



Gaixotasun kronikoen progresioa ulertzeko eta aurrerateko metodoa eta bere aplikazioa datu errealetan

Gradu Amaierako Lana
Matematikako Gradua

Leire Agirretxe Muguruza

Inmaculada Arostegui
Irakasleak zuzendutako lana

Leioa, 2020ko ekainaren 18a

Gaien Aurkibidea

Sarrera	v
1 Gaixotasunaren progresioa estimatzeko metodologia	1
1.1 Sarrera	1
1.2 Estimazio metodoa	2
1 Lehenengo urratsa	2
2 Bigarren urratsa	3
3 Hirugarren urratsa	4
4 Laugarren urratsa	6
1.3 Ziurgabetasunaren estimazioa	8
2 Aplikazioa	15
2.1 Biriketako gaixotasun buxatzaile kronikoa	15
2.2 Programak eta inplementazioa	16
2.3 Datubasearen deskribapena	17
2.4 Gaixotasunaren progresioa	23
1 Ariketa ahalmena	24
2 Kuadrizepseko indarra	27
3 Birikien funtzioa	30
4 Disnea	33
3 Eztabaida	37
Bibliografia	39
A Aldagaien azalpena	41
B Kodea	43
B.1 R programa	43
B.2 Mathematica programa	47

Sarrera

Gaixotasun kronikoen inguruko ikerketa askotan, paziente kronikoen osasuna neurtzen denean, indibiduo bakoitza gaixotasunaren etapa desberdin batean egon ohi da. Batzuk gaixotasunaren hasieran egoten dira, hau da, gaixotasuna diagnostikatu berri, eta beste batzuk, aldiz, etapa aurreratua-goetan, egoera egonkorragoan edo gaixotasunaren une larrietan. Hori dela eta, indibiduoak eta taldeak aztertzean, denbora tarte berdinetan indibiduen eboluzioa desberdina izan daiteke. Gaixotasunaren etaparen arabera izango da eboluzioa, hain zuzen ere. Beraz, garrantzitsua izan daiteke gaixotasun kroniko horien progresioa ezagutzea, denboran zehar zelan egingo duten okerrera ezagutzeko, zenbat denbora beharko duen gaixotasunak etapa batetik bestera igarotzeko edota progresio hori zein abiaduratan egingo duen jakiteko.

Oinarri desberdinak dituzten metodoak garatu izan dira gaixotasun kronikoen progresioa auresateko eta ezagutzeko. Horietako askok ez dute kontuan hartzen epe motzean eta urte gutxitan bildutako datuen bidez, denbora tarte zabal bateko gaixotasunaren progresioa auresan nahi dela eta hori erronka gogorra izan daitekeela. Beste metodo batzuek, berriz, indibiduen datu kopuru handiagoak erabiltzen dituzte.

Budgeon eta kideek (2017) proposatutako metodoa aukeratu dugu guk, epe motzeko datu kopuru txikiak erabiliz gaixotasun kronikoen progresioa auresateko [2]. Gaixotasunaren eboluzioa neurtzeko aukeratutako aldagai adierazle bakoitzerako lau urratsetan oinarritzen den metodoa aplikatzen da: lehenengo urratsean, indibiduo bakoitzarentzat erantzun aldagaiaren itxarotako balioa eta denborarekiko malda kalkulatzeko; bigarrenean, polinomio monotono bat doitzen zaie datu bikote horiei; hirugarrenean, polinomio horren alderantzizkoaren integrala kalkulatzeko da eta laugarrenean, aurreko urratsean lortutako kurba alderantzizkatzen da gaixotasunaren progresioa estimatzeko.

Gure helburua indibiduen epe motzeko datuetan oinarrituz gaixotasun kronikoen progresioa auresango duen metodoa inplementatzea eta metodo hori gaixotasun bati aplikatzea da. Hain zuzen, biriketako gaixotasun buxatzai-

le kronikoari. Eskerrak eman nahi genizkioke Galdakao-Usansoloko Unibertsitate Ospitaleko pneumologia zerbitzuko Cristóbal Esteban González. Batetik, lan akademiko honetarako *Estudio de evolución de pacientes con EPOC estable* ikerketako datuak erabiltzeko aukera emateagatik, eta bestetik, emaitzak interpretatzerako orduan behar izandako laguntza emateagatik.

Hurrengo egitura jarraitu dugu txosten honetan. Lehenengo kapituluan, egindako lana ulertzeko oinarrikoak diren kontzeptu batzuk azaldu ditugu sarreran. Ondoren, erabilitako metodologia azaldu dugu, Budgeon eta kideek (2017) proposatutako metodoa, hain zuzen ere [2]. Bigarren kapituluan, deskribatutako metodologia datu errealetara aplikatu dugu, biriketako gaixotasun buxatzaile kronikoa duten pazienteek osaturiko datubasean oinarritu gara, hain zuzen ere. Datubase hori deskribatzeaz gain, lehenengo kapituluan aurkeztutako metodoa aplikatzean lortutako emaitzak eta egindako lana aurrera eramateko erabilitako programak azaldu ditugu kapitulu horretan. Hirugarren kapituluan, eztabaida txiki bat idatzi dugu. Azkenik, eranskinetan, erabilitako aldagaien esanahia eta lan hau egiteko erabilitako *R*-ko eta *Mathematica*-ko kodea gehitu ditugu.

1. Kapitulua

Gaixotasunaren progresioa estimatzeke metodologia

1.1 Sarrera

Hainbat gaixotasun kronikoren kasuan, ez daude denbora tarte zabal bateko indibiduen datuak eskuragarri. Epe motzeko eta neurketa gutxitako datuak bakarrik daudenean erabilgarri, ez da erraza izaten datu horietan oinarrituz denboran zeharreko gaixotasunaren progresioa auresatea eta gaixotasunaren progresio kurba ezagutzea. Gaixotasun kronikoen kasuan garrantzitsua izan daiteke progresio hori ezagutzea, denboran zehar nola eta zein abiaduratan joko duten okerrera jakiteko. Hori dela eta, metodo asko garatu izan dira gaixotasunen progresioa auresateko. Budgeon eta kideek (2017) proposatutako metodoan oinarrituko gara gu [2]. Bertan, epe motzeko datuak erabiliz gaixotasunen progresioa auresateko metodo bat proposatzen da eta datu simulatuei eta gaixotasun baten errealitateko datuei aplikatzen zaie metodo hori bere baliagarritasuna erakusteko. Hainbat indibiduen epe motzeko datuetan oinarrituz gaixotasunen progresioa auresango duen metodoa implementatzea eta metodo hori gaixotasun bati aplikatzea izango da gure helburua.

Gaixotasunaren progresioa auresateko erabiliko dugun metodoa, itxura sigmoidala duten gaixotasunen kasurako baliagarria izan daiteke. Ondoko itxura duten gaixotasunak dira horiek: gaixotasunaren lehenengo etapatan, motel progresatzen du, gaixotasunaren progresioaren hazkuntza motela da; ondoren, hazkuntza hori handia da, azkarra da progresioa; eta azkenik, hazkuntza hori geldituz doa.

Indibiduo bakoitzaren gaixotasunaren progresio kurbak itxura sigmoidala duela onartuz, Budgeon eta kideek (2017) proposatutako metodoan oinarrituta [2], biriketako gaixotasun buxatzaile kronikoen (BGBK, gazteleraz *enfermedad pulmonar obstructiva crónica*, *EPOC*) progresioa estimatzen eta

aurren saiaturako gara. Metodo horrek indibiduen epe motzeko hainbat denboratako datuak erabiliz, denboraren menpekora den gaixotasunaren progresioaren kurba lortzea du helburu.

1.2 Estimazio metodoa

Gaixotasuna pairatzen duten indibiduoak aztertzen dira mota honetako ikerketetan eta gaixo horiek osatzen duten taldea izango da gure lagina. Gaixotasunaren narriaduran eragina duten hainbat aldagai izango dira aztergai eta banan-banan aztertuko dugu gaixotasunak eta denborak aurrera egin ahala duten eboluzioa. Aldagai horietako bakoitzari erantzun aldagai deritza eta gure kasuan denboran zehar hartzen dituzten balioak nola aldatzen diren ikusi nahi izango dugu.

Defini ditzagun lehenik erabiliko ditugun parametroak:

t \equiv indibiduo bakoitzaren lehenengo neurketatik igarotako denbora, hau da,

ikerketa denbora (t_{ij} i. indibiduoaren j. neurketa egindako unea izango da)

τ \equiv gaixotasunaren progresio denbora (denbora hau etorkizuneko da)

y_{ij} \equiv i. indibiduoaren j. neurketako erantzun aldagaiaren balioa

$\hat{\mu}$ \equiv indibiduo bakoitzarentzat erantzun aldagaiaren itxarotako balioa edo batezbestekoaren estimazioa

$\hat{\mu}'$ \equiv indibiduo bakoitzaren datuei doitutako erregresio lineal bakuneko ereduaren denborarekiko malda

$\hat{\mu}(\tau)$ \equiv τ -ren menpeko gaixotasunaren progresioa

Aipatu bezala, Budgeon eta kideek (2017) proposatutako metodoa erabiliko dugu gaixotasunaren progresioa aurreratzeko [2]. Metodo hau lau urrats nagusitan oinarritzen da eta gaixotasunaren progresioa azaltzen duen erantzun aldagai bat behar dugu. Aldagai askea denbora izango da.

Deskriba dezagun, beraz, metodoa.

1 Lehenengo urratsa

Erantzun aldagaiaren balioetan oinarrituz, gure lagineko indibiduo bakoitzaren itxarotako balioa edo batezbestekoaren estimazioa eta erregresio linealeko ereduaren denborarekiko malda kalkulatzeko da lehenengo urrats honen helburua.

Horretarako, lehenik, erregresio linealeko eredu bat doitu behar zaie indibiduo bakoitzaren datuei. Erantzun aldagai jarraitu bat eta aldagai aske bat izango ditugu. Gure kasuan, gaixotasunaren progresioa azaltzen duen aldagaietako bat izango da erantzun aldagaia, eta aldagai askea, berriz, ikerketa

denbora. Erantzun aldagaiaren balioak denboraren menpe izango ditugu. Indibiduen epe motzeko datuei erregresio linealeko ereduak doitzeko, erregresio lineal bakuneko ereduak erabiliko ditugu, aldagai aske bakarra izango baitugu.

Behaturik datuak (t_{ij}, y_{ij}) bikoteak dira, $i \in \{1, \dots, n\}$ eta $j \in \{1, \dots, n_i\}$ izanik. Gure lagineko indibiduo kopurua n izango da eta n_i , berriz, i . indibiduoaren behaturiko datu kopurua.

Erregresio lineal bakunaren helburua X aldagai askearen (jarraitua) eta Y menpeko aldagaiaren (jarraitua) arteko erlazio lineala aztertzea da. Izan bedi y_{ij} i . indibiduoaren j . neurketako erantzun aldagaiaren balioa eta t_{ij} i . indibiduoaren j . neurketako ikerketa denbora. Demagun i indibiduo bakoitzeko t_{ij} uneetako n_i datu ditugula. Orduan, erregresio lineal bakunaren eredu probabilitikoa ondokoa da (i finko mantenduz):

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \epsilon_{ij} \tag{1.1}$$

β_{0i} eta β_{1i} i . indibiduoaren datuei doitutako erregresio linealeko ereduaren gai askea eta malda izanik, hurrenez hurren. *Osagai probabilitikoak* (ϵ_{ij}) erantzunaren zorizkotasuna deskribatzen du eta *osagai sistematikoak* ($y_{ij} = \beta_{0i} + \beta_{1i}t_{ij}$), berriz, itxarotako erantzunaren eta aldagai askearen arteko erlazioa. β_{0i} -ri erregresio linealeko ereduaren gai askea deritzo eta β_{1i} -ri, aldiz, malda.

Jarraian, indibiduo bakoitzari dagokion itxarotako balioa kalkulatu da erregresio lineal bakuneko eredu horiek erabiliz. Horretarako, gure kasuan, indibiduo bakoitzaren datuei doitutako eredu dagokion ikerketa denboraren erdiko puntuan ebaluatu behar da, $\bar{t}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} t_{ij}$ izanik indibiduo bakoitzaren ikerketa denboraren erdiko puntua. Hau da,

$$\hat{\mu}(\bar{t}_i) = \beta_{0i} + \beta_{1i}\bar{t}_i.$$

Orduan, erregresio lineal bakuneko eredu hauetatik n malda, $\hat{\mu}'_i = \beta_{1i}$ $i \in \{1, \dots, n\}$, eta n itxarotako balio, $\hat{\mu}(\bar{t}_i)$ $i \in \{1, \dots, n\}$, lortuko dira (n indibiduo kopurua izanik). Hau da, honako bikoteak izango ditugu lagineko indibiduo bakoitzarentzat:

$$(\hat{\mu}(\bar{t}_i), \hat{\mu}'_i), i \in \{1, \dots, n\}$$

2 Bigarren urratsa

Estimatutako batezbestekoei eta erregresio lineal bakuneko eredu denborarekiko maldei polinomio monotono bat doitzea da bigarren urratseko eginbeharra. Lehenengo urratsean kalkulatu erregresio lineal bakuneko eredu bidez lortutako itxarotako balioei eta eredu horien denborarekiko maldei, beraz, polinomio bat doitu behar diegu.

Datu bikote horiei doitutako polinomioak monotonoa izateaz gain, positiboa edo negatiboa izan behar du, eta hori erantzun aldagaiaren arabera izango da. Gaixotasunaren progresioa azaltzen duen aldagaiaren bilakaera gorakorra bada, polinomioak positiboa izan behar du. Alderantziz, gaixotasunaren progresioa azaltzen duen aldagaiaren bilakaera beherakorra bada, polinomioak negatiboa behar du izan.

Maila txikiko polinomioak erabili ohi dira batezbesteko estimazioen eta malden arteko lotura azaltzeko:

$$p(x) = p(x; \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \dots + \gamma_q x^q \quad (1.2)$$

non q polinomioaren maila eta $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_q)$ koefiziente bektorea diren. Maila baxuko polinomioak erabiliko ditugu, hirugarren mailako polinomioak aukeratu ditugu, hain zuzen ere. Izan ere, maila altuagoko polinomioekin lan egitea zailagoa da hurrengo urratsetan eta gainera, grafikoki antzeko itxura duten polinomioak lortzen dira hiru baino maila (bakoiti) handiagoko polinomioak erabiltzean [9].

Estimatutako batezbestekoei eta maldei polinomio bat doitzera, kontuan hartu behar dugu polinomioaren erroak aldagai askearen eremuan egon daitezkeela eta egoera horretan arazoak izango genituzkeela hurrengo urratsetan.

Behin indibiduo bakoitzaren erregresio lineal bakuneko ereduaren denborarekiko malda eta itxarotako balioa lortuta, n itxarotako balio eta n malda izango ditugu, hau da, n luzerako bi datu zutabe. Datu bikote horiei hirugarren mailako polinomio monotono bat doitzean, itxurako honetako polinomio bat lortzen da:

$$p(x) = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 \quad (1.3)$$

Hau da, aldagai asketzat $\hat{\mu}(t)$ (itxarotako balioak) eta menpeko aldagaitzat $\hat{\mu}'$ (maldak) dituen hirugarren mailako polinomio monotono bat lortzen da, erantzun aldagaiaren arabera positiboa edo negatiboa izango dena.

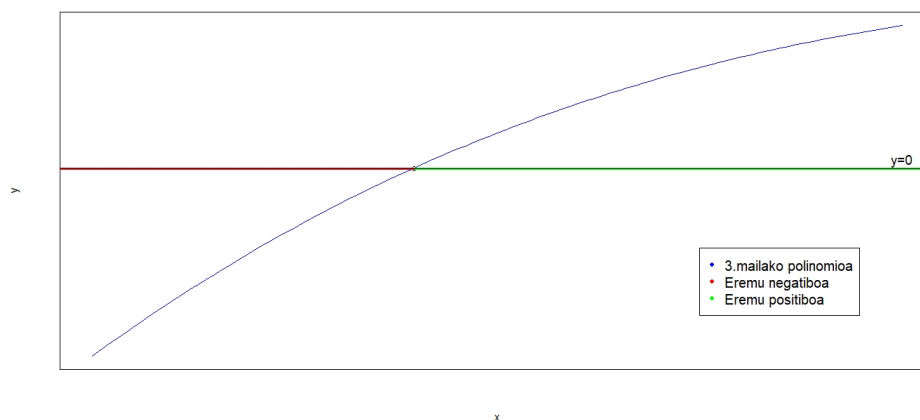
3 Hirugarren urratsa

Bigarren urratsean lortutako polinomioaren alderantzizkoa integratzean datza urrats honek, hau da, $\frac{1}{\hat{\mu}}$ integratu behar da $\hat{\mu}(t)$ aldagaiarekiko.

Datuei doitutako polinomioaren alderantzizkoaren integrala kalkulatzeko, aurretik, polinomioa positiboa edo negatiboa izan behar dela gogoratu behar dugu, eta horretarako, polinomioaren eremu positiboa edo negatiboa kalkulatu behar da, ondoren integrala eremu horretan ebaluatzeko. Eremu horren hautaketa erantzun aldagaiaren arabera izango da.

- a) Denborak aurrera egin ahala, erantzun aldagaiak gero eta balio handiagoak hartzen baditu, polinomioaren eremu positiboarekin gelditu behar da. Hau da, gaixotasunak okerrera egitean erantzun aldagaia gorakorra bada eremu positiboa hautatu behar da. Izan ere, denborak aurrera egin ahala erantzun aldagaiaren balioak handitzen badoaz, progresioaren maldak positiboa izan behar du.
- b) Alderantziz, denborak aurrera egitean, erantzun aldagaiaren balioak gero eta txikiagoak badira, polinomioaren eremu negatiboa hautatu behar da. Beste modu batera esanda, gaixotasunak okerrera egitean erantzun aldagaia beherakorra bada, eremu negatiboarekin gelditu behar da. Progresioaren maldak negatiboa izan behar du denborak aurrera egin ahala erantzun aldagaiaren balioak txikitzen joan daitezen.

Horretarako, polinomioa positibo edo negatibo egiten duten aldagai askearen balioetara mugatuko dugu bere eremua. Hau da, kasuan kasu, polinomioa positiboa edo negatiboa den aldagai askearen balioekin geldituko gara eta 1.1. irudian ikus dezakegu polinomio baten kasurako eremuaren bi aukerak.



1.1. Irudia. Hirugarren mailako polinomio baten eta bere eremu positiboaren eta negatiboaren adibidea.

Behin datuei doitutako polinomioa eta behar den eremua lortuta, hirugarren mailako polinomioaren alderantzizkoaren integrala kalkulatu behar da. Hemendik lortutako emaitza honela adieraz daiteke:

$$(\hat{\mu}(\tau), \tau),$$

$\hat{\mu}(\tau)$ gaixotasunaren progresioa eta τ gaixotasunaren progresio denbora (etorkizuneko denbora) izanik. Hau da, denbora progresioaren menpe lortzen da, gaixotasunaren progresioaren menpekua den kurba bat.

Hurrengo zeregina kurba hau irudikatzea litzateke. Horretarako, bi datu zutabe izango ditugu, bat gaixotasunaren progresioaren balioei dagokiona eta bestea, berriz, gaixotasunaren progresio denboraren balioei dagokiona. Gaixotasunaren progresioa izango da integraleko aldagai askea eta aldagai honen balioak polinomioaren alderantzizkoaren integralean ebaluatuta denboraren balioak lortzen dira. Aurretik aukeratutako eremuan hartuko ditu gaixotasunaren progresioak balioak, hain zuzen ere. Hau da, gaixotasunaren progresioaren lehenengo balioa polinomioa positibo egiten duen lehenengo balioa izango da, eremu positiboa hautatu bada (eta ondorengoak puntu horretatik eskuinetarakoak), eta eremu negatiboa aukeratu bada, aldiz, polinomioa negatiboa egiten duen lehenengo balioa izango da gaixotasunaren progresioak hartuko duen lehenengo balioa (eta ondorengoak puntu horretatik ezkerretarakoak).

Hurrengo urratsa, gaixotasunaren progresio denbora koordenatu jatorrira transladatzea da, etorkizuneko denbora hau 0an hastea nahi baitugu. Horretarako, aldagai aldaketa bat egingo dugu.

Orain, urrats honetan lortutako bi aldagaien kurba irudikatzeko, aldagai askeztat gaixotasunaren progresioa, $\hat{\mu}(\tau)$ (abzisa ardatza) eta menpeko aldagaitzat denbora, τ , (ordenatu ardatza) hartuko dira. Modu honetan, gaixotasunaren progresioaren menpeko kurba lortuko da, hau da, gaixotasunaren progresioak aurrera egin ahala denborak hartzen dituen balioak ikusi ahal izango dira.

4 Laugarren urratsa

Hirugarren urratsean lortutako bi aldagaiak alderantzikatzen dira urrats honetan gaixotasunaren epe luzerako progresioa ezagutzeko. Ondoko bikoteak lortzen dira, beraz:

$$(\tau, \hat{\mu}(\tau))$$

Hau da, kasu honetan denbora izango da aldagai askea eta gaixotasunaren progresioa, berriz, menpeko aldagaia. Gaixotasunaren progresioa denboraren menpe lortzen da, eta beraz, denboraren menpeko kurba bat izango dugu.

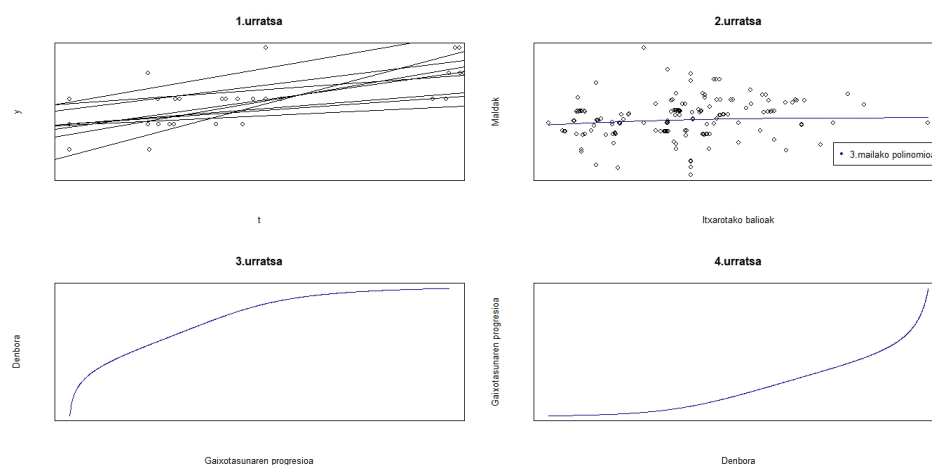
Gaixotasunaren epe luzerako progresioa ezagutzeko eta aurrerako, denboraren menpeko kurba irudikatu behar da. Hirugarren urratsean egindakoaren alderantzizkoa egiten da oraingoan, denbora aldagai askeztat (absiza ardatza) hartu eta gaixotasunaren progresioa menpeko aldagaitzat (ordenatu ardatza). Aldagaiak alderantzikatzen dira.

Bi datu zutabe horiek modu horretan irudikatuta, epe luzerako denboran zeharreko gaixotasunaren progresioa ezagutu ahal izango dugu, hau da, denborak aurrera egitean gaixotasuna azaltzen duen aldagaiak hartuko lituzkeen

balioak ikusi ahal izango ditugu. Grafikoa mugatuta, denbora tarte jakin batean gaixotasunak izango duen progresioa ezagutuko dugu.

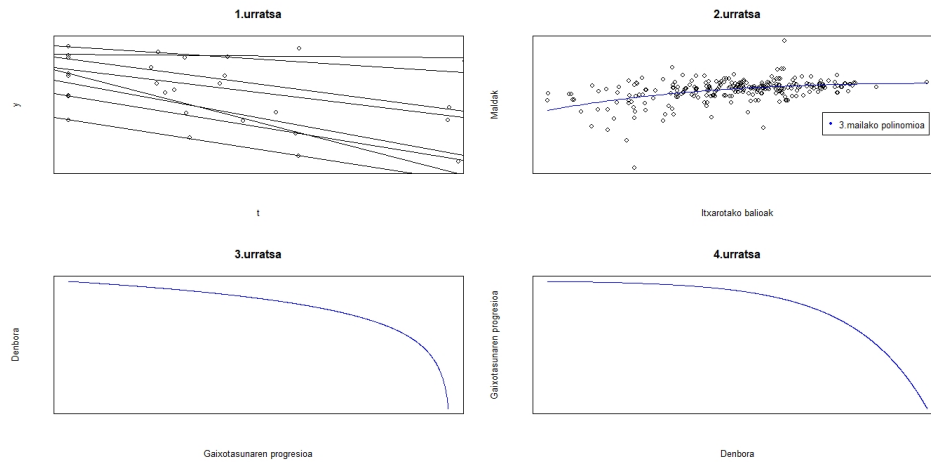
Beraz, batezbestekoen estimazioak eta maldak erabiliz lortutako kurbatik abiatuz, gaixotasunak denboran zehar izango duen eboluzioa ezagutzen lagun diezaguen kurba lortzen da azken bi urrats hauetan.

Budgeon eta kideek (2017) proposatutako metodoa azalduta [2], irudika ditzagun bi aldagaiaren kasurako (batek eboluzio gorakorra du, eta bestearen eboluzioa, berriz, beherakorra da) lau urrats horiek grafikoen bidez 1.2. eta 1.3. irudietan.



1.2. Irudia. 1.urratsa - Indibuo multzo baten ikerketa denborak eta erantzun aldagaiaren balioak, eta doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Erregresio linealeko erduetatik lortutako itxarotako balioak eta denborarekiko maldak, eta datu horiei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala (eremu positiboan ebaluatuta). 4.urratsa - Denboraren menpeko gaixotasunaren progresio kurba.

Lehenengo kasuan, 1.2. irudiko lehenengo urratseko grafikoen ikus daitekeenez, erantzun aldagaiaren balioak handitzen doaz denborak aurrera egin ahala, eta beraz, indibiduen datuei doitutako erregresio lineal bakuneko ereduaren malda positiboa da. Itxarotako balioei eta ereduaren denborarekiko maldei doitutako hirugarren mailako polinomioaren eremu positiboarekin geldituko ginateke kasu honetan, gaixotasunak okerrera egitean erantzun aldagaia gorakorra baita. Irudi honetako azken bi grafikoetan, denboraren eta erantzun aldagaiaren balioak nola aldatuko liratekeen ikus dezakegu.



1.3. Irudia. 1.urratsa - Hainbat indibiduoren ikerketa denborak eta erantzun aldagaiaren balioak, eta doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Itxarotako balioak eta denborarekiko maldak, eta datu bikote horiei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala (eremu negatiboan ebaluatuta). 4.urratsa - Denboraren menpeko gaixotasunaren progresio kurba.

Bigarren kasuan, 1.3. irudiko lehenengo urratseko grafikoan ikus daiteke erantzun aldagaiaren balioak txikitzen doazela denborak aurrera egin ahal, eta beraz, indibiduen datuei doitutako erregresio lineal bakuneko ereduaren maldak negatiboak dira. Gaixotasunaren progresioak okerrera egitean erantzun aldagaia beherakorra denez, datuei doitutako hirugarren mailako polinomioaren eremu negatiboarekin geldituko ginateke. Azken bi grafikoe-tan etorkizuneko denboraren eta erantzun aldagaiaren balioak nola aldatuko liratekeen ikus dezakegu.

1.3 Ziurgabetasunaren estimazioa

Aurreko atalean, indibiduen epe motzeko datuetatik abiatuz gaixotasunaren progresioa ezagutzeko eta auresateko kurba bat lortu dugu, hau da, gaixotasunaren progresioa denboraren menpekoa den kurba baten bidez azaltzen dugu. Baina lortutako kurba hori hainbat indibiduen datuak aztertuz lortutako batezbesteko progresio moduan uler dezakegu. Orduan, ziurgabetasuna edo sakabapena izango du kurba horrek. Sakabanapen hori estimatzeko *bootstrap* metodo estatistikoa erabiliko dugu.

Bootstrapa berlaginketan oinarritzen den estimazio teknika ez parametrikoa da. Lagin bat oinarri izanda, n lagin-tamaina izanik, laginketa egiten da lagin honen gainean, hau da, zoriz n tamainako lagin kopuru handiak sortzen

dira. Lagineko indibiduo bakoitza hautatua izateko probabilitatea berdina da eta hauen errepikapena baimentzen da.

Lagin bakoitzean interesekoa den estatistikoa edo teknika aplikatzen da (hala nola, bariantza, batezbestekoa, alborapena, konfiantza-tarteak...) eta modu honetara, lagin guztiak konbinatuz, aldagaien edo parametroen banaketa estima daiteke.

Azal dezagun metodo estatistikoko honen prozedura orokorra. Behatutako datuen lagin batetik abiatuz, estatistiko baten balioa kalkulatzeko ondorengo urratsak jarraituko lirateke:

- (i) Behatutako lagineko datuak erabiliz, lagin berriak sortu (behatutako datuen laginarekin tamaina bereko laginak sortu ohi dira), lagin bakoitzean datuak errepika daitezkeela onartuta (hau da, lagineko indibiduo bakoitza behin baino gehiagotan izan daiteke hautatua).
- (ii) Lagin bakoitzarekin intereseko estatistikoaren balioa kalkulatu.

Estatistikoaren banaketaren estimazio bat lortzen da modu honetan. Beraz, behatutako datuen lagin baten eta hortik sortutako lagin berrien bidez, estatistikoaren banaketaren itxura estima daiteke.

Gure kasuan, gaixotasunaren progresio denborarako konfiantza-tarteak estimatzeko erabiliko dugu *bootstrapa*, ondoren, denboraren menpeko gaixotasunaren progresio kurbaren ziurgabetasuna estimatzeko eta konfiantza-tartea irudikatzeko.

Behatutako datuetatik abiatuz n tamainako B lagin sortuko ditugu, n gure lagineko indibiduo kopurua izanik eta B *bootstrap*eko iterazio kopurua. Beste modu batera esanda, B iterazioetako bakoitzean n tamainako lagin berri bat sortuko dugu. Lagin berri bakoitzak n indibiduoren datuak izango ditu beraz, eta indibiduo baten datuak behin baino gehiagotan ager daitezke (errepikapenak onartzen dira).

Aztergai den erantzun aldagai bakoitzerako $(\mathbf{t}_1, \mathbf{y}_1), (\mathbf{t}_2, \mathbf{y}_2), \dots, (\mathbf{t}_n, \mathbf{y}_n)$ bikoteak izango ditugu, \mathbf{t}_i lagineko i . indibiduoari dagokion denboraren bektorea eta \mathbf{y}_i lagineko i . indibiduoari dagokion erantzun aldagaiaren bektorea izanik. Horrez gain, y_{ij} eta t_{ij} i . indibiduoaren j . neurketako erantzun aldagaiaren balioa eta j . neurketa egindako unea izango dira, hurrenez hurren, $i \in \{1, \dots, n\}$ eta $j \in \{1, \dots, n_i\}$ izanik.

*Bootstrap*eko iterazio bakoitzean sortutako laginarekin, eta baita gure laginarekin ere, aurreko 1.2 atalean azaldutako metodoa jarraitzen da gaixotasunaren progresioaren azaltzailea den aldagai bakoitzarekin: indibiduo bakoitzaren datuei doitutako erregresio lineal bakuneko eredu baten bidez,

denborarekiko malda eta batezbestekoen estimazioak kalkulatzeko dira lehenik; ondoren, hirugarren mailako polinomio monotono bat doitzen zaie datu bikote horiei, eta azkenik, polinomio honen alderantzizkoa integratzen da. Gaixotasunaren progresio denbora, τ , gaixotasunaren progresioaren, $\mu(\tau)$, menpe lortzen da horrela.

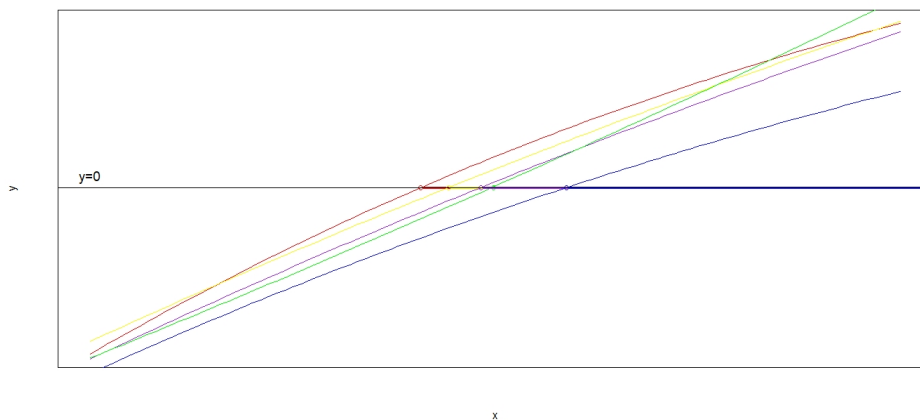
Beraz, *bootstrap* iterazio beste polinomio eta integral izango ditugu. Erantzun aldagaiaren balioen arabera, itzarotako balioei eta maldei doitutako polinomioa positiboa edo negatiboa den eremua aukeratu behar dela gogoratu behar dugu. Hau bermatzeko bi aukera izango ditugu:

- a) Batetik, hirugarren mailako polinomio bakoitza positiboa (edo negatiboa) den eremua kalkulatu dugu, ondoren integrala eremu horretan ebaluatzeko. Hau da, polinomioa positiboa (edo negatiboa) egiten duen aldagai askearen lehenengo balioarekin geldituko gara eta balio hori izango da polinomio honen alderantzizkoaren integrala ebaluatuko dugun lehenengo puntua.

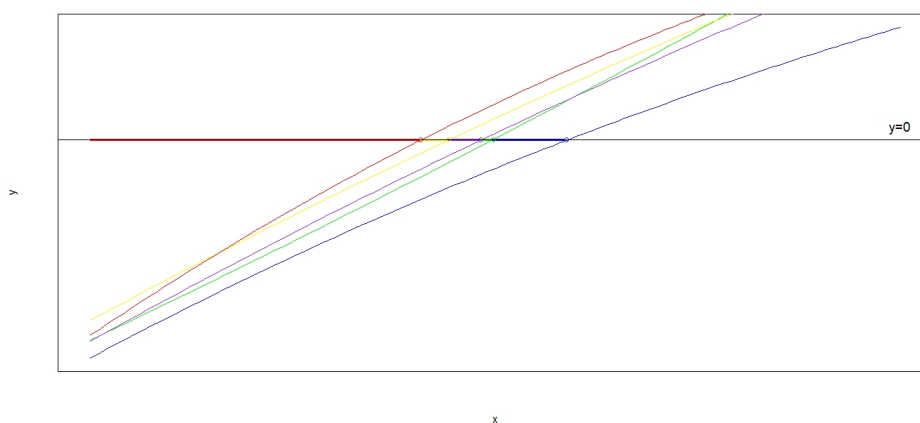
Lagin bakoitzerako bi datu zutabe berri izango ditugu: bat, gaixotasunaren progresioaren balioei dagokiona eta bestea, aldiz, denboraren balioei dagokiona. Progresioaren balio horiek izango dira polinomioaren alderantzizkoaren integraleko aldagai askea eta bertan ebaluatuta, denboraren balioak lortuko ditugu (gaixotasunaren progresioaren lehenengo puntua polinomioa positiboa (edo negatiboa) den eremuko lehenengo puntua izango da, eta ondorengoak puntu horretatik eskuinetarakoak edota ezkerretarakoak). Gure laginarekin lortutako polinomioarekin ere prozesu bera jarraituko dugu eta dagokion eremu positiboan (edo negatiboan) ebaluatuko dugu bere alderantzizkoaren integrala.

Beraz, *bootstrap* iterazio bakoitzeko datu zutabe bikote desberdin bat lortuko dugu, polinomio bakoitzaren eremu positiboa edo negatiboa desberdina izango baita eta ondorioz, baita progresioaren eta denboraren datu zutabeak ere. Hurrengo urratsa, gaixotasunaren progresio denbora koordenatu jatorrira transladatzea da, etorkizuneko denbora hau 0an hastea nahi baitugu. Horretarako, aldagai aldaketa bat egingo dugu.

Hurrengo 1.4. eta 1.5. irudietan ikus dezakegu *bootstrapa* erabiliz sortutako bost lagini doitutako polinomioetako bakoitza positiboa edo negatiboa den eremua. Polinomio hauetako bakoitza positiboa (edo negatiboa) den eremua desberdina da.



1.4. Irudia. Polinomio bakoitza positiboa den eremua



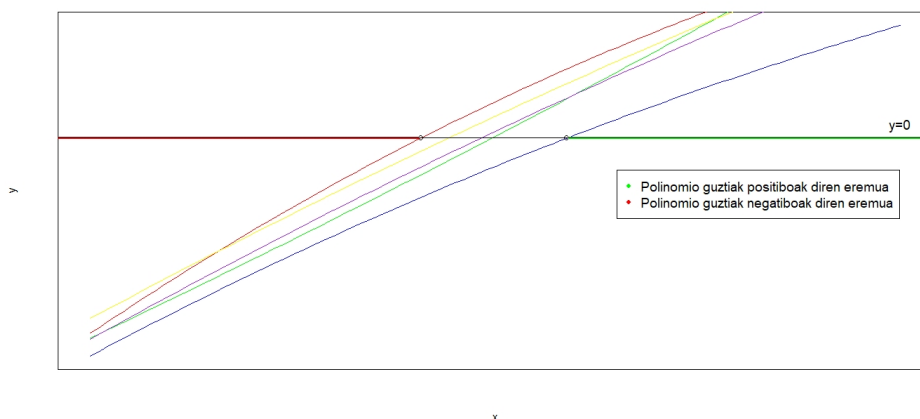
1.5. Irudia. Polinomio bakoitza negatiboa den eremua

- b) Bestetik, hirugarren mailako polinomio guztiak, kasuan kasu, positiboak edo negatiboak diren eremua kalkulatu dugu, hau da, polinomio bakoitza positiboa edo negatiboa den eremua kalkulatu dugu lehenik eta hauen ebakidura ondoren.

Horretarako, polinomio hauetako bakoitza positiboa (edota negatiboa) egiten duen lehenengo balioa lortuko dugu eta jarraian balio hauen artean maximoa edo minimoa hartuko dugu. Balio maximo horretatik aurrera, polinomio guztiak positiboak direla ziurtatzen dugu. Balio minimoetik beherako eremua hartuta, berriz, polinomio guztiak negatiboak

direla ziurtatzen dugu. Berrito, gaixotasunaren progresioaren balioak integralean ebaluatuta denboraren balioak lortuko ditugu eta progresioaren lehenengo puntua aurretik kalkulaturako balio maximo (edo minimo) hori izango da, eremu guztien ebakidurako baliorik handiena (edo txikiena), hain zuzen ere. Gure laginarekin lortutako polinomioaren alderantzizkoaren integrala ebaluatzeko lehenengo puntua ere balio maximo (edo minimo) hori izango da.

Kasu honetan, orduan, lagin bakoitzerako datu zutabe berri bat izango dugu, denborari dagokiona. Izan ere, gaixotasunaren progresioaren balioak berdinak izango dira lagin guztietarako, integral guztiak eremu berean ebaluatzen baititugu. Hurrengo urratsa, gaixotasunaren progresio denbora koordinatu jatorrira transladatzea da, etorkizuneko denbora hau 0an hastea nahi baitugu. Horretarako, aldagai aldaketa bat egingo dugu.



1.6. Irudia. Polinomio guztiak positiboak edo negatiboak diren eremuak

Polinomio guztiak positiboak edo negatibok diren eremuak, hau da, eremu positibo edo negatibo guztien ebakidura, ikus dezakegu 1.6. irudian.

Azken hiru irudi hauei erreparatzen badiegu, diferentzia nabarmenak ikus ditzakegu eremuen artean, integralak ebaluatzeko eremuak erabat desberdinak dira. Lehenengo bi irudietan (1.4. eta 1.5. irudiak) abzisa ardatza eta polinomioa mozten diren puntuaren ezker (eremu negatiboa) edota eskuin aldeko (eremu positiboa) eremuarekin gelditzen gara. Hirugarrenean (1.6. irudia), berriz, mozketa puntu horien guztien artean minimoaren ezkerretarako eremuarekin (eremu negatiboekin ari bagara) edo maximoaren eskuin aldekoarekin (eremu positiboekin kasuan) gelditzen gara.

Bigarren aukera erabili dugu gure ikerketako erantzun aldagaiekin. Izan ere, polinomio guztientzat eremu berdinak mantendu nahi izan ditugu, polinomio guztiak eremu berean ebaluatu eta eskala berdinak mantentzeko kasu guztietan.

Beraz, progresioaren puntu bakoitzeko B denboraren puntu izango ditugu, iterazio eta polinomio bakoitzeko gaixotasunaren progresio denboraren balio bat lortuko baitugu.

Ondoren, $\mu(\tau)$ -ren balio bakoitzerako τ -ren konfiantza-tarteak kalkulatzeko pertzentilen metodoa erabiliko dugu. Progresioaren balio bakoitzeko B denboraren balio ditugu, B tamainako datu bektore bat eta, beraz, balio horien P2,5 eta P97,5 (2,5. pertzentila eta 97,5. pertzentila) kalkulatu ditugu. Datuen %2,5 uzten du bere azpitik 2,5. pertzentilak eta 97,5. pertzentilak, berriz, datuen %2,5 uzten du bere gainetik. Progresioaren balio bakoitzeko denboraren %95eko konfiantza-tartea lortuko dugu modu honetan.

Horrela, $\mu(\tau)$ -ren balio bakoitzeko τ -ren hiru balio izango ditugu: puntu-estimazioa eta %95eko konfiantza-tartearen goi- eta behe-mugak.

Aurreko 1.2 atalean azaldutako metodoaren azken bi urratsetako kurbak irudikatu behar dira jarraian. Batetik, aldagai asketzat gaixotasunaren progresioa, $\mu(\tau)$ (abzisa ardatza), eta menpeko aldagaitzat denbora, τ (ordenatu ardatza), hartuko ditugu. Gaixotasunaren progresioaren puntu bakoitzeko denboraren hiru balio izango ditugunez, puntu horiek elkartuz hiru kurba lortuko ditugu: gure laginari dagokion kurba eta konfiantza-tartearen goi eta behe limitei dagokien kurba, progresioaren puntu bakoitzeko lortutako denboraren P2,5 eta P97,5 balioak elkartuz lortutakoak. Modu horretara estimatutako kurbaren %95eko konfiantza-banda lortzen da. Bestetik, gaixotasunaren progresioa denboraren menpe lortzeko, alderantzikatu egingo ditugu bi aldagaiak, denbora izango da aldagai askea eta gaixotasunaren progresioa, berriz, menpeko aldagaia. Aurreko kasuan bezala, hiru kurba izango ditugu kasu honetan ere: gure laginari dagokion gaixotasunaren progresio kurba eta konfiantza-tartearen goi- eta behe-mugei dagokien kurba. Azken kurba hauen bidez, gaixotasunaren progresioa denboraren menpe eta kurba horren ziurgabetasunaren %95eko konfiantza-tartea lortuko ditugu.

2. Kapituluia

Aplikazioa

2.1 Biriketako gaixotasun buxatzaile kronikoa

Biriketako gaixotasun buxatzaile kronikoa, BGBK (gazteleraz *enfermedad pulmonar obstructiva crónica, EPOC*), birikei eragiten dien eta arnasbidearen hantura kronikoa dakarren gaixotasuna da. Hantura horrek hertsadura sorrarazten eta airea biriketatik pasatzea eragozten du. Espainian populazio helduaren %10i eragiten dion gaixotasuna da, gutxi gorabehera. Hiru milioitik gora lagun hil ziren munduan gaixotasun honen eraginez 2015ean, urte honetako heriotzen %5 inguru. Sendatu ezin den gaixotasuna izan arren, bere aurrerabidea atzeratu dezaketen tratamenduen bidez tratatu daiteke eta estrategia terapeutikoez gaixotasunaren hasierako etapatan eragin handiagoa izan dezakete.

BGBK pairatzen duten gaixoen datu gutxi (eta epe motzekoak) izan ohi dira. Hori dela eta, ez da gaixotasunaren bilakaera osoa ikusten ikerketetan eta datuetan. Gaixotasunaren progresioaren itxura edo forma sigmoidala dela onartzen da, hau da, lehenengo etapan progresio horren hazkuntza txikia eta motela da; bigarren etapan hazkuntza handia da, eta azkenik, hazkuntza hori geldituz doa.

Denborak aurrera egin ahala okertzen doan eta arnasketa zailtasunak dakartzan gaixotasun progresiboa da. Sintomak ere okertu egiten dira denborarekin eta egunerokotasuneko jarduerak galarazten ditu. Horregatik, garrantzitsua izan daiteke bere progresioa ezagutzea, zenbat denbora behar duen gaixotasunak etapa batetik bestera progresatzeko, eta zein abiaduratan eta nola eragiten dien gaixotasun hau pairatzen duten indibiduoek jakiteko.

Lehenengo kapituluian azaldutako metodoa erabiliko dugu biriketako gaixotasun buxatzaile kronikoaren progresioa auresateko eta denboraren menpeko progresioaren kurba estimatzeko.

Gaixotasunaren progresioaren adierazle diren lau aldagai izango ditugu aztergai lan honetan: ariketa ahalmena, birikien funtzioa, kuadripzeseko indarra eta disnea. Denborak aurrera eta gaixotasunak okerrera egin ahala, aldagai hauetako bakoitzak izango duen progresioa aztertzea eta estimatzea izan da gure helburua.

2.2 Programak eta inplementazioa

Bi programa erabili ditugu lana inplementatzeko: batetik, R eta bestetik, *Mathematica*.

Kalkulu estatistikoak egiteko R erabili dugu. Hasteko, indibiduoari erregresio lineal bakuneko ereduak doitzeko eta eredu horiek erabiliz kalkulaturako itxarotako balioei eta eredu horien denborarekiko maldei, hirugarren mailako polinomio monotono bat doitzeko erabili dugu programa hau. Aipaturako datu bikoteari polinomioa doitzeko, `MonoPoly` paketeko `monpol` funtzioa erabili dugu [10]. Funtzio horrek maila bakoitiko polinomio bat doitzen die bi datu zutaberi edo bi datu bektoreri. Horretarako $y \sim x$ formula erabiltzen da. Lortu nahi dugun polinomioaren maila ezar dezakegu eta hau `degree` argumentuaren bidez egiten da. Pakete bereko `ismonotone` funtzioaren bidez datuei doitutako polinomioa monotonoa dela egiazta dezakegu.

Bootstrap egiteko ere programa hau erabili dugu. Lagin-tamaina n duten B lagin sortzeko `rsample` paketeko `bootstraps` funtzioa erabili dugu [6]. Funtzio horren bidez, gure laginaren tamaina bereko lagin berriak sortzen dira eta indibiduo bakoitza behin baino gehiagotan ager daiteke lagin berean (hau da, indibiduoaren errepikapenak onartzen dira). Iterazio kopurua ezar daiteke `times` argumentuaren bidez. Ondoren, lagin hauetako bakoitzari polinomio monotono bat doitu diogu.

Hirugarren mailako polinomioen alderantzizkoen integralak kalkulatzeko, berriz, *Mathematica* programa erabili dugu. Baita polinomio bakoitzaren eremu positiboa (edo negatiboa, kasuan kasu) eta polinomio guztiak positiboak (edo negatiboak) diren eremua kalkulatzeko, eta integralak eremu horretan ebaluatzeko ere. Polinomioen alderantzizkoen integralak kalkulatzeko `Integrate` funtzioa erabili dugu. Behin polinomioak positiboak (edo negatiboak) diren eremuak kalkulatu eta integral bakoitza dagokion eremuan ebaluatuta, gaixotasunaren progresioaren eta denboraren datu zutabeak ere bertan kalkulatu ditugu. Gaixotasunaren progresioaren puntu bakoitzeko *bootstrap* iterazio beste denboraren balio lortuko ditugu modu honetan.

Azkenik, progresioaren puntu bakoitzerako denboraren konfiantza-tarteak lortzeko, pertzentilen metodoa erabili dugu, aipatu bezala. Horretarako,

denboraren datu zutabeen P2,5 eta P97,5 kalkulatu ditugu (2,5. pertzentila eta 97,5. pertzentila), *R* programa erabiliz. Lehenengo kapituluko 1.2 atalean proposatutako metodoaren azken bi urratsetako kurbak eta kurba horien %95eko konfiantza-tarteak irudikatzeko ere programa hau erabili dugu.

Horrez gain, analisi estatistikoak egiteko eta datubasea deskribatzeko ere *R* programa erabili dugu. Batezbestekoak, desbideratze estandarrak, maiztasunak, ehunekoak eta kuartilen arteko heinak kalkulatzeko erabiltzeaz gain, normaltasunerako, bariantzen homogeneotasunerako eta batezbestekoen arteko berdintasunerako hipotesi kontrasteak egiteko ere erabili dugu. Azken hipotesi kontraste hauek egiteko, Shapiro testa, `shapiro.test` funtzioa, Levene testa, `leveneTest` funtzioa, eta Kruskal-Wallis-en testa eta ANOVA erabili ditugu, `kruskal.test` eta `aov` funtzioak.

2.3 Datubasearen deskribapena

Ikerketa honetan parte hartu duten indibiduoak 2003ko urtarrila eta 2004ko urtarrila bitartean Galdakao-Usansolo Ospitaleko arnasketa zerbitzuarekin lotutako bost kliniketako batean tratatuak izandako gaixoak dira. Gaixotasuna diagnostikatuta sei hilabete edo gehiago eta ikerketako lehenengo neurketa egin aurretik gutxienez sei aste egoera egonkorrean zeramatzaten gaixoen datuak bildu ziren. Indibiduen lau denboratako datuak ditugu: ikerketa hasitako unekoak eta handik urtebetera, bi urtetara eta bost urtetara bildutakoak (nahiz eta gerta daitekeen, indibiduo batzuen kasuan, hainbat aldagaitarako lau datu bikoteak ez izatea, hau da, balio galduak izatea). Denbora tarte horretan gaixoen historial klinikoak eta ospitalaratzeak ere kontuan hartu ziren [4].

Gure laginean 543 indibiduo ditugu eta lau taldetan banatuta dago lagin hau. Hainbat aldagai izan dira kontuan banaketa hau egiterakoan. Aipa ditzagun, aldagai soziodemografikoez gain, ikerketa honetan erabili ditugun aldagaiak [4].

Aldagai deskribatzaileekin hasiz, Charlson-en indizeak (*Charlson Comorbidity Index*) [3], indibiduo bakoitzak pairatzen dituen gaixotasunak neurtzen ditu. Gaixotasun bakoitzak pisu desberdina du, hau da, gaixotasun larriek arinenek baino gehiago pisatzen dute indize hau kalkulatzeko eta indize honek zenbat eta balio altuagoak hartu, indibiduo gaixogo dagoela adierazten du. Ospitalaratzeak aldagairen bidez aurreko bi urtetan metatutako ospitalaratze episodioak adierazten dira. Ospitalaratzeren bat izan duten indibiduo kopurua zenbatu dugu. Tratamendua aldagaiak zenbaki osoak hartzen ditu eta botika kopuruaren arabera indibiduoek daukaten tratamendua adierazten du. IMC-ren (*índice de masa corporal*) bidez, gorputz-masa

indizea neurtzen da. Taldeen arteko banaketa egiteko erabilitako beste aldagai bat ariketa fisikoa izan da eta indibiduoek beraien denbora librean ibiltzen (oinez) igarotako denbora adierazten du.

Gaixotasunaren progresioaren adierazleak diren lau aldagai izango ditugu aztergai lan honetan. Ariketa ahalmena aldagaia, *6 minute walking test* testaren bidez neurturiko aldagai jarraitua da eta 6 minututan ibilitako distantzia metrotan adierazten du [1]. Kuadrizepseko indarra aldagaiak, kuadrizepsen indarra kilogramotan neurtuta adierazten du [8]. Hurrengo bi aldagaiak arnasketarekin lotutakoak dira. Birikien funtzioa aldagaia, lehenengo segundoan arnas botatzeko bolumen behartuaren (FEV%1 deritzo) bidez, ehunekotan, neurtzen den aldagai jarraitua da [4]. Azkenik, disnea aldagaiak, indibiduen disnea edo arnasestu maila adierazten du eta hau neurtzeko mMRC eskala erabili da [7]. Bat eta bost bitarteko zenbaki oso baten bidez adierazten da. Bat zenbakiak disnea mailarik baxuena eta arinena adierazten du eta bost zenbakiak, berriz, disnea mailarik altuena eta gogorrena.

Lau clusterretan dago banatuta gure populazioa ikerketaren abiapuntuko osasun egoeraren arabera egindako sailkapetan oinarrituta [4]: A clusterra (164 indibiduo), B clusterra (195 indibiduo), C clusterra (89 indibiduo) eta D clusterra (95 indibiduo).

Deskriba ditzagun cluster hauetako bakoitzaren ezaugarri nagusiak.

A clusterra disnea edo arnasestu gutxien duten inbiduoak dituen taldea da, eta talde honetan daude banako gazteenak. Osasunari dagokionez bizi kalitate onena eta giharretako indar eta ariketa ahalmen handiena dutenak dira. Horrez gain, Charlson indizeko baliorik txikienak dituen taldea da.

C clusterrean erretze historial eta disnea gogorrenak dituzten indibiduoak daude, birikien funtzio eta bizi kalitate eskasena dutenak. Giharretako indar eta egoera fisiko eskasena, eta ariketa ahalmen gutxien duten indibiduoak ere talde honetakoak dira. Cluster honetako indibiduoek A clusterreko indibiduen Charlson indizeko balio antzekoak dituzte.

B clusterra, berriz, A eta C clusterren artean kokatzen da. D clusterreko indibiduen arnasketa ezaugarri antzekoak dituzte.

D clusterreko indibiduoak ospitalaratze tasa altuenak dituztenak dira. Nahiz eta B clusterreko indibiduen arnasketa ezaugarri antzekoak izan, ariketa ahalmen eta egoera fisiko eskasagoak dituzte talde honetako banakoek.

Tratamendu medikoari dagokionez, clusterren arteko diferentzia bakarra A clusterreko indibiduoek iraupen luzeko bronkodilatadore gutxiago erabiltzen dituztela da.

Ikus ditzagun, hasteko, ikerketa honetan aztertu diren hainbat aldagaien balioak lau taldeetan.

2.1. Taula. Indibiduen ezaugarri nagusien deskribapena lau clusterretan sailkatuta. Taulan agertzen diren estatistikoak ondokoak dira: aldagai jarraituetarako, batezbestekoa eta desbideratze estandarra eta aldagai diskretuetarako, maiztasuna eta ehunekoa.

Aldagaiak	Kategoriak	A clusterra	B clusterra	C clusterra	D clusterra
Kopurua		164	195	89	95
Adina		63, 41(8, 94)	71, 12(5, 94)	69, 86(8, 01)	69, 62(8, 07)
Sexua	<i>Gizonak</i>	160(%97, 56)	188(%96, 41)	83(%93, 26)	91(%95, 79)
	<i>Emakumeak</i>	4(%2, 44)	7(%3, 59)	6(%6, 74)	4(%4, 21)
Charlson indizea	1	92(%56, 1)	36(%18, 46)	44(%49, 44)	0
	2	46(%28, 05)	88(%45, 13)	28(%31, 46)	0
	3	23(%14, 02)	65(%33, 33)	13(%14, 61)	1(%1, 05)
	4	3(%1, 83)	5(%2, 56)	3(%3, 37)	50(%52, 63)
	5	0	1(%0, 51)	1(%1, 12)	26(%27, 37)
	≥ 6	0	0	0	18(%18, 95)
Ospitalaratzeak		12(%7, 32)	30(%15, 38)	38(%42, 7)	36(%37, 89)
Tratamendua	0	13(%7, 93)	3(%1, 54)	3(%3, 37)	5(%5, 26)
	1	19(%11, 58)	16(%8, 2)	6(%6, 74)	8(%8, 42)
	2	44(%26, 83)	38(%19, 49)	18(%20, 22)	21(%22, 11)
	3	88(%53, 66)	138(%70, 77)	62(%69, 66)	61(%64, 21)
IMC		28, 27(3, 73)	28, 38(4, 15)	26, 65(4, 88)	29, 6(5, 19)

Ondoko 2.1. taulan indibiduen hainbat ezaugarriren deskribapena egiten da, indibiduoak aipatutako lau clusterretan sailkatuta.

Ikus ditzagun, orain, lau taldeen arteko diferentziak adierazgarriak diren gaixotasunaren progresioaren adierazle eta lan honetan aztergai diren lau aldagaietan. Horretarako, neurketa bakoitzeko aldagai hauen batezbestekoen berdintasunerako hipotesi kontrastea planteatu dugu.

Lehenik, aldagaien normaltasuna eta bariantzen homogeneotasuna begiratu behar ditugu. Shapiro testa erabili dugu normaltasuna aztertzeko eta Levene testa, berriz, bariantzen homogeneotasuna aztertzeko. Aldagaien normaltasuna bermaturik ez dugunean, hondarren normaltasuna begiratu dugu.

Honakoak dira normaltasunerako eta bariantzen homogeneotasunerako planteatutako hipotesi kontrasteak:

$$\begin{cases} H_0 : Y \simeq N(\mu, \sigma) \\ H_1 : Y \not\simeq N(\mu, \sigma) \end{cases} \quad \begin{cases} H_0 : \sigma_A = \sigma_B = \sigma_C = \sigma_D \\ H_1 : \exists i, j \in \{A, B, C, D\} / \sigma_i \neq \sigma_j \end{cases} \quad (2.1)$$

Progresioaren adierazleak diren aldagaien lau neurketen kasuan normaltasuna bermatuta dugun kasu bakarrak hirugarren eta laugarren neurketetako birikien funtzioari dagozkion aldagaiak dira (YFEV1_, ZFEV1_).

Hondarren normaltasuna begiratu dugu, beraz. Kasu honetan, hondarren normaltasuna bermatuta dugun aldagaiak, aurretik aipatutako bi aldagaiez gain, lehenengo eta bigarren neurketetako birikien funtzioari dagozkion aldagaiak (WFEV1_, XFEV1_) eta hirugarren eta laugarren neurketetako kuadrupeseko indarrari dagozkion aldagaiak dira (YFcuad, ZFcuad).

Bariantzen homogeneotasunari dagokionez, aipatutako aldagaietan bermatuta dugu baldintza hau, birikien funtzioari dagokion lehenengo neurketeko aldagaien izan izak (WFEV1_).

Neurketa bakoitzeko aldagaien batezbestekoen berdintasunerako hipotesi kontrastea planteatzeko bi metodo erabili ditugu, normaltasuna eta bariantzen homogeneotasuna bermatuta ditugun edo ez kontuan hartuta. Bi baldintza horiek betetzen diren kasuetan *ANOVA* erabili dugu, eta gainontzekoetan, berriz, Kruskal-Wallis-en testa.

Honakoa da planteatutako hipotesi kontrastea:

$$\begin{cases} H_0 : \mu_A = \mu_B = \mu_C = \mu_D \\ H_1 : \exists i, j \in \{A, B, C, D\} / \mu_i \neq \mu_j \end{cases} \quad (2.2)$$

Ikus ditzagun aldagai eta neurketa bakoitzaren kasurako lortutako batezbestekoak eta desbideratze estandarrak. Baita kasu bakoitzerako dagokion testaren bidez batezbestekoen berdintasunerako hipotesi kontrastea planteatuta lortutako emaitzak ere.

Gaixotasunaren progresioaren adierazle diren aldagaien deskribapena egiten da, beraz, 2.2. taulan, indibiduoak lau clusterretan sailkatuta.

Aztergai diren aldagaiak izendatzeko hainbat laburdura erabili ditugu 2.2. taulan. Ariketa laburduraren bidez, ariketa ahalmena aldagaia azaldu nahi da, indarraren bidez kuadrupeseko indarra aldagaia eta birikiak hitzaren bidez, berriz, birikien funtzioa aldagaia. Batezbestekoen berdintasunerako hipotesi kontrastea planteatuta lortutako p-balioa adierazten du taulan agertzen den p-balioak.

2.2. Taula. Indibiduen gaixotasunaren progresioa adierazten duten aldagaien deskribapena lau clusterretan sailkatuta. Taulan ageritzen diren estatistikoak batezbestekoa eta desbideratze estandarra dira. Clusterren arteko diferentziak adierazteko erakusten den p-balioa bariantza-analisia edo Kruskal-Wallis testa erabiliz kalkulatu da.

Aldagaiak	A clusterra	B clusterra	C clusterra	D clusterra	p-balioa
<i>Lehenengo neurketako datuak</i>					
Kopurua	164	195	89	95	
Ariketa	484,05(59,8)	403,2(67,95)	339,81(99,6)	355,52(83,18)	< 0,0001
Indarra	35,63(7,57)	28,42(8,05)	24,4(8,67)	27,19(8,95)	< 0,0001
Birikiak	59,54(11,06)	56,18(12,9)	42,98(12,59)	56(11,89)	< 0,0001
Disnea	1,86(0,63)	2,35(0,73)	3,28(1,03)	2,48(0,78)	< 0,0001
<i>Bigarren neurketako datuak</i>					
Kopurua	157	177	71	75	
Ariketa	504(62,69)	412,55(83,7)	358,28(117,47)	373(96,96)	< 0,0001
Indarra	34,85(8,31)	27,5(8,75)	24,2(10,17)	27,66(8,96)	< 0,0001
Birikiak	59,8(13,54)	57,37(13,7)	42,47(13,06)	56,8(13,99)	< 0,0001
Disnea	1,78(0,6)	2,2(0,71)	3(1,01)	2,28(0,76)	< 0,0001
<i>Hirugarren neurketako datuak</i>					
Kopurua	149	156	61	62	
Ariketa	491,72(67,83)	395,3(91,1)	335,48(128,85)	359,81(107,34)	< 0,0001
Indarra	32,67(8,65)	25,36(8,82)	23,29(10,5)	25,96(9,36)	< 0,0001
Birikiak	61,89(14,25)	58,88(12,56)	43,1(13,34)	59,18(13,7)	< 0,0001
Disnea	1,85(0,72)	2,35(0,72)	3,13(1)	2,37(0,85)	< 0,0001
<i>Laugarren neurketako datuak</i>					
Kopurua	137	114	39	34	
Ariketa	469,02(87,31)	358,28(111,32)	353,38(106)	343,91(93,75)	< 0,0001
Indarra	29,83(9,58)	21,07(9,96)	20,58(7,72)	22,03(9,22)	< 0,0001
Birikiak	55,99(14,46)	54,39(13,69)	44,5(13,36)	56,59(15,96)	0,000157
Disnea	1,98(0,83)	2,6(0,94)	2,92(0,96)	2,47(0,9)	< 0,0001

Kasu guztietan p-balioa 0,05 baino txikiagoa denez, H_0 errefusatzeko dugu 0,05eko adierazgarritasun-mailaz eta, beraz, aztergai diren aldagaien lau neurketetako eta lau taldeetako batezbestekoen arteko ezberdintasunak adierazgarriak direla esan genezake.

Ikus dezagun, orain, hainbat estatistikok gaixoen lau taldeetan hartzen dituzten balioak disnea aldagaiari dagokionez.

Izan ere, disnea aldagaia indibiduen disnea edo arnasestu maila neurtzen duen aldagai ordinala da. Nahiz eta ez izan aldagai jarraitua, klinikan maiz hartu ohi dira aldagai ordinalak jarraitu modura. Beraz, aldagai honen jarraitasuna onartuko dugu denboran zehar dituen aldaketak ikusteko. Agian

ez da aldagai hau aztertzeko modurik egokiena jarraitu modura hartzea, baina klinikoek hala esan digute egiteko. Hori dela eta, aldagai kategoriko modura ere deskribatu nahi izan dugu.

2.3. Taula. Disnea aldagaiaren deskribapena indibiduoak lau clusterretan sailkatuta. Taulan disnea aldagairako mediana, kuartil arteko heina eta kategoria bakoitzerako maiztasuna eta ehunekoa erakusten dira.

	Kategoriak	A clusterra	B clusterra	C clusterra	D clusterra
<i>Lehenengo neurketako datuak</i>					
Kopurua		164	195	89	95
Maiztasunak	1	42(%25,61)	19(%9,74)	1(%1,12)	7(%7,37)
	2	105(%64,02)	97(%49,74)	19(%21,35)	43(%45,26)
	3	16(%9,76)	72(%36,92)	39(%43,82)	39(%41,05)
	4-5	1(%0,61)	7(%3,59)	30(%33,71)	6(%6,32)
Mediana		2(1)	2(1)	3(1)	2(1)
<i>Bigarren neurketako datuak</i>					
Kopurua		157	177	71	75
Maiztasunak	1	49(%31,21)	25(%14,12)	1(%1,41)	10(%13,33)
	2	93(%59,24)	95(%53,67)	23(%32,39)	37(%49,33)
	3	15(%9,55)	54(%30,51)	32(%45,07)	26(%34,67)
	4-5	0	3(%1,69)	15(%21,13)	2(%2,67)
Mediana		2(1)	2(1)	3(1)	2(1)
<i>Hirugarren neurketako datuak</i>					
Kopurua		149	156	61	62
Maiztasunak	1	48(%32,21)	17(%10,9)	1(%1,64)	9(%14,52)
	2	77(%51,68)	71(%45,51)	14(%22,95)	26(%41,93)
	3	23(%15,44)	65(%41,67)	31(%50,82)	23(%37,1)
	4-5	1(%0,67)	3(%1,92)	15(%24,59)	4(%6,45)
Mediana		2(1)	2(1)	3(0)	2(1)
<i>Laugarren neurketako datuak</i>					
Kopurua		137	114	39	34
Maiztasunak	1	41(%29,93)	8(%7,02)	2(%5,13)	6(%17,65)
	2	62(%45,25)	52(%45,61)	11(%28,21)	9(%26,47)
	3	31(%22,63)	37(32,46)	16(%41,02)	16(%47,06)
	4-5	3(%2,19)	17(%14,91)	10(%25,64)	3(%8,82)
Mediana		2(1)	2(1)	3(1,5)	3(1)

Indibiduoak lau clusterretan sailkatuta disnea aldagaiaren (aldagai kategoriko moduan hartuta) deskribapena egiten da 2.3. taulan.

2.4 Gaixotasunaren progresioa

Azal ditzagun, lehenik, gure erantzun aldagaietan 1.2 atalean azaldutako metodoa eta *bootstrap* aplikatzerakoan kontuan hartutako hainbat puntu.

Hasteko, indibiduo bakoitzaren erantzun aldagaiaren eta denboraren datu bikoteei erregresio lineal bakuneko eredu bat doitzerako orduan, kontuan izan dugu indibiduo bakoitzaren datu kopurua. Datu bakarra duten indibiduoak baztertu egin ditugu eta bi, hiru edo lau datu dituzten indibiduoek datu kopuru hori erabiliz doitu diegu erregresio linealeko eredu.

Erregresio lineal bakuneko eredu bidez lortutako itzarotako balioei eta eredu horien denborarekiko maldei hirugarren mailako polinomio monotono bat doitu eta gero, polinomio honen eremu positiboa edo negatiboa aukeratu dugu. Gogora dezagun hautaketa hori garrantzitsua dela aldagai bakoitzaren eboluzioa ezagutzeko eta erantzun aldagaiaren arabera dela: denborak eta gaixotasunak okerrera egitean erantzun aldagaia beherakorra bada, eremu negatiboarekin gelditu gara (ariketa ahalmena, kuadripzeseko indarra eta birikien funtzioa aldagaien kasuan), eta erantzun aldagaia gorakorra bada, berriz, polinomioaren eremu positiboarekin gelditu gara (disnea aldagaiaren kasua).

Aldagai bakoitzarekin gure laginarentzat gaixotasunaren progresio kurbaren eta kurba horren %95eko konfiantza-tartearen estimazioak kalkulatu nahi izan ditugu. Horretarako, alde batetik, gure laginaren progresio kurba lortzeko 1.2 atalean azaldutako metodoaren urratsak jarraitu ditugu. Bestalde, progresio kurbaren %95eko konfiantza-tartea lortzeko *bootstrap* metodo estatistikoa erabili dugu, 1.3 atalean azalduta bezala. Horretarako, $B=1000$, iterazio kopurua, eta $n=543$, lagin-tamaina, hartu ditugu, hau da, 543 indibiduo dituzten 1000 lagin sortu ditugu (indibiduoaren errepikapenak onartuta). *Bootstrap*eko iterazio bakoitzean sortutako laginarekin, 1.2 atalean azaldutako metodoa errepikatu dugu.

Gaixotasunaren progresio kurba eta bere konfiantza-tartea irudikatzeko, *bootstrap*eko iterazio bakoitzean sortutako laginari doitutako polinomioaren alderantzizkoaren integrala ebaluatzeke, bi aukera posible ditugu 1.3 atalean azaldutako era berean: bat, polinomio hori positiboa (edo negatiboa) den eremuan ebaluatzea; eta bi, polinomio guztiak positiboak (edo negatiboak) diren eremuan ebaluatzea, hau da, eremu guztien ebakiduran. Bigarren aukera hau erabili dugu guk. Izan ere, polinomio guztientzat eremu berdinak mantendu nahi izan ditugu, polinomio guztiak eremu berean ebaluatu eta eskala berdinak mantentzeko kasu guztietan.

Beraz, *bootstrap* iterazio beste polinomio eta integral izango ditugu eta progresioaren puntu bakoitzerako B denboraren puntu lortuko ditugu. Gaixotasunaren progresio kurbaren %95eko konfiantza-tartea lortzeko, azkenik,

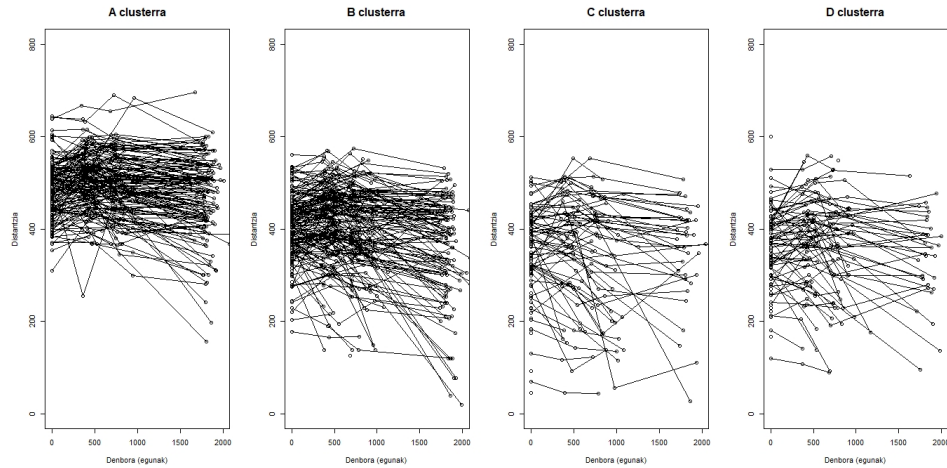
pertzentilen metodoa erabili dugu, 1.3 atalean azaldutako metodoa jarraituz.

Azter dezagun aldagai bakoitza eta ikus ditzagun lortutako emaitzak.

1 Ariketa ahalmena

Ariketa ahalmena *6 minute walking test* testaren bidez neurturiko aldagai jarraitua da [1].

Ikus dezagun aldagai honek denboran zehar lau taldeetan hartze dituen balioak:



2.1. Irudia. Cluster bakoitzeko indibiduen ikerketa denbora bakoitzeko ariketa ahalmenaren balioak.

Talde bakoitzeko indibiduen ariketa ahalmenaren balioak denboran zehar nola aldatzen diren ikus dezakegu 2.1. irudian. Taldeak alderatuz, A clusterrako indibiduoek ariketa ahalmenaren balio altuenak dituzte eta orokorrean denborak aurrera egin ahala, txikitzen doaz erantzun aldagai honen balioak. Hau da, gaixotasunaren progresioak aurrera egin ahala aldagai honek gero eta balio txikiagoak hartzen ditu.

Aldagai asketzat ikerketa denbora eta menpeko aldagaitzat ariketa ahalmena hartuta lehenengo kapituluko 1.2 atalean azaldutako metodoa jarraitu dugu gaixotasunaren progresio kurba ezagutzeko. Gaixotasunak okerrera egitean erantzun aldagai hau beherakorra denez, hirugarren mailako polinomioaren eremu negatiboarekin gelditu gara.

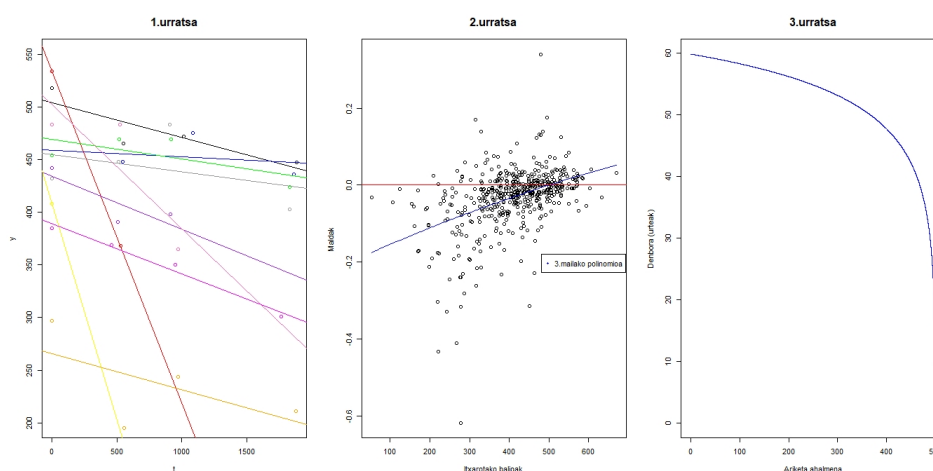
Honakoak dira lehenengo hiru urratsetan lortutako emaitzak eta kurbak:

- Hirugarren mailako polinomioa:

$$p(x) = 1,679 \cdot 10^{-11}x^3 - 1,542 \cdot 10^{-7}x^2 + 4,721 \cdot 10^{-4}x - 0,1997$$

- Hirugarren mailako polinomioaren alderantzizkoaren integrala:

$$\int \frac{1}{p(x)} dx = 5240,79 \cdot \arctan(0,00225381(-8683,73 + 2x)) \\ + 3026,51 \cdot \log(500,306 - x) - 1513,26 \cdot \log(2,37734 \cdot 10^7 - 8683,73x + x^2)$$



2.2. Irudia. 1.urratsa - Hainbat indibiduoren denboraren eta ariketa ahalmena aldagaiaren datuen puntu-hodeia eta doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Erregresio lineal bakuneko ereduetatik lortutako itzarotako balioak eta denborarekiko maldak, eta datu bikote horiei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala, gaixotasunaren progresioaren menpeko kurba.

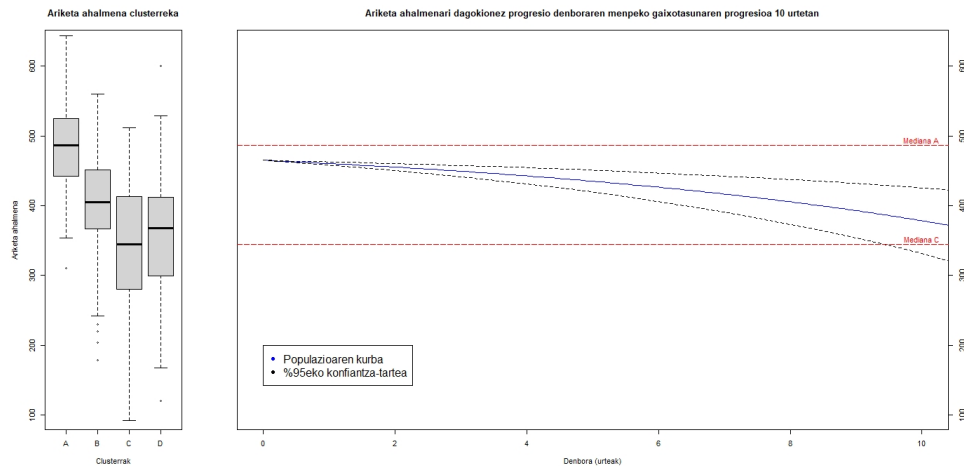
Gaixotasunaren progresio kurba estimatzeko erabilitako metodoaren lehenengo hiru urratsak ikus ditzakegu 2.2. irudiko hiru grafikoetan.

Lehenengo urratsari dagokion grafikoak, hainbat indibiduoren erantzun aldagaiaren eta denboraren datuen puntu-hodeia eta datu horiei doitutako erregresio lineal bakuneko ereduak erakusten ditu. Zuzen hauek beherakorrak dira, malda negatiboa dute, denborak aurrera egitean aldagaiaren balioak txikitzen baitoaz. Bigarren grafikoan eredu horien denborarekiko maldak eta itzarotako balioak, eta datu bikote horiei doitutako polinomioa ikus ditzakegu. Azken grafikoan, polinomioaren alderantzizkoaren integralaren kurba irudikatu dugu, $(\hat{\mu}(\tau), \tau)$ bikoteak, hain zuzen ere. Polinomioaren

eremu negatiboa hautatu dugu urrats honetarako. Horrela, progresioaren menpekoa den kurba bat lortu dugu.

Denboraren menpeko gaixotasunaren progresioa ezagutzeko eta auresateko, hirugarren urratsean lortutako bi aldagaiak alderantzikatu ditugu eta horrela, 1.2 atalean azaldutako metodoaren azken urratseko kurba irudikatu dugu. Horrez gain, estimatutako progresio kurbaren %95eko konfiantza-tartea lortzeko *bootstrap*etik eta pertzentilen metodotik lortutako emaitzak erabili ditugu, 1.3 atalean azaldutako urratsak jarraituz.

Ikus ditzagun gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak:



2.3. Irudia. Lehenengo neurketako ariketa ahalmena aldagaiaren datuen kutxa diagramak clusterreka banatuta eta gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak.

Ondoko 2.3. irudiaren ezker aldean lehenengo neurketako ariketa ahalmenena aldagaiaren datuekin ikerketa talde bakoitzerako lortutako kutxa diagramak ikus ditzakegu, eta eskuin aldeko grafikoan, berriz, ariketa ahalmenari dagokionez gaixotasunaren progresioa 10 urtetan eta kurba horren ziurgabetasunaren konfiantza-tartea. A taldeko indibiduoek dituzte daturik onenak, ariketa ahalmenaren baliorik altuenak, eta C taldekoek, berriz, okerrenak. Hori dela eta, bi talde hauen medianak irudikatu ditugu lerro horizontalean gaixotasunaren progresio kurbarekin batera, etapa batetik bestera igarotzeko beharrezkoa den batez besteko denbora ezagutzeko eta aldaketa hori zein abiaduratan egingo den ikusteko.

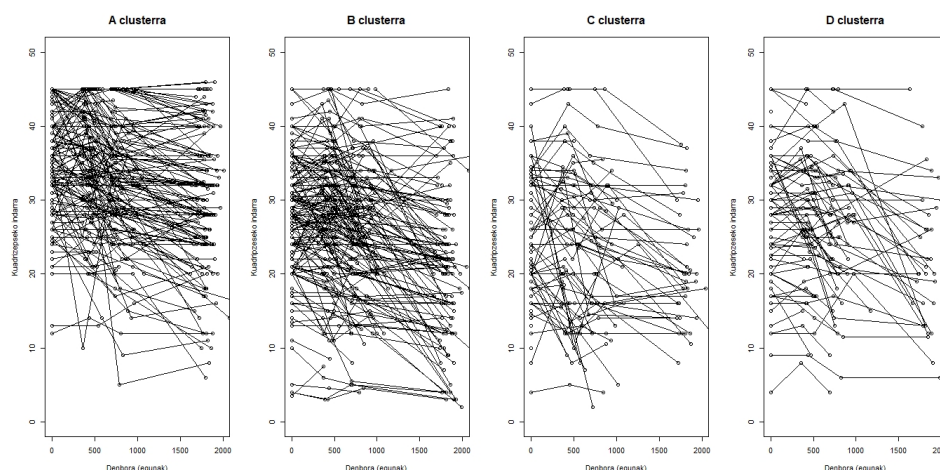
Gaixotasunaren progresioa aztertzen hasitako unean ariketa ahalmenak 465 balioa hartzen duen indibiduo baten kasuan, 450etik 400era jaisteko, hau

da, 50 unitateko jauzia emateko, bost urte eta erdi behar dituela estimatzen da. Bestalde, 10 urtetan ariketa ahalmena 465 izatetik 375 izatera jaistea espero da, balio horren %95eko konfiantza-tartearen goi- eta behe-mugetako balioak 425 eta 331 izanik, hurrenez hurren. Hau da, 10 urtetan 90 unitateko jauzia ematea espero da, jauzi horren %95eko konfiantza-tartearen goi- eta behe-mugak 134 eta 40 izanik.

2 Kuadrizepseko indarra

Lan honetan aztertuko dugun bigarren aldagaia kuadrizepseko indarra izango da, kilogramotan neurtzen den aldagai jarraitua [8].

Ikus dezagun aldagai honek denboran zehar lau clusterretan hartzen dituen balioak:



2.4. Irudia. Lagineko indibiduen ikerketa denboraren eta kuadrizepseko indarraren balioak clusterreka banatuta.

Gure lagineko indibiduen ikerketa denboraren eta kuadrizepseko indarraren balioak taldeka banatuta ikus ditzakegu 2.4. irudian, hau da, talde bakoitzeko kideen erantzun aldagaiaren balioak denborak aurrera egitean nola aldatzen diren. A taldeko indibiduoak pilatzen dira balio altuenetan. Gero eta balio txikiagoak hartzen ditu aldagai honek denborak aurrera egitean eta, beraz, gaixotasunak okerrera egitean erantzun aldagai hau beherakorra da.

Aldagai asketzat ikerketa denbora eta menpeko aldagaitzat kuadrizepseko indarra hartuta, lehenengo 1.2 atalean azaldutako metodoa jarraitu dugu denboran zehar aldagai honek izango duen progresioa ezagutzeko. Denborak aurrera egitean erantzun aldagaiaren balioak txikitzen doazenez, hirugarren mailako polinomioa negatiboa den eremuarekin gelditu gara.

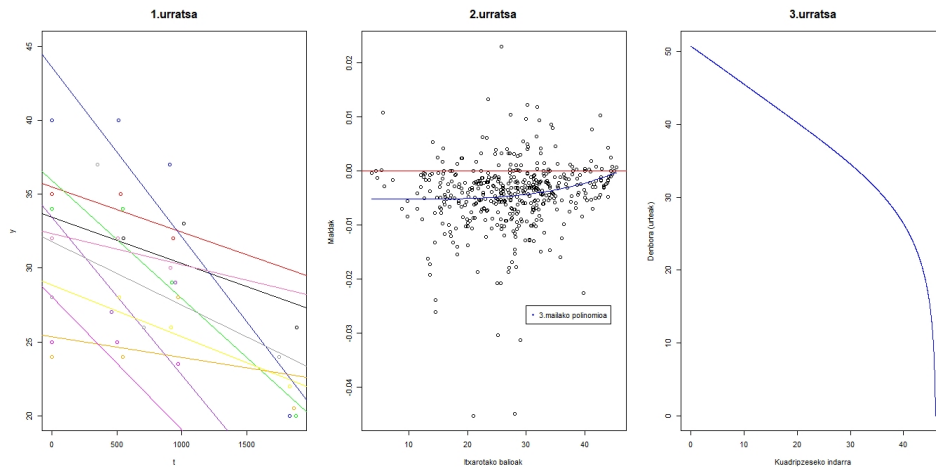
Ikus ditzagun metodoaren lehenengo hiru urratsetan lortutako emaitzak eta kurbak:

- Hirugarren mailako polinomioa:

$$p(x) = 1,172 \cdot 10^{-7}x^3 - 3,752 \cdot 10^{-6}x^2 + 4,003 \cdot 10^{-5}x - 5,356 \cdot 10^{-3}$$

- Hirugarren mailako polinomioaren alderantzizkoaren integrala:

$$\int \frac{1}{p(x)} dx = -3923,82 \cdot \arctan(0,0162945(14,091 + 2x)) \\ + 2265,33 \cdot \log(46,1047 - x) - 1132,67 \cdot \log(991,215 + 14,091x + x^2)$$

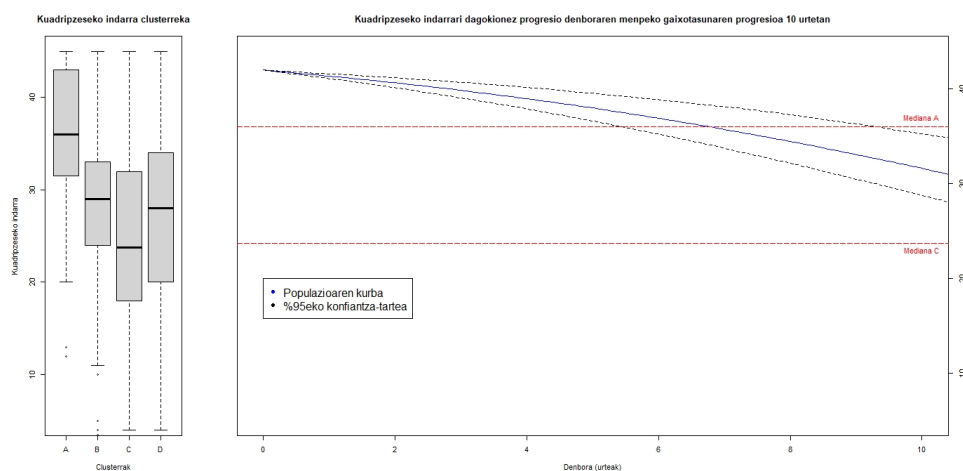


2.5. Irudia. 1.urratsa - Hainbat indibiduoren denboraren eta kuadrizpeseko indarra aldagaiaren datuei doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Itzarotako balioak eta denborarekiko maldak, eta datu bikote horiei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala, gaixotasunaren progresioaren menpeko kurba.

Lehenengo kapituluko 1.2 atalean azaldutako metodoaren lehenengo hiru urratsak irudikatu ditugu 2.5. irudian. Lehenengo grafikoan, hainbat indibiduoren denboraren eta erantzun aldagaiaren datu bikoteak, eta doitutako erregresio lineal bakuneko ereduak. Zuzen hauen malda negatiboa da, denborak aurrera egitean aldagaia beherakorra baita. Bigarren grafikoan, eredu horiek erabiliz lortutako itzarotako balioei eta denborarekiko maldei doitutako hirugarren mailako polinomioa. Eta azkenik, polinomio horren alderantzizkoaren integralaren kurba irudikatu dugu, hau da, gaixotasunaren progresioaren menpekoa den kurba. Polinomioaren eremu negatiboa hautatu dugu urrats honetarako.

Gaixotasunaren progresioa denboraren menpe lortzeko azken urrats honetan lortutako bi aldagaiak alderantzikatu behar ditugu. Horrez gain, gaixotasunaren progresio kurbaren %95eko konfiantza-tartea lortzeko, *bootstrapa* eta pertzentilen metodoa aplikatu ditugu, lehenengo kapituluko 1.3 ataleko urratsak jarraituz.

Horrela, gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak lortu ditugu:



2.6. Irudia. Lehenengo neurketako kuadrripzeseko indarra aldagaia-
ren datuen kutxa diagramak clusterreka banatuta eta gaixotasunaren
progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak.

Azken 2.6. irudi honetan, kuadrripzeseko indarrari dagokionez gaixotasuna-
ren progresioaren kurba 10 urtetan eta kurba horren konfiantza-tartea ikus
ditzakegu. Baita lehenengo neurketan lau clusterretako indibiduen datue-
kin lortutako kutxa diagramak ere. Azken grafiko honi erreparatzen ba-
diogu, A taldeko indibiduoak dira balio eta mediana altuenak dituztenak,
eta C taldekoak, berriz, balio eta mediana baxuenak dituztenak. Hori dela
eta, gaixotasunaren progresio kurbarekin batera bi talde hauen medianak le-
rro horizontalen bidez irudikatu ditugu, etapa batetik bestera progresatzeko
batez beste igaroko den denbora ikusteko.

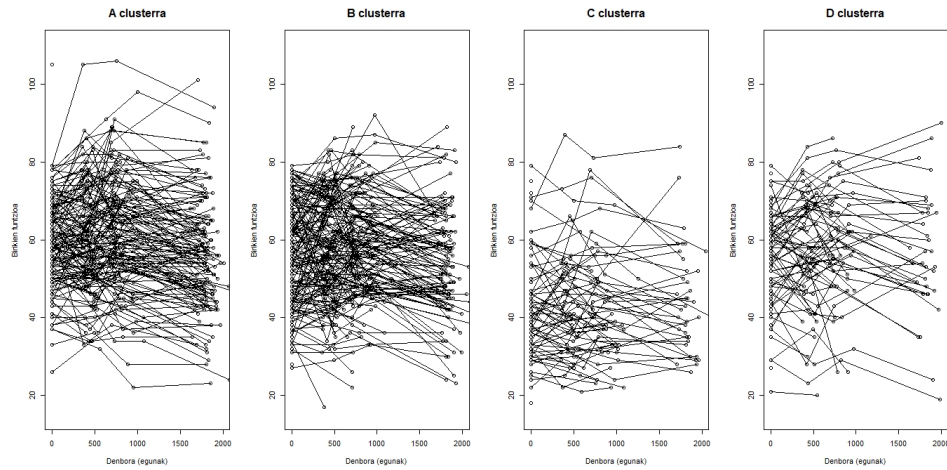
Gaixotasunaren progresioa aztertzen hasitako unean kuadrripzeseko indarrak
42 balioa hartzen duen indibiduo baten kasuan, 36tik (A taldeko indibiduen
lehenengo neurketako datuen medianatik) 31ra jaisteko, hau da, 5 unitate-
ko jauzia emateko, hiru urte eta erdi behar dituela estimatzen da. Beste
aldetik, 10 urtetan kuadrripzeseko indarra 42 izatetik 31,5 izatera igarotzea
espero da, azken balio honen %95eko konfiantza-tartearen goi eta behe mu-
turretako balioak 35,3 eta 28,8 izanik. Beraz, progresioaren hasierako unean

42 kilogramoko indarra duen indibiduo batek handik 10 urtetara 31,5 kilogramoko indarra izatea espero da (10,5 unitatetako jauzia), jauzi horren konfiantza-tartearen goi- eta behe-mugetako balioak 13,2 eta 6,7 izanik, hurrenez hurren.

3 Birikien funtzioa

Hirugarren aldagaiaren bidez birikien funtzioa aztertuko dugu. FEV%1-en bidez, ehunekotan, neurtzen den aldagai jarraitua da [4].

Ikus dezagun aldagai honek denboran zehar lau ikerketa taldeetan hartzen dituen balioak:



2.7. Irudia. Cluster bakoitzeko indibiduen ikerketa denboraren eta birikien funtzioa aldagaiaren balioak.

Lau taldeetako indibiduen birikien funtzioa denboran zehar nola aldatzen den ikus dezakegu 2.7. irudian, ikerketa puntu bakoitzeko aldagaiaren balioak, hain zuzen ere. Talde hauek konparatuz, C taldean pilatzen dira baliorik baxuenak. Gainontzeko hiru taldeek antzeko balioak hartzen dituztela esan genezake, B eta D taldeetako indibiduoek balio apur bat baxuagoak hartzen dituzte A taldekoek baino. Denborak aurrera egin ahala, aldagai honek ere gero eta balio txikiagoak hartzen ditu.

Aldagai asketzat ikerketa denbora eta menpeko aldagaitzat birikien funtzioa hartuta 1.2 atalean azaldutako metodoa jarraitu dugu gaixotasunaren progresio kurba ezagutzeko. Gaixotasunak okerrera egitean erantzun aldagai hau beherakorra denez, hirugarren mailako polinomioaren eremu negatiboarekin gelditu gara.

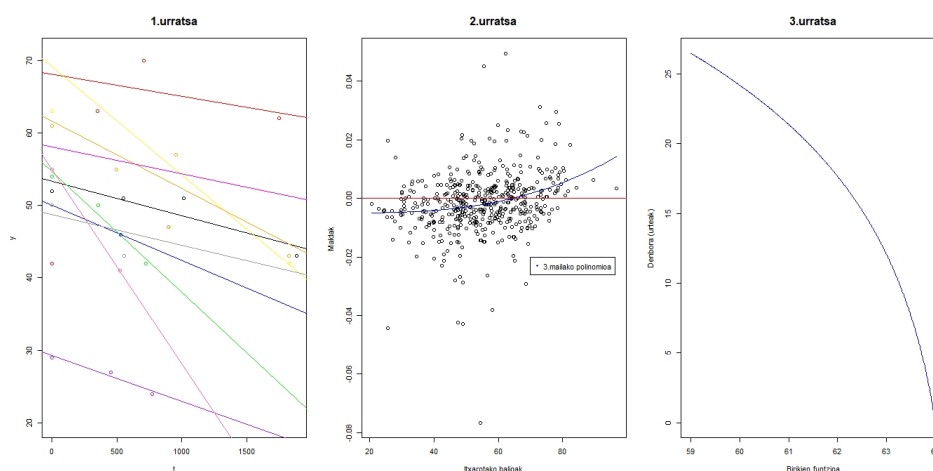
Idatz ditzagun lortutako emaitzak eta ikus ditzagun lortutako kurbak:

- Hirugarren mailako polinomioa:

$$p(x) = 2,725 \cdot 10^{-8}x^3 - 6,235 \cdot 10^{-7}x^2 + 4,756 \cdot 10^{-6}x - 5,016 \cdot 10^{-3}$$

- Hirugarren mailako polinomioaren alderantzizkoaren integrala:

$$\int \frac{1}{p(x)} dx = -6558,08 \cdot \arctan(0,0101576(41,5852 + 2x)) \\ + 3786,33 \cdot \log(64,4659 - x) - 1893,16 \cdot \log(2855,36 + 41,5852x + x^2)$$



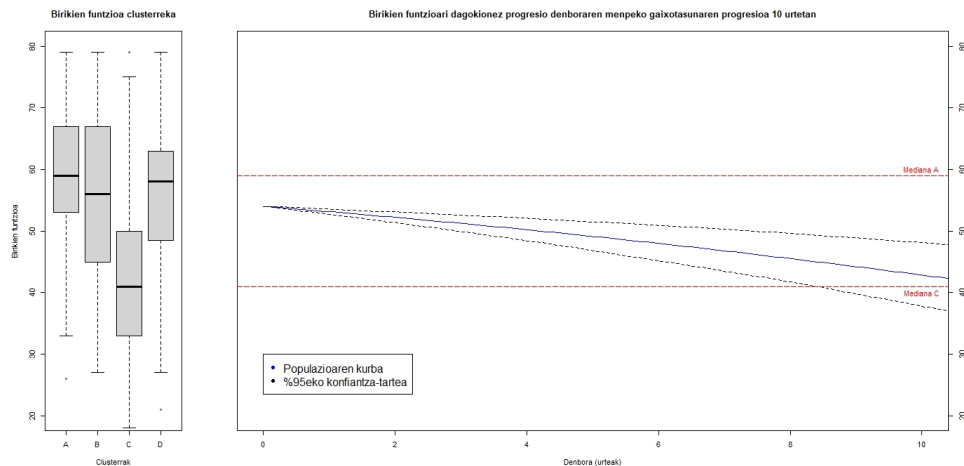
2.8. Irudia. 1.urratsa - Hainbat indibiduoren denboraren eta birikien funtzioa aldagaiaren datuei doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Itxarotako balioei eta denborarekiko maldei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala, gaixotasunaren progresioaren menpeko kurba.

Lehenengo kapituluko 1.2 atalean azaldutako metodoaren lehenengo hiru urratsetan lortutako grafikoak irudikatu ditugu 2.8. irudian. Batetik, hainbat indibiduoren erantzun aldagaiaren eta denboraren balioen puntu-hodeia, eta datu horiei doitutako erregresio lineal bakuneko ereduak. Malda negatiboa dute ereduak, denborak aurrera egitean aldagaiaren balioak txikitzen baitoaz. Bestetik, eredu horiek erabiliz lortutako itxarotako balioei eta maldei doitutako polinomioa. Eta azkenik, polinomio horren alderantzizkoaren integralaren kurba irudikatu dugu. Polinomioaren eremu negatiboa aukeratu dugu urrats honetarako. Modu honetan, gaixotasunaren progresioaren menpekoa den kurba lortu dugu.

Gure laginaren gaixotasunaren progresio kurba denboraren menpe lortzeko, azken urrats honetako bi aldagaiak alderantzikatu behar ditugu. Horrez

gain, gaixotasunaren progresio kurbaren %95eko konfiantza-tartea lortzeko *bootstrap* eta pertzentilen metodoa erabili ditugu, 1.3 atalean azaldutakoari jarraituz.

Beraz, azkenik, gaixotasunaren progresio kurbaren eta kurba horren %95eko konfiantza-tartearen estimazioak lortu ditugu. Honakoak dira lortutako kurbak:



2.9. Irudia. Lehenengo neurketako birikien funtzioa aldagaiaren datuen kutxa diagramak clusterreka banatuta eta gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak.

Alde batetik, lehenengo neurketako birikien funtzioa aldagaiaren datuak erabiliz lau taldeetako indibiduen kutxa diagramak, eta bestetik, birikien funtzioari dagokionez gaixotasunaren progresio kurba 10 urtetan eta kurba horren ziurgabetasunaren konfiantza-tartea irudikatu ditugu 2.9. irudian. A taldeko indibiduoak dira birikien funtzio onena dutenak eta C taldekoak, berriz, birikien funtzio txarrena dutenak. Bi talde hauen medianak lerro horizontalaren bidez irudikatu ditugu gaixotasunaren progresio kurbarekin batera, batetik bestera igarotzeko batez beste behar den denbora eta aldaketa horren abiadura ezagutzeko.

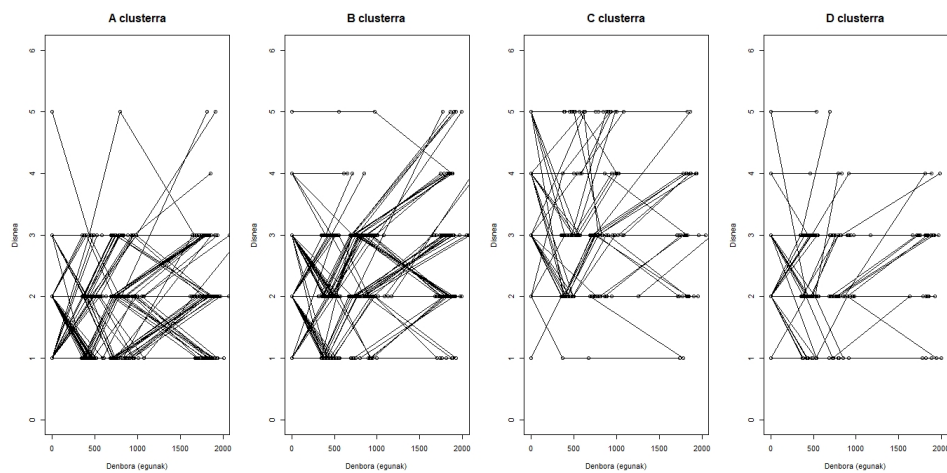
Gaixotasunaren progresioa aztertzen hasitako unean birikien funtzioak 54 balioa hartzen duen indibiduo baten kasuan, 50etik 45era jaisteko, hau da, 5 unitateko jauzia emateko, 4 urte behar direla estimatzen da. Kasu honetan, FEV%1 50 baino txikiagoa izatera iristean, gaixotasuna arina izatetik moderatua izatera igarotzen delako aukeratu dugu puntu hori [5]. Bestalde, 10 urtetan birikien funtzioak 11 unitate inguruko jauzia ematea espero da, 54 izatetik 43 izatera igarotzea espero da, hain zuzen ere. Azken balio horren %95eko konfiantza-tartearen muturretako balioak 48 eta 38 dira eta

jauzi horren konfiantza-tartearen goi- eta behe-mugetako balioak 16 eta 6 inguru, hurrenez hurren.

4 Disnea

Laugarren aldagai honek indibiduen disnea edo arnasestu maila neurtzen du eta horretarako, mMRC eskala erabiltzen da (*modified Medical Research Council Dyspnea Scale*) [7].

Ikus ditzagun aldagai honek lau taldeetan hartzen dituen balioak denboran zehar:



2.10. Irudia. Cluster bakoitzeko indibiduen ikerketa denboraren eta disnearen balioak.

Ondoko 2.10. irudiari erreparatuz, A taldeko indibiduoek dituzte daturik onenak, hau da, disnea mailarik txikiena. C taldeko indibiduoak pilatzen dira, berriz, disnea mailarik altuenetan. Disnearen balioek gora egiten dute denborak aurrera egin ahala. Beraz, gaixotasunak okerrera egitean erantzun aldagai hau gorakorra da, aurreko hiru aldagaiak ez bezala.

Aldagai asketzat ikerketa denbora eta menpeko aldagaitzat disnea hartuta 1.2 atalean azaldutako metodoa jarraitu dugu denboran zehar aldagai honek izango duen progresioa ezagutzeko. Denborak aurrera eta gaixotasunak okerrera egitean disnearen balioak handitzen doazenez, hirugarren mailako polinomioaren eremu positiboa hautatu dugu.

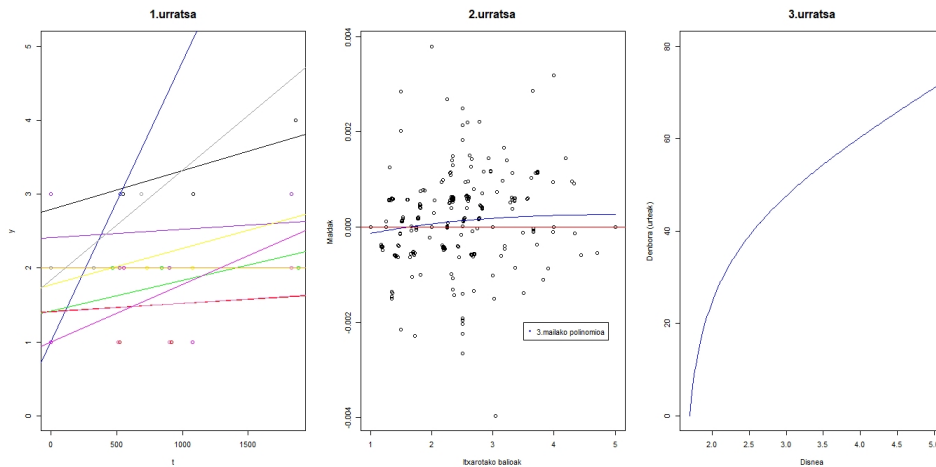
Idatz ditzagun lortutako emaitzak eta kurbak:

- Hirugarren mailako polinomioa:

$$p(x) = 3,430 \cdot 10^{-6}x^3 - 6,059 \cdot 10^{-5}x^2 + 3,568 \cdot 10^{-4}x - 4,345 \cdot 10^{-4}$$

- Hirugarren mailako polinomioaren alderantzizkoaren integrala:

$$\int \frac{1}{p(x)} dx = 9255,96 \cdot \arctan(0,135376(-16,0398 + 2x)) \\ + 5345,73 \cdot \log(1,62488 - x) - 2672,87 \cdot \log(77,9605 - 16,0398x + x^2)$$

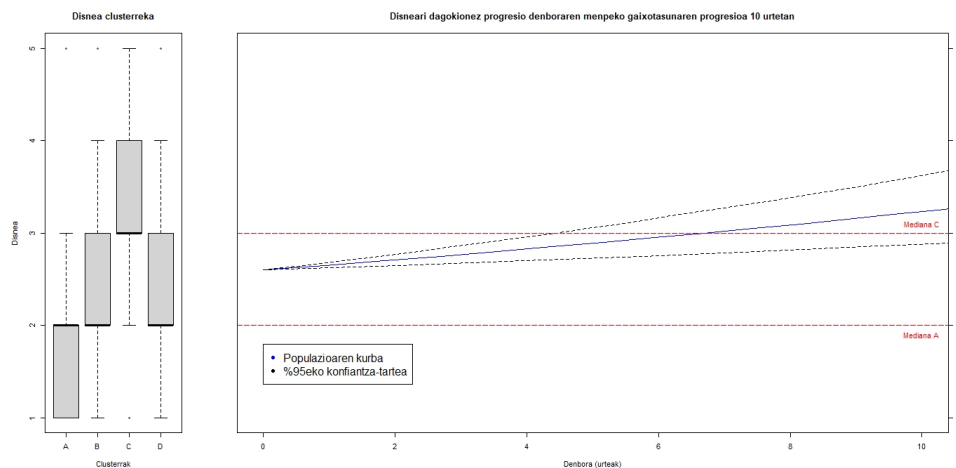


2.11. Irudia. 1.urratsa - Hainbat indibiduoren denboraren eta disnea aldagaiaren datuei doitutako erregresio lineal bakuneko ereduak. 2.urratsa - Itxarotako balioei eta denboarekiko maldei doitutako hirugarren mailako polinomioa. 3.urratsa - Hirugarren mailako polinomioaren alderantzizkoaren integrala, gaixotasunaren progresioaren menpeko kurba.

Lehenengo kapituluko 1.2 atalean azaldutako metodoaren lehenengo hiru urratsetan lortutako emaitzak irudikatu ditugu 2.11. irudian. Lehenengo, hainbat indibiduoaren datuei doitutako erregresio lineal bakuneko ereduak. Malda positiboa dute oraingoan eredu hauek, aldagai honen bilakaera gorakorra baita. Bigarren grafikoan, itxarotako balioei eta eredu horien denboarekiko maldei doitutako hirugarren mailako polinomioa. Eta azkenik, polinomio horren alderantzizkoaren integralaren kurba. Polinomioaren eremu positiboa hautatu dugu urrats honetarako. Urrats hauen bidez, hirugarren grafikoan gaixotasunaren progresioaren menpekoa den kurba lortu dugu.

Laugarren urratseko kurba irudikatzeko, hau da, gure laginaren gaixotasunaren progresio kurba denboraren menpe irudikatzeko, hirugarren urratsean lortutako bi aldagaiak alderantzizatu behar ditugu. Gainera, gaixotasunaren progresio kurbaren %95eko konfiantza-tartea lortzeko, lehenengo kapituluko 1.3 atalean azaldutako *bootstrap* metodoa eta pertzentilen metodoa erabili ditugu.

Horrela, gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak lortu ditugu. Ikus ditzagun azken urrats honetan lortutako kurbak.



2.12. Irudia. Lehenengo neurketako disnea aldagaiaren datuen kutxa diagramak clusterreka banatuta eta gaixotasunaren progresio kurbaren eta bere %95eko konfiantza-tartearen estimazioak.

Ondoko 2.12. irudian, batetik, talde bakoitzeko indibiduen disnea aldagaiaren lehenengo neurketako datuekin lortutako kutxa diagramak irudikatu ditugu, eta bestetik, disneari dagokionez gaixotasunaren progresioa 10 urtetan eta progresio horren ziurgabetasunaren konfiantza-tartea. A eta C taldeetako indibiduoak dira disnea mailarik baxuenak eta altuenak dituztenak. Beraz, bi talde horien medianak lerro horizontal baten bidez irudikatu ditugu, gaixotasunak etapa batetik bestera igarotzeko batez beste behar duen denbora ezagutzeko.

Gaixotasunaren progresioa denboran zehar aztertzen hasitako unean disneak 2,5 inguruko balioa hartzen duen indibiduo baten kasuan, disnearen balioa 2,5etik 3ra igarotzeko, hau da, unitate erdiko jauzia emateko, sei urte eta erdi behar direla estimatzen da. Beste aldetik, 10 urtetan disnea 2,5 izatetik 3,25 izatera igarotzea espero da, azken balio honen %95eko konfiantza-tartearen goi eta behe muturretako balioak 3,65 eta 2,95 izanik. Hau da, 10 urtetan 0,75 unitateko jauzia ematea espero da, jauzi honen konfiantza-tartearen goi- eta behe-mugetako balioak 1,05 eta 0,35 izanik, hurrenez hurren.

3. Kapituluia

Eztabaida

Hainbat indibiduoren epe motzeko datuetan oinarrituz gaixotasunen progresioa auresango duen metodoa inplementatu dugu eta metodo hori gaixotasun bati, biriketako gaixotasun buxatzaile kronikoari, hain zuzen ere, aplikatu diogu. Horretarako, lehenengo atalean azaldutako prozedura jarraitu dugu eta gaixotasunaren 10 urtetako progresioa auresatea eta progresio horren kurbaren ziurgabetasunaren konfiantza-tartea lortu ditugu. Honekin, gaixotasunaren adierazleak diren aldagaiak (arnasketa ezaugarriez gain, indarra eta ariketa ahalmena) denboran zehar nola aldatuko diren ikus dezakegu. Eta hau garrantzitsua izan daiteke gaixotasun hau pairatzen duten gaixoen eboluzioa ezagutzen laguntzeko.

Budgeon eta kideek (2017) gaixotasunen progresioa auresateko proposatutako metodoa funtzio sigmoidearen antzeko itxura duten gaixotasunen kasurako proposatutako metodoa da [2]. Honako itxura duten gaixotasunak dira horiek: gaixotasunaren lehenengo etapatan, motel progresatzen du, gaixotasunaren progresioaren hazkuntza motela da; ondoren, hazkuntza hori handia da, azkarra da progresioa; eta azkenik, hazkuntza hori geldituz doa. Gure kasuan ez dakigu biriketako gaixotasun buxatzaile kronikoaren kurbaren itxura sigmoidala den. Ez dugu erreferentziarik aurkitu eta, beraz, metodo hau aplikatzeko BGBKren bilakerak funtzio sigmoidearen itxura duela onartu dugu. Ez dugu argudiorik aurkakoa pentsatzeko eta gaixotasunak aurretik aipatutako itxura izan dezakeela esan digute klinikoei. Hala ez balitz, ez dakigu aplikatutako metodoak balioko lukeen ala ez.

Oinarritzat hartu dugun artikuluan, denborak aurrera eta gaixotasunak okerrera egin ahala balioak handitzen doazen erantzun aldagai bat aztertzen da. Indibiduo bakoitzaren datuei erregresio lineal bakuneko eredu bat doitu eta itxarotako balioa eta malda kalkulatu ondoren, hirugarren mailako polinomio monotono bat doitzen zaie datu bikote horiei. Polinomio horrek ez-negatiboa izan behar du eta, beraz, polinomioa positiboa den aldagai askearen balioetara mugatzen dugu bere eremua, ondoren polinomio honen

alderantzizkoaren integrala ere eremu berean ebaluatzeko. Aztergai diren gure hiru aldagaien kasuan, ordea, denborak aurrera eta gaixotasunak okerrera egitean erantzun aldagaiaren balioak txikitzen doaz. Hori dela eta, datuei doitutako polinomioaren eremu negatiboa hautatzea erabaki dugu, gaixotasunaren progresio kurbaren malda negatiboa izan dadin.

Aurretik lan honetan aipatu bezala, *bootstrap* iterazio bakoitzean sortutako laginari hirugarren mailako polinomio monotono bat doitu eta bere eremu positiboa edo negatiboa kalkulatu dugu erantzun aldagaiaren arabera. Hauen alderantzizkoen integralak ebaluatzeko bi eremu posible bururatu zaizkigu: polinomio bakoitza dagokion eremu positiboan (edo negatiboan) ebaluatzea eta polinomio bakoitza polinomio guztiak positiboak (edo negatiboak) diren eremuan ebaluatzea. Bigarren eremu honekin lan egitea erabaki dugu, polinomio guztiak eremu berean ebaluatzeko eta eskala bera mantentzeko.

Epe motzeko eta ikerketa denbora puntu gutxitako datuak dituzten gaixotasunen kasuan, ez da erraza izaten gaixotasunaren progresioa auresatea. Ikerketetan eta datuetan ez da ikusten nolakoa izango den gaixotasunaren progresioa. Bost urtetako lau ikerketa denboratako datuak ditugu gure indibiduen kasuan. Beraz, epe motzeko gaixotasunaren jarraipena dago jasota datu hauetan, bost urtetako jarraipena bakarrik. Bost urte behatuz, erronka izan da gaixotasunaren hamar urtetako progresioa auresatea, nahiz eta literaturak egoera hau ohikoa dela dioen. Izan ere, artikuluan alzheimer gaixotasunerako bost urtetako datuak erabiliz 50 urtetako gaixotasunaren progresioa auresaten da.

Biriketako gaixotasun buxatzaile kronikoaren 10 urtetako progresioa auresatea eta progresio kurba horren ziurgabetasunaren konfiantza-tartea lortu ditugu, bost urtetako datuetan oinarrituta. Beraz, gaixotasunaren progresioaren inguruko ezagutza zabalagoa eta gaixotasunaren adierazleak diren hainbat aldagaien progresioaren ezagutza izateko baliagarria izango dela diote klinikoek metodo hau.

Bibliografía

- [1] ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories (2002). ATS statement: guidelines for the six-minute walk test. *American Journal of Respiratory and Critical Care Medicine*; 166(1):111-117.
- [2] Budgeon, C.A., Murray, K., Turlach, B.A., Baker, S., Villemagne, V.L., Burnham, S.C., Alzheimer's Disease Neuroimaging Initiative (2017). Constructing longitudinal disease progression curves using sparse, short-term individual data with an application to Alzheimer's disease. *Statistics in Medicine*; 36(17): 2720-2734.
- [3] Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 40(5):373-383.
- [4] Esteban, C., Arostegui, I., Aburto, M., Moraza, J., Quintana, J.M., García-Loizaga, A., Basualdo, L.V., Aramburu, A., Aizpiri, S., Uranga, A., Capelastegui, A. (2016). Chronic Obstructive Pulmonary Disease Subtypes. Transitions over Time. *PLoS One*;11(9):e0161710.
- [5] Global Initiative for Chronic Obstructive Lung Disease. (2020). Pocket Guide to COPD Diagnosis, Management, and Prevention. A Guide for Health Care Professionals. Global Initiative for Chronic Obstructive Lung Disease, Inc. URL https://goldcopd.org/wp-content/uploads/2020/03/GOLD-2020-POCKET-GUIDE-ver1.0_FINAL-WMV.pdf.
- [6] Kuhn, M., Chow, F., Wickham, H. (2020). rsample: General Resampling Infrastructure. R package version 0.0.7. <https://CRAN.R-project.org/package=rsample>
- [7] Mahler, D.A., Ward, J., Waterman, L.A., McCusker, C., ZuWallack, R., Baird, J.C. (2009). Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest*; 136(6):1473-1479.

- [8] Mathiowetz, V., Kashman, N., Volland, G., Weber, K., Dowe, M., Rogers, S. (1985). Grip and pinch strength: normative data for adults. *Archives of Physical Medicine and Rehabilitation*; 66(2):69-74.
- [9] Murray, K., Müller, S., Turlach, B.A. (2016). Fast and flexible methods for monotone polynomial fitting. *Journal of Statistical Computation and Simulation*; 86(15):2946–2966.
- [10] Murray, K., Müller, S. and Turlach, B.A. (2013). Revisiting fitting monotone polynomials to data, *Computational Statistics* 28(5): 1989-2005.

A. Eranskina

Aldagaien azalpena

Biriketako gaixotasun buxatzaile kronikoaren progresioaren adierazleak diren aldagaiak:

1. **WDist, XDist, YDist, ZDist**: ariketa ahalmena aldagaiaren lehen, bigarren, hirugarren eta laugarren neurketak adierazten dituzten aldagaiak dira, hurrenez hurren. *6 minute walking test* testaren bidez neurturiko aldagai jarraitua da eta 6 minututan ibilitako metroak adierazten du [1].
2. **WFcuad, XFcuad, YFcuad, ZFcuad**: kuadripzeseko indarra aldagaiaren lehen, bigarren, hirugarren eta laugarren neurketak adierazten dituzten aldagaiak dira, hurrenez hurren. Kuadripzesen bidez egindako indarra kilogramotan neurtuta adierazten du [8].
3. **WFEV1₋, XFEV1₋, YFEV1₋, ZFEV1₋**: birikien funtzioa aldagaiaren lehen, bigarren, hirugarren eta laugarren neurketak adierazten dituzten aldagaiak dira, hurrenez hurren. Lehenengo segundoan arnas botatzeko bolumen behartuaren (FEV%1 deritzo) bidez, ehunekotan, neurtzen da aldagai jarraitu hau [4].
4. **WCV23, XCV23, YCV23, ZCV23**: disnea aldagaiaren lehen, bigarren, hirugarren eta laugarren neurketak adierazten dituzten aldagaiak dira, hurrenez hurren. Indibiduen disnea edo arnasestu maila neurtzen du eta bat eta bost bitarteko zenbaki osoen bidez adierazten da. Horretarako, mMRC eskala erabiltzen da (*modified Medical Research Council Dyspnea Scale*) [7].

Beste hainbat aldagai deskribatzaile:

1. **Edad**: indibiduen adina adierazten duen aldagaia da.
2. **Sexo**: 1 balioa hartzen du indibidua gizona bada eta 2 balioa, berriz, indibidua emakumea bada.
3. **IMC**: gorputz masa indizea adierazten duen aldagaia da.

4. **Charlson:** Charlson-en indizeak (*Charlson Comorbidity Index*) [3], indibiduo bakoitzak pairatzen dituen gaixotasunak neurtzen ditu, gaixotasun larriek arinenek baino pisu handiagoa dute eta indize honek zenbat eta balio altuagoak hartu, indibidua gaixoago dagoela adierazten du.
5. **IngPre:** azken bi urteetan metatutako ospitalaratze episodioak adierazten dira aldagai honen bidez.
6. **Tto:** botika kopuruaren arabera indibiduoek daukaten tratamendua adierazten du.

B. Eranskina

Kodea

Bi programa erabili ditugu gure lana inplementatzeko, *R* eta *Mathematica*. Gure lau erantzun aldagaietarako antzeko programazioa erabili dugu eta programazio hori idazterakoan kontuan izan dugu, alde batetik, gaixotasunak okerrera egitean aldagaia gorakorra edo beherakorra den, eta bestetik, lagin bakoitzaren datuei doitutako polinomioa ebaluatzeko eremuaren eta eremu horien guztien ebakiduraren kalkulua. Hori dela eta, birikien funtzioa aldagaiaren kasurako erabilitako programa bakarrik idatzi dugu eranskinean, adibide gisa.

B.1 R programa

```
#Sor ditzagun bootstrapa erabiliz 1000 laginak
set.seed(4) #hasiera finkatuta, zoriz beti lagin berdinak
sortuko ditu
boot<-bootstraps(datuak, times = 1000)
lagina1<-as.data.frame(boot$splits[[1]]) #datu base gisa
irakurtzeko

#####
##### BIRIKIEN FUNTZIOA - FEV #####
#####

#Indibiduo bakoitzaren itxarotako balioaren eta maldaren
#kalkulua. Matrize batean gordeko ditugu gure lagineko
#indibiduo bakoitzarentzat lortutako itxarotako balioa
#eta erregresio linealeko ereduaren denborarekiko malda
matrizea_fev<-matrix(0,nrow=length(datuak$ENCUESTA_N_),
ncol=3)
for(i in (1:length(datuak$ENCUESTA_N_))){
denbora<-c(0,datuak$denbora1[i],datuak$denbora2[i],
datuak$denbora3[i])
```

```

menpekoa<-c(datuak$WFEV1_[i], datuak$XFEV1_[i], datuak$
  YFEV1_[i], datuak$ZFEV1_[i])
if((is.na(menpekoa[2])==TRUE && is.na(menpekoa[3])==
  TRUE && is.na(menpekoa[4])==TRUE) || (is.na(denbora
  [2])==TRUE && is.na(denbora[3])==TRUE && is.na(
  denbora[4])==TRUE)){
  matrizea_fev[i, ]<-c(NA,NA,NA)
}
else if((is.na(menpekoa[2])==FALSE && is.na(menpekoa[3]
  )==FALSE && is.na(menpekoa[4])==FALSE) && (is.na(
  denbora[2])==FALSE && is.na(denbora[3])==FALSE &&
  is.na(denbora[4])==FALSE)){
  eredua<- lm(menpekoa~denbora)
  bbdenbora<- datuak$denbora3[i]/2
  itxarotakobalioa<- eredua$coefficients[1] + eredua$
    coefficients[2]*bbdenbora
  matrizea_fev[i, ]<-c(itxarotakobalioa, eredua$
    coefficients[2], eredua$coefficients[1])
}
else{
  na_m<-which(is.na(menpekoa))
  na_d<-which(is.na(denbora))
  na_guztiak<-unique(c(na_m,na_d))
  denbora_nagabe<-denbora[-na_guztiak]
  menpekoa_nagabe<-menpekoa[-na_guztiak]
  eredua<-lm(menpekoa_nagabe~denbora_nagabe)
  bbdenbora<-denbora_nagabe[length(denbora_nagabe)
    ]/2
  itxarotakobalioa<- eredua$coefficients[1] +
    eredua$coefficients[2]*bbdenbora
  matrizea_fev[i, ]<-c(itxarotakobalioa, eredua$
    coefficients[2], eredua$coefficients[1])
}
}

#Itxarotako balioei eta maldei doitutako polinomio
  monotonoaren kalkulua
polinomioa3_fev<-monpol(matrizea_fev[,2]~matrizea_fev[,1
  ], degree=3)
ismonotone(polinomioa3_fev)

plot(matrizea_fev[,1],matrizea_fev[,2], xlab="Itxarotako_
  balioak", ylab="Maldak")
curve(0.0000002725*x^3-0.000006235*x^2+0.000004756*x-0.
  005016, add = TRUE, col = "blue")
legend(80,-0.05,col=c('blue'),legend = c('3.mailako_
  polinomioa'), pch=20)

#Polinomioaren eremu negatiboaren kalkulua

```

```

polfev<- function(x) {0.00000002725*x^3-0.0000006235*x^2+
  0.000004756*x-0.005016}
polfev(64) #negatiboa da funtzioa
polfev(65) #positiboa da funtzioa

#Polinomioaren integrala Mathematican kalkulatur
integrala_fev<-function(x){-6558.08*atan(0.0101576*(41.58
  52+2*x))+3786.33*log(as.complex(64.4659-x))-1893.16*
  log(as.complex(2855.36+41.5852*x+x^2))}

#Eremu negatiboan ebaluatuta
ufev_neg<-seq(59,64, by=0.01)
vfev_neg<-rep(0, length(ufev_neg))
for(i in (1:length(ufev_neg))){
  vfev_neg[i]=integrala_fev(ufev_neg[i])
}

#Koordenatu jatorrira transladatzeke
vfev_neg=vfev_neg-vfev_neg[length(vfev_neg)]
#Denbora urtetan neurtuta izateko
vfev_neg_urteak=vfev_neg/365

#### BOOTSTRAP ####

#Bootstrap iterazio bakoitzean sortutako laginari
#doitutako polinomioaren eta polinomioaren koefizienteak
#gordetzeko matrizearen kalkulua.
matrizea_boot_fev<-matrix(0, nrow=1000, ncol=4)
for (j in (1:1000)){
  lagina<-as.data.frame(boot$splits[[j]])
  m_fev<-matrix(0,nrow=length(lagina$ENCUESTA_N_),ncol=3)
  for(i in (1:length(lagina$ENCUESTA_N_))){
    denbora<-c(0,lagina$denbora1[i],lagina$denbora2[i],
      lagina$denbora3[i])
    menpekoa<-c(lagina$WFEV1_[i], lagina$XFEV1_[i],
      lagina$YFEV1_[i], lagina$ZFEV1_[i])
    if((is.na(menpekoa[2])==TRUE && is.na(menpekoa[3])==
      TRUE && is.na(menpekoa[4])==TRUE) || (is.na(
      denbora[2])==TRUE && is.na(denbora[3])==TRUE &&
      is.na(denbora[4])==TRUE)){
      m_fev[i, ]<-c(NA,NA,NA)
    }
    else if((is.na(menpekoa[2])==FALSE && is.na(menpekoa[
      3])==FALSE && is.na(menpekoa[4])==FALSE) && (is.
      na(denbora[2])==FALSE && is.na(denbora[3])==FALSE
      && is.na(denbora[4])==FALSE)){
      eredia<- lm(menpekoa~denbora)
      bbdenbora<- lagina$denbora3[i]/2
    }
  }
}

```

```

    itxarotakobalioa<- eredua$coefficients[1]+eredua$
      coefficients[2]*bbdenbora
    m_fev[i, ]<-c(itxarotakobalioa, eredua$coefficients
      [2],eredua$coefficients[1])
  }
  else{
    na_m<-which(is.na(menpekoa))
    na_d<-which(is.na(denbora))
    na_guztiak<-unique(c(na_m,na_d))
    denbora_nagabe<-denbora[-na_guztiak]
    menpekoa_nagabe<-menpekoa[-na_guztiak]
    eredua<-lm(menpekoa_nagabe~denbora_nagabe)
    bbdenbora<-denbora_nagabe[length(denbora_nagabe)]/2
    itxarotakobalioa<- eredua$coefficients[1]+eredua$
      coefficients[2]*bbdenbora
    m_fev[i, ]<-c(itxarotakobalioa, eredua$coefficients
      [2], eredua$coefficients[1])
  }
}
polinomioa3_fev<-monpol(m_fev[,2]~m_fev[,1], degree=3)
ismonotone(polinomioa3_fev)
koef_fev<-polinomioa3_fev$beta.raw
names(koef_fev)<-NULL
matrizea_boot_fev[j, ]<-c(koef_fev[1], koef_fev[2],
  koef_fev[3], koef_fev[4])
}

#Polinomioen koefizienteak dituen matrizea excel batean
#gorde, gero Mathematican irekitzeko
write.table(matrizea_boot_fev, file = "matrizeaboot_fev.
  xls", row.names = FALSE, col.names=FALSE)

#Lagin bakoitzari doitutako polinomioaren integrala
#Mathematican kalkulatu ondoren, polinomio guztiak
#negatiboak diren eremua kontuan hartuz
bektoreak_fev <- read_excel("C:/Users/leire/Desktop/MATE/
  4.MAILA/TFG/Rprograma/bektoreakfev_guztiak.xls")
ilarak_fev<-as.matrix(bektoreak_fev)

#Denboraren pertzentilak gordetzeko bektoreak sortu
p1_fev<-rep(0, length(ilarak_fev[1,]))
p2_fev<-rep(0, length(ilarak_fev[1,]))

#Pertzentilen metodoa aplikatzen dugu
#Denboraren bektore bakoitzarentzat P2,5 eta P97,5
  kalkultzen ditugu
for(i in(1:length(ilarak_fev[1,]))) {
  p1_fev[i]=quantile(ilarak_fev[,i], prob=c(0.025,0.975))
  [1]

```

```

    p2_fev[i]=quantile(ilarak_fev[,i], prob=c(0.025,0.975))
      [2]
}

```

B.2 Mathematica programa

```

(*Gure lagineko indibiduen datuei doitutako
  polinomioaren alderantzizkoaren integrala*)
polinomioa3fev[x_] = 0.00000002725*x^3 - 0.0000006235*x^2
  + 0.000004756*x - 0.005016;
polinomioa3fevald[x_] = 1/polinomioa3fev[x]
integrala3fev[x_] = Integrate[polinomioa3fevald[x], x]

(*Eremu negatiboaren kalkulua*)
polinomioa3fev[64](*Polinomioa negatiboa da*)
polinomioa3fev[65](*Polinomioa positiboa da*)

(*Lagin bakoitzaren datuei doitutako polinomioen
  alderantzizkoen integralak polinomio guztiak
  negatiboak diren eremuan ebaluatuta*)

matrizeafev = Import["C:/Users/leire/Desktop/MATE/4.MAILA
  /TFG/Rprograma/matrizeaboot_fev.xls"];

bektoreakfev = ConstantArray[0, Length[matrizeafev] + 1];
ufev = {0};
For[j = 1, j <= 250, j++,
  AppendTo[ufev, 0.1*j]];
bektoreakfev[[1]] = ufev;

azkenakfev = {};
For[i = 1, i <= Length[matrizeafev], i++,
  pol[x_] = matrizeafev[[i, 1]] + matrizeafev[[i, 2]]*x +
    matrizeafev[[i, 3]]*x^2 + matrizeafev[[i,4]]*x^3;
  neg = {};
  For[j = 1, j <= 1000, j++,
    If[pol[j] < 0, AppendTo[neg, j]]];
  AppendTo[azkenakfev, neg[[Length[neg]]]]]

minimoafev = Min[azkenakfev]

For[i = 1, i <= Length[matrizeafev], i++,
  polinomioa[x_] = 1/(matrizeafev[[i, 1]] + matrizeafev[[
    i, 2]]*x + matrizeafev[[i, 3]]*x^2 + matrizeafev[[i
    , 4]]*x^3);
  integrala[x_] = Integrate[polinomioa[x], x];
  u = {minimoafev};
  For[j = 1, j <= 250, j++,

```

```
AppendTo[u, minimoafev - 0.1*j]];
u = Reverse[u];
v = ConstantArray[0, Length[u]];
For[k = 1, k <= Length[u], k++,
  v[[k]] = integrala[u[[k]]]];
v = v - v[[Length[v]]];
bektoreakfev[[i + 1]] = v]

Export["C:/Users/leire/Desktop/MATE/4.MAILA/TFG/Rprograma
/bektoreakfev_guztiak.xls", bektoreakfev2]
```