites del intervalo

Límite inferior > 17.000	Límite superior ≤ 18.000
-------------------------------	----------------------------------

Hemos marcado mediante flechas rojas el primer y el último intervalo. En ambos casos podemos ver que sólo se indica uno de los dos límites. En estos casos, al faltar uno de los límites, se dice que el intervalo es abierto. El significado de esto parece evidente:

≤ 8.000	5	→	Pertenecen a este intervalo o clase todos los elementos cuyo valor inferior a 8.000 €.
LA AMPLITUD NO ESTÁ DEFINIDA			
> 20.000	1	→	Pertenecen a este intervalo o clase todos los elementos cuyo valor sea superior a 20.000€.

En la tabla superior podemos ver la distribución de frecuencias resultante de dividir el rango en intervalos de 1.000 €. Pero podríamos haberlo dividido el rango en intervalos o fragmentos más pequeños, como se muestra en el gráfico siguiente.



En esta ocasión hemos dividido el rango en intervalos de 500 €. La distribución que obtenemos es distinta de la anterior, no sólo porque los datos estén más desglosados, sino porque el punto del rango en el que la frecuencia de valores aumenta bruscamente es en el paso del intervalo $>8.000 - \leq 8.500$ al intervalo $>8.500 - \leq 9.000$.

Intervalo	Frecuencia
<7.500	2
$\geq 7.500 \leq 8.000$	2
$\geq 8.000 \leq 8.500$	5
$\geq 8.500 \leq 9.000$	14
$\geq 9.000 \leq 9.500$	24
$\geq 9.500 \leq 10.000$	36
$\geq 10.000 \leq 10.500$	55
$\geq 10.500 \leq 11.000$	42
$\geq 11.000 \leq 11.500$	23
$\geq 11.500 \leq 12.000$	27
$\geq 12.000 \leq 12.500$	5
$\geq 12.500 \leq 13.000$	4
$\geq 13.000 \leq 13.500$	5
$\geq 13.500 \leq 14.000$	0
$\geq 14.000 \leq 14.500$	4
$\geq 14.500 \leq 15.000$	0
$\geq 15.000 \leq 15.500$	2
$\geq 15.500 \leq 16.000$	0
$\geq 16.000 \leq 16.500$	0
$\geq 16.500 \leq 17.000$	0
$\geq 17.000 \leq 17.500$	0
$\geq 17.500 \leq 18.000$	0
> 18.000	1

A la hora de elaborar tablas de distribución de frecuencias de los valores de la variable tenemos que tener claros unos cuantos principios fundamentales:

- Las tablas de distribución de frecuencias constituyen una simplificación con respecto a la serie de datos inicial. Como hemos visto en el ejemplo de la renta personal disponible, de una larga lista de valores, de la que es muy difícil extraer conclusiones, pasamos a una tabla de dimensiones mucho más reducidas. La ventaja de la tabla de distribución de frecuencias es que nos permite, de un simple vistazo, identificar los intervalos que acumulan un mayor número de valores y los que contienen pocos o, incluso, están vacíos. La conclusión parece evidente: **a cambio de la ventaja que nos proporciona la observación de los datos agrupados en intervalos perdemos información; una vez agrupados los datos ya no vemos los valores individuales ni a que elemento de la población pertenece cada uno de ellos.**

Serie original de datos (fragmento)		Tabla de distribución de frecuencias	
Elementos de la población	Valor	Intervalo	Frecuencia
Lanestosa	7.192	< 7.500	2
Karrantza Harana	7.275		
Lapuebla de Labarca	7.800	$\geq 7.500 \leq 8.000$	2
Turtzioz	7.813		
Kripan	8.131	$\geq 8.000 \leq 8.500$	5
Ekora	8.138		
Lantziego	8.185		
Harana	8.218		
Erandio	8.492		

Al pasar de la serie inicial de datos a la tabla de distribución de frecuencias, las nueve filas de la serie inicial se han convertido en tres

PÉRDIDA DE INFORMACIÓN: sabemos que hay dos municipios con rentas inferiores a 7.500 €, pero no sabemos ni cuáles son ni sus valores concretos de renta.

- ▣ El tamaño o amplitud de los intervalos lo elegimos nosotros. Los intervalos o clases puede ser más grandes o más pequeños, como en los dos ejemplos que hemos mostrado. La cuestión es entonces cuántos grupos, clases o intervalos hacer. Por el momento no debe preocuparnos. **Para decidir la amplitud de los intervalos lo más conveniente es hacer diferentes pruebas hasta encontrar una clasificación que muestre de la forma más adecuada posible la distribución de los valores de la variable en el rango.**

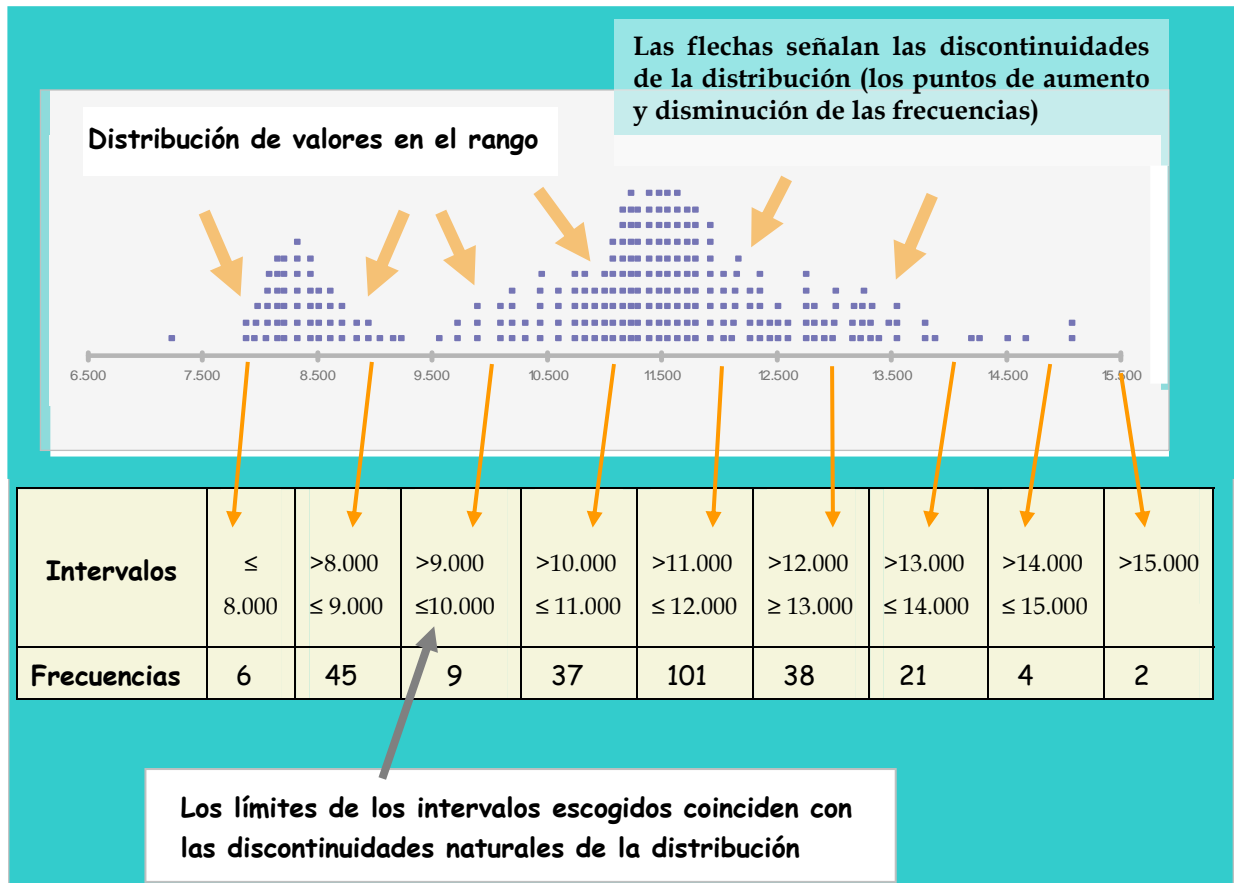
Una clasificación de los valores de la variable en intervalos será adecuada cuando refleje las discontinuidades naturales de la distribución de valores, es decir, cuando los límites de los intervalos coincidan con puntos del rango en los que la frecuencia aumenta o disminuye significativamente. En función de este criterio, de las dos pruebas que hemos realizado anteriormente con los valores de renta disponible sería la que corresponde a la tabla con intervalos de menor tamaño la que mejor refleja las discontinuidades naturales de la distribución

Veremos a continuación un gráfico que representa una distribución imaginaria. En dicho gráfico indicaremos cuáles son las discontinuidades naturales de la distribución, es decir, los puntos del rango en los que deberíamos situar, en la medida de lo posible, los límites de los intervalos. Trataremos, por último, de adecuar la distribución en intervalos de la tabla de frecuencias a la distribución propia de los valores de la variable.

El objetivo es doble:

- ▶ Obtener una tabla cuyo tamaño o número de clases permita visualizar de forma rápida la distribución de valores de la variable.
- ▶ Descubrir en qué punto o sector del rango se produce un aumento o descenso brusco de la frecuencia de valores.

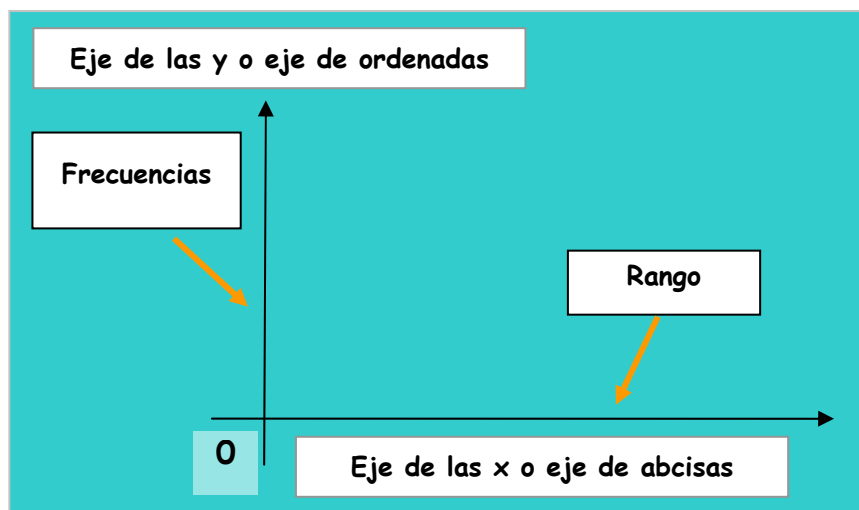
Distribución de los valores en el rango



Representación gráfica de la distribución: el histograma

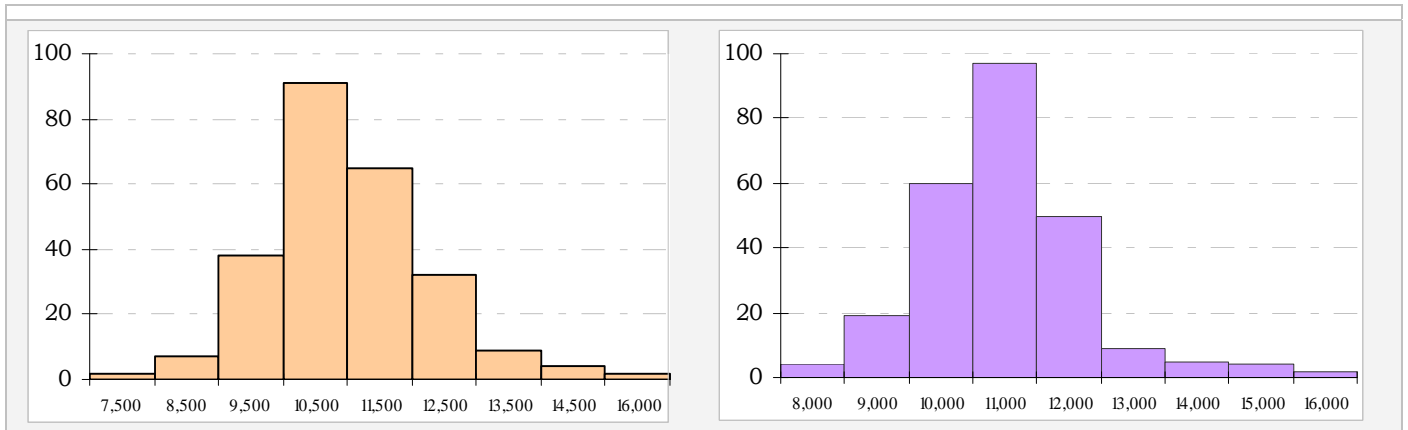
El histograma es un tipo de gráfico que se utiliza para analizar y mostrar la distribución interna de los valores de la variable. Se trata de una herramienta gráfica extremadamente útil para representar de modo adecuado las características más destacables de la distribución de valores de una variable. En definitiva, mediante el histograma se busca una lectura ágil, lo más visual posible, de los rasgos de la distribución.

Como hemos dicho anteriormente, para elaborar el histograma se utilizan las tablas de distribución de frecuencias. Los valores de la tabla se representan en un eje de coordenadas cartesianas:



Se puede afirmar que elaborar un histograma es una tarea fácil siempre y cuando, al hacer la tabla de distribución de frecuencias, se hayan tomado las decisiones adecuadas. Como hemos visto anteriormente, para mostrar de la forma más adecuada posible las características de la distribución, lo más importante es una buena elección del número de clases o intervalos y de los límites de éstos. Repetiremos que nos corresponde a nosotros tomar estas decisiones y que, en función de las elecciones realizadas, la imagen de la distribución que ofrece el histograma puede ser distinta. De hecho, hemos visto los distintos resultados que obteníamos para los valores de renta personal disponible al elegir intervalos de diferente amplitud. A continuación veremos los distintos histogramas que se pueden realizar con los mismos datos.

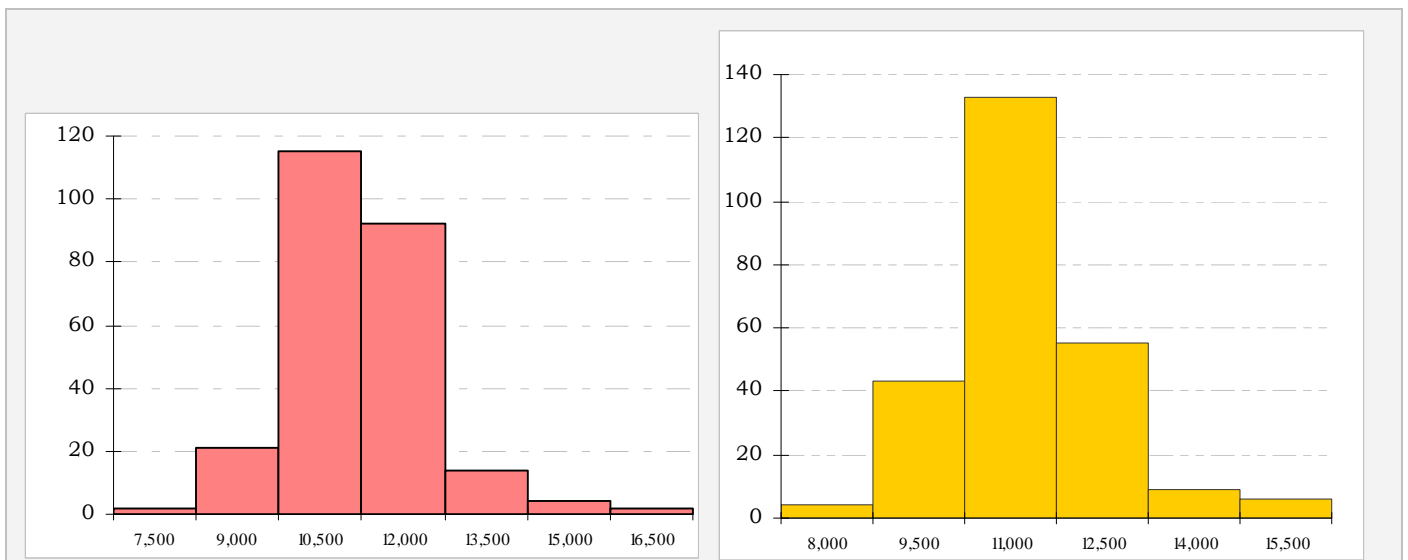
Renta personal disponible de los municipios de Euskadi. 2001



Amplitud de los intervalos: 1.000

Histograma nº 1. Primer intervalo: 6.500 - 7.500 € Histograma nº 2. Primer intervalo: 7.000- 8.000 €

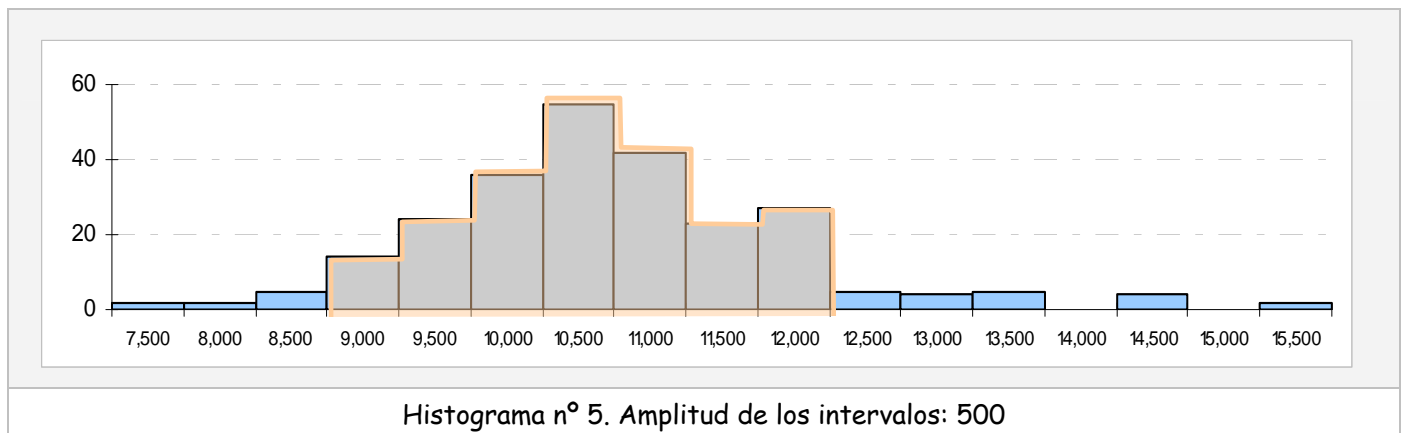
Al cambiar el punto de inicio (límite inferior del primer intervalo) la imagen que ofrece el histograma cambia: en el de la izquierda la sensación es que se acumula un mayor número de valores en la parte derecha del gráfico; en el histograma de la derecha ocurre lo contrario.



Amplitud de los intervalos: 1.500

Histograma nº 3. Primer intervalo: 6.000- 7.500 € Histograma nº 4. Primer intervalo: 6.500- 8.000 €

En este caso, al haber aumentado la amplitud de los intervalos, se obtiene la sensación de que hay una mayor acumulación de valores en la parte central del gráfico. Esta sensación es todavía más acentuada en el histograma de la izquierda, en el que las dos barras centrales, muy destacadas sobre el resto, acaparan la atención.



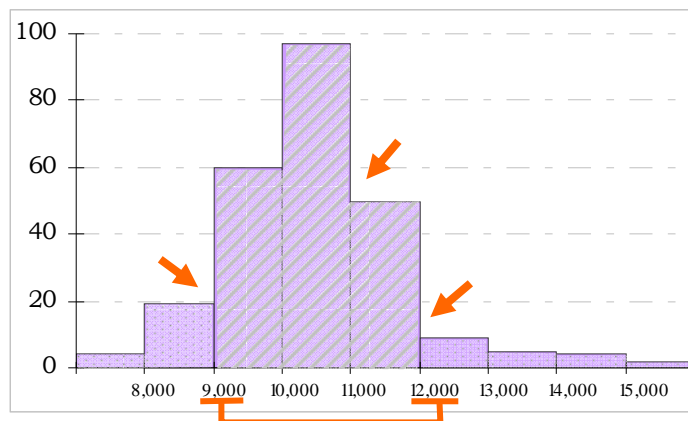
En los histogramas realizados hemos utilizado distinto número de intervalos y, por tanto, distintas amplitudes. Hemos variado también el límite inferior del primer intervalo, empezando en algunos en 6.000 € y en otros en 6.500 €. Es evidente que la impresión visual resultante varía de unos a otros. A la vista de las diferencias cabe preguntarse cuál de los histogramas resultaría más adecuado para representar gráficamente la distribución que estamos analizando; nos podemos preguntar también qué es mejor, si utilizar pocos o muchos intervalos. Lo cierto es que no hay una respuesta única válida para todas las ocasiones. Lo único que se puede decir, a modo de principio general, es que se debería optar por una división en intervalos que, siendo lo más sencilla posible, no enmascare los rasgos propios distribución. Aplicando este principio a los histogramas que acabamos de realizar podríamos concluir lo siguiente:

- El histograma realizado con intervalos de amplitud de 500 muestra de forma impecable las discontinuidades propias de la distribución (los puntos en los que se produce un aumento o un descenso notable de las frecuencias). Nos permite ver de forma clara que la renta de la mayoría de los municipios se sitúa entre los 8.500 y los 12.000 € y que hay un mayor número de valores en la parte izquierda del histograma (mayor número de municipios del lado de las rentas menores) Pese a todo, revisaremos el resto de gráficos para ver si existe la posibilidad de acortar el número intervalos sin perder información esencial.
- El histograma nº 2, (amplitud de intervalo 1.000 €) mucho más simple que el

anterior, muestra también las características esenciales de la distribución de frecuencias, fundamentalmente el brusco descenso que se produce a partir de 12.000 €. Es igualmente perceptible la mayor acumulación de municipios en el lado de las rentas más bajas.

- Los intervalos de los histogramas nº 3 y 4 son excesivamente amplios, de modo que distorsionan los rasgos propios de la distribución.

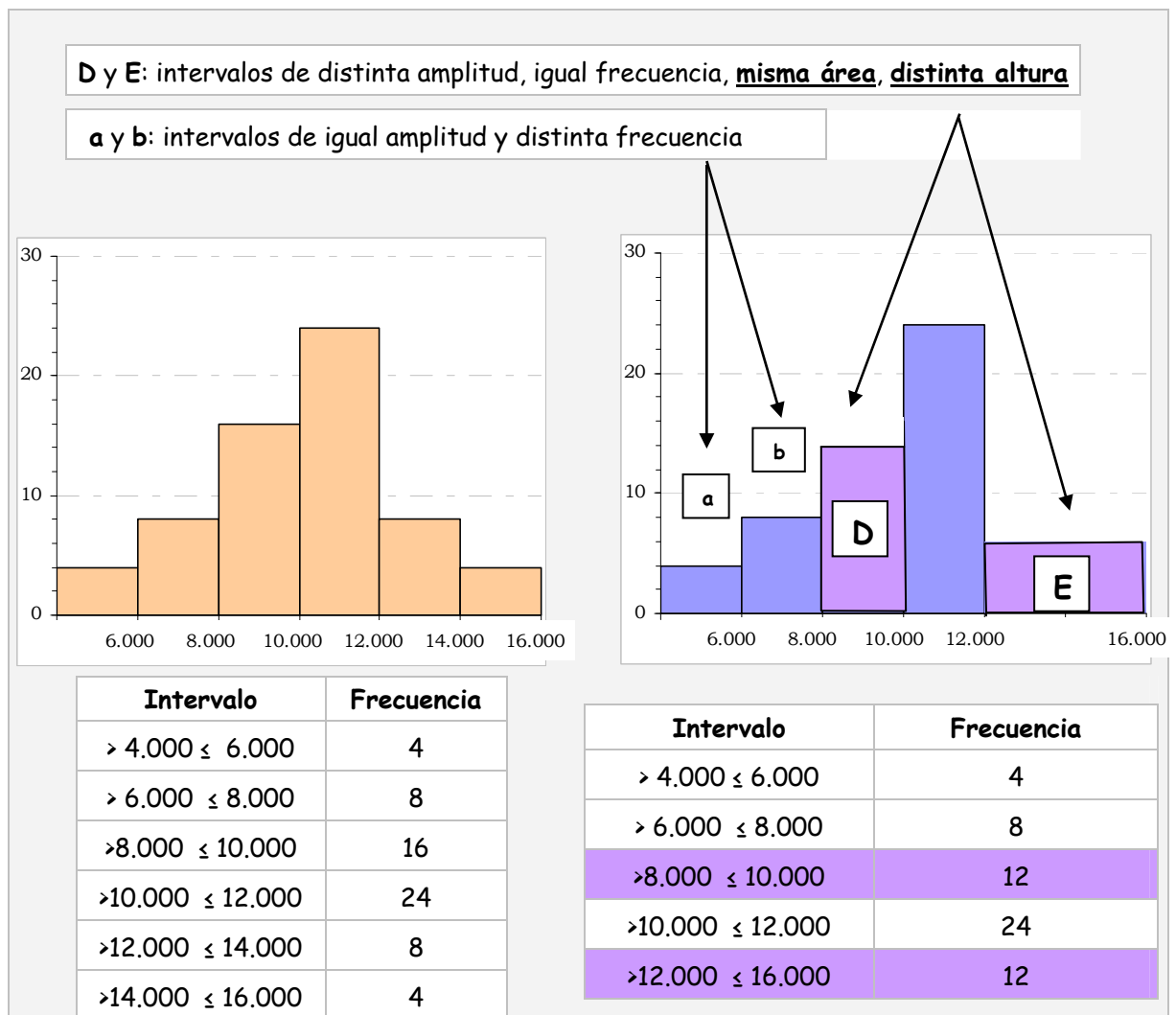
Finalmente pensamos que el histograma nº 2 proporciona una imagen idónea de la distribución de los valores de renta pesonal media disponible en los municipios de Euskadi, en el año 2001. Permite observar de forma clara:



- Que la renta media de los habitantes de los municipios de Euskadi oscila en un rango de valores muy amplio
- Que pese a la gran variabilidad general, existe una mayoría de municipios con rentas más similares entre ellos (el área rallada), ya que son muy pocos los que tienen rentas muy bajas o muy altas (áreas punteadas).
- Que en la mayoría de municipios la renta media de sus habitantes se encuentra entre 9.000 y 12.000 euros. Podemos entonces hablar de que las rentas se concentran fundamentalmente en este intervalo.
- Que hay más municipios con rentas en la mitad inferior del rango que en la mitad superior.
- Una buena aproximación a las “discontinuidades” naturales: los puntos en los que se produce un descenso o aumento brusco de las frecuencias. Las flechas rojas en el gráfico señalan los límites de los intervalos escogidos que

coinciden con las discontinuidades naturales que podemos observar en el [gráfico de la página 27](#).

Los cinco histogramas que hemos realizado tienen una característica en común: los intervalos o clases de cada uno de ellos son de igual amplitud. Esto no es una condición obligatoria pero sí facilita la lectura e interpretación del gráfico debido al hecho de que **en los histogramas la frecuencia de cada clase viene representada por el área del rectángulo y no por su altura**. Cuando todos los intervalos son de idéntica amplitud para comparar las distintas frecuencias de cada uno basta con mirar la altura de los rectángulos o barras. Ante intervalos de diferente amplitud hay que fijarse no en las alturas sino en las áreas. Una ilustración permitirá comprender mejor esta idea:



En el histograma de la izquierda está representada una distribución mediante intervalos de igual amplitud. Siendo así, nos fijaremos solamente en las alturas, que son proporcionales a la frecuencia: el segundo intervalo agrupa dos veces más valores que el primero y, por tanto, su altura es el doble; el tercer intervalo, también con una frecuencia doble a la del segundo es dos veces más alto que éste.

En el histograma de la derecha observamos un intervalo de amplitud diferente a la del resto (E, de 12.000 a 16.000). Vemos también que este intervalo distinto, de doble amplitud, tiene la misma frecuencia que el intervalo D. Si nos fijamos en las áreas de los intervalos D y E veremos que son iguales, pero si atendemos a la altura comprobaremos que son diferentes.

Aunque la lectura del histograma es más ágil y directa cuando utilizamos intervalos de igual amplitud, en algunas ocasiones se aconseja realizar intervalos distintos. Este sería el caso, por ejemplo, de las distribuciones en las que muchos de los valores se concentran en una pequeña parte del rango y sólo unos pocos valores se diseminan en el resto. Ante distribuciones de este tipo, puede resultar adecuado utilizar intervalos amplios para las zonas del rango con valores dispersos y utilizar intervalos más pequeños para las zonas que acumulan gran número de valores.

Al presentar las tablas de distribución de frecuencias hemos hablado de los intervalos abiertos y cerrados. Hemos visto que es posible, y frecuente también, realizar tablas en las que el primer y el último intervalo se dejan abiertos. Ahora bien, en el caso del histograma todos los intervalos tienen que ser cerrados. Esto es así porque, como acabamos de ver, el área de cada uno de los rectángulos que forman el histograma se define en función de la amplitud del intervalo y de la frecuencia de los valores que acumula. Está claro entonces que sólo se puede calcular el área para los intervalos cerrados.

Para finalizar el tema haremos un cuadro resumen de las principales características del histograma:

- ❖ El histograma es una herramienta gráfica extremadamente útil no sólo en el momento de presentar los resultados del análisis. Resulta de gran ayuda también durante la fase de exploración de los datos, ya que nos permite identificar la estructura natural de distribución de los valores de la variable.

- ❖ Dependiendo de las decisiones que se tomen en relación al número y amplitud de los intervalos del histograma, la imagen que éste ofrece sobre la distribución de los valores de la variable, puede ser muy diferente.
- ❖ No existe ninguna norma que permita decidir cuál es el número y la amplitud idónea de los intervalos de un histograma. Son decisiones que debe tomar el autor del histograma. Lo importante es reflejar de la manera más sencilla posible la estructura propia de los datos.
- ❖ Cuando se elabora un histograma con el fin de presentar una imagen gráfica de la distribución, es conveniente hacer varios ensayos, analizar las diferencias y escoger el que consideremos más idóneo.

TERCERA PARTE:
CÓMO ANALIZAR Y DESCRIBIR LA VARIABILIDAD
CUANDO LOS DATOS DIBUJAN UNA FORMA DE
CAMPANA

Tema 5:

Análisis del sector central de la distribución: media y
desviación estándar

Introducción	2
El punto central en las distribuciones campaniformes	3
Promedio o media aritmética	4
▣ Cálculo del promedio: fundamentos de la notación estadística.....	4
▣ Utilización de la media	8
Otra medida para determinar el punto central de la distribución : la mediana.....	12
El promedio en distribuciones con valores extremos	18
El intervalo de concentración de los valores. La desviación estandar	24
▣ Introducción	24
▣ El concepto de desviación estandar y su procedimiento de cálculo.....	27
▣ La utilización de la desviación estandar.....	34
▣ Otros usos de la desviación estandar	38
▣ El coeficiente de variación	44

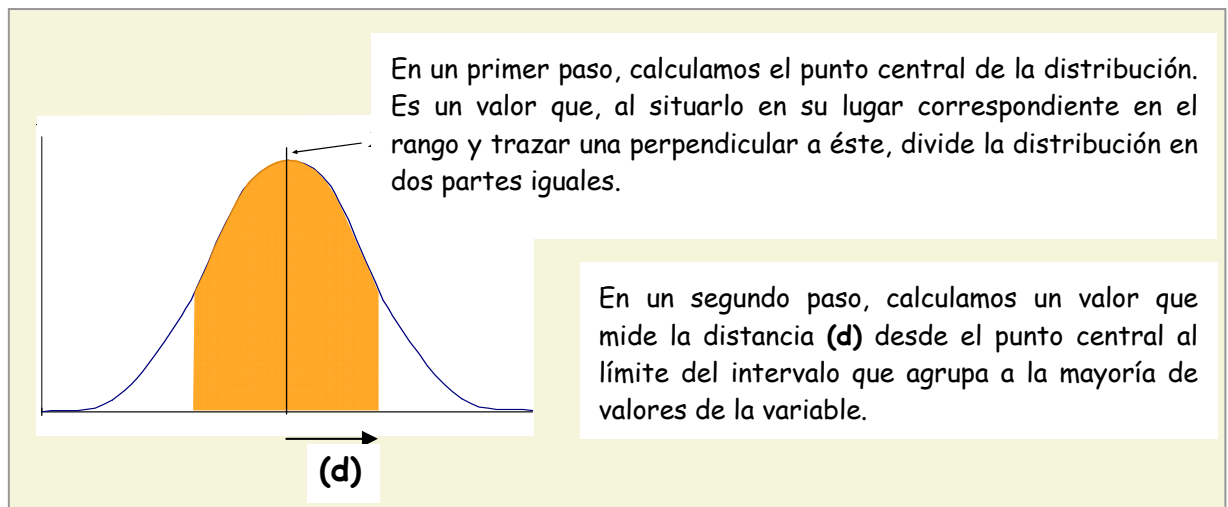
Tema 5:

Análisis del sector central de la distribución: media y desviación estándar

INTRODUCCIÓN

Analizar el sector central en las distribuciones campaniformes significa determinar la variabilidad del conjunto de datos que aparecen agrupados en el sector central del rango. En los capítulos anteriores nos hemos referido a ellos como el grupo mayoritario de valores, para diferenciarlos del resto de valores que aparecen más dispersos hacia los extremos de la distribución.

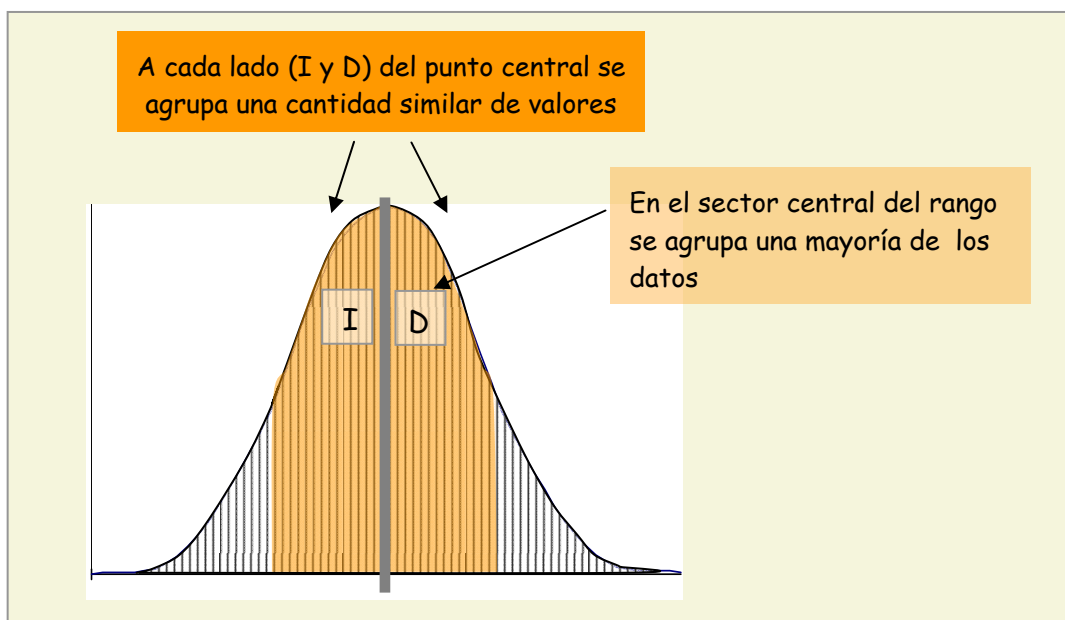
Conceptualmente, el procedimiento para calcular la variabilidad del grupo mayoritario de valores es bien sencillo. Consta de dos pasos:



EL PUNTO CENTRAL EN LAS DISTRIBUCIONES CAMPANIFORMES

El primer paso para hacer la descripción de la zona del rango que concentra la mayoría de los valores es calcular el punto central de la distribución, puesto que es el punto en torno al cual se agrupa la mayoría de los datos. Se trata de un valor que, en las distribuciones campaniformes, se sitúa en el sector central de la distribución.

En el gráfico adjunto, hemos dibujado una franja que se corresponde con el sector central de la distribución. La línea en su interior, perpendicular a la línea del rango, representa el valor central al que nos referimos.



Existen diversos procedimientos para calcular el punto central de una distribución de valores. De todos los métodos existentes, el más ampliamente utilizado es, sin duda, la media aritmética o promedio. La facilidad de su cálculo explicaría el uso tan extendido de este procedimiento, incluso cuando dicho uso no es aconsejable.

Aunque su cálculo es una tarea muy sencilla que, además, hoy en día realizan los ordenadores, lo que no parece tan sencillo es comprender que no siempre es aconsejable usar el promedio para calcular el punto central de una distribución.

Como punto de partida, y como principio general también, podemos decir que el promedio sólo debe utilizarse para la descripción de variables cuya distribución sea campaniforme. En las páginas que siguen se darán las explicaciones que permiten entender esta afirmación.

PROMEDIO O MEDIA ARITMÉTICA

El promedio o media aritmética es el valor que se obtiene al sumar todos los datos y dividir el resultado entre el número total de datos.

Cálculo del promedio: fundamentos de la notación estadística

En todas las explicaciones que hemos presentado hasta ahora no hemos utilizado ninguno de los símbolos estadísticos habituales en cualquier manual. De aquí en adelante, será inevitable recurrir a dichos símbolos a la hora de presentar determinados procedimientos y fórmulas. Por este motivo, y porque entender y utilizar los símbolos estadísticos es algo mucho más fácil de lo que puede parecer en un principio, utilizaremos el espacio dedicado al promedio para explicar con detenimiento el significado de los símbolos estadísticos más comunes.

Para referirnos de forma genérica a cualquier variable que analizamos mediante las herramientas estadísticas, utilizamos el símbolo de la X , en mayúscula. Lo que se hace, realmente, es utilizar una letra para mencionar de forma abreviada un concepto. Supongamos, a modo de ejemplo, la siguiente expresión:

Se deben sumar todos los valores de la variable

Ahora, sustituiremos la palabra *variable* por su correspondiente símbolo:

Se deben sumar todos los valores de X

Expresado de este modo, se entiende que, sea cual sea la variable de la que estemos hablando, se deben sumar todos sus valores

Para mencionar los valores de la variable de cualquier población estadística utilizamos también como símbolo la x , esta vez minúscula, seguido de un subíndice que indica que se trata de cualquier valor de la variable: x_i

Si tomamos cualquier valor de la variable...

Si tomamos x_i

x_i

Cuando queremos concretar la mención y referirnos a un grupo de valores de cualquier variable se sustituye el subíndice (i) por números. Supongamos la expresión:

Tomaremos los seis primeros valores de la variable y...

Tomaremos $x_1 x_2 x_3 x_4 x_5 x_6$ y...

$x_1 x_2 x_3 x_4 x_5 x_6$

Siguiendo esta lógica, cuando queremos decir que una operación estadística consiste en sumar los valores de la variable, podemos escribirlo así:

$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \dots\dots$

Supongamos ahora que debemos expresar mediante símbolos que hay que sumar todos los valores de la variable. Hacerlo como hasta ahora sería imposible porque cuando nos estamos refiriendo a cualquier variable, de forma genérica, el número de valores puede ser también cualquiera. Necesitamos, por tanto, una expresión que indique, de forma abreviada, el concepto de todos los valores de la variable.

Un modo de hacerlo es escribir, como anteriormente, unos cuantos valores de la variable con el subíndice correspondiente. Después se añaden puntos suspensivos y una última x, pero esta vez con el subíndice n, para indicar que sumaremos todos los valores de la variable que estamos analizando:

$x_1, + x_2 + x_4 + x_5 \dots x_n$

Es la expresión $\dots x_n$, la que indica que la suma continuará hasta llegar al último valor de la variable. La letra **n** alude siempre al número de valores que estamos analizando. Nos permite mencionar mediante un símbolo el

concepto de *número total de valores*, sin importar cuál sea esta cantidad.

Una vez explicado el significado de los símbolos más habituales, podemos presentar la expresión matemática de la media aritmética o promedio, que sustituye a su definición:

$$\overline{X} = \frac{x_1 + x_2 + x_4 + x_5 \dots x_n}{n}$$

La expresión matemática, que ahora debe resultarnos comprensible, se puede abreviar más todavía, ya que existen otros símbolos que lo permiten.

Tenemos, en primer lugar el símbolo Σ denominado *sumatorio*. Colocado delante de una serie de números, indica que se deben sumar todos los números que hay a su derecha. Lo que nos permite es sustituir todos los símbolos de suma (+) por uno único.

Podemos simplificar la operación, y sustituir todos los símbolos de la suma por un único símbolo, y las menciones individuales de los valores por una única mención:

$$x_1 + x_2 + x_4 + x_5 \dots x_n \longrightarrow \Sigma x_i$$

A la expresión que hemos escrito le faltaría indicar que la suma debe incluir todos los valores de la variable.

$$\sum_{i=1}^n x_i$$



$$\sum x_i$$

El símbolo del sumatorio seguido del símbolo de variable, indican que se sumarán los valores de la variable

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

Debajo del sumatorio se detalla el significado del subíndice i que acompaña a la x . Se indica, concretamente, que el subíndice alude a todos los valores de la variable, desde el primero ($i=1$) al último ($i=n$)

Finalmente, la fórmula matemática de la media queda así:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

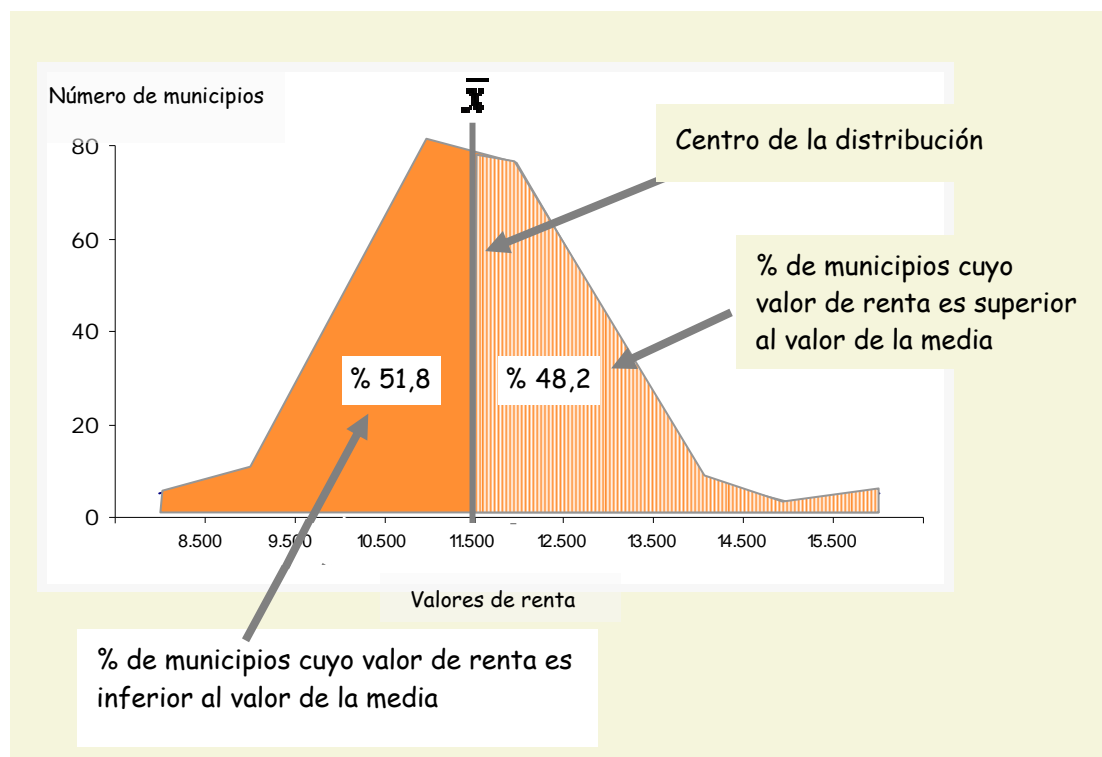
Utilización de la media

La media de los valores de una variable se utiliza para definir el punto central de distribuciones campaniformes. Como se ha indicado anteriormente, cuando la distribución tiene forma de campana una mayoría de los valores se agrupa a cada uno de los lados del punto central de la distribución.

Para entender mejor el significado y la utilización de la media veremos, a continuación dos ejemplos: uno de ellos se corresponde con una distribución campaniforme y, el otro, con una distribución de forma distinta, no campaniforme.

Veremos en primer lugar el ejemplo correspondiente a la distribución campaniforme. Los datos corresponden a los valores de renta personal disponible de los municipios de Euskadi en 2003.

La gráfica nos muestra lo que decíamos al inicio del tema:



En las distribuciones campaniformes una mayoría de los valores se concentra en el sector central del rango; en el centro de este sector, también, se sitúa el valor de la media

En relación a la variable de renta disponible, las características que comentamos se concretan en el hecho de que el valor de la media, 11.506 €, se sitúa en el

centro de la distribución; el porcentaje de municipios con valores inferiores y superiores a la media es similar.

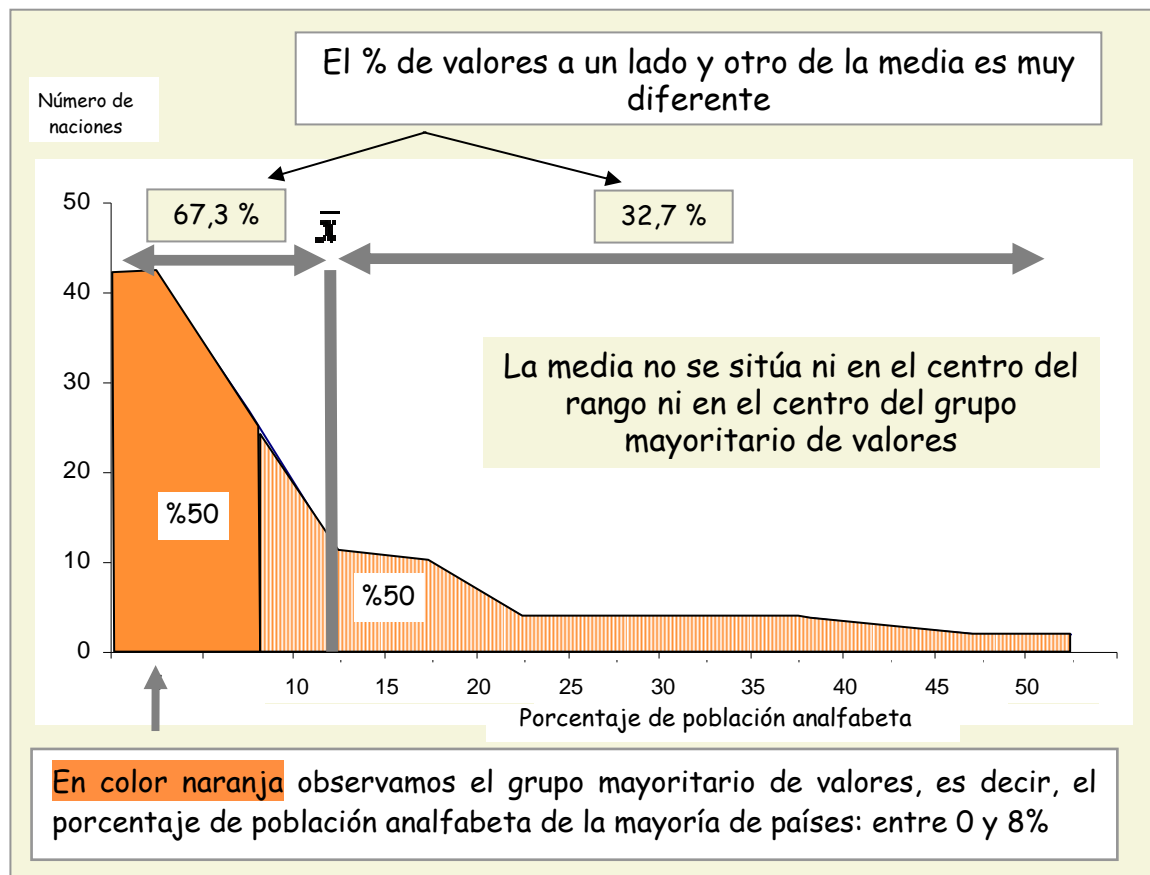
La conclusión fundamental que se extrae de lo que acabamos de mostrar es que:

El valor de la media en las distribuciones campaniformes es un dato significativo y adecuado como descriptor de la distribución.

El valor de renta media de los municipios de Euskadi en 2003, 11.506 €, es un dato que representa de forma idónea al conjunto de valores de renta analizados: un número importante de municipios tuvieron en 2003 un valor de renta próximo al de la media. En la tabla anexa podemos observar que 193 municipios tuvieron un valor de renta comprendido entre los 10.000 y los 13.000 €. Estas 193 entidades constituyen el 77% del total de los municipios de la Comunidad Autónoma.

Intervalos	Nº de municipios	Porcentaje	Porcentaje acumulado
>10.000 ≤10.500	29	12	12
> 10.500 ≤ 11.000	27	11	23
> 11.000 ≤ 11.500	47	19	41
> 11.500 ≤ 12.000	37	15	56
> 12.000 ≤ 12.500	32	13	69
> 12.500 ≤ 13.000	21	8	77
	193	77 %	

A continuación veremos el ejemplo de una distribución no campaniforme. La población en este caso está constituida por 113 naciones del mundo, con niveles de desarrollo muy distintos. La variable que se analiza es el porcentaje de población analfabeta existente en cada uno de dichos países, según datos de 2006 recogidos por la UNESCO. (Ver Anexo nº 8) Como en el caso anterior, la gráfica nos muestra la distribución de valores dentro del rango. Mediante colores distintos hemos señalado también los dos lados de la distribución, cada uno de los cuales agrupa el 50% de los valores.



Resulta evidente que ahora no estamos ante una distribución campaniforme. La cuestión es ahora definir cuáles son las diferencias, además de la forma, entre las características de las dos distribuciones. Dicho de otro modo, se trata de relacionar las distintas formas de las distribuciones con las características de las poblaciones a las que representan.

- En la distribución no campaniforme, el grupo mayoritario de valores no se sitúa en el centro de la distribución sino en el extremo inferior de ésta. Esto significa, como podemos ver en la gráfica, que la mitad de los países tienen un porcentaje de población analfabeta inferior al 8%.
- El valor de la media, 12,2%, no se sitúa en el centro del grupo mayoritario de valores sino desplazado hacia la derecha en relación a aquellos. El hecho es que, mientras según el valor del promedio el porcentaje medio de población analfabeta se sitúa en un 12,2%, la gráfica muestra claramente que en una mayoría de países este porcentaje es claramente inferior. Del total de

naciones analizadas (113), un 67,3% tenían en 2006 un porcentaje de población analfabeta inferior al que indica la media.

La conclusión más importante que podemos extraer del ejemplo sobre la distribución del porcentaje de población analfabeta en los distintos países es que:

En las distribuciones no campaniformes, la media no es un valor representativo de la población, no proporciona una imagen idónea del centro de la distribución.

Intervalos	Nº de naciones	Porcentaje	Porcentaje acumulado
$>0 \leq 5$	42	37,17	37,17
$> 5 \leq 10$	27	23,89	61,06
$> 10 \leq 15$	11	9,73	70,80
$> 15 \leq 20$	10	8,85	79,65
$> 20 \leq 25$	4	3,54	83,19
$> 25 \leq 30$	4	3,54	86,73
	98		

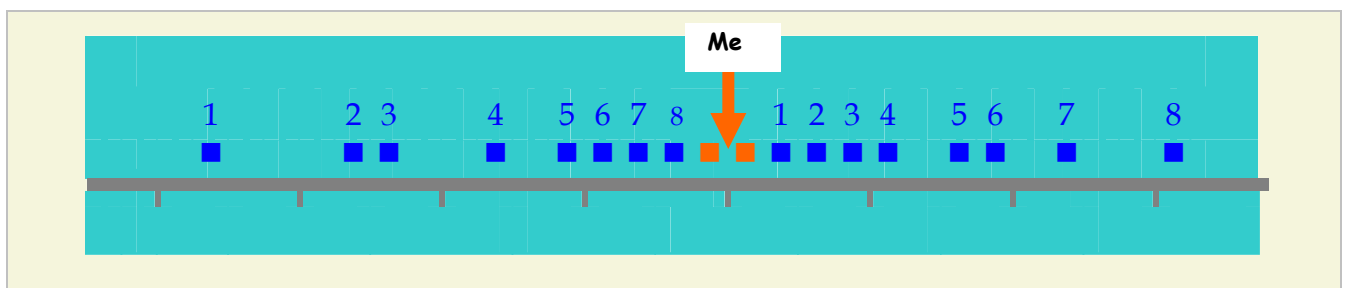
En un 61,06 % de los países, la población analfabeta no alcanza el 10%

¿Qué ocurriría entonces si utilizáramos el promedio como descriptor del porcentaje de población analfabeta en el mundo?. La respuesta es que proporcionaríamos una imagen distorsionada de la realidad que llevaría a pensar que, en la mayoría de los países, el porcentaje de personas analfabetas es muy superior al real.

OTRA MEDIDA PARA DETERMINAR EL PUNTO CENTRAL DE LA DISTRIBUCIÓN : LA MEDIANA

En el apartado anterior hemos podido comprobar que en las distribuciones campaniformes la media se sitúa en el centro de la distribución. Hemos podido ver igualmente que, cuando la distribución no es campaniforme, la media aparece desplazada con respecto al centro de la distribución.

Cuando la media no es una medida representativa de los valores de la variable se puede utilizar la mediana, que es otra medida distinta para determinar el punto central de la distribución. Para conocer el valor de la mediana se ordenan los datos de la variable, de forma creciente o decreciente; a continuación se busca el valor central que divide la distribución en dos mitades; a dicho valor se le denomina mediana. Cuando el número de valores de la variable es par, para obtener la mediana se debe calcular el promedio de los dos valores centrales de la distribución.



En la tabla siguiente podemos ver los datos correspondientes a las cantidades de cartón y papel que se recogieron en las distintas comunidades autónomas españolas en 2004 mediante procesos de recogida selectiva de basuras. Los datos reflejan, para cada una de las comunidades autónomas, la cantidad media, medida en kg., de papel y cartón producida por persona y año. Para obtener la mediana de estos datos, puesto que el número de comunidades autónomas es par, calculamos el promedio de los dos valores centrales de la distribución.

Comunidades autónomas	Papel y cartón (Kg.)
Extremadura	8,3
Ceuta y Melilla	9,3
Castilla y León	10
Andalucía	11
Castilla la Mancha	12,6
Valencia	13
Murcia	13,6
Galicia	14
Canarias	14,5
Cantabria	14,8
Aragón	15,8
Madrid	20,3
Cataluña	21,2
La Rioja	21,2
Asturias	22,8
Islas Baleares	25,7
Navarra	26
Euskadi	38,9

$$Me = \frac{14,5 + 14,8}{2} = 14,65$$

Fuente: Centro Regional de Estadística de Murcia. Anuario estadístico de la ciudad de Murcia. 2007.

<http://www.carm.es/econet/anuario/a2007/anuario.html>

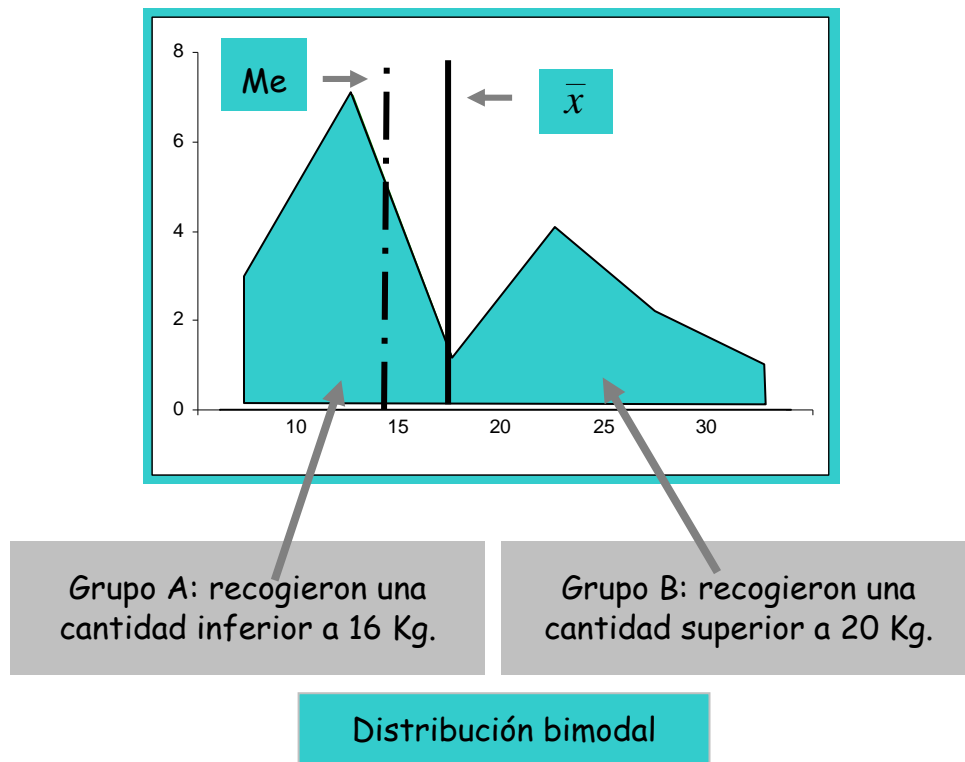
Además de la mediana, calcularemos ahora también la media, ya que el objetivo es mostrar la diferencia que existe entre los dos valores:

$$\frac{\sum_{i=1}^n X_i - \bar{X}}{n} = \frac{313}{18} = 17,39$$

Veamos ahora la diferencia entre las dos medidas:

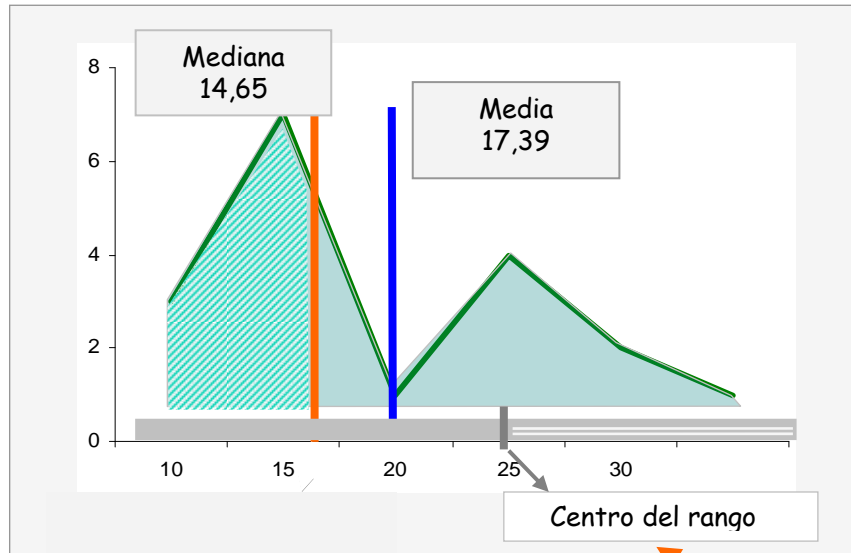
	Media	Mediana
	17,39	14,65
Diferencia	2,74	

Comparados los resultados de las dos medidas -media y mediana- podemos concluir que la diferencia entre los valores de ambas es considerable: 2,7 Kg. de papel y cartón por habitante al año supone un aumento del 18,7% en relación a la mediana. Si miramos la gráfica del polígono de frecuencias en seguida comprenderemos por qué los valores de media y mediana son tan diferentes:



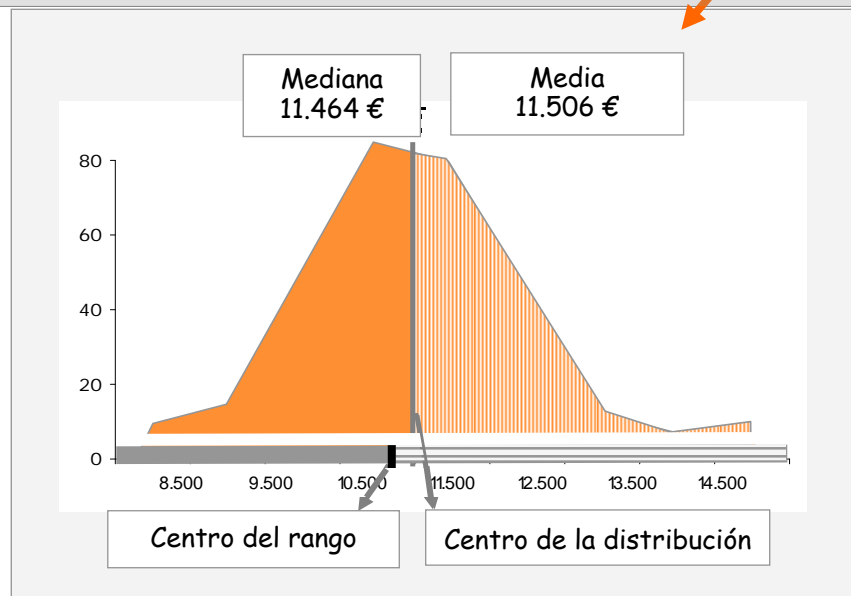
El polígono de frecuencias nos muestra una imagen reveladora del diferente comportamiento de las 18 comunidades autónomas en relación a la cantidad de papel y cartón que produjeron por persona y año en 2004. La gráfica nos permite observar también que existen dos grupos de comunidades. El primero, y más numeroso, está formado por 11 comunidades, en las cuales la cantidad de papel y cartón recogidos por persona no llegó a los 16 kg. El segundo de los grupos está formado por las 7 comunidades restantes; en ellas las cantidades de papel y cartón recogidas fueron muy superiores, alcanzándose en un caso los 38,9 Kg..

Resulta fácil comprender ahora que son los valores del segundo grupo de comunidades autónomas los responsables de que el resultado de la media sea sensiblemente superior al de la mediana. De hecho, el valor de la media, que se sitúa en medio de los dos subgrupos, no es un valor representativo de ninguno de ellos. 17,39 Kg. por persona y año es bastante más de lo que se recogió en ninguna de las primeras once comunidades y mucho menos también de lo que se recogió en las siete comunidades restantes.



Cantidad de papel y cartón recogido en 2004 en las CCAA españolas.
 Distribución no campaniforme
 Mediana, media y centro del rango: **valores muy diferentes**

Mediana, media y centro del rango: valores **muy similares**
 Renta personal disponible. Municipios de Euskadi. 2001. Distribución campaniforme



En Estadística, la media aritmética y la mediana, junto con otras medidas, se clasifican en el grupo denominado medidas de tendencia central. Para ahora, ya es evidente la razón que justifica esta denominación, que no es otra que su *tendencia* a situarse en el centro de la distribución. Precisamente cuando esto no ocurre, es decir, cuando la media se aleja sensiblemente del sector central, deja de resultar un valor idóneo para representar a los datos.

EL PROMEDIO EN DISTRIBUCIONES CON VALORES EXTREMOS

El promedio es una medida muy sensible a la presencia de valores extremos, que pueden llegar a influir notablemente en su resultado. Para comprender esta idea hay que tener en cuenta que para calcular el promedio utilizamos todos los valores de la variable que estamos analizando y por ello, todos ellos tienen influencia en el resultado final. El problema surge cuando alguno de estos valores ejerce una influencia desmesurada, elevando o descendiendo notablemente el resultado del promedio.

La presencia de valores extremos es una situación habitual. En ocasiones, dentro de una serie de datos cuya distribución es campaniforme, existe algún dato especialmente alto o bajo con respecto a los demás. Esto es, por ejemplo, lo que ocurre con los datos del precio medio de la vivienda nueva en las principales ciudades aragonesas, en 2007. En la tabla adjunta podemos ver la gran diferencia que existe en el precio medio de la vivienda entre la ciudad más cara -Zaragoza- y la segunda de la lista -Jaca-. El precio de la vivienda en Zaragoza, más de 100.000 euros mayor que el de Jaca, constituye un valor extremo que eleva notablemente el valor del promedio.

	Ciudades	Precio	Incremento
1	Caspe	103.900	
2	Fraga	127.000	23.100
3	Barbastro	128.500	1.500
4	Alcañiz	129.700	1.200
5	Ejea de Los Caballeros	155.400	25.700
6	Calatayud	158.600	3.200
7	Teruel	170.600	12.000
8	Huesca	194.900	24.300
9	Jaca	197.600	2.700
10	Zaragoza	301.500	103.900

¹La columna *incremento* nos permite ver la diferencia en el precio medio de la vivienda entre una ciudad y la siguiente más cara de la lista

¹ Precio medio, nominal, en euros, de la vivienda nueva en junio de 2007. Sociedad de Tasación. SA. <http://web.st-tasacion.es/html/index.php>. Última consulta 10-09-2008. Valor nominal: En el valor se incorpora el IPC (unidad monetaria habitual a nivel de consumo). Valor Real: Valor en el supuesto de que el incremento del IPC hubiese sido 0 desde diciembre de 1985.

Si calculamos el promedio con el dato de Zaragoza y sin él, comprobamos hasta qué punto influye este valor en el resultado: la diferencia llega prácticamente hasta los 15.000 euros.

	Sin Zaragoza	Con Zaragoza
Promedio	166.770	151.800

Diferencia: 14.970 €

A la vista de los resultados, es preciso concluir que cuando tenemos una serie de datos en la que existe algún valor especialmente alto o bajo tenemos que prestar especial atención a la influencia que estos valores pueden tener en el promedio.

Veremos ahora nuevos datos sobre el precio medio de la vivienda en 2007, pertenecientes también a la Comunidad autónoma de Aragón, pero esta vez sólo de la provincia de Zaragoza.² Se trata de los precios medios de 2007, a nivel comarcal. A diferencia de los datos anteriores, estos incluyen la vivienda nueva y la de segunda mano. El objetivo no es hacer ninguna comparación con los datos anteriores ni obtener conclusiones con respecto al precio de la vivienda. Se trata exclusivamente de mostrar que los valores extremos no siempre afectan de forma significativa el valor del promedio.

² Precio de la vivienda por metro cuadrado en las comarcas de Zaragoza, 2007. Fuente: Mercado Inmobiliario de Aragón 2007. Caja de Ahorros de la Inmaculada de Aragón.

	Comarca	Precio	Incremento
1	Bajo Martín	52.647	
2	Bajo Aragón-Caspe	78.960	26.313
3	Campo De Daroca	90.090	11.130
4	Sierra de Albarracín	92.584	2.494
5	Campo de Belchite	95.184	2.600
6	Campo de Borja	97.468	2.284
7	La Ribagorza	100.152	2.684
8	Matarraña	100.386	234
9	Sobrarbe	104.958	4.572
10	Andorra-Sierra de Arcos	105.840	882
11	Jiloca	107.334	1.494
12	Ribera Alta Del Ebro	108.864	1.530
13	Bajo Aragón	109.896	1.032
14	Valdejalón	112.800	2.904
15	Cinca Medio	121.869	9.069
16	Cinco Villas	122.400	531
17	Comunidad de Calatayud	123.172	772
18	Bajo Cinca	127.846	4.674
19	Gúdar-Javalambre	142.389	14.543
20	Somontano de Barbastro	145.376	2.987
21	Tarazona Y El Moncayo	153.510	8.134
22	La Jacetania	167.014	13.504
23	Ribera Baja Del Ebro	176.157	9.143
24	Los Monegros	178.830	2.673
25	Hoya de Huesca	181.764	2.934
26	Comunidad de Teruel	185.472	3.708
27	La Litera	186.400	928
28	Campo de Cariñena	186.944	544
29	Alto Gállego	192.975	6.031
30	D.C. Zaragoza	236.672	43.697

En este caso, también, es el dato de Zaragoza el que constituye un valor muy superior al del resto de la población. En la columna de la derecha de la tabla, en la que hemos calculado los incrementos de precio que se producen de cada comarca a la siguiente, podemos ver que el precio de Zaragoza supera en 43.697 euros al de la comarca anterior. Expresado en porcentajes, se puede decir que el precio medio en la comarca Zaragoza es un 22,6% superior al de la comarca del Alto Gállego.

Veamos ahora cuál es la influencia del valor de Zaragoza en el promedio:

	Con el dato de Zaragoza	Sin el dato de Zaragoza
Promedio	132.865,1	129.285,55
Diferencia	3.579,55	

En este caso también comprobamos que el valor de Zaragoza eleva el resultado del promedio en 3.579,55 euros. La cantidad no es despreciable, pero sí es muy inferior a los 15.000 euros que hemos obtenido anteriormente. Es evidente que, en este último caso, el valor de Zaragoza no es tan extremo como en el ejemplo anterior. Pero hay otro factor que explica la menor influencia en el promedio del elevado valor de Zaragoza: el tamaño de la población. Mientras la población de ciudades está integrada por 10 elementos, la población de comarcas cuenta con 30 elementos: el elevado valor de Zaragoza se reparte entre ellos de modo que se difumina su influencia en el promedio.


Imaginemos ahora que el precio medio de la vivienda en la comarca de Zaragoza es todavía mayor, por equipararlo con el dato extremo del precio medio de la vivienda en las ciudades. Imaginemos que en lugar de 236.672 es de 290.000 y veamos cómo repercute este incremento -ficticio- en el resultado del promedio:

	Con el dato de Zaragoza		Sin el dato de Zaragoza
	Ficticio	Real	
Promedio	134.643	132.865,1	129.285,5
Diferencia	1.779,9		3.579,6

Los resultados muestran que el dato ficticio que hemos creado incrementa la diferencia entre las dos comarcas más caras casi hasta 100.000 €. Pese a este notable aumento, el promedio se eleva sólo en 1.779 €

A la vista de los resultados, podemos concluir que, incluso elevando de forma notable el valor correspondiente a la comarca de Zaragoza, su influencia en el promedio sigue siendo discreta.

La influencia variable de los valores extremos en función del tamaño de la población, es decir, del número de valores que analizamos, se aprecia muy claramente en el caso de la renta personal disponible de los municipios de Euskadi en 2001. Uno de los 251 municipios tuvo un valor de renta muy superior al del resto. El valor de renta del municipio de Laukiz, con 20.627 €, constituye un valor extremo. La diferencia entre el valor de Laukiz y el valor del siguiente municipio en la lista, Lanestosa, es de 5.281 €.

Renta personal disponible de los municipios de Euskadi en 2001 (€)		
Valor mínimo	Lanestosa	7.192
Valor máximo	Laukiz	20.627
Segundo valor más alto	Leintz Gatzaga	15.346
	Incluido el valor extremo	Excluido el valor extremo
Promedio	10.514,47	10.474
Diferencia de 40,45 euros 		

En la tabla superior podemos observar que la influencia del valor extremo de Laukiz en el promedio es mínima: la diferencia entre el promedio calculado con todos los valores y el promedio calculado excluyendo el dato de Laukiz se reduce a 40,45 €. Una vez más, podemos comprobar que en el caso de poblaciones con un gran número de elementos, la influencia de los valores extremos en el promedio es muy leve.

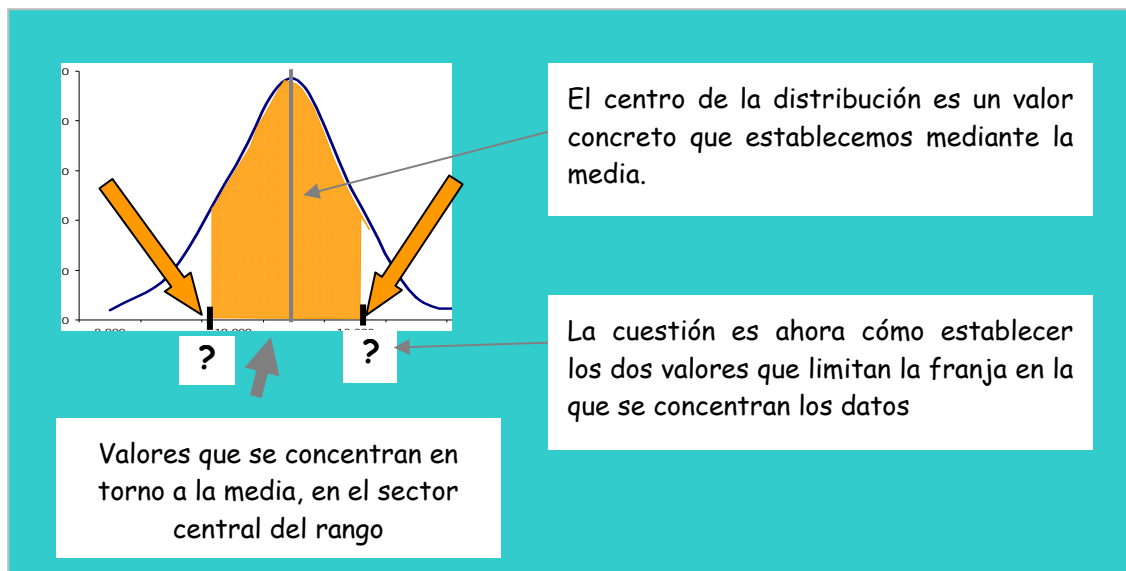
¿Qué podemos concluir, finalmente, sobre el modo de analizar una serie de datos que contiene valores extremos?

- ❖ En general, cuando el número de elementos de la población no es pequeño, la existencia de un elemento extremo no altera sustancialmente el valor del promedio.
- ❖ No hay ninguna norma o convención que permita determinar en qué momento el tamaño de una población deja de ser pequeño. Por esta razón la influencia de los valores extremos en el cálculo del promedio debe ser valorada en cada caso, es decir, siempre que detectemos su presencia.
- ❖ Cuando consideremos que la influencia de algún valor extremo en el promedio es significativa, podemos tomar la opción de realizar el análisis excluyendo el o los valores extremos. Es una práctica habitual.
- ❖ En cualquier caso, los valores extremos, si no son errores o anomalías, en tanto que en realidad existen, no se pueden olvidar o simplemente apartar. Cuando excluimos los valores extremos del cálculo del promedio, no los estamos excluyendo del análisis estadístico. Lo que hacemos es calcular el punto central de la distribución sin elementos distorsionantes, pero en nuestro análisis y conclusiones les damos la importancia que tienen.

EL INTERVALO DE CONCENTRACIÓN DE LOS VALORES. LA DESVIACIÓN ESTANDAR

Introducción

Al hablar de las distribuciones campaniformes decíamos que la media es un valor que se sitúa en el centro del rango, y en el centro también del grupo mayoritario de valores.



De ese grupo de valores en torno a la media nos ocuparemos a continuación, es decir, del intervalo dentro del rango en el que se concentra una mayoría de los valores de la variable que analizamos. Para conocer dicho intervalo lo que hacemos es calcular los valores o puntos del rango que lo limitan.

Es importante comprender bien la lógica de los procedimientos de cálculo que explicaremos a continuación. Pero antes de eso, lo fundamental es tener muy claro por qué y con qué fin se utilizan estos procedimientos:

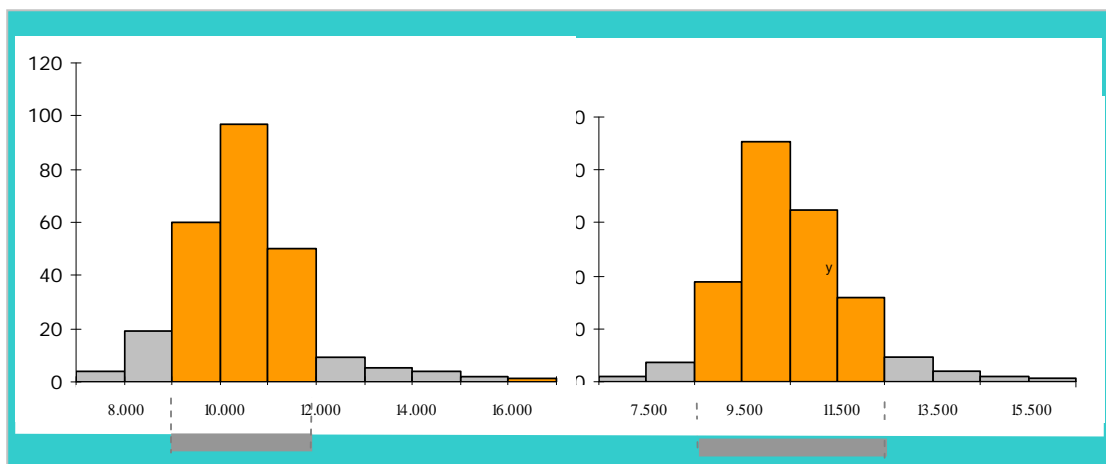
- El promedio y el intervalo que agrupa una mayoría de los datos de una variable se calculan cuando queremos conocer hasta qué punto, o en qué medida, son variables (diferentes entre sí) los valores de una variable.
- Las variables son características o propiedades de los elementos de una

población. De una población compuesta por todas las empresas registradas, por ejemplo, en Guipúzcoa, una característica o variable podría ser el número de empleados que tiene cada una.

- ▣ Dado que los valores de las variables suelen ser distintos para cada uno de los elementos de la población, una forma de conocer, y de describir, su variabilidad es calcular en qué rango de valores se encuentra la mayor parte de los miembros de la población. Aplicado al ejemplo anterior, intentaríamos conocer cuáles son el número máximo y mínimo de trabajadores entre los que se encuentra la mayoría de las empresas.

PERO, es preciso recordar también que sólo podemos realizar este tipo de análisis si la distribución es campaniforme.

Al explicar los histogramas y polígonos de frecuencia hemos podido ver que este tipo de gráficos permite generar la imagen de cualquier distribución de valores y destacar en ella los intervalos que acumulan la mayor parte de los valores de una variable. En base a esta idea podríamos concluir que bastaría con utilizar el histograma de una distribución para calcular el intervalo que concentra una mayoría de los valores. La conclusión sería errónea puesto que la imagen que proporciona un histograma puede ser muy variable en función del número y de la amplitud de los intervalos que se elijan. Lo que necesitamos es un método que nos permita calcular de forma inequívoca el intervalo que buscamos.



Los dos histogramas que vemos en la imagen están elaborados en base a los mismos datos. Los dos contienen un número idéntico de intervalos, de igual amplitud. Lo que cambia entre ambos es el punto de inicio del histograma. El resultado es que la imagen que proporciona cada uno sobre el intervalo que acumula una mayoría de los

valores es distinta. En el gráfico de la izquierda la mayor parte de los valores de la variable se acumula entre 9.000 y 12.000; en el de la derecha la percepción de agrupamiento es menor y el intervalo se desarrolla entre 8.500 y 12.500. (Municipios de Euskadi. Renta personal disponible en 2001)

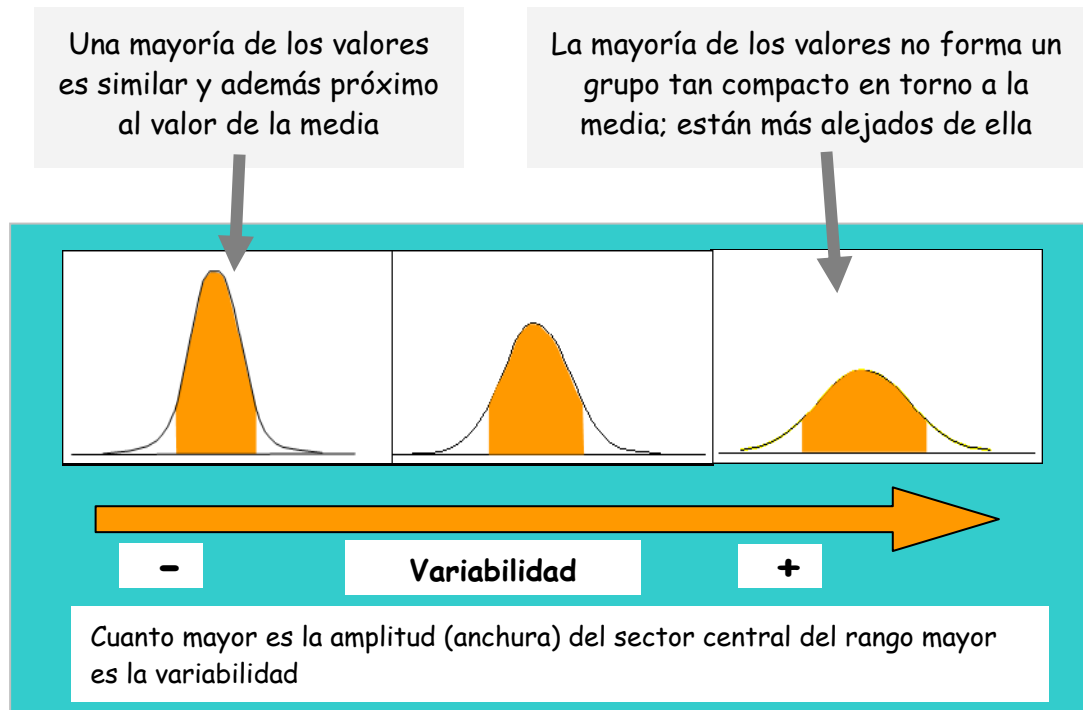
Para calcular el intervalo que agrupa una mayoría de los valores necesitamos un método que cumpla una serie de condiciones:

- ▣ Que su procedimiento de cálculo sea independiente de las decisiones que pueda tomar el usuario.
- ▣ Que, utilizados los mismos datos, proporcione siempre idénticos resultados.
- ▣ Que sea conocido, aceptado y utilizado por la comunidad científica.

La utilización de métodos estándar para el análisis estadístico permite no sólo que los resultados de nuestros estudios puedan ser comprendidos por todos aquellos que conocen los métodos sino que dichos resultados puedan ser comparados con los de otros estudios.

Antes de abordar la explicación sobre el procedimiento para calcular los límites del intervalo que agrupa la mayoría de los datos, es preciso introducir otra idea.

Aunque la concentración de valores en el sector central del rango es una característica de las distribuciones campaniformes, el grado de concentración puede ser muy variable. En algunas distribuciones una mayoría de los valores son muy semejantes y se encuentran muy próximos al valor de la media; en otras distribuciones el grupo mayoritario de valores no se sitúa tan cercano al valor central y conforma un intervalo más amplio dentro del rango.



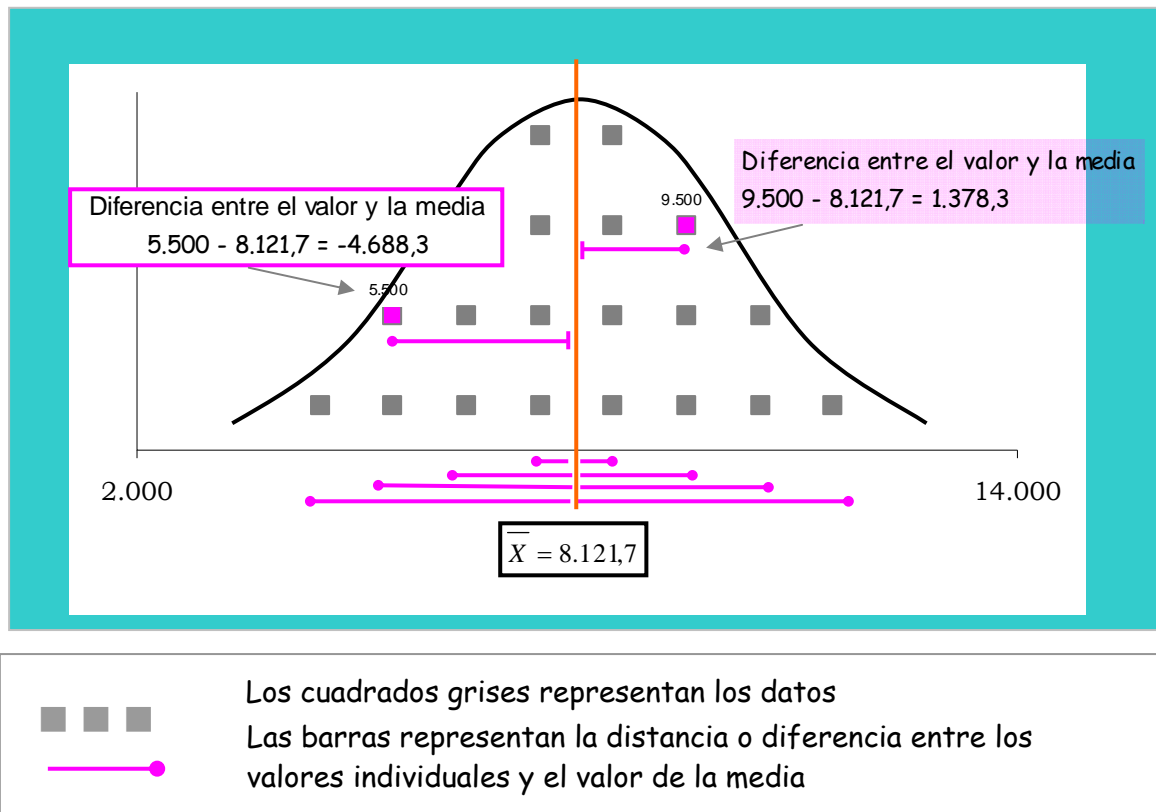
Las diferencias en el grado de concentración de los datos son en realidad diferencias en la variabilidad. De hecho, cuando intentamos determinar el sector del rango en el que se concentran buena parte de los valores de la variable, lo que hacemos en realidad es una valoración de la variabilidad que presenta esa mayoría de valores que se agrupa, en mayor o menor medida en torno a la media.

El concepto de desviación estándar y su procedimiento de cálculo

La desviación estándar es un procedimiento estadístico que se utiliza para determinar en qué medida se aleja la mayoría de los datos de la media. En el lenguaje estadístico, para referirnos a este alejamiento o distancia que existe entre los valores y la media, se utiliza el concepto de dispersión. Cuando los valores están alejados de la media, repartidos por todo el rango, se dice que están muy dispersos o que la dispersión es muy elevada. Se entiende así que la desviación estándar forme parte de un conjunto de herramientas que se conocen con el nombre genérico de medidas de dispersión.

Lo esencial, una vez más, es entender la idea o el principio que subyace al procedimiento de cálculo de la desviación estándar: la medida de dispersión de

los valores de la variable en torno a la media se basa en el cálculo de la distancia (diferencia) que existe entre los distintos valores y el valor de la media. A mayor lejanía, mayor dispersión y viceversa.



Siguiendo el principio mencionado, mediante la desviación estándar lo que se calcula es la diferencia entre cada uno de los valores de la variable y el valor de la media.

Este procedimiento nos proporciona un conjunto de valores y lo que a nosotros nos interesa es disponer de un valor global de la diferencia o distancia que hay entre los valores de la variable y la media. Para lograrlo lo que se hace es calcular un valor medio de las distancias de todos los valores a la media.

Para ganar claridad en la explicación expondremos el procedimiento de cálculo mediante un ejemplo. Utilizaremos para ello los datos del precio de la vivienda

en distintas ciudades de Euskadi en 2007³. En la tabla que mostramos a continuación podemos ver el listado de las 13 ciudades escogidas con sus correspondientes precios. Puesto que el objetivo es calcular la diferencia entre los valores individuales y el valor de la media, hemos dividido las ciudades en dos grupos, separando así las ciudades en las que el precio de la vivienda es inferior al de la media y las ciudades en las que es superior a ésta.

Valores de la variable: precio del m ² en viviendas nuevas, €/m ²				
Valores inferiores a la media: diferencias negativas		Valores superiores a la media: diferencias positivas		
Basauri	- 2.120	Promedio 2.881,31	+ 2.916	Barakaldo
Hernani	- 2.255		+ 2.988	Gasteiz
Erandio	- 2.571		+ 3.097	Irun
Arrasate	- 2.577		+ 3.268	Bilbo
Portugalete	- 2.753		+ 3.272	Getxo
Santurtzi	- 2.781		+ 4.061	Donostia
Leioa	- 2.798			

A continuación calcularemos las diferencias individuales entre los valores y la media. Los valores que están a la izquierda del promedio tendrán diferencias negativas con respecto a éste; los que están a la derecha, tendrán diferencias positivas.

³ Precio medio nominal, en euros por metro cuadrado, de la vivienda en junio de 2007. Sociedad de Tasación, SA. <http://web.st-tasacion.es/html/index.php>

Diferencias negativas		Diferencias positivas	
Ciudades	$x_i - \bar{x}$	$x_i - \bar{x}$	Ciudades
Basauri	-761,31	34,69	Barakaldo
Hernani	-626,31	106,69	Gasteiz
Erandio	-310,31	215,69	Irun
Arrasate	-304,31	386,69	Bilbo
Portugalete	-128,31	390,69	Getxo
Santurtzi	-100,31	1.179,69	Donostia
Leioa	-83,31		
$\sum x_i - \bar{x}$	-2.314,1	2.314,1	

Como decíamos anteriormente, el cálculo de las diferencias entre los valores de la variable y la media nos proporciona un nuevo listado de datos. Sin embargo, repetimos que lo que nos interesa a nosotros es disponer de un valor global, de una única medida de dispersión. Lo ideal sería poder calcular el valor medio de todas las diferencias individuales. Sería lo ideal pero no es posible porque si sumamos todas las diferencias el resultado es 0.

$$\begin{array}{r}
 \boxed{\text{Diferencias negativas}} + \boxed{\text{Diferencias positivas}} = 0 \\
 \boxed{-2.314,1} + \boxed{2.314,1} = 0
 \end{array}$$

La suma de las diferencias entre los valores de la variable y el promedio, en este caso, y en todos, es siempre cero, ya que se trata de una propiedad de la media aritmética:

$$\sum (x_i - \bar{x}) = 0$$

Una vez que hemos comprendido la imposibilidad de calcular el promedio de las diferencias entre los valores de la variable y el promedio, nos resultará más sencillo entender el procedimiento estadísticos existentes para obtener un valor medio de las diferencias.

La estadística nos ofrece al menos dos alternativas para obtener un valor medio de la distancia de los valores de la variable a la media. Una de las alternativas es utilizar lo que se denomina *Desviación media*. Su cálculo consiste en eliminar el signo de las diferencias entre los valores y el promedio. Al eliminar el signo se suman las cantidades absolutas⁴ y se calcula el promedio:

$$D\bar{x} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Aunque el cálculo de la desviación media es muy sencillo, en la práctica se usa muy poco como medida de desviación precisamente porque exige trabajar con valores absolutos.

Cálculo de la *Desviación media* aplicada a los datos sobre el precio medio nominal de la vivienda en 13 ciudades de Euskadi:

⁴ El valor absoluto de un número real es su valor numérico, prescindiendo del signo.

$$\frac{\sum_{i=1}^n x_i - \bar{x}}{n} = \frac{4.628,31}{13} = 356,02$$

La segunda alternativa, y en la práctica la más utilizada, es la *desviación estándar*. En este caso también el procedimiento se inicia mediante el cálculo de las diferencias entre los valores de la variable y el promedio. A partir de aquí, la solución para evitar que se compensen las diferencias positivas y las negativas consiste en elevar las diferencias al cuadrado. Desaparecen de este modo los valores negativos y es posible entonces calcular el promedio de las diferencias.

Para ilustrar el procedimiento continuaremos analizando los datos sobre el precio de la vivienda:

Ciudades de Euskadi	Valores de la variable	Diferencias entre los valores y el promedio	Cuadrado de las diferencias
	x_i	$(x - \bar{x})$	$(x - \bar{x})^2$
Basauri	2.120	-761,31	579.592,92
Hernani	2.255	-626,31	392.264,22
Erandio	2.571	-310,31	96.292,30
Arrasate	2.577	-304,31	92.604,58
Portugalete	2.753	-128,31	16.463,46
Santurtzi	2.781	-100,31	10.062,10
Leioa	2.798	-83,31	6.940,56
Barakaldo	2.916	34,69	1.203,40
Gasteiz	2.988	106,69	11.382,76
Irun	3.097	215,69	46.522,18
Bilbo	3.268	386,69	149.529,16
Getxo	3.272	390,69	152.638,68
Donostia	4.061	1.179,69	1.391.668,50
$\sum_{i=1}^n (x_i - \bar{x})^2$			2.947.164,77

Después de calcular las diferencias, elevarlas al cuadrado y realizar la suma, calculamos el promedio:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{2.974.164,77}{13} = 226.704,98 \text{ €}$$

Al hacer el promedio del cuadrado de las desviaciones para el ejemplo, nos damos cuenta de que obtenemos un resultado cuyo valor no guarda relación alguna con los valores de la variable. Mientras el precio por metro cuadrado de las viviendas estudiadas oscila entre 2.120 y 4.061 €, el valor de la *desviación estándar* se eleva hasta 226.704, 98 €. La razón es evidente: se trata de una cifra tan alta porque hemos elevado al cuadrado las diferencias entre los valores individuales y el valor de la media. Para compensar la elevación obtenemos la raíz cuadrada del promedio:

$$S = \sqrt{\frac{2.947.164,77}{13}} = 476,14$$

La última operación que hemos realizado, que nos proporciona un resultado cuya cifra es acorde al precio de la vivienda en las ciudades estudiadas, constituye el último paso en el cálculo de la desviación estándar, cuya fórmula queda como sigue:

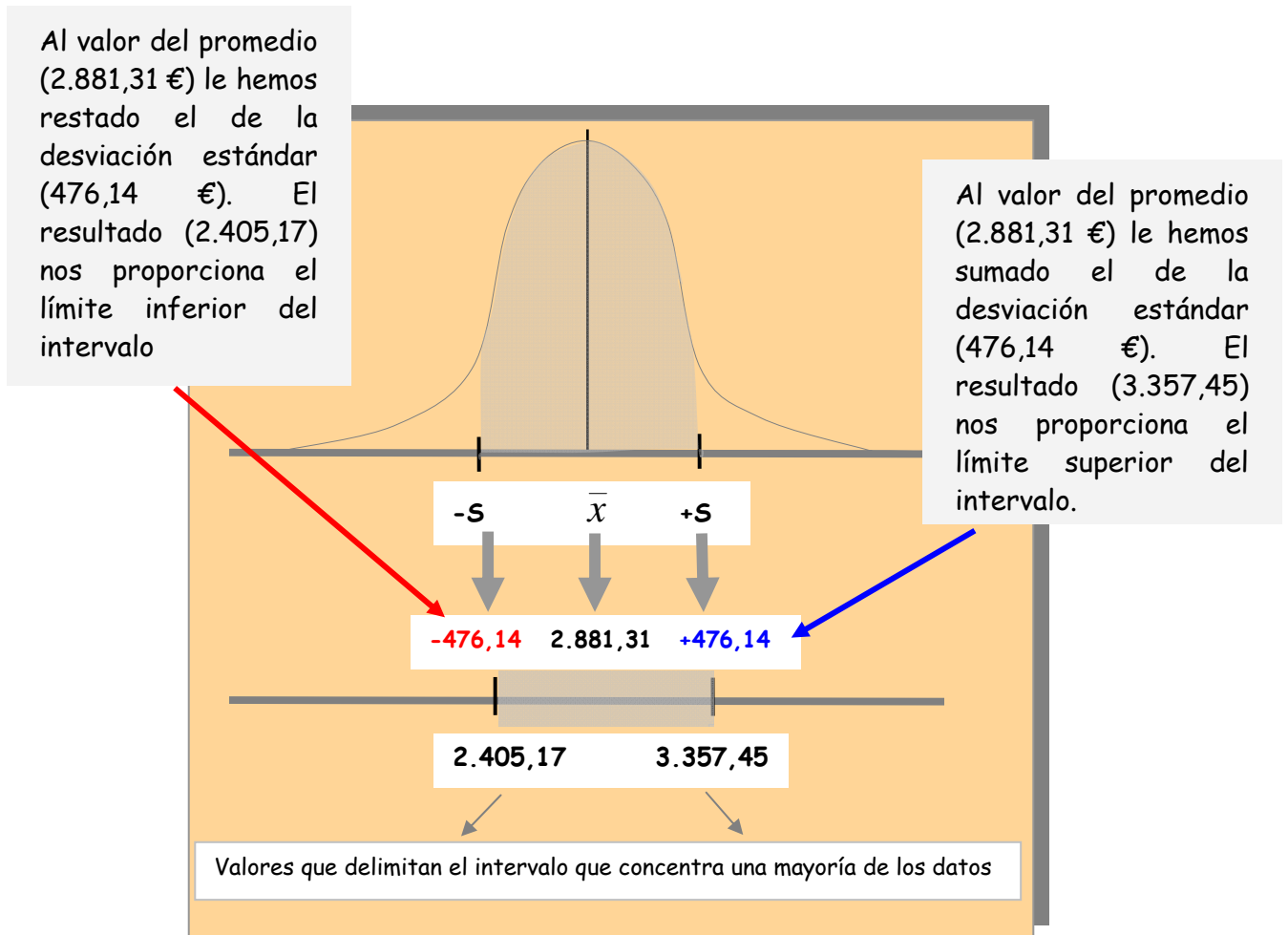
$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

La utilización de la desviación estándar

Al principio del apartado decíamos que nuestro objetivo era buscar el modo de calcular los dos puntos del rango entre los que se concentra una gran parte de los valores de la variable. Hemos dicho también que para calcular esos dos puntos se utiliza la *Desviación estándar* y hemos mostrado cómo se calcula. Ahora sólo nos falta explicar cómo se utiliza el valor de la *desviación estándar* para calcular el intervalo que concentra una mayoría de los valores de la variable.

Si aceptamos que la *desviación estándar* es una medida del grado en que los valores de la variable se alejan del promedio, nos resultará sencillo entender que una forma de calcular el intervalo que concentra una mayoría de los valores

consiste en sumar y restar al promedio el valor de la *desviación estándar*. Los dos valores resultantes de estas operaciones son los límites del intervalo que buscamos. Para ilustrar el procedimiento seguiremos utilizando los datos sobre el precio medio de la vivienda en trece ciudades de Euskadi:



En la ilustración aparecen reflejadas las operaciones que hemos realizado :

Una vez calculados los límites del intervalo estamos en situación de afirmar que en una mayoría de las ciudades analizadas el precio de la vivienda por metro cuadrado oscila entre los 2.405,17 y los 3.357,45 euros.

Todavía existe otra forma más de utilizar el valor de la desviación estándar para describir la variabilidad de los valores de una variable. En lugar de calcular los límites del intervalo que agrupa una mayoría de los valores, podemos mencionar directamente el resultado de la desviación estándar como medida

del alejamiento o dispersión de los valores con respecto a la media. Aplicado al caso de los precios de la vivienda, podríamos afirmar lo siguiente: la variabilidad con respecto a la media del precio medio de la vivienda en trece ciudades de Euskadi es de 476,14 €. Esto significa que los valores de la variable se dispersan una media de 476,14€ con respecto a la media aritmética.

A estas alturas resulta evidente que, cuanto mayor es el resultado de la desviación estándar, mayor es el grado de variabilidad o de dispersión de los valores con respecto al promedio. Ahora bien, no olvidemos que, para calificar la dispersión de alta o de baja, no debemos fijarnos en el valor en bruto de la desviación sino compararlo con el valor del promedio. Una dispersión, supongamos, de 500 € referida a los precios de ropa de abrigo cuyo promedio es de 800 €, es, sin lugar a duda, un valor muy elevado. El mismo valor de dispersión referido al precio de vehículos cuyo promedio es de 24.000 €, es, por el contrario, muy bajo.

De todos modos sólo a veces es posible calificar de alto o bajo un valor de desviación estándar. En la mayoría de situaciones no disponemos de referentes que nos permitan añadir un calificativo de grado al valor de la desviación. Fijémonos en los resultados obtenidos para los datos del precio medio de la vivienda en las ciudades de Euskadi. ¿Cómo podríamos calificar el resultado de la desviación (476,14€) en relación al valor del promedio (2.881,31 €)?

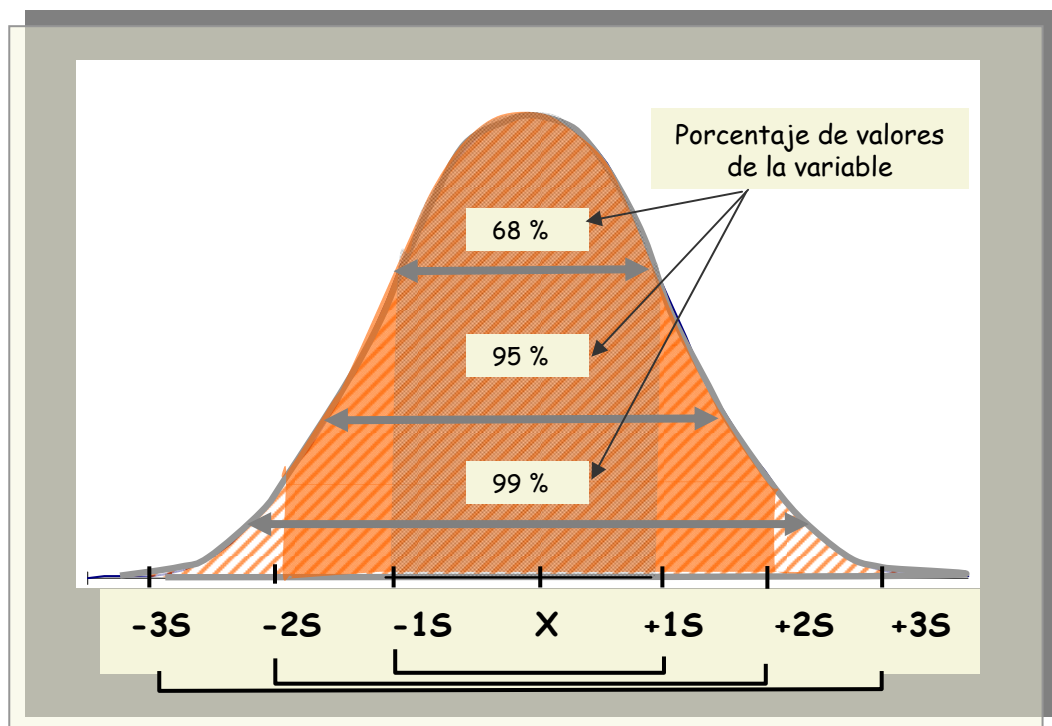
Las profesoras Neer y Lopetegui proporcionan comentarios interesantes en relación a la valoración de la desviación estándar:

Solemos preguntarnos cuándo una desviación es grande y cuándo pequeña. Por un lado la interpretación de grande o pequeña depende de la variable que estemos analizando, por ejemplo una desviación de 5 será grande si estamos hablando de la dispersión de la edad en que los niños aprenden a leer y será pequeña si nos estamos refiriendo a la dispersión del salario de los habitantes de La Plata. Por otro lado grande o pequeño tienen un significado relativo, es más grande o más pequeño que el encontrado para algún otro grupo o algún otro test. (Neer, Lopetegui, 2003)

Tal como hemos explicado, a la hora de elaborar una descripción de la variabilidad de los datos podemos utilizar el valor de la desviación estándar para calcular el intervalo de agrupamiento de valores o como un valor medio de variabilidad o dispersión. Podemos también utilizar ambos modos en distintos momentos de la explicación. En cualquier caso, sea cual fuere la forma

de expresión que adoptemos, es importante que incluyamos en nuestro trabajo una pequeña tabla con los resultados de todas las medidas y técnicas que hemos utilizado y, dentro de ella junto al promedio, el resultado de la desviación estándar.

Cuando en los resultados de un estudio proporcionamos el valor de la desviación estándar o los límites del intervalo $\bar{X} \pm S$, estamos afirmando, aunque no sea de forma explícita, que una gran mayoría de los valores de la variable estudiada se encuentra dentro de los límites de dicho intervalo. El posible lector de nuestro trabajo sabe que esto debe ser así. La razón es que en todas las distribuciones campaniformes se cumple la condición de que esa mayoría de valores de la que hablamos se sitúa en el intervalo de la media +/- la desviación. De hecho, las distribuciones campaniformes cumplen las siguientes condiciones:



El 68% de los valores de la variable se encuentra dentro del intervalo formado por la media +/- la desviación.

El 95% de los valores de la variable se encuentra dentro del intervalo formado por la media +/- dos veces el valor de la desviación.

- El 99% de los valores de la variable se encuentra dentro del intervalo formado por la media +/- tres veces el valor de la desviación.

Otros usos de la desviación estándar

El valor de la desviación estándar se utiliza frecuentemente también como medida para identificar valores de la variable especialmente alejados del promedio, es decir, notablemente más altos o más bajos que aquel. El objetivo suele ser destacar la existencia de algunos elementos de la población cuyo comportamiento respecto a la variable estudiada se pueda considerar especial. En este caso no estaríamos hablando de los valores extremos (outliers) a los que nos referíamos en el tema tres.

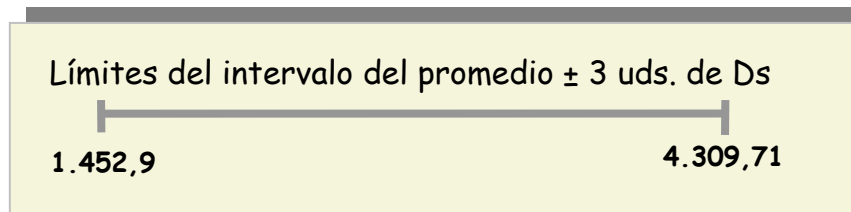
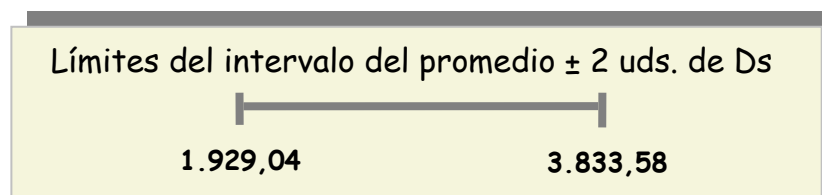
El procedimiento al que nos referimos consiste en sumar y restar al promedio el valor de la *desviación estándar* multiplicado por dos o por tres. Mediante la suma y la resta obtenemos los límites de un intervalo; estos límites se utilizan para separar los valores que se consideran especialmente altos o bajos. Los límites del intervalo marcan los puntos del rango a partir de los cuales se sitúan los valores destacables.

Una vez explicado el procedimiento es preciso aclarar la forma correcta de aludir al mismo. Cuando se utiliza la doble desviación estándar hablamos del intervalo que surge del promedio más/menos *dos unidades de desviación estándar* y lo expresamos como sigue: $\bar{X} \pm 2S$. Siguiendo esta lógica, cuando utilizamos la triple desviación hablamos del intervalo que surge del promedio más/menos *tres unidades de desviación estándar* y lo expresamos así: $\bar{X} \pm 3S$

Calcularemos ahora los intervalos que surgen del promedio más/menos *dos y tres unidades de desviación estándar* para los datos referentes al precio medio de la vivienda en las trece ciudades de Euskadi:

Promedio	2.281,31
Desviación estándar	476,14

	Promedio	Unidades de desviación estándar					Resultado		
			1		2			3	
$\bar{X} \pm 2S$	2.881,31	-	476,14	-	476,14	=	1.929,04		
		+	476,14	+	476,14	=	3.833,58		
$\bar{X} \pm 3S$	2.881,31	-	476,14	-	476,14	-	476,14	=	1.452,90
		+	476,14	+	476,14	+	476,14	=	4.309,71



▶ En ninguna de las ciudades analizadas el precio medio de la vivienda supere los límites del intervalo $\bar{X} \pm 3S$.

▶ En una de las ciudades analizadas, En Donosti, el precio medio de la vivienda supera el limite superior del intervalo $\bar{X} \pm 2S$.

Desde el principio hemos visto que el precio medio de la vivienda en Donosti, 4.061€/m², es muy superior al del resto de las ciudades analizadas. Utilizar el

intervalo $\bar{X} \pm 2S$ puede ser una forma idónea de describir hasta qué punto es elevado dicho precio.

Son muchos los estudios en los que se quiere particularizar la situación de los valores situados en los extremos del rango y que utilizan para ello los límites del intervalo $\bar{X} \pm 2S$ o $\bar{X} \pm 3S$: los elementos de la población con valores por encima o por debajo tendrán una consideración especial.

En un informe de septiembre de 2008, el Servicio Meteorológico Nacional mexicano, para destacar el elevado número de tormentas intensas ocurridas de enero a agosto de 2008, se expresaba en los siguientes términos:

Durante el mes de agosto de 2008 se registraron a nivel nacional un total de **126 tormentas intensas** superando el promedio de 85.75 tormentas en el mes. El record de agosto de 2008, representa un valor superior al promedio más una desviación estándar. (CONAGUA, 2008)

En este caso, los autores del estudio han utilizado el intervalo del promedio más/menos una unidad de desviación estándar. Aunque es más frecuente que se utilicen los intervalos de dos o tres unidades de desviación, cuando el valor de la desviación es elevado, el intervalo del promedio más/menos una unidad de desviación también es muy amplio y se usa entonces como límite para detectar valores especialmente elevados.

En el Atlas de la industria en la Comunidad de Madrid, podemos ver un ejemplo concreto en el que el límite definido por el promedio más dos unidades de desviación estándar se ha aplicado para detectar los municipios especializados en actividades industriales concretas:

CÁLCULO DE LOS VALORES DE ESPECIALIZACIÓN

El nivel de especialización en las diferentes actividades de la industria manufacturera se ha calculado mediante el índice de Nelson. El índice se basa en las propiedades de la desviación típica como medida de dispersión de los valores de una distribución, para discriminar aquellos que sobrepasan anormalmente determinados umbrales. Utilizando los sectores urbanos como unidad de referencia espacial, se obtiene en primer lugar el porcentaje de cada una de las once actividades industriales. A continuación se calculan para cada una la media aritmética y desviación típica. Los valores límite para considerar una zona como especializada se definen como el valor medio más una, dos, tres o más desviaciones típicas. En nuestro caso hemos elegido como umbral

mínimo el valor medio más dos desviaciones típicas. Puede ocurrir que algunas zonas no sobrepasen los valores límite establecidos, y se consideren que no están especializados en ninguna de las actividades. También pueden darse zonas especializadas en dos o más actividades. En este caso se ha asignado la actividad de mayor grado de especialización.

Debe tenerse en cuenta que esta forma de cálculo de especialización funcional es relativa, en el sentido de que no da una medida de la importancia o peso total de una actividad, sino la importancia en relación al resto de las zonas consideradas. (Comunidad de Madrid, 2007)

Tal como explica el texto, el punto de partida ha sido el cálculo, para cada área urbana de la Comunidad de Madrid de la importancia, medida en el porcentaje que representa cada una de las once ramas industriales con respecto al total de la industria en dicha área. Este cálculo produce una estructura de datos similar a la que mostramos a continuación:

Áreas urbanas	Ramas de actividad industrial										
	1	2	3	4	5	6	7	8	9	10	11
A	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
B	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
C	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11

En la primera línea tendríamos los porcentajes de cada una de las actividades industriales en el área urbana A. La suma de los porcentajes en las 11 ramas de actividad (de A1 a A11), constituye el 100%.

Si lo miráramos por columnas, en la número 1 estarían los porcentajes de la rama industrial 1 de todas las áreas urbanas de la Comunidad de Madrid. La suma de los porcentajes de la actividad industrial nº 1 en todas las áreas urbanas (A1, B1, C1, etc.) constituye el 100%.

Partiendo de los porcentajes obtenidos, los autores del atlas calcularon la media y la desviación estándar de los datos por columnas, es decir, de los porcentajes por rama de actividad industrial.

Áreas urbanas	Ramas de actividad	
	1	2
A	A1	A2
B	B1	B2
C	C1	C2

Se calcula el promedio y la desviación estándar de los valores, en porcentajes, para cada una de las ramas de actividad industrial: A1, B1, C1, etc.

Una vez obtenidos estos datos pudieron calcular el valor del promedio más dos unidades de desviación estándar. El valor definido sirvió a modo de umbral a partir del cual se podía considerar que un área urbana estaba especializada en determinada rama de actividad industrial. Se pueden considerar especializadas en una rama concreta de actividad industrial aquellas áreas urbanas en las que el porcentaje de dicha rama de actividad supere el valor establecido por el promedio más dos unidades de desviación estándar.

La zona A se considera especializada en la rama de actividad 1 cuando su porcentaje de actividad industrial en dicha rama sea superior al promedio (de A1, B1, C1, etc.) más dos unidades de desviación estándar

En un estudio sobre la evolución de la precipitación anual en Andalucía, su autor utiliza las unidades de desviación estándar, sumadas o restadas al promedio, para generar una clasificación del grado de humedad de una serie de años en distintas regiones pluviométricas de Andalucía.

Como vemos en la tabla adjunta, para clasificar los años con una humedad superior a la media utiliza los límites que marcan el promedio más una, dos o tres unidades de desviación estándar.

- ▣ Los años cuya precipitación supera el límite del promedio más dos unidades de desviación estándar forman la clase de años extremadamente húmedos.
- ▣ El año se considera hiperhúmedo si la precipitación recogida supera el límite formado por el promedio más tres unidades de desviación estándar.
- ▣ En el extremo opuesto, para clasificar los años especialmente secos ha creado un único umbral, definido por el promedio menos una unidad de desviación estándar. (Castillo Requena, 2000)

AÑO MUY SECO	$R \text{ año} < \bar{X} - S$
AÑO SECO	$R \text{ año} < \bar{X} - 1/4 S \quad y \quad > \bar{X} - S$
AÑO NORMAL/SECO	$R \text{ año} < \bar{X} \quad y \quad > \bar{X} - S$
AÑO NORMAL/HÚMEDO	$R \text{ año} < \bar{X} \quad y \quad > \bar{X} + S$
AÑO HÚMEDO	$R \text{ año} < \bar{X} \quad y \quad > \bar{X} + S$
AÑO MUY HÚMEDO	$R \text{ año} < \bar{X} + 1/4 S \quad y \quad > \bar{X} + S$
AÑO EXTREMADAMENTE HÚMEDO	$R \text{ año} < \bar{X} + S \quad y \quad > \bar{X} + 2S$
AÑO HIPERHÚMEDO	$R \text{ año} < \bar{X} + 3S$

R: Media de la región pluviométrica

Para terminar con las explicaciones sobre la desviación estándar haremos un resumen de las características fundamentales de esta técnica:

- ❖ La desviación estándar (o desviación típica) es una herramienta para obtener una medida de la variabilidad de una serie de datos. Dentro de las herramientas estadísticas se incluye en la categoría de técnicas destinadas a la medición de la dispersión de los datos.
- ❖ La desviación estándar se puede utilizar como complemento al promedio, con el fin de proporcionar una medida de dispersión de los datos. Se puede utilizar también para calcular los límites del intervalo que concentra una mayoría de los valores. En cualquier caso, el valor de la desviación estándar es un complemento obligado del promedio.
- ❖ Como medida de dispersión el valor de la desviación estándar se interpreta en relación al valor del promedio.
- ❖ La desviación estándar se expresa en las mismas unidades de medida que la variable que analizamos. (Si la variable se refiere al precio en euros de la vivienda, la desviación estándar nos da un valor de la dispersión, también en euros)

El coeficiente de variación

Como hemos visto, la desviación estándar mide la desviación de los valores de la variable en las mismas unidades de la variable. En el ejemplo que veíamos sobre el precio del metro cuadrado de las viviendas en las distintas ciudades de Euskadi, hemos obtenido un precio medio de 2.281,31 euros por metro cuadrado. El resultado de la desviación estándar -476,14- también está expresado en euros por metro cuadrado. Esta característica de la desviación estándar resulta muy cómoda a la hora de interpretar la variabilidad de una distribución. Sin embargo no resulta útil cuando queremos comparar el grado de variabilidad de dos o más distribuciones. Veremos mediante un ejemplo cuál es el problema.

En la tabla siguiente podemos ver el precio por metro cuadrado de vivienda, correspondiente al año 2007, en los principales municipios de dos comunidades autónomas.

Ciudades de Euskadi	Precio de la vivienda €/m ²	Ciudades de Galicia	Precio de la vivienda €/m ²
Basauri	2.120	Monforte de Lemos	898
Hernani	2.255	O Barco de Valdeorras	1.080
Erandio	2.571	Ames	1.206
Arrasate	2.577	Ferrol	1.391
Portugalete	2.753	Pontevedra	1.477
Santurtzi	2.781	Viveiro	1.526
Leioa	2.798	Lugo	1.540
Barakaldo	2.916	Ourense	1.742
Gasteiz	2.988	Santiago de Compostela	1.786
Irun	3.097	Coruña	1.999
	Bilbo	3.268	
	Getxo	3.272	
	Donostia	4.061	

Puesto que los datos de la tabla están ordenados es fácil darse cuenta de que la variabilidad entre los municipios en las dos comunidades es importante. En las dos comunidades autónomas el precio de la vivienda en las ciudades más caras es el doble, o casi el doble, del precio en las ciudades más baratas. Ahora bien: ¿en cuál de las dos comunidades autónomas es mayor la variabilidad?. Con el fin de intentar dar respuesta a la pregunta calcularemos los promedios y las desviaciones de los valores de la variable en cada comunidad.

	Euskadi	Galicia
Promedio	2.881,31	1.513
Desviación estándar	476,14	340,51

Las diferencias entre las dos comunidades son importantes, fundamentalmente en sus promedios. El País Vasco es la comunidad con el promedio más elevado

y también con la mayor desviación estándar. La cuestión es ahora decidir si con estos datos podemos saber en cuál de las dos comunidades es mayor la variación entre municipios.

A primera vista podríamos pensar que la mayor desviación estándar se corresponde con la mayor variabilidad y que es el País Vasco la comunidad con mayor variación interna de los precios; tal conclusión sería un error. Recordemos que la variabilidad medida mediante la desviación estándar se corresponde con la distancia o diferencia media que existe entre el promedio y los valores de la variable. Cuanto más elevados sean los valores de la variable más elevado será, en general, el valor del promedio y el de la desviación estándar.

Lo que ocurre en el ejemplo que hemos puesto es que los precios en el País Vasco son mucho más elevados que los de Galicia. Consecuentemente, los valores del promedio y de la desviación en el País Vasco tenderán a ser mayores que los de Galicia, incluso aunque la dispersión de los datos sea menor.

La conclusión parece clara: el valor de la desviación estándar no sirve para comparar el grado de variación interna de dos o más series de datos si los valores de dichas series no son de magnitud muy similar.

Existe, sin embargo, una herramienta que sí nos permite hacer comparaciones entre la variabilidad de diferentes conjuntos de datos, aunque los valores de unos sean mucho mayores o menores que los de los otros. La herramienta a la que nos referimos es el coeficiente de variación. Se trata de una herramienta de fácil manejo mediante la cual se calcula la magnitud de la desviación estándar en relación a la magnitud del promedio.

$$cv = \frac{S}{\bar{X}} \cdot 100$$

El coeficiente de variación se calcula mediante una simple división entre el promedio y la desviación estándar. El resultado se multiplica por 100 y se obtiene así el porcentaje que representa la desviación típica con respecto a la media

Podemos calcular ahora los coeficientes de variación para los valores del precio de la vivienda en las dos comunidades y comprobaremos que no es el País Vasco la comunidad en la que existe mayor variación de los precios entre municipios:

$$CV_{\text{País Vasco}} = \frac{S}{\bar{x}} = \frac{476,14}{2.881,31} = 0,1652 \rightarrow 16,52\%$$

$$CV_{\text{Galicia}} = \frac{S}{\bar{x}} = \frac{340,51}{1.513} = 0,2251 \rightarrow 22,51\%$$

Los resultados del coeficiente de variación muestran que, para los datos del País Vasco, el valor de la desviación estándar supone un 16,52% del valor de la media. En el caso de Galicia, el valor de la desviación estándar es mayor y supone concretamente un 22,51% del valor de la media.

A la vista de los resultados del coeficiente de variación podemos afirmar que, de las dos comunidades autónomas, es la de Euskadi la que presenta una menor dispersión o variabilidad de los valores en torno a la media. Es cierto que, de media, los pisos son más baratos en Galicia pero también es cierto que el precio presenta mayores variaciones entre las ciudades gallegas que entre las vascas.

En el estudio que hemos mencionado anteriormente sobre la evolución de la precipitación anual de Andalucía, su autor utiliza también el coeficiente de variación para destacar la elevada variabilidad interanual de las precipitaciones en Andalucía, en comparación con el resto de España.

	Precipitaciones anuales		
	Media	Desviación estándar	Coefficiente de variación
Cuenca del Guadalquivir	600,47	180,45	30,05
Cuenca Sur	537,82	165,72	30,81
Unidad Central	605,26	127,55	21,07
Fachada Mediterránea	527,69	116,60	22,10
España peninsular	667,09	119,25	17,88

Pese a tener un valor de desviación superior al de la Cuenca Sur, el coeficiente de variación es menor

La variabilidad de la precipitación en la España peninsular es notablemente inferior a la de las regiones andaluzas

Castillo Requena, J.M. (2000)

Tema 6:

Análisis de los extremos de la distribución. La simetría

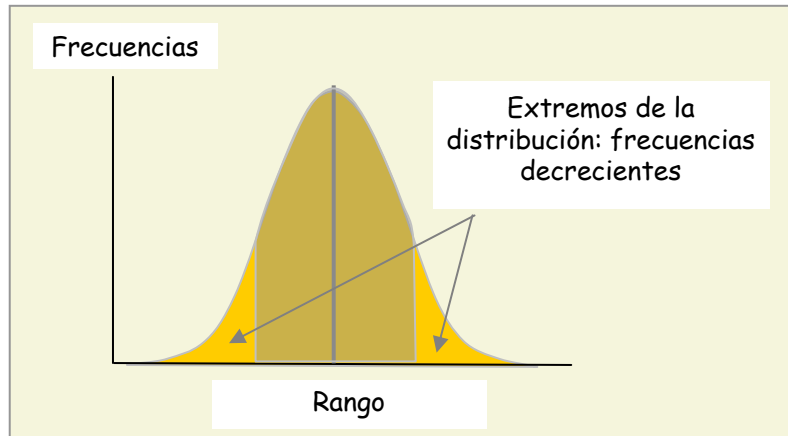
El concepto de simetría	3
Análisis gráfico de la simetría	4
Qué aporta el análisis de simetría	10
Cuantificación de la simetría	12

Tema 6:

Análisis de los extremos de la distribución. La simetría

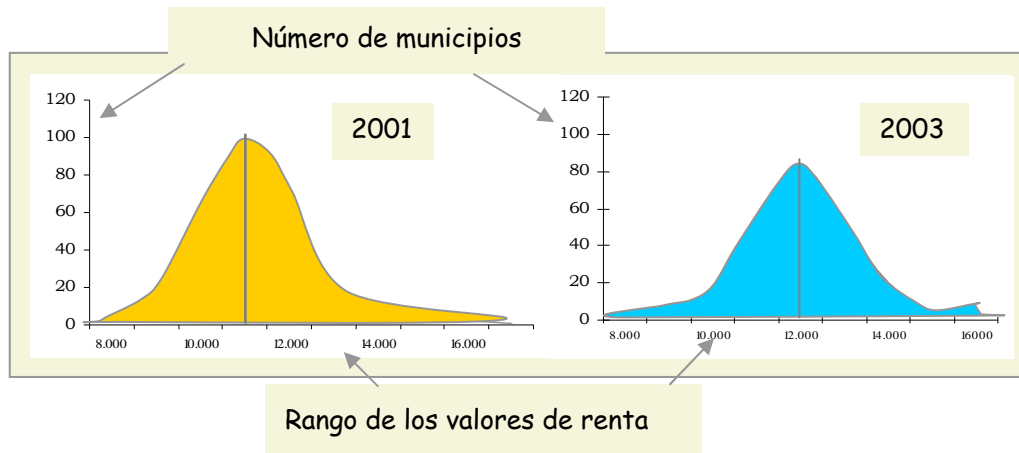
En el tema anterior hemos explicado el procedimiento para analizar el sector central de la distribución. Hemos visto que para realizar dicho análisis se utiliza la media aritmética junto con la desviación estándar. En este tema nos centraremos en el procedimiento para estudiar los lados de la distribución, para lo cual se utiliza el análisis de simetría.

Ya hemos visto, en repetidas ocasiones, que en las distribuciones campaniformes los extremos de la distribución representan al conjunto de los valores más altos y más bajos de la variable. Se trata de dos grupos porcentualmente minoritarios cuyas bajas frecuencias producen un descenso más o menos brusco de la curva de frecuencias:



Aunque esta es la forma prototípica de las distribuciones campaniformes -su representación ideal- muchas de las variables que se ajustan a este tipo de distribución presentan ciertas variantes o diferencias con respecto a ella.

EL CONCEPTO DE SIMETRÍA



Los polígonos de frecuencias de la ilustración se corresponden con la distribución de la variable¹ Renta personal disponible en los municipios de Euskadi. El polígono de frecuencias de la izquierda está elaborado con los datos de 2001 y el de la derecha con los datos de 2003.

Podemos observar que las dos distribuciones son de tipo campaniforme: ambas tienen un sector central que agrupa una mayoría de los datos; desde el centro las frecuencias van disminuyendo progresivamente a medida que desciende el número de municipios con rentas particularmente altas o bajas.

Las dos distribuciones son de tipo campaniforme pero presentan entre ambas diferencias de forma evidentes.

- En la gráfica de la izquierda la forma de la distribución a cada uno de los lados del eje central es diferente. El lado derecho es más ancho que el izquierdo y, sobre todo, en el tramo final de la curva se produce un alargamiento pronunciado de la misma que no está ni en su lado izquierdo ni en la distribución de los datos de 2003. Debido a la diferencia de forma entre los dos lados de la distribución, decimos que esta distribución es asimétrica.

¹ Ver el listado original de valores en el apartado de ANEXOS

- ▣ En la gráfica de la derecha los dos lados de la distribución son muy similares y, por ello, decimos que es simétrica.

Para entender el concepto de simetría acudiremos al diccionario:

Simetría: Arreglo equilibrado de partes de una figura en lados opuestos de un punto, línea, o plano. Los tipos más comunes incluyen la simetría con respecto a un punto, simetría con respecto a una línea y simetría rotacional.²

Geom. Correspondencia exacta en la disposición regular de las partes o puntos de un cuerpo o figura con relación a un centro, un eje o un plano³.

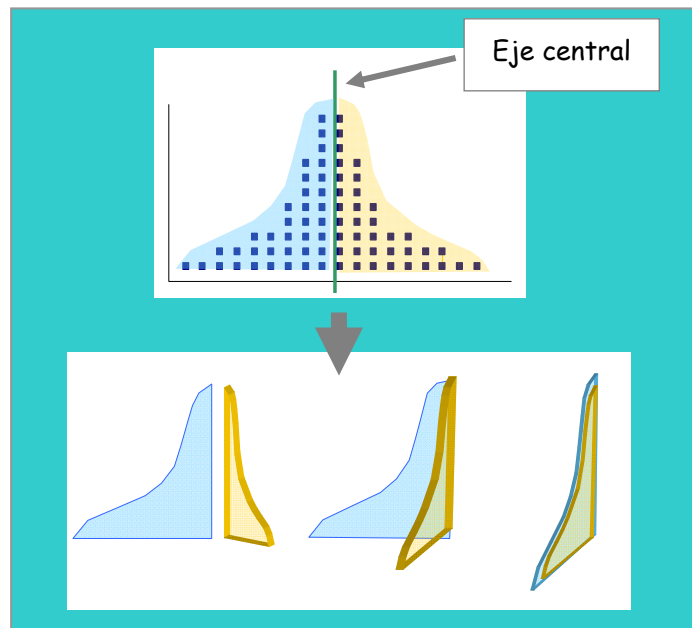
Vemos que se trata de un concepto perteneciente al ámbito de la geometría, que se utiliza en el análisis de datos y que resulta una valiosa herramienta para poner de relieve características significativas de las distribuciones.

ANÁLISIS GRÁFICO DE LA SIMETRÍA

Mediante la ilustración de la página anterior hemos visto a qué se refiere el concepto de *simetría* cuando hablamos de una distribución de valores de la variable. Podremos afirmar que una distribución es simétrica si al dividirla en dos a partir del eje de simetría, la forma de los dos lados es igual.

² <http://www.mathematicsdictionary.com/spanish/vmd/full/s/symmetry.htm>

³ Diccionario de la Real Academia Española de la lengua. RAE

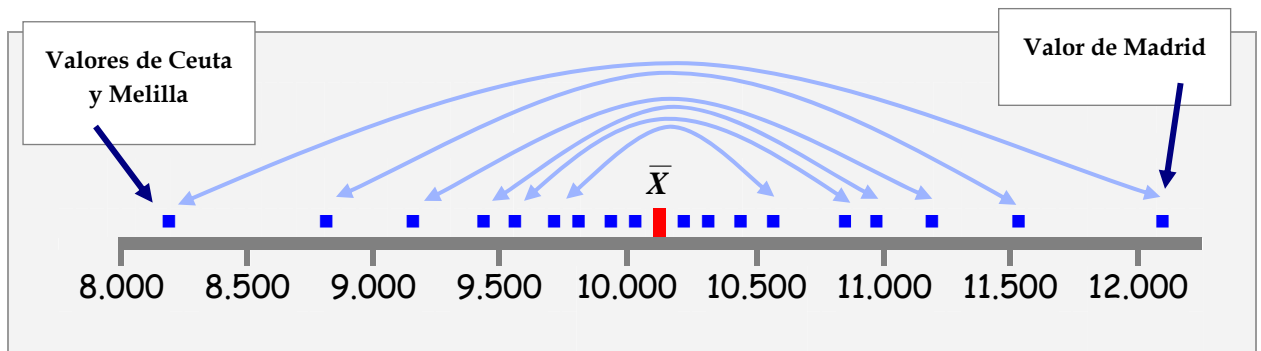


Cuando la distribución es simétrica, al doblarla por la mitad y juntar los dos lados de la distribución, ambos coinciden.

La similitud entre ambas partes nos permite identificar el carácter simétrico o asimétrico de una distribución. A este respecto, lo fundamental es tener claro lo que significa la simetría, o la asimetría, con respecto a un conjunto de datos y de qué modo podemos valorarla o cuantificarla.

El hecho de que los dos lados de una distribución sean iguales o similares significa que el número de valores a un lado y al otro del promedio es similar y se encuentran también a distancias similares de aquel. Esto significa a su vez que la dispersión de los datos a ambos lados del promedio es también similar.

Para verlo más claro, representaremos la distribución de un conjunto hipotético de datos sobre una recta que representa el rango. Podemos suponer que los datos se corresponden con la cantidad media de dinero que gastaron un grupo de personas en cada una de las comunidades autónomas del estado español, durante 2006.



La marca roja en la recta representa el promedio, cuyo valor es de 10.150,98 euros. Las marcas azules representan el valor del gasto medio por persona en cada comunidad. Las líneas azules unen los valores equidistantes del promedio a uno y otro lado del mismo.

En la tabla siguiente podemos la serie original de datos (ficticia). En ella también hemos enlazado mediante líneas azules algunas de las parejas de variables cuya distancia al promedio es igual.

Comunidades autónomas	Gasto €	Diferencia con respecto a la media
Ceuta y Melilla	8.190,13	-1.960,85
Extremadura	8.565,99	-1.359,99
Castilla - La Mancha	8.945,31	-980,67
Canarias	9.257,72	-740,26
La Rioja	9.320,97	-563,22
Murcia	9.391,17	-438,06
Andalucía	9.554,63	-312,90
Cantabria	9.668,21	-187,74
Castilla y León	10.000,27	-62,58
Galicia	10.083,93	62,58
Asturias	10.662,51	187,74
Aragón	10.684,19	312,90
Comunidad Valenciana	10.720,13	438,06
Islas Balears	11.208,57	563,22
Navarra	11.542,70	740,26
País Vasco	11.855,74	980,67
Cataluña	11.994,94	1.359,99
Madrid	12.111,83	1.960,85

Si observamos la ilustración precedente y la tabla de valores de la variable podemos comprobar que los que gastaron más que la media y los que gastaron menos lo hicieron en la misma medida: vgr. el grupo de la Comunidad de Madrid gastó 600,86 euros más que la media y el grupo de las ciudades de

Ceuta y Melilla gastó 600,86 euros menos que la media. Si comparamos ahora Cataluña con Extremadura, veremos que los grupos gastaron 379,32 euros más que la media el primero y menos que la media el segundo.

Si hacemos la comprobación con todos los valores de la tabla obtendremos 9 parejas de comunidades. Los dos valores de cada pareja se sitúan a la misma distancia del promedio, uno por encima y el otro por debajo.

Parejas de comunidades equidistantes		Distancia al promedio
Ceuta y Melilla	Madrid	± 1.960,85
Extremadura	Cataluña	± 1.359,99
Castilla - La Mancha	País Vasco	± 980,67
Canarias	Navarra	± 740,26
La Rioja	Islas Baleares	± 563,22
Murcia	Comunidad Valenciana	± 438,06
Andalucía	Aragón	± 312,90
Cantabria	Asturias	± 187,74
Castilla y León	Galicia	± 62,58

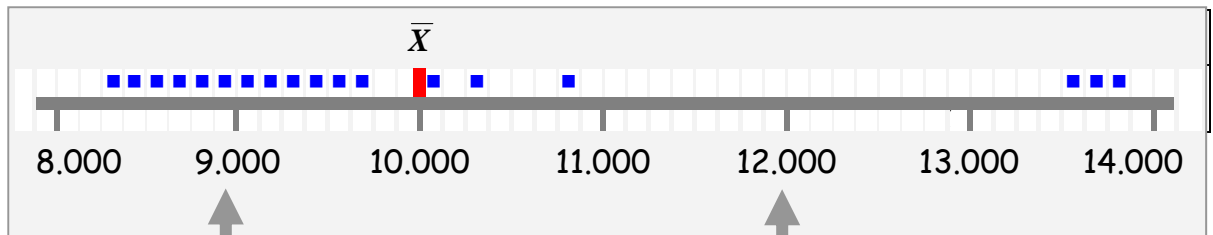
Finalmente, las conclusiones que se pueden obtener sobre la distribución de valores que hemos creado son las siguientes:

- ▣ Hay el mismo número de valores a la izquierda y a la derecha del promedio
- ▣ Los valores a ambos lados del promedio presentan la misma dispersión con respecto a éste.
- ▣ Se trata de una distribución totalmente simétrica.

Ahora bien, ¿qué significado tiene la simetría con respecto a la variable que estamos estudiando?

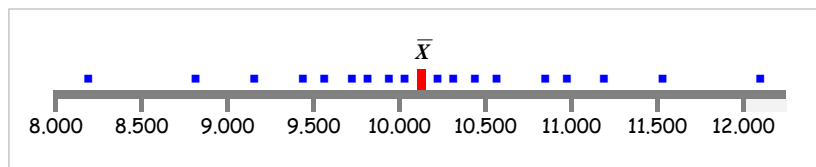
- ▣ Que no hay un grupo que destaque sobre el resto en su comportamiento con respecto al gasto. Quienes gastaron más y quienes gastaron menos estuvieron igual de cerca o de lejos del grupo mayoritario en torno al promedio.

Para comprender mejor el significado de la simetría el siguiente paso será plantearnos cómo sería una distribución asimétrica. Para ello utilizaremos también datos ficticios, relacionados con el mismo tema, pero modificando las cantidades del gasto asignadas a los distintos grupos y comunidades. (Promedio 10.051,61) (Ver tabla de datos en la última página del tema)



Valores poco dispersos:
cercaos al promedio

Valores muy dispersos: alejados del promedio y
distantes entre sí



Véase la diferencia con la distribución del ejemplo anterior:

Un vistazo rápido basta para comprobar que la distribución en este caso es radicalmente diferente:

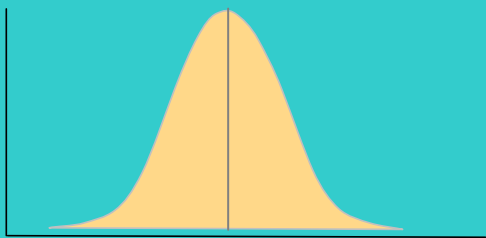
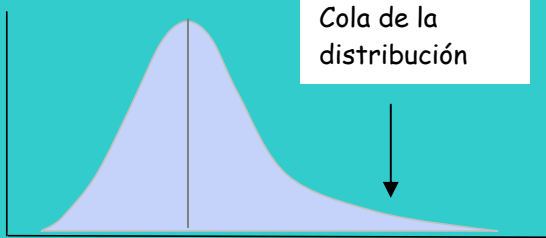
- ▣ 12 de los 18 valores se acumulan a la izquierda del promedio, muy cercanos a éste.
- ▣ A la derecha del promedio encontramos sólo un tercio de los valores. Tres de ellos se sitúan en el extremo derecho del rango, a mucha distancia del resto.
- ▣ El promedio, 10.051,61 euros, se sitúa muy desplazado hacia la izquierda dentro del rango.
- ▣ La distribución es claramente asimétrica: la dispersión a ambos lados del promedio es muy diferente. A su izquierda se concentra un grupo de valores, próximos a él y próximos también entre sí. A la derecha del

¿Qué significa en este caso la asimetría con respecto a la variable que estamos estudiando? ¿Cómo podemos interpretar el hecho de que la distribución sea asimétrica?:

Ya no podemos hablar, como en el ejemplo anterior, de un comportamiento muy homogéneo de los grupos con respecto al gasto. En aquel, veíamos a la mayoría de los grupos con gastos muy similares, próximos al promedio, y el resto de los grupos con gastos ligeramente inferiores o superiores.

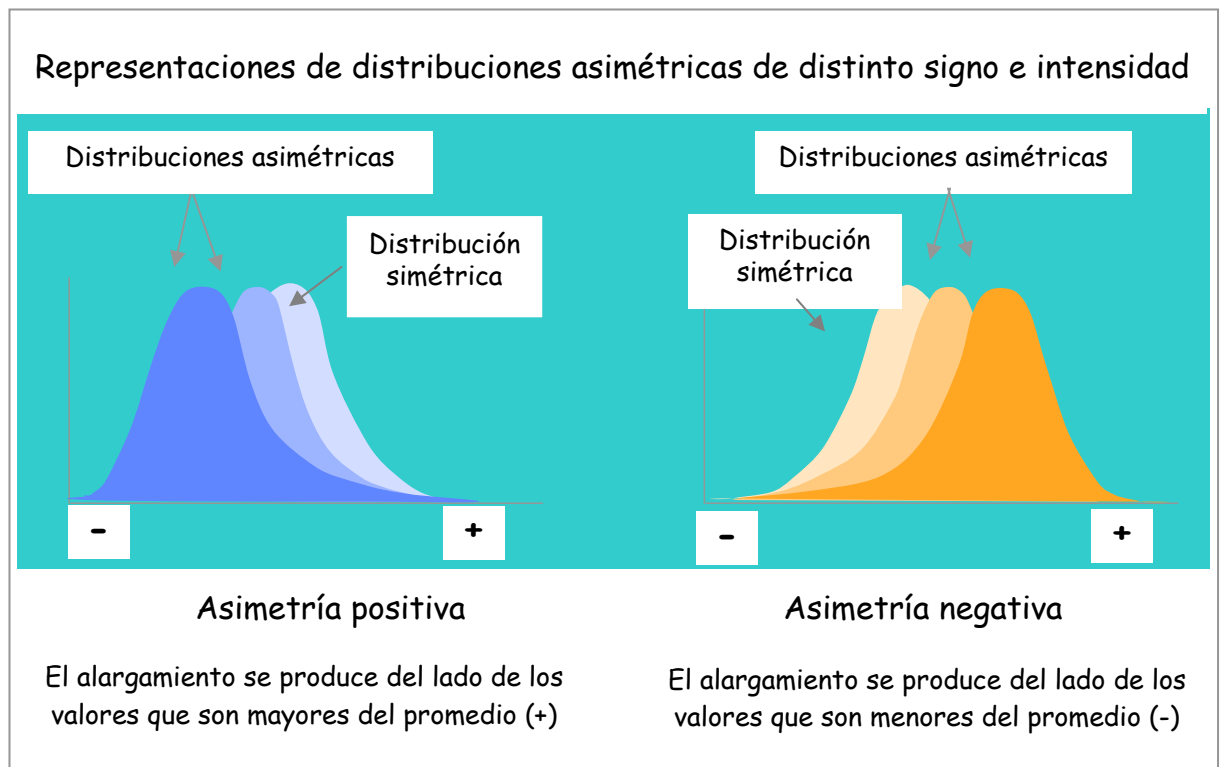
En este caso, dentro del grupo de los 18 elementos que constituyen la población existe un subgrupo de 3 elementos que ha tenido un gasto notablemente superior al de los demás. No existe, sin embargo, ningún subgrupo que destaque por haber tenido un gasto especialmente inferior al del resto.

Si representáramos las distribuciones de frecuencias de los dos ejemplos que hemos utilizado, obtendríamos gráficas muy similares a las que veremos a continuación:

Gráfica de una distribución simétrica	Gráfica de una distribución asimétrica
	
Los dos lados de la distribución son iguales y, por tanto, simétricos	Debido a la presencia de valores mucho más elevados que la mayoría, el lado derecho de la gráfica se alarga, formando lo que se denomina cola de la distribución.

En el ejemplo de asimetría que hemos utilizado, la distribución se alarga hacia la derecha porque los elementos con fuerte dispersión se corresponden con valores elevados, esto es, con valores que se ubican en el extremo derecho del rango. Cuando esto ocurre, hablamos de que la asimetría positiva.

Lo mismo puede ocurrir con respecto al lado izquierdo de la distribución. En este caso los elementos con mayor dispersión que el resto se corresponderían con los valores más bajos, es decir, con aquellos que se ubican en el extremo izquierdo del rango de la variable. En este caso hablamos de asimetría negativa.



El alargamiento de uno de los lados de la distribución puede ser grande o pequeño, dependiendo también del grado de dispersión de los valores de la variable.

Hemos visto que las distribuciones son asimétricas cuando la dispersión de los valores a uno y otro lado del promedio es distinta. Hemos visto igualmente que cuando hablamos de la asimetría de una distribución nos estamos refiriendo a una característica del conjunto de datos, a un aspecto significativo de la realidad que es preciso analizar y describir. Veremos ahora qué aporta el análisis de simetría a la descripción de una distribución y, por último, cómo calcular una medida de simetría de las distribuciones.

QUÉ APORTA EL ANÁLISIS DE SIMETRÍA

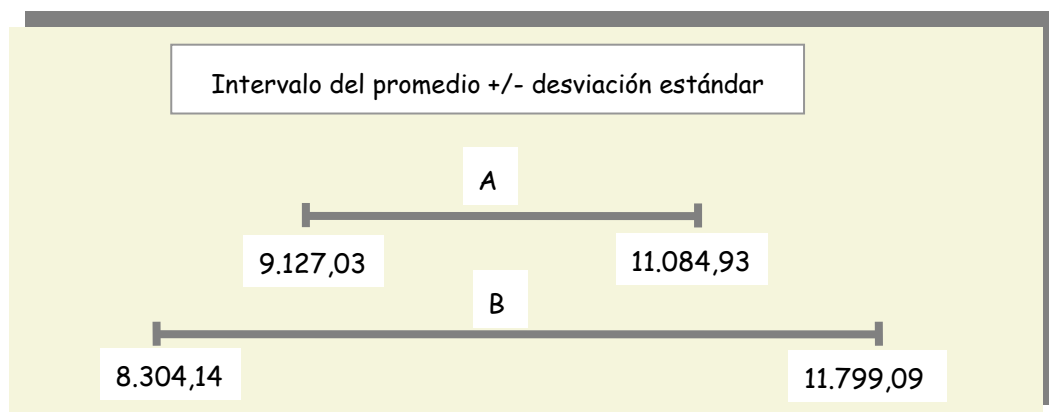
Con el fin de que se entienda mejor la necesidad de calcular el grado de simetría de las distribuciones y lo que esta medida nos proporciona veremos, mediante un ejemplo, las características de la distribución que se explican mediante el promedio y la desviación estándar. Esto nos servirá para evidenciar los rasgos de la distribución que dichas medidas no explican, rasgos que se pueden descubrir mediante el análisis de simetría. Utilizaremos los mismos datos que hemos usado anteriormente como ejemplos de distribuciones simétricas y asimétricas. Recordemos que se trata de datos ficticios, creados a partir de la

modificación de los datos pertenecientes a la encuesta de presupuestos familiares del INE en 2006. (Ver tabla de datos en la última página del tema)

Gasto total medio por persona. €. Comunidades autónomas.				
	Distribución simétrica A		Distribución asimétrica B	
Promedio	10.150,98		10.051,61	
Desviación estándar	933,95		1.747,47	
Promedio \pm desviación est.	9.217,03	11.084,93	8.304,14	11.799,09
$\bar{X} \pm S$ amplitud del intervalo	1.868		3.495	
Valor mayor y menor	8.190,13	12.111,83	8.400,13	13.836,83
Rango	3.921,7		5.436,7	

Si atendemos al promedio de las dos distribuciones que figuran en la tabla, las dos pueden parecer similares; la diferencia entre los promedios de ambas apenas supera los 100 €. La conclusión será diferente si nos fijamos en los valores de la desviación estándar; la distribución B tiene un valor de desviación estándar que casi duplica al de la distribución A.

La diferencia en la variabilidad de las dos distribuciones es evidente. El intervalo formado por el promedio +/- la desviación estándar nos da una idea clara:



Las medidas obtenidas (promedio, desviación estándar, rango) han puesto en evidencia la mayor variabilidad de la distribución A con respecto a la B. Sin embargo, no nos dicen nada sobre las características de los datos que generan la mayor variabilidad de la distribución B. Nosotros, que hemos analizado y revisado los datos una y otra vez, sabemos que la mayor variabilidad de la distribución B se genera debido al gasto particularmente elevado de tres comunidades autónomas. El análisis y descripción de la variable que hemos estudiado resultaría ambiguo, y podría además generar interpretaciones equivocadas, si no incluimos en el mismo una medida que de cuenta de la asimetría de la distribución o, lo que es lo mismo, de la presencia de elementos de la población con valores de gasto muy por encima del promedio.

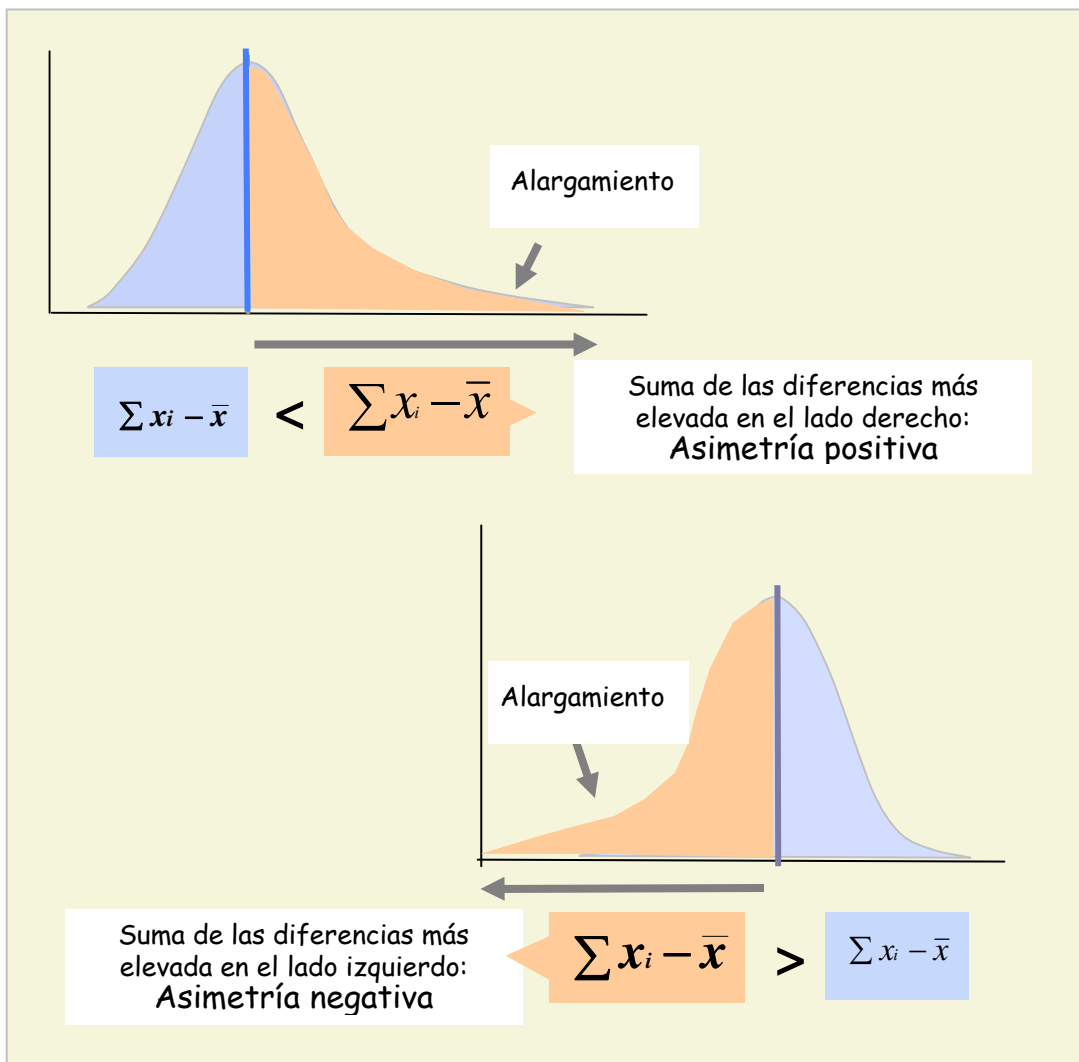
CUANTIFICACIÓN DE LA SIMETRÍA

El análisis gráfico de una distribución nos permite comprobar la existencia de asimetría en la distribución de los valores de la variable. Un vistazo al histograma o al polígono de frecuencias basta para descubrir el alargamiento de la gráfica a uno de los lados del rango. Sin embargo, a la hora de describir la asimetría de un conjunto de datos resulta mucho más cómodo disponer de una medida numérica.

La Estadística nos proporciona herramientas diferentes para obtener una medida de la simetría de las distribuciones. Una de ellas es el llamado coeficiente de asimetría de Fisher que, además de ser muy utilizada, resulta muy sencilla en su cálculo.

$$G_1 = \frac{\sum (x_i - \bar{x})^3}{N \cdot S^3}$$

El coeficiente de asimetría de Fisher se basa, una vez más, en el cálculo de las diferencias entre los valores de la variable y el promedio. En esta ocasión, las diferencias se elevan al cubo de modo que se mantiene el signo de las mismas; los valores por debajo del promedio proporcionan resultados negativos y los valores por encima del promedio dan resultados positivos. Consecuentemente, el resultado del coeficiente puede ser positivo o negativo:





- ▣ Cuando las diferencias de los valores a la izquierda de la media son superiores a las diferencias de los valores a su derecha, el resultado del coeficiente es negativo. Hablamos en ese caso de asimetría negativa que, gráficamente, se refleja en un alargamiento de la curva en el lado izquierdo de la distribución.
- ▣ Cuando las diferencias de los valores a la derecha de la media son superiores a las diferencias de los valores a su izquierda, el resultado del coeficiente es positivo. Hablamos en ese caso de asimetría positiva que, gráficamente, se refleja en un alargamiento de la curva en el lado derecho de la distribución

- ▣ Cuando las diferencias de los valores a ambos lados del promedio son iguales o muy similares, el resultado del coeficiente es próximo a cero. En este caso decimos que la distribución es simétrica.

Calcularemos ahora los coeficientes de asimetría de las dos distribuciones correspondientes al gasto medio total por persona en las comunidades autónomas del estado español. Reproduciremos nuevamente la tabla con los valores del promedio, desviación, etc.

Gasto total medio por persona. €. Comunidades autónomas.				
	Distribución simétrica A		Distribución asimétrica B	
Promedio	10.150,98		10.051,61	
Si miramos exclusivamente el promedio pensaremos que las dos distribuciones son muy similares				
Desviación estándar	933,95		1.747,47	
El valor de la desviación estándar nos pone en la pista de las diferencias de variabilidad entre las dos distribuciones				
Promedio \pm desviación est.	9.217,03	11.084,93	8.304,14	11.799,09
$\bar{X} \pm S$ amplitud del intervalo	1.868		3.495	
Lógicamente, la amplitud del intervalo es muy superior en el caso de la distribución asimétrica				
Valor mayor y menor	8.190,13	12.111,83	8.400,13	13.836,83
Rango	3.921,7		5.436,7	
Coeficiente de asimetría	0		1,1	
	Simétrica		Asimétrica positiva	
Finalmente, es el coeficiente de asimetría el que nos aclara que es la presencia de algunos valores elevados en el extremo de la distribución la responsable de la diferencia de variabilidad entre las dos distribuciones. Esto significaría que, si exceptuamos las tres comunidades con gastos especialmente elevados, la variabilidad de los dos grupos de datos sería bastante similar.				

Gasto total medio por persona. €. Comunidades autónomas					
Diferencias negativas con respecto a la media $(x_i - \bar{x})^3$		Diferencias positivas con respecto a la media $(x_i - \bar{x})^3$			
Ceuta y Melilla	-5367188169	<p>Cada uno de los resultados se ha obtenido restando al promedio el valor del gasto en la comunidad. El resultado se ha elevado al cubo.</p>  			
Extremadura	-4581326379				
Castilla - La Mancha	-3622817076				
Canarias	-2808293712				
Rioja (La)	-2126037535				
Murcia (Región de)	-1564329795				
Andalucía	-1111451743				
Cantabria	-755684628				
Castilla y León	-485309700,9			Balears (Illes)	2694322,836
Galicia	-288608211,2			Navarra	261100906,6
Asturias (Principado de)	-153861409	País Vasco	40677415776		
Aragón	-69350544,29	Cataluña	45275986328		
Comunidad Valenciana	-1362094,039	Madrid (Comunidad de)	50074079137		
-22.935.620.996,429		136.291.276.470,44			

Las diferencias positivas con respecto a la media son superiores a las diferencias negativas: la distribución es asimétrica positiva.

Gasto total medio por persona. €. Comunidades autónomas		
	Distribución simétrica A	Distribución asimétrica B
Ceuta y Melilla (Ciudades Aut.)	8.190,13	8.400,13
Extremadura	8.790,99	8.490,13
Castilla - La Mancha	9.170,31	8.615,13
Canarias	9.410,72	8.740,13
Rioja (La)	9.587,77	8.865,13
Murcia (Región de)	9.712,93	8.990,13
Andalucía	9.838,09	9.115,13
Cantabria	9.963,25	9.240,13
Castilla y León	10.088,41	9.365,13
Galicia	10.213,57	9.490,13
Asturias (Principado de)	10.338,73	9.615,13
Aragón	10.463,89	9.740,13
Comunidad Valenciana	10.589,05	10.040,13
Balears (Illes)	10.714,21	10.290,13
Navarra (Comunidad Foral de)	10.891,24	10.790,13
País Vasco	11.131,65	13.590,13
Cataluña	11.510,97	13.715,13
Madrid (Comunidad de)	12.111,83	13.836,83

Valores responsables del incremento de la asimetría de B con respecto a A

La diferencia fundamental entre las dos tablas de datos reside en los tres últimos valores que, en el caso de la distribución asimétrica, significan que el gasto en tres de las comunidades autónomas ha sido sensiblemente superior al del resto de las comunidades. Este comportamiento particularizado de las tres comunidades con mayor gasto sólo se puede identificar utilizando el coeficiente de asimetría.

Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis

Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis	1
El concepto de curtosis y su aportación al análisis de variabilidad.....	1
Análisis gráfico de la curtosis.....	4
Cuantificación de la curtosis.....	5

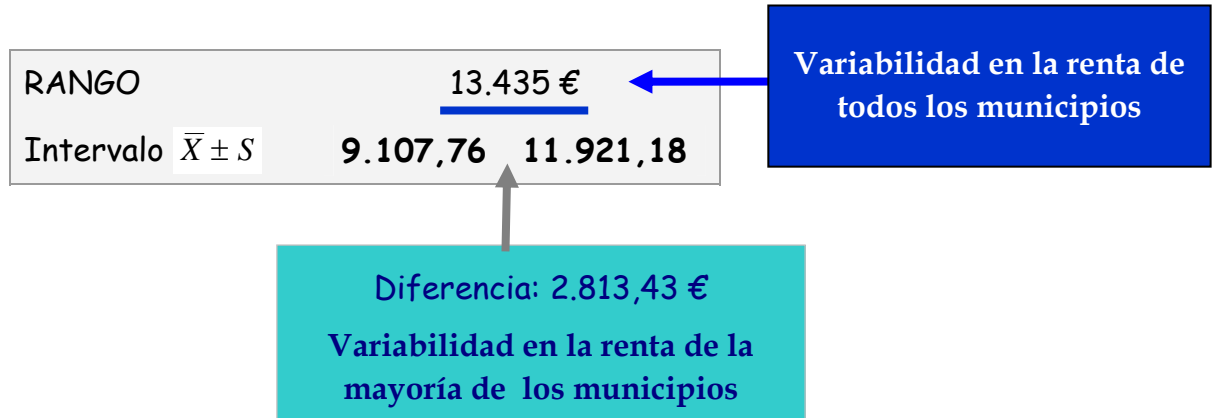
Tema 7:

La variabilidad total frente a la variabilidad de la mayoría: la curtosis

EL CONCEPTO DE CURTOSIS Y SU APORTACIÓN AL ANÁLISIS DE VARIABILIDAD

En el tema 6 hemos podido comprobar que mediante el intervalo que surge al sumar y restar a la media el valor de la desviación estandar logramos una buena descripción sobre la variabilidad del colectivo mayoritario de los valores de la variable.

Veamos de nuevo los resultados obtenidos para el intervalo $\bar{X} \pm S$ en el caso de la variable *Renta personal disponible* de los municipios de Esukadi en 2001:



Puesto que la distribución de los valores de la variable es de tipo campaniforme, sabemos que en un porcentaje de municipios superior al 60%¹ la

¹ En realidad son 200 municipios los que tienen valores comprendidos dentro del intervalo $\bar{X} \pm S$, lo que representa un 80% del total.

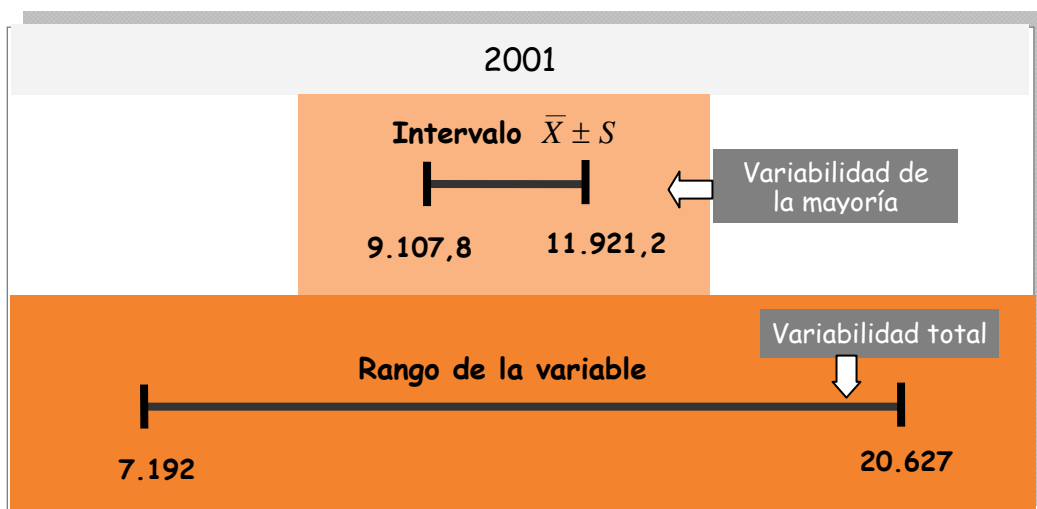
renta personal media de sus habitantes oscila entre 9.107,76 y 11.921,18 euros, lo que supone una variabilidad entre municipios de 2.813,43 euros. Esta variabilidad próxima a los 3.000 euros es pequeña, comparada con la variabilidad total, es decir, con el rango, que asciende a 13.435 €.

A la vista de semejante diferencia entre los dos intervalos (el de $\bar{X} \pm S$ y el del rango) podemos concluir que:

Aunque es cierto que existen algunos municipios cuyas rentas son sensiblemente más altas o más bajas que el promedio, una mayoría de los municipios tiene valores de renta bastante similares, que oscilan entre 9.107,76 y 11.921,18 €. La diferencia máxima entre sus rentas es de 2.813,43 €.

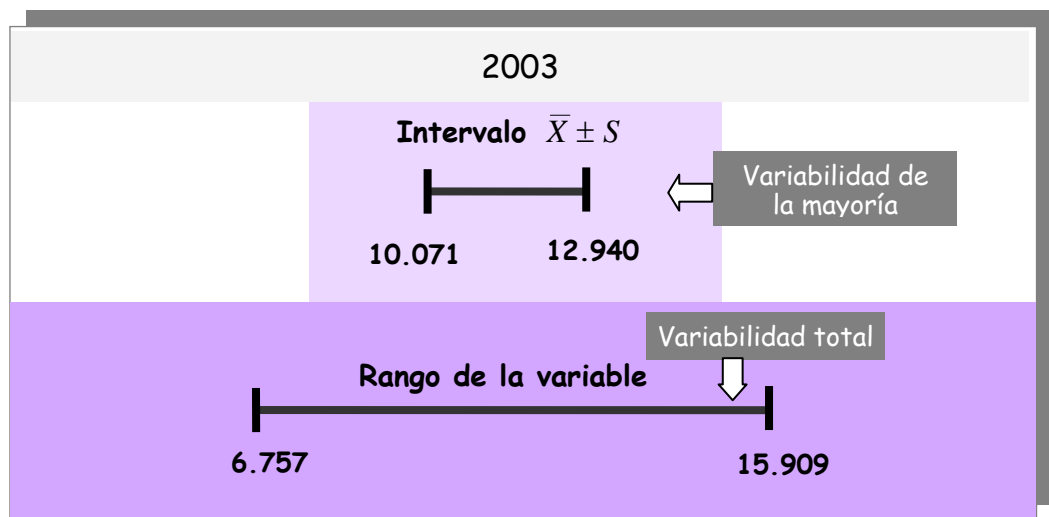
El grado de variabilidad de la mayoría no tiene nada que ver con el grado de variabilidad de toda la población.

La cuestión ahora es encontrar el modo de describir la diferencia de variabilidad entre el grupo mayoritario de municipios y el total de ellos, o lo que es lo mismo, de dejar constancia del grado de concentración de los valores del intervalo $\bar{X} \pm S$ con respecto al rango total de la variable. Volveremos a utilizar el gráfico que representa el rango para ilustrar la idea:



pequeña comparada con la segunda. Esto significa que la mayoría de los valores, agrupados dentro del intervalo $\bar{X} \pm S$ se encuentran verdaderamente concentrados en el sector central del rango. Esto equivale a decir que la variabilidad de esa mayoría es francamente inferior a la variabilidad del total de municipios. (Entre la mayoría de municipios no existe una diferencia de rentas superior a 2.813,42)

Veremos ahora una ilustración igual, pero esta vez con los datos de renta personal disponible de los municipios de Euskadi correspondientes al año 2003. (Ver tabla de datos en ANEXO)



Las diferencias entre las dos ilustraciones, y por tanto las diferencias en la variabilidad de las dos distribuciones, están a la vista: el rango de la variable correspondiente a 2003 es muy inferior al de 2001. También el agrupamiento de valores en el sector central del rango es ligeramente inferior en la distribución correspondiente a 2003.

Si quisiéramos hacer un análisis comparativo entre la renta personal disponible de los municipios en 2001 y 2003 sería necesario dejar constancia de las diferencias con respecto a la variabilidad que hay entre ambos años. Existe para ello una herramienta estadística denominada curtosis, que relaciona precisamente la amplitud del intervalo $\bar{X} \pm S$ y la amplitud total del rango.

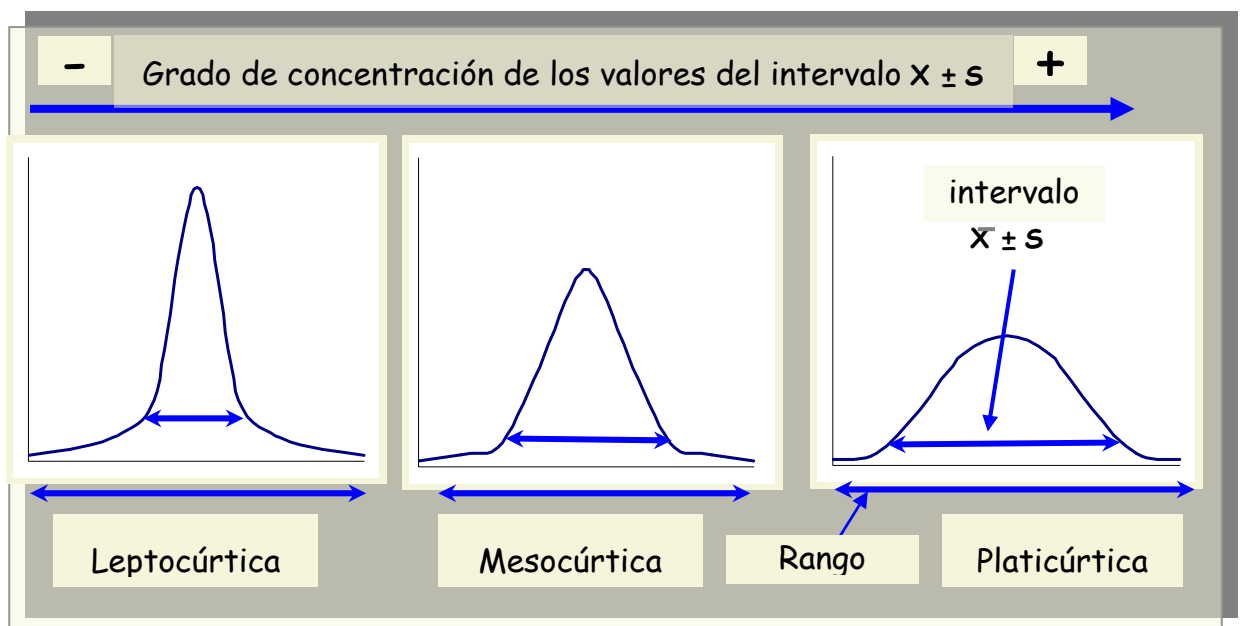
Al igual que ocurre con la simetría, la curtosis de la distribución de valores de una variable se puede analizar de dos modos:

- ▣ Mediante simple observación del histograma o del polígono de frecuencias
- ▣ Mediante la aplicación de un algoritmo que permite asignar un valor a la relación entre el rango de la variable y el del intervalo $\bar{X} \pm S$.

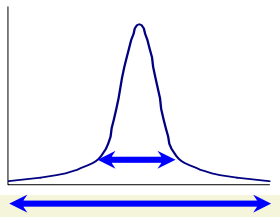
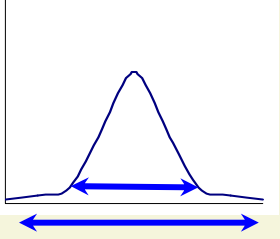
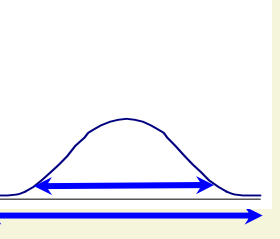
ANÁLISIS GRÁFICO DE LA CURTOSIS

La diferente relación entre el rango de la variable y el intervalo $\bar{X} \pm S$ produce distribuciones de formas muy variadas: cuando la concentración del intervalo $\bar{X} \pm S$ con respecto al rango total de la variable es muy elevada, obtenemos distribuciones con forma de campana apuntada. Por el contrario, cuanto mayor es la variabilidad del intervalo $\bar{X} \pm S$, la forma de la distribución es más achatada.

Habitualmente se clasifican las distribuciones, en función del grado de apuntamiento de la campana en tres tipos:



La curtosis es una medida que relaciona la amplitud del rango de la variable con la amplitud del intervalo $\bar{X} \pm S$. En la imagen, las flechas azules simbolizan la amplitud de los dos intervalos.

<p>Se dice que la distribución es leptocúrtica cuando una mayoría de los valores se concentra en torno a los valores centrales de la distribución. Resultado de esta elevada concentración, este tipo de distribuciones se caracterizan por la diferente amplitud del rango de la variable y del intervalo $\bar{X} \pm S$</p>	
<p>Se dice que la distribución es mesocúrtica cuando presenta un grado de concentración medio alrededor de los valores centrales de la variable.</p>	
<p>Se dice que la distribución es platicúrtica cuando los valores presentan un reducido grado de concentración y se distribuyen prácticamente a lo largo de todo el rango. Resultado de esta baja concentración, este tipo de distribuciones se caracterizan por la similitud en la amplitud del rango de la variable y del intervalo $\bar{X} \pm S$.</p>	

CUANTIFICACIÓN DE LA CURTOSIS

Existen además diferentes fórmulas estadísticas que permiten obtener una medida de la curtosis, es decir, cuantificar el grado de apuntamiento de una distribución. Se conocen con el nombre de *medidas del apuntamiento de una distribución*. Se denominan así debido al hecho de que cuanto mayor es el agrupamiento de los datos en torno al sector central de la distribución, más puntiaguda o apuntada es la forma de la *campana*. Aplicando alguna de las fórmulas existentes para medir la curtosis podemos cuantificar el grado de apuntamiento y definir así, con mayor precisión, el tipo de forma de la distribución.

De entre las fórmulas existentes presentaremos aquí una de las más conocidas: el coeficiente de Curtosis G2 de Fisher:

$$G_2 = \frac{\sum (x_i - \bar{x})^4}{N \cdot S^4} - 3$$

Los criterios para interpretar el resultado del coeficiente de curtosis son los siguientes:

- ▣ $G_2 > 0$. Cuando el resultado es positivo, es decir, cuando es superior a cero, significa que la distribución es leptocúrtica.
- ▣ $G_2 < 0$. Cuando el resultado es negativo, es decir, cuando es inferior a cero, significa que la distribución es platicúrtica.
- ▣ $G_2 \approx 0$. Cuando el resultado es cero, o muy próximo a cero, significa que la distribución es mesocúrtica.

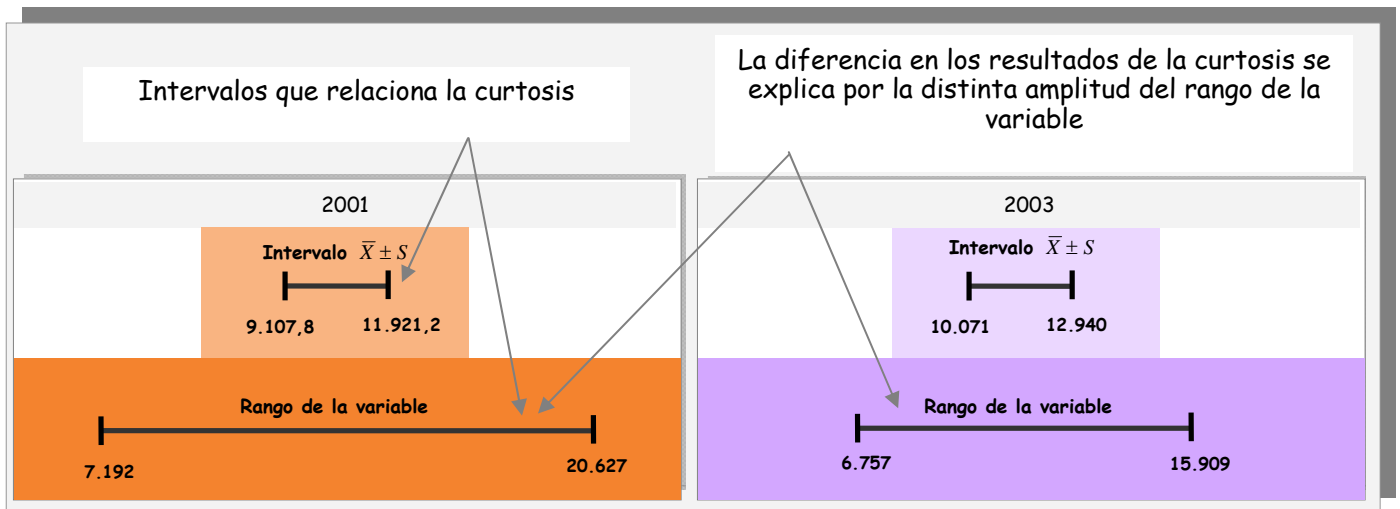
A continuación calcularemos los valores de curtosis de la variable Renta personal disponible de los municipios de Euskadi en los años 2001 y 2003.

	2001	2003
Suma de las diferencias elevadas a la 4ª potencia $\sum (x_i - \bar{x})^4$	13.315.132. 10 ⁹	4.352.128. 10 ⁹
Desviación estándar	1.403,91	1.461,65
Desviación estándar elevada a la 4ª potencia	3.884.685. 10 ⁶	4.200.998. 10 ⁶

2001
$G_2 = \frac{\sum (x_i - x)^4}{N \cdot S^4} - 3 = \frac{13.315.132 \cdot 10^9}{251 \cdot 3.884.685 \cdot 10^6} - 3 = \frac{13.315.132 \cdot 10^9}{975.056 \cdot 10^9} - 3 = 10,65$
2003
$G_2 = \frac{\sum (x_i - x)^4}{N \cdot S^4} - 3 = \frac{4.352.128 \cdot 10^9}{251 \cdot 4.200.998 \cdot 10^6} - 3 = \frac{4.352.128 \cdot 10^9}{1.054.450 \cdot 10^9} - 3 = 1,27$

Los resultados obtenidos ponen en evidencia las diferencias entre las dos distribuciones. En 2001 el rango de la variable es más amplio que en 2003, de ahí que el agrupamiento de los valores centrales de la distribución resulte también mayor en 2001 que en 2003 y que, consecuentemente, para el primero de los años se obtenga una curtosis más elevada. De todas formas, aunque el valor de la curtosis de las dos distribuciones es diferente,

las dos distribuciones pertenecen al tipo de distribuciones leptocúrticas ya que, en ambas, los valores centrales de la distribución se encuentran bastante concentrados.. Expresado en términos de renta personal disponible, el hecho de que la distribución sea leptocúrtica significa que en una mayoría de los municipios las rentas consideradas fueron bastantes similares, es decir, se situaron dentro de un rango pequeño de variación.

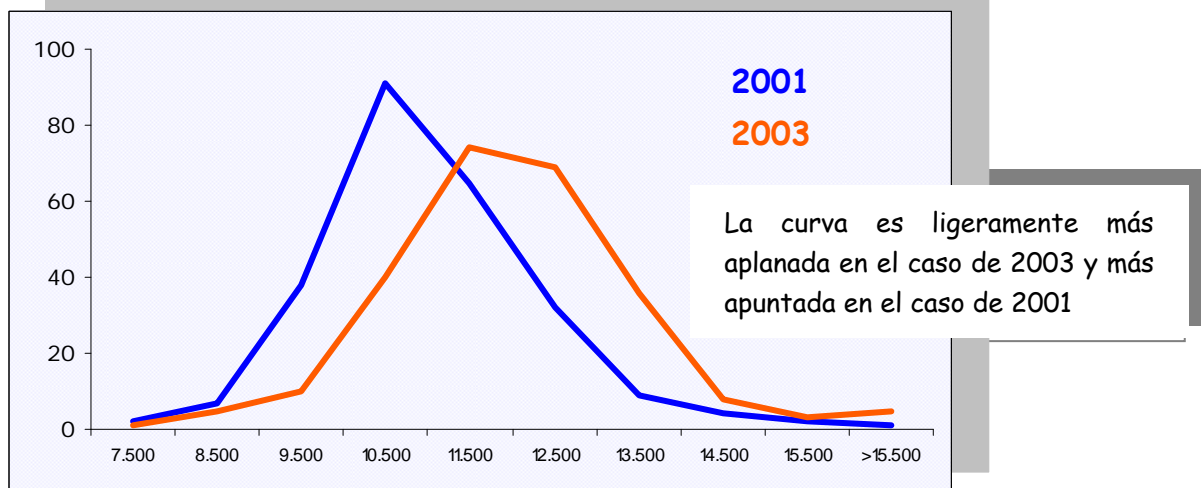


Puesto que la curtosis mide la relación entre los dos intervalos, aunque las dos distribuciones presentan un grado de concentración similar en torno a los valores centrales, la diferencia en el rango hace que el resultado de la curtosis sea distinto en las dos.

Por otra parte, ya hemos visto anteriormente que, en el caso de los datos de 2001, la amplitud del rango de la variable se debe a la presencia de un valor extremo (Laukiz con 20.627 €). Este valor eleva el rango de 15.346€ a 20.627 € y es responsable también del elevado resultado de la curtosis. Para comprobarlo, realizaremos de nuevo los cálculos de la curtosis pero eliminando esta vez el valor extremo:

$$G_2 = \frac{\sum (x_i - \bar{x})^4}{N \cdot S^4} - 3 = \frac{2.857.357.374 \cdot 10^6}{250 \cdot 2.459.110 \cdot 10^6} - 3 = \frac{2.857.357.374 \cdot 10^6}{614.776.103 \cdot 10^6} - 3 = 1,65$$

Tal como esperábamos, el resultado de la curtosis ha descendido notablemente, acercándose mucho al valor obtenido para el año 2003. De todos modos, aun eliminado el valor extremo, la curtosis de los datos de 2001 sigue siendo algo superior que la correspondiente a 2003. La propia gráfica nos permite comprobar que la concentración en torno a los valores centrales es algo mayor en los datos de 2001.



En la gráfica podemos ver la pequeña diferencia que existe en la curtosis entre la distribución correspondiente al año 2001 y la de 2003: la primera, marcada en azul (2001), muestra un apuntamiento ligeramente superior a la segunda (2003), marcada en rojo. (Atención! la gráfica no permite ver la diferencia en el rango de las dos distribuciones, ya que el eje de las abcisas sólo llega hasta los 15.500 €)

PRIMERA PARTE

[Tema 1. La estadística: un método para conocer la realidad](#)

SEGUNDA PARTE: EL ANÁLISIS DE VARIABILIDAD ES UNO DE LOS OBJETIVOS FUNDAMENTALES DE LA ESTADÍSTICA

[Tema 2: En qué consiste la variabilidad](#)

[Tema 3. Cómo descubrir la variabilidad de un conjunto de datos](#)

[Tema 4: El tipo de variabilidad condiciona el modo de analizarla](#)

TERCERA PARTE: CÓMO ANALIZAR Y DESCRIBIR LA VARIABILIDAD CUANDO LOS DATOS DIBUJAN UNA FORMA DE CAMPANA

[Tema 5: Análisis del sector central de la distribución: media y desviación estándar](#)

[Tema 6: Análisis de los extremos de la distribución. La simetría](#)

[Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis](#)

Usted no se ha autenticado. ([Entrar](#))

Estadística aplicada a la Geografía