eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Estatistika: SPSS praktikak

María Merino Maestre

EUSKARA ETA ELEANIZTASUNEKO ERREKTOREORDETZAREN SARE ARGITALPENA

Liburu honek UPV/EHUko Euskara eta Eleaniztasuneko Errektoreordetzaren dirulaguntza jaso du

María Merino Maestre

Matematika Aplikatua eta Estatistika eta Ikerkuntza Operatiboa

Zientzia eta Teknologia Fakultatea

maria.merino@ehu.es

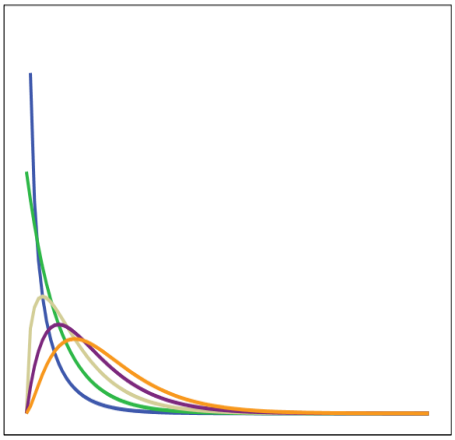
http://www.ehu.es/mae/html/prof/Maria.html

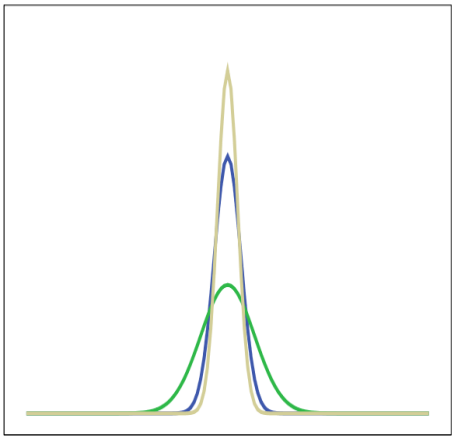
Euskara eta Eleaniztasuneko Errektoreordetzaren Sare Argitalpena

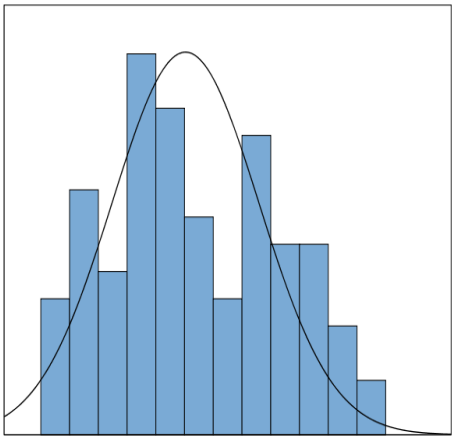
ISBN 978-84-694-0972-5

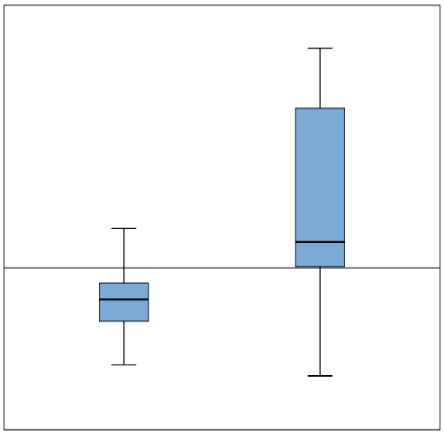
ESTATISTIKA: SPSS PRAKTIKAK

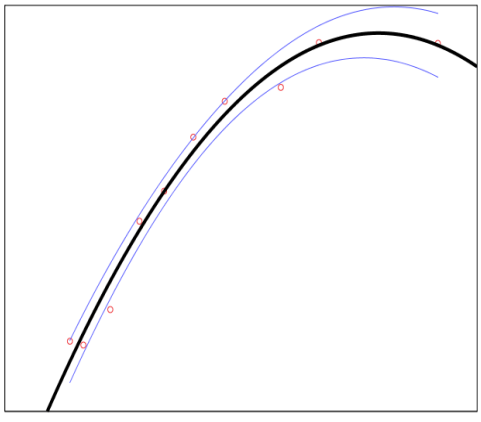
María Merino Maestre

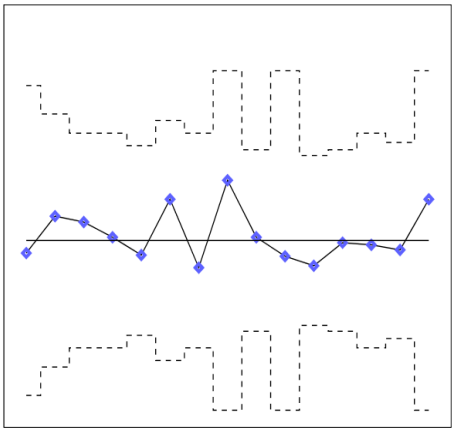












Aurkibidea

Sarrera v

1. Datu-baseak **1**

1.1. Datu-baseen irakurketa 1

1.2. Datu-baseen eraikuntza 2

2. Probabilitate-teoria **5**

2.1. Probabilitate-legea, dentsitate-funtzioa eta banaketa-funtzioa 5

2.2. Probabilitate-teoriaren oinarrizko teorema batzuk 6

3. Estatistika deskribatzailea **9**

3.1. Maiztasun-taulak 10

3.2. Grafikoak 11

3.3. Estatistikoak 14

3.4. Normaltasuna aztertzeko metodo deskribatzaileak 15

4. Konfiantza-tartezko zenbatespena **17**

4.1. Lagin bakar baten zenbatespen-tarteak 17

4.2. Lagin birako zenbatespen-tarteak 19

4.3. Populazio binomialetarako zenbatespen-tarteak 20

5. Hipotesi-kontraste parametrikoak **23**

5.1. Lagin bakar baten hipotesi-kontrasteak ($H_0 : \mu = \mu_0$) 23

5.2. Lagin birako hipotesi-kontrasteak 24

5.2.1. Populazio askeak ($H_0 : \mu_1 = \mu_2$) 24

5.2.2. Binakako datuak ($H_0 : \mu_1 = \mu_2$ edo $H_0 : \mu_D = d_0$) 25

5.3. Populazio binomialetarako hipotesi-kontrasteak 27

5.3.1. Lagin bakar baterako ($H_0 : p = 0{,}5$ edo $H_0 : p \geq p_0$ edo $H_0 : p \leq p_0$) 27

5.3.2. Lagin birako ($H_0 : p_1 = p_2$) 28

5.4. Batezbestekoak konparatzeko metodo grafikoa 29

6. Hipotesi-kontraste ez-parametrikoak **33**

6.1. Doikuntza-egokitasunerako kontrastea 33

6.2. Independentzia-kontrastea eta homogenotasun-kontrastea 34

6.3. Zorizkotasun-kontrastea 36

6.4. Populazioak konparatzeko kontrasteak 37

6.4.1. Bi lagin askeren konparaketa 37

6.4.2. Bi lagin askeren baino gehiagoren konparaketa 37

6.4.3. Binakako datuen bi laginen konparaketa 37

6.4.4. Binakako datuen bi laginen baino gehiagoren konparaketa 38

7. Erregresioa **41**

7.1. Populazio-eredua proposatzea 41

7.2. MKB metodoa erabiliz parametroak zenbatestea 43

7.3. Ereduaren erabilgarritasuna zehaztea 43

7.3.1. Doikuntza-egokitasuna 43

7.3.2. $\mathbf{B_j}$ aldaparekiko inferentzia 44

7.4. Korrelazioa 44

7.5. Hondarren azterketa 45

7.6. Doikuntza egokia bada, **Y**-ren iragarpenak egiteko erabiltzea 46

A eranskina: Eredu anizkoitzaren eraikuntza urratsez urrats 47

B. eranskina: Zer egin daiteke *EXCEL* programarekin? 48

8. Bariantza-analisia **49**

8.1. Faktore bakarreko bariantza-analisia 49

8.2. Faktore biko bariantza-analisia ($\mathbf{n > 1}$) 51

8.3. Faktore biko bariantza-analisia ($\mathbf{n = 1}$) 51

A. eranskina: Zer egin daiteke *EXCEL* programarekin? 53

9. Kalitatearen kontrol estatistikoa **55**

9.1. Aldagaien grafikoak 55

9.1.1. Batez besteko balioak eta heinak ($\overline{x}$ eta R kontrol-grafikoak) 55

9.1.2. Batez besteko balioak eta desbideratze estandarrak ($\overline{x}$ eta s kontrol-grafikoak) 56

9.2. Atributuen grafikoak 56

9.2.1. Akastun unitateen ehunekoa (p kontrol-grafikoa) 56

9.2.2. Akastun unitateen kopurua (np kontrol-grafikoa) 57

9.2.3. Batez besteko akatsen kopurua unitateko (u kontrol-grafikoa) 58

9.2.4. Akatsen kopurua artikuluko (c kontrol-grafikoa) 59

9.3. Paretoren diagrama 59

Eranskina: ariketak **61**

Bibliografia **74**

Sarrera

Zalantza barik, Estatistika arlo anitzetan agertzen den interes handiko irakasgaia da. Are gehiago, jakintza-arlo guztietan aintzat hartzen da, Zientzia Esperimentaletan, Medikuntza eta Osasun Zientzietan, Irakaskuntza Teknikoetan, Gizarte eta Ekonomia Zientzietan, Zientzia Juridikoetan eta Giza Zientzietan, hain zuzen ere.

Bestalde, gaur egun gauza jakina da softwarea erabiltzea ezinbestekoa dela heziketa-prozesuan. *SPSS (Statistical Package for Social Sciences) Statistics* izeneko IBMren programaren bidez, arlo horretan espezializatua, Estatistikako oinarrizko kontzeptuak aplika daitezke era erraz eta zabalean. Nahiz eta programa ugari egon, erabilerraztasuna kontuan izan dugu hautua egitean. Aipatzekoa da programaren zehaztapenak 17.0 bertsioari dagozkiola (informatika-geletan 2009/2010 ikasturtean instalatuta zegoena); hori dela eta, lengoaiaren hitz batzuk gaztelaniaz agertuko dira.

Nahiz eta Estatistika irakasgaiari buruzko material ugari egon, tamalez, euskarazko bibliografia urria da, bereziki aplikazio informatikoei dagokienez. Materiala bost ikasturtetan osatu dut, Ingeniaritza Kimikako hirugarren mailako irakasle-lanetan aritu naizen bitartean, Zientzia eta Teknologia Fakultatean. Osatutako materiala lagungarria izan daiteke edozein Fakultate eta Eskolatan, Estatistika Deskribatzailearekin zein Inferentzia Estatistikoarekin lotutako irakasgaietan lantzeko.

Ikasmaterial honek, bi ataletan antolatuta, Estatistikaren oinarrizko kontzeptuak jorratzen ditu SPSS softwarearen bidez. I. atalean, bederatzi gairi buruzko bederatzi praktika antolatzen dira ariketa batzuen bidez: datu-baseak, probabilitate-teoria, estatistika deskribatzailea, konfiantza-tartezko zenbatespena, hipotesi-kontraste parametriko eta ez-parametrikoak, erregresioa, bariantza-analisia eta kalitatearen kontrola. Praktika bakoitza kapitulu bati dagokio. Praktikotan kontzeptuak birpasatzen dira hainbat adibideren laguntzaz (programa barneko fitxategiak zein liburu klasikoen ariketak) eta autoebaluaziorako ariketak proposatzen dira. II. atalean, eranskin moduan, ariketak eta bibliografia azaltzen dira.

Azkenik, eskerrak eman nahi dizkiet ikasleei eta Matematika Aplikatua, Estatistika eta Ikerkuntza Operatiboa Saileko irakasleei, haien ondoan ikasi dudan guztiagatik, eta bereziki, Euskara eta Eleaniztasuneko Errektoreordetzari emandako laguntzagatik, ikasmaterialgintza proiektu honen hizkuntza egokitzeko.

María Merino Maestre

2010eko apirilean

1. SPSS PRAKTIKA

Datu-baseak

Helburua

Praktika honen helburua da *SPSS* programaren bidez datu-baseak irakurtzea eta eraikitzeko gai izatea. Beste jarduera batzuen artean, ondokoak egingo dira: fitxategi zaharrak ireki, fitxategi berriak osatu, aldaketak gorde, informazioa bilatu, iragazkiak egin.

1.1. Datu-baseen irakurketa

Ireki ezazu `1991 U.S. General Social Survey.sav (Encuesta general USA 1991.sav)` izeneko fitxategia. Horretarako, `Archivo` → `Abrir` → `Datos` → `Escritorio` → `C:` → `Archivos de Programa` → `...` → `Samples` → `Spanish` → `fitxategia.sav`

Fitxategiari buruzko informazio orokorra bilatzeko: `Utilidades` → `Comentarios del archivo de datos`. Eta aldagai bakoitzari buruzkoa: `Utilidades` → `Variables` edo `Vista de variables` atalean, errenkadetan aldagaiak eta zutabeetan aldagaien ezaugarriak daude.

- `Nombre:` aldagaia identifikatzeko hitz labur bat.
- `Tipo: numerico` (zenbakizkoa dela adierazteko).
- `Anchura:` aldagaiaren balioen zifra-kopurua komaren ezkerraldean.
- `Decimales:` aldagaiaren balioen zifra-kopuru maximoa.
- `Etiqueta:` aldagaiaren izen luzea edo azalpena.
- `Valores:` balio bereziak definitu. Adibidez, 1=gizonezkoa, 2=emakumezkoa.

- `Perdidos:` balio bereziak. Adibidez, `0=no procede, 8=no sabe, 9=no contesta` honako hau adierazteko: 0=ez da egokia, 8=ez daki, 9=ez du erantzun.
- `Columnas:` zutabearen luzera.
- `Alineación:` lerrokatzea.
- `Medida:`
  - `Nominal:` kualitatiboa eta ordenarik gabea. Esate baterako, sexua.
  - `Ordinal:` kualitatiboa eta ordenatua. Esate baterako, iritzi graduala.
  - `Escala:` kuantitatiboa. Esate baterako, adina.

`Vista de datos` atalean, errenkadetan gizabanakoen erantzunak eta zutabeetan aldagaiak daude. Hemendik, balioen adierazpenak ikusteko: `Ver` → `Etiquetas de valor`

1.2. Datu-baseen eraikuntza

*Energiaren kontsumoa*ri buruzko inkesta bat egin da, eta pertsona batzuen erantzunak jaso dira. Osa ezazu datu-basea SPSSren bidez.

INKESTA:

Sexua: ☐ Gizonezkoa ☐ Emakumezkoa
Eg. zibila: ☐ Ezkongabea ☐ Ezkondua ☐ Alarguna ☐ Dibortziatua ☐ Banandua
Adina: ☐

1. galdera: Non bizi zara?
a) Apartamentu batean
b) Etxebizitza familiabakar batean
c) Beste batean

2. galdera: Zenbat kilometro egiten dituzu egunero automobilean edo motoan?
a) Ez dut inoiz erabiltzen
b) 5 km baino gutxiago
c) 5 eta 25 km bitartean
d) 25 eta 50 km bitartean
e) 50 km baino gehiago

3. galdera: Zenbat kilometro egiten dituzu egunero autobusean edo trenean?
a) Ez dut inoiz erabiltzen
b) 5 km baino gutxiago
c) 5 eta 25 km bitartean
d) 25 eta 50 km bitartean
e) 50 km baino gehiago

4. galdera: Behar ez izanda ere martxan izaten al dituzu gailu elektrikoak etxean?
a) Ez, inoiz ez
b) Bai, batzuetan telebista edo argia piztuta izaten ditut, etxean inor ez egon arren
...

ERANTZUNAK:

1. pertsona: Emakumezkoa, Ezkongabea, 22 urte, 1. a), 2. a), 3. c), 4. a), ...

2. pertsona: Emakumezkoa, ez du erantzuten, 47 urte, 1. b), 2. c), 3. a), 4. b), ...

3. pertsona: Gizonezkoa, Ezkongabea, ez du erantzuten, 1. a), 2. b), 3. c), 4. b), ...

4. pertsona: Emakumezkoa, Alarguna, 65 urte, 1. a), 2. a), 3. d), 4. a), ...

5. pertsona: Gizonezkoa, Ezkondua, 31 urte, 1. b), 2. c), 3. c), 4. b), ...

...

• Datu-basea osatu, aldagai bakoitzaren `medida` zehaztuz eta irizpide hauei jarraituz:

`Etiqueta [Nombre] (Valores, Perdidos)`

Sexua [sexua] (1=gizonezkoa, 2=emakumezkoa)

Egoera zibila [egoera] (1=ezkongabea, 2=ezkondua, 3= alarguna, 4=dibortziatua, 5=banandua, 9=faltako balioa)

Adina [adina] (faltako balioak: 97=ez da egokia, 98=ez daki, 99=ez du erantzuten)

Etxebizitza [galdera1] (1=a, 2=b, 3=c)

Automobilean edo motoan egunero egindako distantzia [galdera2] (1=a, 2=b, 3=c, 4=d, 5=e)

Autobusean edo trenean egunero egindako distantzia [galdera3] (1=a, 2=b, 3=c, 4=d, 5=e)

Etxeko gailu elektrikoen erabilera [galdera4] (0=ez, 1=bai)

• Bi iragazki (`filtro`) prestatuko ditugu: (i) 2. galderari lotutako kontsumo ertaina adierazteko (hots, b), c) eta d) erantzunak batzeko) eta (ii) 3. galderari lotutako muturreko kontsumoa adierazteko (hau da, a) eta e) erantzunak batzeko). Horretarako:

`Datos` → `Seleccionar casos` → `Si se satisface la condicion` →

(i) kasuan, → `(galdera2 >= 2)` & `(galdera2 <= 4)` → `Aceptar`

(ii) kasuan, → `(galdera3 = 1)` | `(galdera3 = 5)` → `Aceptar`

Kontuan hartu & ikurrak ebakidura adierazten duela, eta | ikurrak, bildura.

2. SPSS PRAKTIKA

Probabilitate-teoria

Helburua

Praktika honen jomuga da Probabilitate-teorian landutako kontzeptu batzuk jorratzea. Adibidez, zorizko aldagaien probabilitate-legea, dentsitate-funtzioa eta banaketa-funtzioak kalkulatzea, grafikoki irudikatzea, teorema batzuen aplikazioak zenbakiz eta grafikoki egiaztatzea.

2.1. Probabilitate-legea, dentsitate-funtzioa eta banaketa-funtzioa

Izan bitez X z.a., f funtzioa haren probabilitate-legea (z.a.d. kasuan) edo dentsitate-funtzioa (z.a.j. kasuan), eta F haren banaketa-funtzioa. $f(x)$ eta $F(x) = P(X \leq x)$ balioak SPSSren bidez kalkula daitezke PDF eta CDF funtzioak erabiliz, hurrenez hurren. Horretarako, `Archivo` $\rightarrow$ `Nuevo` $\rightarrow$ `Datos` $\rightarrow$ `1` $\rightarrow$ `Transformar` $\rightarrow$ `Calcular` bideari jarraitu behar zaio.

Baldin $X : Bin(n,\ p)$ bada, $f(x) =$ `PDF.BINOM(x, n, p)` eta $F(x) =$ `CDF.BINOM(x, n, p)`

Baldin $X : \mathcal{P}(\lambda)$ bada, $f(x) =$ `PDF.POISSON(x, λ)` eta $F(x) =$ `CDF.POISSON(x, λ)`

Baldin $X : \mathcal{U}[a,\ b]$ bada, $f(x) =$ `PDF.UNIFORM(x, a, b)` eta $F(x) =$ `CDF.UNIFORM(x, a, b)`

Baldin $X : \mathcal{E}(\lambda)$ bada, $f(x) =$ `PDF.EXP(x, λ)` eta $F(x) =$ `CDF.EXP(x, λ)`

Baldin $X : \mathcal{N}(\mu,\ \sigma)$ bada, $f(x) =$ `PDF.NORMAL(x, μ, σ)` eta $F(x) =$ `CDF.NORMAL(x, μ, σ)`

Baldin $X : \chi^2_n$ bada, $f(x) =$ `PDF.CHISQ(x, n)` eta $F(x) =$ `CDF.CHISQ(x, n)`

Baldin $X : t_n$ bada, $f(x) =$ `PDF.T(x, n)` eta $F(x) =$ `CDF.T(x, n)`

Baldin $X : \mathcal{F}_{m,n}$ bada, $f(x) =$ `PDF.F(x, m, n)` eta $F(x) =$ `CDF.F(x, m, n)`

Izan bedi $\alpha \in [0,\ 1]$ esangura-mailari dagokion θ_α puntu kritikoa, non $\theta_\alpha \in \{z_\alpha,\ \chi^2_{\alpha;n},\ t_{\alpha;n},\ F_{\alpha;m,n}\}$. Dakigunez, $P(X > \theta_\alpha) = \alpha$ denez, $P(X \leq \theta_\alpha) = 1-\alpha \Leftrightarrow F(\theta_\alpha) = 1-\alpha$; beraz, $\alpha = 1 - F(\theta_\alpha)$ eta $\theta_\alpha = F^{-1}(1-\alpha)$. Hau da, banaketa-funtzioaren alderantzizko funtzioa behar dugu puntu kritikoak kalkulatzeko, eta honela egiten da:

z_α=`IDF.NORMAL (`$1-\alpha$`, `μ`, `σ`)`, $\chi^2_{\alpha;n}$=`IDF.CHISQ (`$1-\alpha$`, `n`)`,
$t_{\alpha;n}$= `IDF.T (`$1-\alpha$`, `n`)` eta $F_{\alpha;m,n}$= `IDF.F (`$1-\alpha$`, `m`, `n`)`.

Autoebaluaziorako ariketak

Egin itzazu 2. ariketa-zerrendatik 2.1, 2.2, 2.3, 2.4 eta 2.5 ariketak, banaketen taulak erabili barik.

Soluzioa:

1. Izan bedi $X : Bin(5,\ 0{,}6)$ z.a.d, a) $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - F(2) =$ `1 - CDF.BINOM (2, 5, 0.6)=0.68256`
   b) $P(X \leq 2) = F(2) =$ `CDF.BINOM (2, 5, 0.6)=0.31744`
2. Izan bedi $X : \mathcal{P}(3{,}4)$ z.a.d, $P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) =$ `1 - CDF.POISSON (10, 3.4)=0.00081`
3. Izan bedi $X : Bin(5000, 0{,}00004)$ $P(X \leq 2) = F(2) =$ `CDF.BINOM (2, 5000, 0.00004)=0.99885`
   $Bin(5000, 0{,}00004) \approx \mathcal{P}(0{,}2)$ denez, $P(X \leq 2) = F(2) \approx$ `CDF.POISSON (2, 0.2)=0.99885`
4. Izan bedi $\Sigma_n : \mathcal{N}(1500, \sqrt{1500})$ z.a.j, a) $P(\Sigma_n > 1480) = 1 - P(\Sigma_n \leq 1480) = 1 - F(1480) =$ `1 - CDF.NORMAL (1480, 1500, SQRT(1500))=0.69721`
   b) $P(1480 < \Sigma_n < 1520) = F(1520) - F(1480) =$
   =`CDF.NORMAL (1520, 1500, SQRT(1500)) - CDF.NORMAL (1480, 1500, SQRT(1500))=0.39442`
5. a) Izan bedi $X : \mathcal{N}(1500, 38{,}7)$ z.a.j, $F(a) = 0{,}5 \Rightarrow a = F^{-1}(0{,}5) =$ `IDF.NORMAL (0.5, 1500, 38.7) )=1500`
   b) $F(b) = 0{,}25 \Rightarrow b = F^{-1}(0{,}25) =$ `IDF.NORMAL (0.25, 1500, 38.7) )=1473.87712`

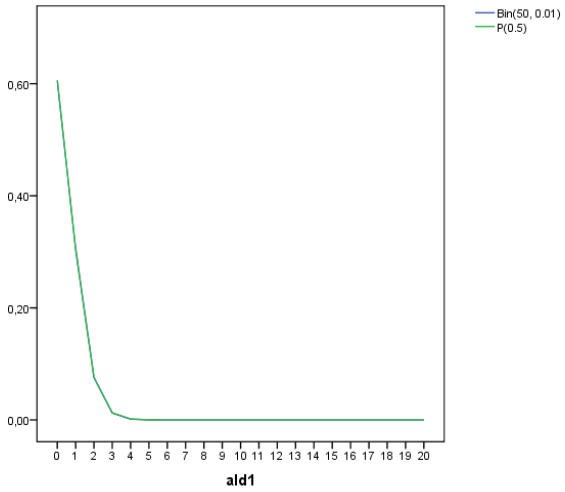
2.2. Probabilitate-teoriaren oinarrizko teorema batzuk

Ikus ditzagun grafikoki aztertutako banaketen hurbilketak.

• Baldin $n > 50$ eta $p < 0{,}1$ badira, $Bin(n,\ p) \approx \mathcal{P}(np)$.

Adibidez, ikus dezagun $Bin(50,\ 0{,}01) \approx \mathcal{P}(0{,}5)$. Osatu ondoko aldagaiak: `ald1`= {0, 1, 2, ..., 20}, `bin0.01=PDF.BINOM(ald1, 50, 0.01), poisson=PDF.POISSON(ald1, 0.5)`

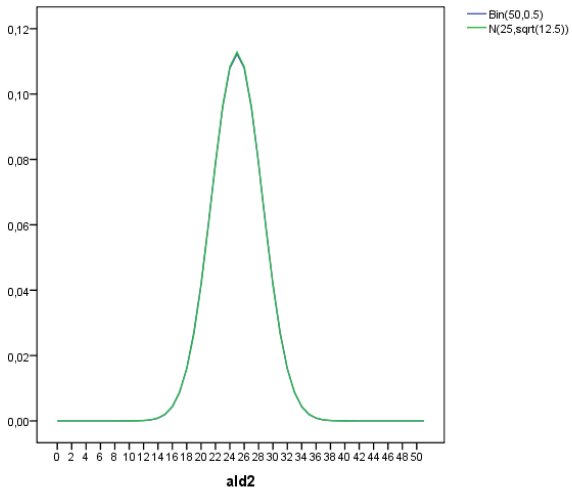
`Gráficos` → `CDA` → `Líneas`→ `Múltiple. Valores individuales de los casos` →
`Las lineas representan [bin0.01,poisson]. Variable [ald1]`



• Baldin $n > 30$ eta $0{,}1 < p < 0{,}9$ badira, $Bin(n,\ p) \approx \mathcal{N}(np,\ \sqrt{npq})$ (Moivreren Teorema).

Adibidez, ikus dezagun $Bin(50,\ 0{,}5) \approx \mathcal{N}(25,\ \sqrt{12{,}5})$. Osatu ondoko aldagaiak: `ald2`= $\{0,\ 1,\ 2,\ ...,\ 50\}$, `bin0.5=PDF.BINOM(ald2, 50, 0.5)` eta `normal=PDF.NORMAL(ald2, 25, SQRT(12.5))`.

`Gráficos` → `CDA` → `Líneas`→ `Múltiple. Valores individuales de los casos` →
`Las lineas representan [bin0.5,normal]. Variable [ald2]`

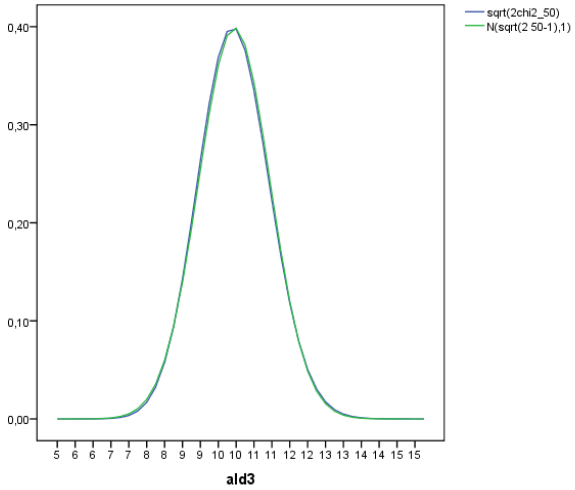


• Baldin $n > 30$ bada, $\chi^2_{\alpha;n} \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$.

Baliokideki, $\sqrt{2 \cdot \chi^2_{\alpha;n}} \approx z_\alpha + \sqrt{2n-1}$. Beraz, $\sqrt{2 \cdot \chi^2_n} \approx \mathcal{N}(\sqrt{2n-1}, 1)$ denez, $P(\chi^2_n \leq x^2/2) = P(\mathcal{N}(2n-1, 1) \leq x)$, hau da, $F_\chi(x^2/2) = F_\mathcal{N}(x)$ eta $x \cdot f_\chi(x^2/2) = f_\mathcal{N}(x)$.

Adibidez, ikus dezagun $\sqrt{2 \cdot \chi^2_{50}} \approx \mathcal{N}(\sqrt{2 \cdot 50 - 1}, 1)$. Osatu ondoko aldagaiak: `ald3`= $\{5,\ 5{,}25,\ 5{,}5,\ ...,\ 15\}$, `chi2f=ald3*PDF.CHISQ(ald3*ald3/2, 50)` eta `normalf=PDF.NORMAL(ald3, SQRT(2*50-1), 1)`.

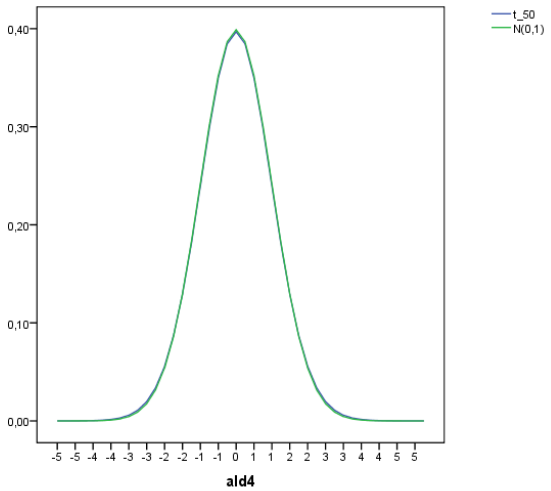
`Gráficos` → `CDA` → `Líneas`→ `Múltiple. Valores individuales de los casos` →
`Las lineas representan [chi2f,normal2f]. Variable [ald3]`



• Baldin $n > 30$ bada, $t_n \approx \mathcal{N}(0, 1)$.

Adibidez, ikus dezagun $t_{50} \approx \mathcal{N}(0, 1)$. Osatu ondoko aldagaiak: `ald4`= $\{-5, -4{,}75,\ -4{,}5, ...,\ 5\}$, `t=PDF.T(ald4, 50)` eta `normal01=PDF.NORMAL(ald4, 0, 1)`.

`Gráficos` → `CDA` → `Líneas`→ `Múltiple. Valores individuales de los casos` →
`Las lineas representan [t,normal01]. Variable [ald4]`



3. SPSS PRAKTIKA

Estatistika deskribatzailea

Helburua

Praktika honen xedea da jasotako datu esperimentalak aztertzea, laburtzea eta deskribatzea, metodo grafiko eta zenbakizkoen bidez. Bai elkartu gabeko datuak bai datu elkartuak erabiliz, honako hauek egiten ikasiko dugu:

1. Maiztasun-taulak
2. Grafikoak
3. Estatistikoak

Horretarako, klasean ikusitako adibide hau zein 3.9 ariketa erabiliko ditugu.

Adibidea: Birus baten latentzia-aldia ikertzeko, 90 txitari inokulatu zitzaien birusa. Bakoitzarengan gaixotasunaren lehenengo sintomak agertu arte pasatutako egun-kopurua aztertu zen. Beraz, X = egun-kopurua izeneko aldagaia da. Lortutako datuak honako hauek ziren:

8	*10*	*8*	*14*	*16*	*9*	*12*	*13*	*9*	*12*	*12*	*10*	*15*	*8*	*6*
5	*9*	*11*	*13*	*5*	*9*	*12*	*13*	*8*	*14*	*8*	*5*	*14*	*6*	*13*
7	*8*	*12*	*12*	*8*	*6*	*8*	*9*	*9*	*15*	*8*	*9*	*8*	*13*	*7*
9	*12*	*8*	*6*	*9*	*14*	*13*	*8*	*12*	*9*	*11*	*8*	*16*	*10*	*6*
10	*13*	*6*	*5*	*14*	*12*	*14*	*6*	*11*	*12*	*10*	*12*	*6*	*7*	*10*
6	*15*	*7*	*9*	*5*	*9*	*7*	*10*	*7*	*10*	*8*	*11*	*11*	*14*	*15*

3.1. Maiztasun-taulak

• Osa ezazu elkartu gabeko datuen maiztasun-taula estatistikoa.

`Analizar` → `Estadísticos descriptivos` → `Frecuencias`

Egun-kopurua (EGD)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	5	5	5,6	5,6	5,6
	6	9	10,0	10,0	15,6
	7	6	6,7	6,7	22,2
	8	14	15,6	15,6	37,8
	9	12	13,3	13,3	51,1
	10	8	8,9	8,9	60,0
	11	5	5,6	5,6	65,6
	12	11	12,2	12,2	77,8
	13	7	7,8	7,8	85,6
	14	7	7,8	7,8	93,3
	15	4	4,4	4,4	97,8
	16	2	2,2	2,2	100,0
	Total	90	100,0	100,0	

• Osa ezazu datu elkartuen maiztasun-taula estatistikoa. Esate baterako, kontsidera itzazu 2 luzeradun sei tartea, lehenengoa [4,5, 6,5] eta azkena (14,5, 16,5] izanik.

`Transformar` → `Recodificar` → `En distintas variables` → `Valores antiguos y nuevos` → `Rango` → `Valor` → `Añadir` → `Cambiar` → `Aceptar`

Egun-kopurua (DE)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	5,50	14	15,6	15,6	15,6
	7,50	20	22,2	22,2	37,8
	9,50	20	22,2	22,2	60,0
	11,50	16	17,8	17,8	77,8
	13,50	14	15,6	15,6	93,3
	15,50	6	6,7	6,7	100,0
	Total	90	100,0	100,0	

3.2. Grafikoak

• Egin ezazu barra-grafikoa (EGD):

`Gráficos` → `CDA` → `Barras` → `Simple` → `Resumen para grupos de casos` → `N`^{o} `de casos [`f_i`]` `edo % de casos [`h_i`]` → `Eje de categorias [aldagaia]`

• Osa ezazu grafiko metakorra (EGD):

`Gráficos` → `CDA` → `Barras` → `Simple` → `Resumen para grupos de casos` → `N`^{o} `acum. [`F_i`]` `edo % acum. [`H_i`]` → `Eje de categorias [aldagaia]`

• Egin ezazu zurtoin- eta hosto-grafikoa (EGD) eta kutxa-diagrama:

`Analizar` → `Estadísticos descriptivos` → `Explorar` → `Dependientes [aldagaia].`
`Visualización gráficos. Gráficos` → `Descriptivos` → `Tallo y hojas`

• Osa ezazu histograma (DE):

`Gráficos`→ `CDA` → `Histograma` → `Variable [aldagaia]`

Oharra: OY ardatzean f_i baino ez du erabiltzen; beraz, zabalera konstantedun a_i tarteak eraikitzea komenigarria da.

• Osa ezazu maiztasun-poligonoa (DE):

`Gráficos` → `CDA` → `Líneas` → `Simple` → `Resumen para grupos de casos` → `N`^{o} `de casos [`f_i`]` `edo % de casos [`h_i`]` → `Eje de categorias [aldagaia]`

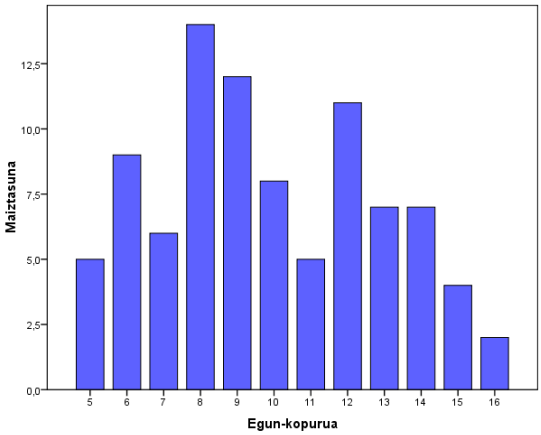
Oharra: OY ardatzean f_i eta h_i baino ez ditu erabiltzen.

• Egin ezazu maiztasun-poligono metakorra (DE):

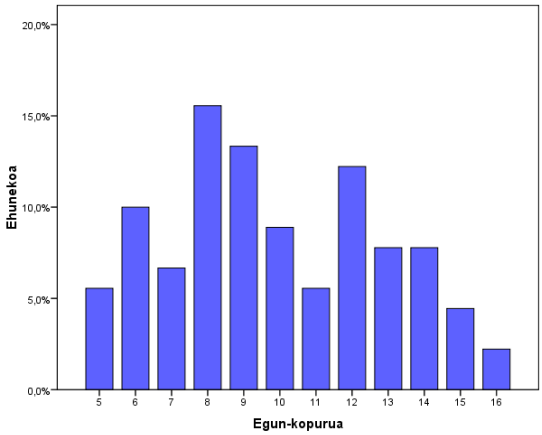
`Gráficos` → `CDA` → `Líneas` → `Simple` → `Resumen para grupos de casos` → `N`^{o} `acum. [`F_i`]` `edo % acum. [`H_i`]` → `Eje de categorias [aldagaia]`

Oharrak: Barra- eta lerro-grafikoetan grafikoetan aldagai kuantitatibo baten grafikoak ere eska ditzakegu beste aldagai kualitatibo baten bidez. Horretarako, `Panel mediante[kualitatiboa]`. Kutxa-diagrama lortzeko beste bide bat: `Graficos` → `CDA` → `Diagramas de caja` → `Simple. Resúmenes para distintas variables.` → `Las cajas representan: [aldagaia]`

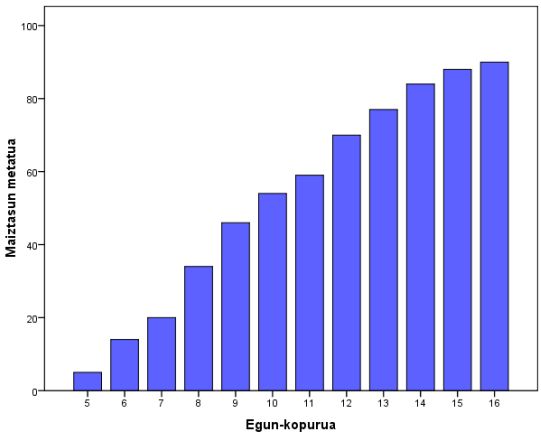
Barra-grafikoa (f_i)



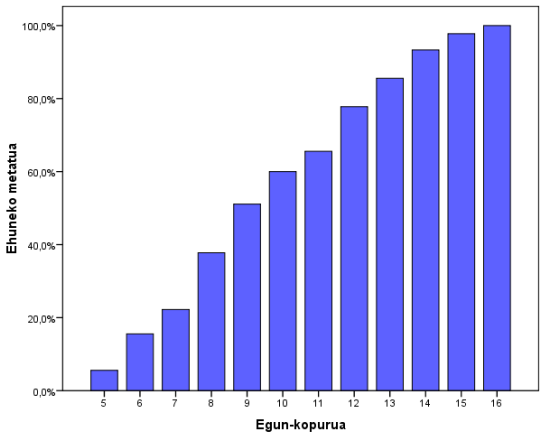
Barra-grafikoa (h_i)



Grafiko metakorra (F_i)



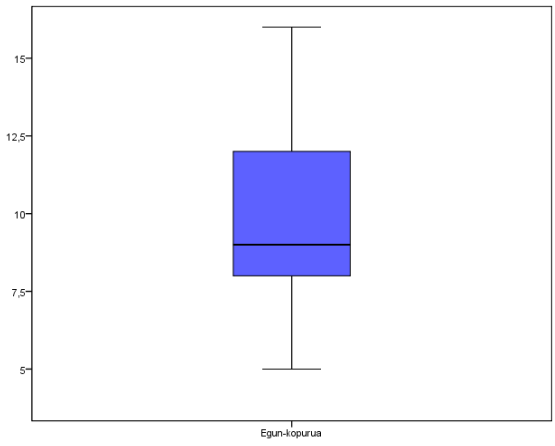
Grafiko metakorra (H_i)



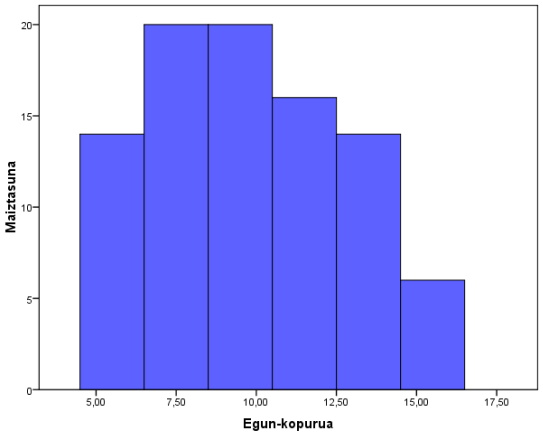
Zurtoin- eta hosto-grafikoa

```
Egun-kopurua Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     5,00        0 .  55555
    15,00        0 .  666666666777777
    26,00        0 .  88888888888888999999999999
    13,00        1 .  0000000011111
    18,00        1 .  222222222223333333
    11,00        1 .  44444445555
     2,00        1 .  66

 Stem width:  10
 Each leaf:       1 case(s)
```

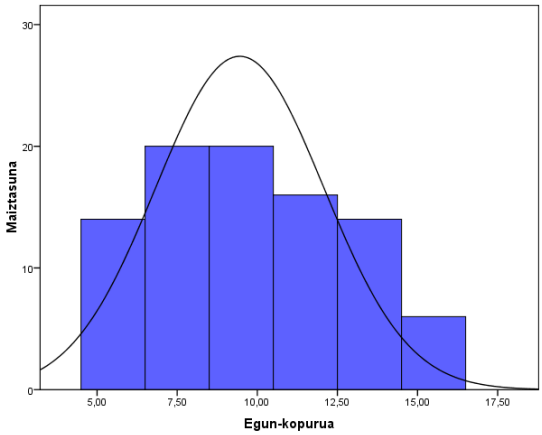
Kutxa-diagrama



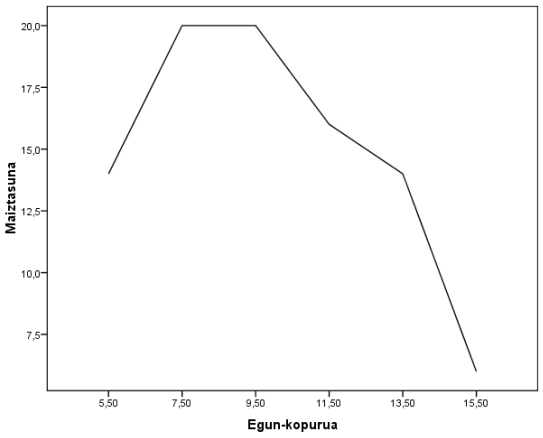
Histograma (f_i/a_i)



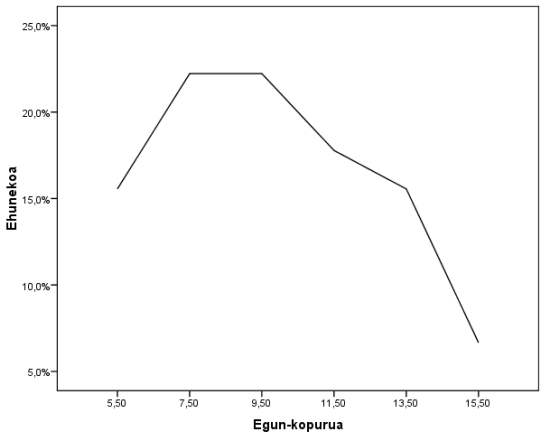
Histograma (f_i/a_i)



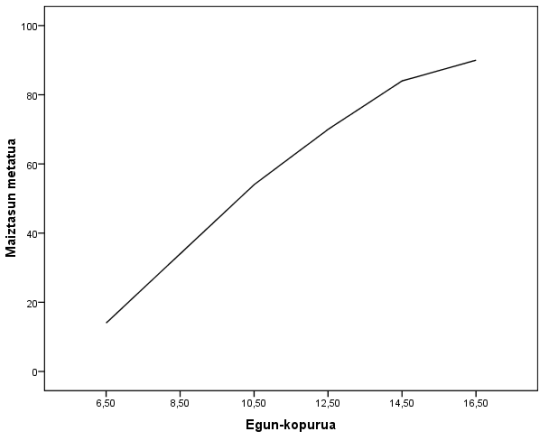
Maiztasun-poligonoa (f_i/a_i)



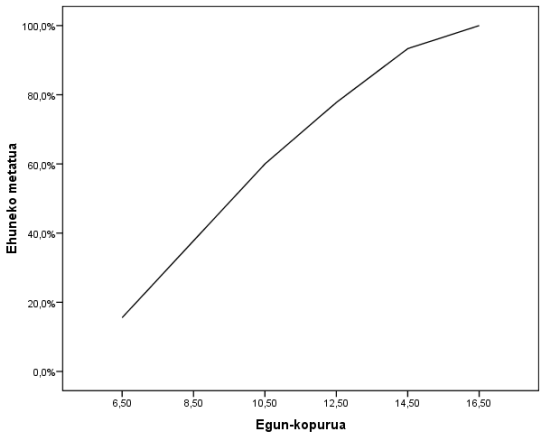
Maiztasun-poligonoa (h_i/a_i)



Ojiba edo maiztasun-poligono metakorra (F_i)



Ojiba edo maiztasun-poligono metakorra (H_i)



3.3. Estatistikoak

Aldagai kuantitatiboen edozein estatistiko kalkula daiteke, hots, joera zentraleko estatistikoak ($\overline{x}$, Me eta Mo); posizio-estatistikoak (p_i, q_i eta d_i); sakabanatze-estatistikoak (R, RI, S_n^2, S_n, S_{n-1}^2 eta CV) eta forma-estatistikoak (ν, g_3 eta g_4). Hala ere, aldagai kualitatibo ordinaletan, soilik Me, Mo, p_i, q_i, d_i, R eta RI kalkula daitezke. Azkenik, aldagai kualitatibo nominaletan Mo baino ezin da kalkulatu.

Oro har ez da komenigarria SPSSn datu elkartuak erabiltzea. Izan ere, alde batetik SPSS datukopuru handi batekin lan egiteko gai da, eta, bestalde, erdiko puntuak besterik ez du kontsideratzen, dagozkien jatorrizko tarteak *ahaztuz*.

• Estatistikoak kalkulatzeko era:

`Analizar` → `Estadísticos descriptivos` → `Frecuencias. Estadísticos` →
`Valores percentiles [Posizio-estatistikoak:` q_i, p_i`].`
`Tendencia central [Joera-zentraleko estatistikoak:` $\overline{x}$, Me, Mo`].`
`Dispersión [Sakabanapen estatistikoak:` S_{n-1}, S_{n-1}^2, R, x_{max}, x_{min}`].`
`Distribución [Forma-estatistikoak:` g_3, g_4`]`

Estadísticos

Egun-kopurua (EGD)

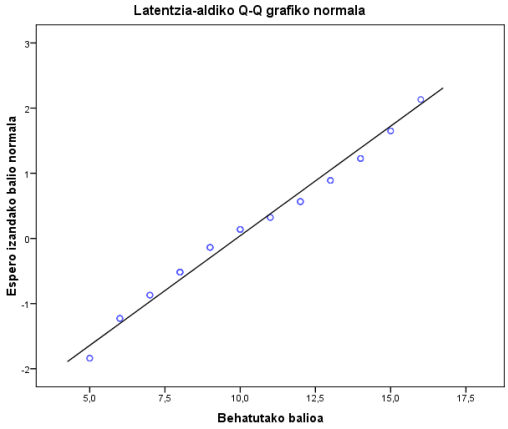
N	Válidos	90
	Perdidos	0
Media		9,88
Mediana		9,00
Moda		8
Desv. típ.		2,971
Varianza		8,828
Asimetría		,203
Error típ. de asimetría		,254
Curtosis		-,954
Error típ. de curtosis		,503
Rango		11
Mínimo		5
Máximo		16
Percentiles	25	8,00
	50	9,00
	75	12,00

Ikasi ditugun beste estatistikoak aurrekoak erabiliz kalkula daitezke; hauek dira: $RI = 4$, $S_n^2 = 8{,}73$, $S_n = 2{,}95$, $CV = 29{,}91\,\%$ eta $\nu = 0{,}64$. RI ere hemendik lor daiteke: `Analizar` → `Estadísticos descriptivos` → `Explorar`

3.4. Normaltasuna aztertzeko metodo deskribatzaileak

1. Histograma eta zurtoin- eta hosto-grafikoa osatu ditugu jadanik.
2. $\frac{RI}{s_n} = \frac{12-8}{2{,}9546} = 1{,}3538 \approx 1{,}3$
3. Probabilitate normalaren grafikoa egiteko:

`Analizar` → `Estadísticos descriptivos` → `Explorar` → `Gráficos` → `De tallo y hojas. Histograma. Gráficos con pruebas de normalidad`



Grafikoei begiratuz eta Kolmogorov-Smirnov-en hipotesi-kontrasteari jarraituz ($p = 0{,}001$), normaltasuna errefusatzen dugu.

Autoebaluaziorako ariketak

Egin ezazu 3. ariketa-zerrendatik 3.9 ariketa. Aldagaiaren balioak banan-banan ez sartzeko aukera duzu; horretarako bi aldagai eratu behar dira: aldagaiaren balioak (x_i) eta maiztasunak (f_i), hain zuzen ere.

x_i	0	100	200	300	400	1000	
f_i	19	4	5	4	2	1	35

Hori egin ondoren, urrats hauei jarraitu:

`Datos` → `Ponderar casos` → `Ponderar casos mediante [maiztasuna` f_i`]`

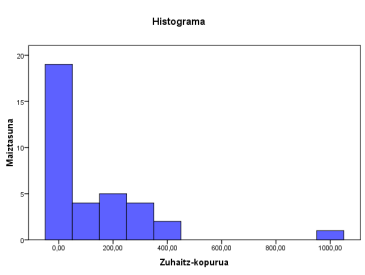
Estadísticos

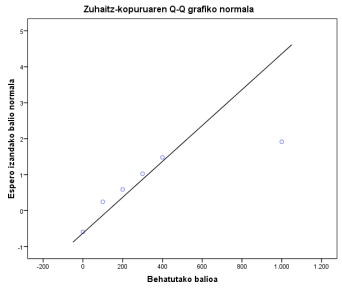
Zuhaitz-kopurua

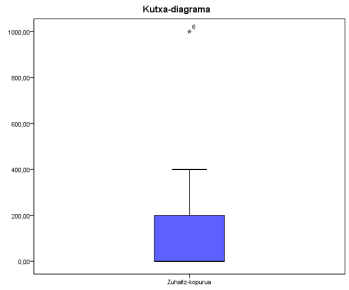
N	Válidos	35
	Perdidos	0
Media		125,71
Mediana		,00
Moda		0
Desv. típ.		200,504
Varianza		40201,681
Asimetría		2,689
Error típ. de asimetría		,398
Curtosis		9,835
Error típ. de curtosis		,778
Rango		1000
Mínimo		0
Máximo		1000
Percentiles	25	,00
	50	,00
	75	200,00

$RI = 200$, $S_n^2 = 39053{,}06$, $S_n = 197{,}62$, $CV = 157{,}20\,\%$ eta $\nu = 0{,}64$.

Normaltasuna aztertzean, $\frac{RI}{s_n} = \frac{200}{197{,}62} = 1{,}0120$ eta histograma, probabilitate normalaren grafikoa eta kutxa-diagrama:







Azkenik, ez dugu normaltasunik, eta balio arraro bakarra 1000 da.

4. SPSS PRAKTIKA

Konfiantza-tartezko zenbatespena

Helburua

Praktika honen bidez, konfiantza-tarteak kalkulatzeko `Pruebas T` izeneko 3 proba orokorrenak ikusiko ditugu (bariantza ezezaguna daukaten populazio normalentzat, lagin-tamainak edozein izanda edo populazio binomialentzat, lagin-tamainak handiak izanik):

1. `Prueba T para una muestra`: $I_{\mu}^{1-\alpha},\ I_{p}^{1-\alpha}$

2. `Prueba T para muestras independientes`: $I_{\mu_1-\mu_2}^{1-\alpha},\ I_{p_1-p_2}^{1-\alpha}$ (populazio askeak izanik).

3. `Prueba T para muestras relacionadas`: $I_{\mu_1-\mu_2}^{1-\alpha}$ (binakako datuak).

4.1. Lagin bakar baten zenbatespen-tarteak

• Egin dezagun 4.1 ariketa: $I_{\mu}^{0,90}$ tartea kalkulatu behar dugu. Zutabe batean asteko erabilitako ur litro-kopuruaren aldagaien datuak idazten ditugu.

`Analizar` → `Comparar medias` → `Prueba T para una muestra` → `Contrastar variables [aldagaia]. Valor de prueba [0]` → `Opciones` → `Intervalo de confianza [90%]. Excluir casos según análisis` → `Continuar` → `Aceptar`

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
Ur-litro kopurua	25	175,76	20,793	4,159

Irakurketa: $n = 25$, $\hat{\mu} = \overline{x} = 175{,}76$, $\hat{\sigma} = s_{n-1} = 20{,}793$ eta $\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s_{n-1}}{\sqrt{n}} = 4{,}159$

Prueba para una muestra

	Valor de prueba = 0					
					90% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Ur-litro kopurua	42,264	24	,000	175,760	168,65	182,87

Interpretazioa: populazioaren batez besteko ur-litro kopurua $I_{\mu}^{0,90} = (168{,}65,\ 182{,}87)$ tartean dago, eta %90 da konfidantza-maila; beraz, ur-depositua (160 litrokoa) ez da nahikoa izango.

• 4.14 ariketa erabiliz zenbatespen-tarte batzuk osatuko ditugu. Lehenengo zutabean, idatz ezazu aldagai kuantitatibo hau: denbora-tartea (segundotan). Bigarrenean aldagai kualitatiboa: sexua. Osa dezagun batezbestekorako puntu-zenbatespena eta %90eko konfiantza-tartea.

`Analizar` → `Estadísticos descriptivos` → `Explorar` → `Dependientes [aldagai kuantitatiboa]` → `Estadísticos` → `Intervalo de confianza para la media [90%]` → `Continuar` → `Aceptar`

			Estadístico	Error típ.
Denbora-tartea	Media		105,58	7,263
	Intervalo de confianza para la media al 90%	Límite inferior	92,54	
		Límite superior	118,63	

Hots, $\hat{\mu} = 105{,}58$ eta $I_{\mu}^{0,90} = (92{,}54,\ 118{,}63)$, non μ denbora-tartearen batezbestekoa den.

Osa dezagun gizonezkoen eta emakumezkoen batezbestekorako puntu-zenbatespenak eta %90eko konfiantza-tarteak.

`Analizar` → `Estadísticos descriptivos` → `Explorar` → `Dependientes [aldagai kuantitatiboa]. Factores [aldagai kualitatiboa]` → `Estadísticos` → `Intervalo de confianza para la media [90%]` → `Continuar` → `Aceptar`

	Sexua		Estadístico	Error típ.
Denbora-tartea	gizonezkoa		110,71	12,165
		Límite inferior	87,08	
		Límite superior	134,35	
	emakumezkoa		98,40	3,906
		Límite inferior	90,07	
		Límite superior	106,73	

Hots, $\hat{\mu_1} = 110{,}71$ seg, $\hat{\mu_2} = 98{,}40$ seg, $I_{\mu_1}^{0,90} = (87{,}08,\ 134{,}35)$ eta $I_{\mu_2}^{0,90} = (90{,}07,\ 106{,}73)$, non μ_1 eta μ_2 gizonezkoen eta emakumezkoen denbora-tartearen batezbestekoak diren, hurrenez hurren.

4.2. Lagin birako zenbatespen-tarteak

• Egin dezagun 4.14 ariketa: $I_{\mu_1-\mu_2}^{0,90}$ tartea kalkulatu behar dugu, non datuak askeak diren.

`Analizar` → `Comparar medias` → `Prueba T para muestras independientes` → `Contrastar variables [aldagai kuantitatiboa]. Variable de agrupación [aldagai kualitatiboa]. Definir grupos [Grupo 1: 1, Grupo2 : 0]` → `Opciones` → `Intervalo de confianza [90%]. Excluir casos según análisis` → `Continuar`

	Sexua	N	Media	Desviación típ.	Error típ. de la media
Denbora-tartea	emakumezkoa	5	98,40	8,735	3,906
	gizonezkoa	7	110,71	32,185	12,165

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									90% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Denbora-tartea	Se han asumido varianzas iguales	3,887	,077	-,824	10	,429	-12,314	14,952	-39,414	14,786
	No se han asumido varianzas iguales			-,964	7,187	,366	-12,314	12,777	-36,427	11,798

Bariantzen berdintasuna aztertzeko, Levene-ren proba erabiltzen da. Baldin $sig \leq \alpha$, bariantzak berdinak direla baztertzen da. Hala ere, $sig > \alpha$ bada, bariantzak berdinak direla ezin da errefusatu. Kasu honetan $sig = 0{,}077 < \alpha = 0{,}10$ denez, bariantzak desberdinak direla ondorioztatzen dugu. Horrela, $I_{\mu_1-\mu_2}^{0,90} = (-36{,}427,\ 11{,}798)$ batezbestekoen diferentziarako %90eko konfiantza-mailako zenbatespen-tartea da. $0 \in I_{\mu_1-\mu_2}^{0,90}$ dagoenez, ezin da ondorioztatu batezbesteko bat bestea baino hobea denik.

• Egin dezagun 4.15 ariketa: $I_{\mu_1-\mu_2}^{0,98}$ tartea kalkulatu behar dugu, non binakako datuak dauden. Bi zutabetan tomate fresko eta ontziraturikoen kobre-kopuruaren datuak idazten ditugu.

`Analizar` → `Comparar medias` → `Prueba T para muestras relacionadas` → `Variables emparejadas: Variable 1 [Freskoa], Variable 2 [Lataraturikoa].` → `Opciones` → `Intervalo de confianza [98%]. Excluir casos según análisis` → `Continuar` → `Aceptar`

Prueba de muestras relacionadas

	Diferencias relacionadas					t	gl	Sig. (bilateral)
				98% Intervalo de confianza para la diferencia				
	Media	Desviación típ.	Error típ. de la media	Inferior	Superior			
Par 1 Tomate freskoa - Latakoa	-,011700	,008394	,002654	-,019189	-,004211	-4,408	9	,002

Irakurketa: $\overline{d} = -0{,}0117$, $s_{n-1,D} = 0{,}008394$, $\frac{s_{n-1,D}}{\sqrt{n}} = 0{,}002654$

Interpretazioa: populazioen kobre-kopuruaren batezbestekoen arteko diferentzia $I_{\mu_D}^{0,98} = I_{\mu_1-\mu_2}^{0,98} = (-0{,}019189, \ -0{,}004211)$ tartean dago %98 konfidantza-mailarekin; beraz, $\mu_1 - \mu_2 < 0$, hau da, tomate freskoen kobre-kopurua ontziraturikoena baino baxuagoa dela ondoriozta daiteke.

4.3. Populazio binomialetarako zenbatespen-tarteak

• Egin dezagun 4.17 ariketa: $I_p^{0,95}$ tartea kalkulatu behar dugu. Zutabe batean 6 huts (jaurtiketaren porrotei dagozkienak) eta 34 bat (jaurtiketaren arrakastei dagozkienak) idazten ditugu. 4.1 ariketan egin dugun bideari jarrai diezaiokezu.

	N	Media	Desviación típ.	Error típ. de la media
Jaurtiketa	40	,85	,362	,057

Prueba para una muestra

	Valor de prueba = 0					
					95% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Jaurtiketa	14,866	39	,000	,850	,73	,97

• Egin dezagun 4.18 ariketa: $I_{p_1-p_2}^{0,94}$ tartea kalkulatu behar dugu. Zutabe batean 0,1,0,1 (A markako eta B markako porrota-arrakasta adierazirik, hurrenez hurren) idazten ditugu. Bigarren zutabean aurreko zutabeari dagokion maiztasuna idazten dugu, alegia, $n_1 - x_1 = 158$, $x_1 = 42$, $n_2 - x_2 = 132$, $x_2 = 18$. Hirugarren zutabean, aldagai kualitatiboen balioak idazten ditugu, adibidez, 1 A markakoentzat eta 2 B markakoentzat.

`Datos` → `Ponderar casos mediante` → `Variable de frecuencia [maiztasuna]`

Azkenik, 4.14 ariketan egin dugun bideari jarrai diezaiokezu.

	Marka	N	Media	Desviación típ.	Error típ. de la media
Artikuluak	marka A	200	,21	,408	,029
	marka B	150	,12	,326	,027

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									94% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Artikuluak	Se han asumido varianzas iguales	21,360	,000	2,220	348	,027	,090	,041	,014	,166
	No se han asumido varianzas iguales			2,292	346,610	,023	,090	,039	,016	,164

Autoebaluaziorako ariketak

1. Ireki ezazu `world95.sav` fitxategia. a) Jaiotza-tasaren eta heriotza-tasaren artean desberdintasun nabarmenik dagoela onar daiteke? b) Eskualde ekonomikoa kontuan hartuta, Asian eta Latinoamerikan jaiotza-tasa desberdina dela ondoriozta daiteke? c) Asia eta Latinoamerika artean, jaiotza-tasa eta heriotza-tasaren arteko diferentzia aldatzen dela onar daiteke?

2. Egin ezazu 4.19 ariketa.

Emaitzak:

a) $I^{0,95}_{\mu_1-\mu_2} = (14{,}2641,\ 18{,}6583)$ denez, $\mu_1 \neq \mu_2$ onar daiteke %95eko konfiantza-mailarekin eta $\mu_1 - \mu_2 > 0$, hau da, $\mu_1 > \mu_2$. Batez besteko jaiotza-tasa heriotza-tasa baino handiagoa dela onar daiteke.

b) $I^{0,95}_{\mu_3-\mu_6} = (-6{,}8292,\ 5{,}5609)$ denez, $\mu_3 = \mu_6$ onar daiteke %95eko konfiantza-mailarekin eta ezin da onartu jaiotza-tasa altuagoa dela eskualde batean bestean baino.

c) $I^{0,95}_{\mu_A-\mu_L} = (-6{,}54986,\ 3{,}58498)$ denez, $\mu_A = \mu_L$ onar daiteke %95eko konfiantza-mailarekin eta ezin da onartu tasen arteko diferentzia aldatzen denik.

5. SPSS PRAKTIKA

Hipotesi-kontraste parametrikoak

Helburua

Praktika honetan, hipotesi-kontraste parametrikoak kalkulatzeko, 5 proba ikusiko ditugu. Lehenengo `Pruebas T` izeneko 3 probak batezbestekoak aztertzeko kasu orokorrenak dira (bariantza ezezaguna daukaten populazio normalak, lagin-tamainak edozein izanda, edo populazio binomialak, lagin-tamainak handiak izanik); 4. proba binomiala da eta 5. proba grafikoki konparatzeko bidea:

1. `Prueba T para una muestra`: $H_0 : \mu = \mu_0$
2. `Prueba T para muestras independientes`: $H_0 : \mu_1 = \mu_2,\ H_0 : p_1 = p_2$ (pop. askeak).
3. `Prueba T para muestras relacionadas`: $H_0 : \mu_1 = \mu_2$ (binakako datuak).
4. `Prueba no paramétrica (binomial)`: $H_0 : p = 0{,}5$ edo $H_0 : p \leq p_0$ edo $H_0 : p \geq p_0$.
5. Batezbestekoak konparatzeko metodo grafikoa

5.1. Lagin bakar baten hipotesi-kontrasteak ($H_0 : \mu = \mu_0$)

• Egin dezagun 5.7 ariketa. Hipotesi nulua $H_0 : \mu_0 = 160$ da. Zutabe batean asteko erabilitako ur litro-kopuruaren aldagaien datuak idazten ditugu.

`Analizar` → `Comparar medias` → `Prueba T para una muestra` → `Contrastar variables [aldagaia]. Valor de prueba [160]` → `Opciones` → `Intervalo de confianza [90%]. Excluir casos según análisis` → `Continuar` → `Aceptar`

Prueba para una muestra

	Valor de prueba = 160					
	t	gl	Sig. (bilateral)	Diferencia de medias	90% Intervalo de confianza para la diferencia	
					Inferior	Superior
Ur-litro kopurua	3,790	24	,001	15,760	8,65	22,87

$t = 3{,}790$ estatistikoa da (t_p)

$gl = 24$ askatasun-graduak dira $(n-1)$

$sig(bilateral) = 2P(t_{n-1} > |t_p|) = 2P(t_{24} > 3{,}790) = 0{,}001$ $p - balioa$ da

Interpretazioa: $p < \alpha = 0{,}10$ denez, H_0 errefusatzen da, hots, $\mu \neq 160$ litro/aste. Gainera, konfiantza-tartea $I_\mu^{0,90} = 160 + (8{,}65,\ 22{,}87) = (168{,}65,\ 182{,}87)$ denez, $\mu > 160$ litro/aste ondoriozta daiteke %10eko esangura-mailarekin.

$H_0 : \mu \leq 160$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} > t_p) = P(t_{24} > 3{,}790) = 0{,}001/2 = 0{,}0005 < \alpha$. Beraz, $\mu > 160$.

$H_0 : \mu \geq 160$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} < t_p) = P(t_{24} < 3{,}790) = 1 - 0{,}0005 = 0{,}9995 > \alpha$. Beraz, $\mu \geq 160$. Kasu hau beherehalakoa da eskualde kritikoa eta estatistikoa konparatzen badugu; izan ere, $S_1 = (-\infty, -t_{0,1;24}]$ eta $t_p > 0$ direnez, $t_p \notin S_1 \Rightarrow H_0$.

5.2. Lagin birako hipotesi-kontrasteak

5.2.1. Populazio askeak ($H_0 : \mu_1 = \mu_2$)

• Egin dezagun 5.11 ariketa. Hipotesi nulua $H_0 : \mu_1 = \mu_2$ da. Zutabe batean denbora-tartea aldagaiaren bi laginen datuak idazten ditugu, bata bestearen atzean. Bigarren zutabean aldagai kualitatiboen balioak idazten ditugu, adibidez, 1 emakumezkoentzat eta 2 gizonezkoentzat.

`Analizar` → `Comparar medias` → `Prueba T para muestras independientes` → `Contrastar variables [aldagai kuantitatiboa]. Variable de agrupación [aldagai kualitatiboa]. Definir grupos [Grupo 1: 1, Grupo2 : 2]` → `Opciones` → `Intervalo de confianza [90%]. Excluir casos según análisis` → `Continuar` → `Aceptar`

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									90% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Denbora-tartea	Se han asumido varianzas iguales	3,887	,077	-,824	10	,429	-12,314	14,952	-39,414	14,786
	No se han asumido varianzas iguales			-,964	7,187	,366	-12,314	12,777	-36,427	11,798

Levene-ren probaren bidez ikusten dugu $sig = 0{,}077 < 0{,}1$; beraz, bi populazioen bariantzak desberdinak direla onar daiteke %90eko konfiantza-mailarekin. Orduan, bigarren errenkadari begiratuz,

$t = -0{,}964$ estatistikoa da (t_p)

$gl = 7$ askatasun-graduak dira

$sig(bilateral) = 2P(t_g > |t_p|) = 2P(t_7 > 0{,}964) = 0{,}366$ $p - balioa$ da

Interpretazioa: $p > \alpha = 0{,}10$ denez, H_0 ezin da errefusatu, hots, $\mu_1 = \mu_2$. Gizonezkoen batezbestekoa eta emakumezkoena berdina denik ezin da ukatu %10eko esangura-mailarekin.

$H_0 : \mu_1 \geq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_g < t_p) = P(t_7 < -0{,}964) = 0{,}366/2 = 0{,}183 > \alpha$. Beraz, $\mu_1 \geq \mu_2$.

$H_0 : \mu_1 \leq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_g > t_p) = P(t_7 > -0{,}964) = 1-0{,}183 = 0{,}817 > \alpha$. Beraz, $\mu_1 \leq \mu_2$. Kasu hau beherehalakoa da eskualde kritikoa eta estatistikoa konparatzen badugu; izan ere, $S_1 = [t_{0{,}1;7}, \infty)$ eta $t_p < 0$ direnez, $t_p \notin S_1 \Rightarrow H_0$.

Beraz, berdintasuna besterik ezin da ondorioztatu.

5.2.2. Binakako datuak ($H_0 : \mu_1 = \mu_2$ edo $H_0 : \mu_D = d_0$)

• Egin dezagun 5.13 ariketa. Hipotesi nulua $H_0 : \mu_1 = \mu_2$. 1. eta 2. zutabeetan tomate freskoaren eta ontziraturikoen kobre-kopuruaren datuak idazten ditugu, hurrenez hurren.

`Analizar → Comparar medias → Prueba T para muestras relacionadas → Selecciones actuales: Variable 1 [Freskoa], Variable 2 [Lataraturikoa]. Variables relacionadas [freskoa--lataraturikoa] → Opciones → Intervalo de confianza [98%]. Excluir casos según análisis → Continuar → Aceptar`

Prueba de muestras relacionadas

	Diferencias relacionadas					t	gl	Sig. (bilateral)
				98% Intervalo de confianza para la diferencia				
	Media	Desviación típ.	Error típ. de la media	Inferior	Superior			
Par 1 Tomate freskoa - Latakoa	-,011700	,008394	,002654	-,019189	-,004211	-4,408	9	,002

Baliokideki, $D = X_1 - X_2$ aldagaia osa daiteke eta 5.1. atalari jarraituz (`valor de prueba=0`).

$t = -4{,}408$ estatistikoa da (t_p)

$gl = 9$ askatasun-graduak dira $(n-1)$

$sig(bilateral) = 2P(t_{n-1} > |t_p|) = 2P(t_9 > 4{,}408) = 0{,}002$ $p - balioa$ da

Interpretazioa: $p < \alpha = 0{,}02$ denez, H_0 errefusatzen da, hots, $\mu_1 \neq \mu_2$. Tomate ontziraturikoen eta freskoen batez besteko kobre-kopurua berdina denik ezin da ondorioztatu %2ko esanguramailarekin. Zenbatespen-tarteari begiratuz, $I^{0,98}_{\mu_1-\mu_2} = (-0{,}0192, \ -0{,}0042) \Rightarrow \mu_1 - \mu_2 < 0$, hots $\mu_2 > \mu_1$. Beraz, tomate ontziraturikoen batez besteko kobre-kopurua freskoena baino altuagoa dela onar daiteke %98ko konfiantza-mailarekin.

$H_0 : \mu_1 \geq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} < t_p) = P(t_9 < -4{,}408) = 0{,}002/2 = 0{,}001 > \alpha$. Beraz, $\mu_1 < \mu_2$.

$H_0 : \mu_1 \leq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} > t_p) = P(t_9 > -4{,}408) = 1-0{,}001 = 0{,}999 > \alpha$. Beraz, $\mu_1 \leq \mu_2$. Kasu hau beherehalakoa da eskualde kritikoa eta estatistikoa konparatzen badugu; izan ere, $S_1 = [t_{0{,}02;9}, \infty)$ eta $t_p < 0$ direnez, $t_p \notin S_1 \Rightarrow H_0$.

• Azkenik, hipotesi nulua $H_0 : \mu_2 = \mu_1 + 0{,}0030$ kontrastatu nahi dugu, hots, $H_0 : \mu_2 = \mu_1'$ non $\mu_1' = \mu_1 + 0{,}0030$ den. Horretarako, $X_1' = X_1 + 0{,}0030$ aldagai berria osatuko dugu, adibidez, freskoa003 izendatuz.

`Datos` → `Insertar variable`

`Transformar` → `Calcular` → `Variable de destino [freskoa003]` → `Expresión numérica [Freskoa+0.0030]` → `Aceptar`

`Analizar` → `Comparar medias` → `Prueba T para muestras relacionadas` → `Selecciones actuales: Variable 1 [freskoa003], Variable 2 [Lataraturikoa]. Variables relacionadas [freskoa003--lataraturikoa]` → `Opciones` → `Intervalo de confianza [98%]. Excluir casos según análisis` → `Continuar` → `Aceptar`

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	98% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Freskoa+0.0030 - Latakoa	-,008700	,008394	,002654	-,016189	-,001211	-3,278	9	,010

Baliokideki, $D = X_1 - X_2$ aldagaia osa daiteke eta 5.1. atalari jarraituz (`valor de prueba=-0.003`).

$t = -3{,}278$ estatistikoa da (t_p)

$gl = 9$ askatasun-graduak dira $(n-1)$

$sig(bilateral) = 2P(t_{n-1} > |t_p|) = 2P(t_9 > 3{,}278) = 0{,}010$ $p - balioa$ da

Interpretazioa: $p < \alpha = 0{,}02$ denez, H_0 errefusatzen da, hots, $\mu_1' \neq \mu_2$. Zenbatespen-tarteari begiratuz, $I^{0,98}_{\mu_1'-\mu_2} = (-0{,}016189,\ -0{,}001211) \Rightarrow \mu_1' - \mu_2 < 0$, hots $\mu_2 > \mu_1' = \mu_1 + 0{,}0030$. Beraz, tomate ontziraturikoen batez besteko kobre-kopurua freskoena baino gutxienez 0.0030 altuagoa dela onar daiteke %98ko konfiantza-mailarekin.

$H_0 : \mu_1' \geq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} < t_p) = P(t_9 < -3{,}278) = 0{,}010/2 = 0{,}005 < \alpha$. Beraz, $\mu_1' < \mu_2$.

$H_0 : \mu_1' \leq \mu_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_{n-1} > t_p) = P(t_9 > -3{,}278) = 1 - 0{,}005 = 0{,}995 > \alpha$. Beraz, $\mu_1' \leq \mu_2$. Kasu hau beherehalakoa da.

5.3. Populazio binomialetarako hipotesi-kontrasteak

5.3.1. Lagin bakar baterako ($H_0 : p = 0{,}5$ edo $H_0 : p \geq p_0$ edo $H_0 : p \leq p_0$)

Lagin bakar baten kasuan, SPSSk erabakitzen du egingo duen kontraste mota. Baldin $p_0 = 0{,}5$ bada, aldebiko kontrastea egingo du. Bestela, hipotesi alternatiborako laginean behatutako aldean egingo du. Hots, $\hat{p} > p_0$ bada, $\begin{cases} H_0: & p \leq p_0 \\ H_1: & p > p_0 \end{cases}$ eta $\hat{p} < p_0$ bada, $\begin{cases} H_0: & p \geq p_0 \\ H_1: & p < p_0 \end{cases}$. Bigarren kasuan, honako ohar hau adierazirik: `La hipótesis alternativa establece que la proporción de casos del primer grupo sea` $< p_0$.

Gainera, p, arrakasta-proportzioa, 1. behaketaren taldeari (`Grupo 1`) lotuta doa.

• Egin dezagun 5.17 ariketa. Zutabe batean, lehenik, 34 bat (jaurtiketaren arrakastei dagozkienak) eta bigarrenik, 6 huts (jaurtiketaren porrotei dagozkienak) idazten ditugu. $p_0 = 0{,}8$ denez eta

$\hat{p} = 0{,}85 > 0{,}8$, orduan, hipotesi alternatiboa $H_1 : p > 0{,}8$ eta hipotesi nulua $H_0 : p \leq 0{,}8$ dira.

`Analizar` → `Pruebas no paramétricas` → `Binomial` → `Contrastar variables [aldagaia]. Contrastar proporción [0.8]` → `Continuar` → `Aceptar`

Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
Jaurtiketa arrakastatsua	Grupo 1	arrakasta	34	,9	,8	,286[a]
	Grupo 2	porrota	6	,2		
	Total		40	1,0		

a. Basado en la aproximación Z.

Interpretazioa: p-balioa $Sig.asintot.(unilateral) = P(Z > z_p) = 0{,}286 > \alpha = 0{,}05$ denez, H_0 ez da errefusatzen; hots, $p \leq 0{,}8$, ezin da onartu $p > 0{,}8$. Ondoriozta daiteke %5eko esangura-mailarekin jaurtiketaren arrakasta-proportzioa 0.8 baino hobea ez dela.

Beste kontrasteak ezin dira egin, baina zenbatespen-tartea kalkulatzean, $I_p^{0{,}95} = (0{,}7393,\ 0{,}9607)$, hau da, ondorioztatu ahal dugu $p = 0{,}8$, baina ez $p > 0{,}8$ edo $p > 0{,}8$ denik.

$H_0 : p \geq 0{,}8$ hipotesi nulua duen kontrastearen p-balioa: $p = P(Z < z_p) = 1 - 0{,}286 = 0{,}0714 > \alpha$. Beraz, ezin da $p < 0{,}8$ onartu.

$H_0 : p = 0{,}8$ hipotesi nulua duen kontrastearen p-balioa: $p = 2P(Z > |z_p|) = 2P(Z > z_p) = 2 \cdot 0{,}286 > \alpha$. Beraz, $p = 0{,}8$.

5.3.2. Lagin birako ($H_0 : p_1 = p_2$)

• Egin dezagun 5.20 ariketa. Kontrastatu nahi dugun hipotesi nulua $H_0 : p_1 - p_2 = 0{,}08$ da, hots, $H_0 : p_1 = p_2'$ non $p_2' = p_2 + 0{,}08$ den. Horretarako, zutabe batean 0, 1, 0, 1 (A markako eta B markako porrota-arrakasta adierazirik, hurrenez hurren) idatzi beharrean, 0, 1, 0.08, 1.08 idatziko ditugu. Bigarren zutabean, aurreko zutabeari dagokion maiztasuna idazten dugu, alegia, $n_1 - x_1 = 158$, $x_1 = 42$, $n_2 - x_2 = 132$, $x_2 = 18$. Hirugarren zutabean, aldagai kualitatiboen balioak idazten ditugu, adibidez, 1 A markakoentzat eta 2 B markakoentzat.

`Datos` → `Ponderar casos mediante` → `Variable de frecuencia [maiztasuna]`

Azkenik, 5.11 ariketan egin dugun bideari jarrai diezaiokezu. Berez, estatistikoaren banaketa $\mathcal{N}(0,1)$ den arren, programak bi populaziorako proba binomiala ez daukanez, Student-en t banaketa erabiliko dugu. $a.g. > 30$ kasuan, hurbilketa ontzat eman dezakegu.

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									94% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Erostea	Se han asumido varianzas iguales	21,360	,000	,247	348	,805	,01000	,04054	-,06649	,08649
	No se han asumido varianzas iguales			,255	346,610	,799	,01000	,03927	-,06411	,08411

$t = 0{,}255$ estatistikoa da (t_p)

$gl = 346{,}610$ askatasun-graduak dira (g)

$sig(bilateral) = 2P(t_g > |t_p|) = 2P(t_{347} > 0{,}255) = 0{,}799$ $p - balioa$ da

Interpretazioa: $p > \alpha = 0{,}06$ denez, H_0 ez da errefusatzen, $p_1 = p_2'$; hots, $p_1 - p_2 = 0{,}08$. Hau da, ondoriozta daiteke %6ko esangura-mailarekin A markako saldutako artikulu-proportzioaren eta B markakoenaren arteko diferentzia %8koa dela.

$H_0 : p_1' \leq p_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_g > t_p) = P(t_{347} > 0{,}255) = 0{,}799/2 = 0{,}3995 > \alpha$. Beraz, ezin da $p_1' > p_2$ onartu.

$H_0 : p_1' \geq p_2$ hipotesi nulua duen kontrastearen p-balioa: $p = P(t_g < t_p) = P(t_{347} < 0{,}255) = 1 - 0{,}3995 = 0{,}6005 > \alpha$. Beraz, ezin da $p_1' < p_2$ onartu.

5.4. Batezbestekoak konparatzeko metodo grafikoa

Bi laginen batezbestekoak konparatzeko haien kutxa-diagramak erabil ditzakegu. Bi kasu desberdindu behar ditugu: bi aldagaiak bi zutaberen bidez adierazita daudenean (binakako datuak direnean) edo zutabe batean bi aldagaiak eta beste zutabe batean sailkatzeko aldagai kualitatiboa daudenean (populazio askeak direnean).

Egin dezagun apunteetan ikusitako adibidea:

Lur mota berberean 20 zuhaitz landatu ziren, jasotako eguzki- eta ur-kantitatea berbera zelarik. Landatzean, zuhaitzen erdiak ez zuen nitrogenorik jaso (kontrol moduan) eta beste erdiak bai. 140 egun igaro ondoren, enborren pisuen balioak (gramotan) honako hauek izan ziren. Egin ezazu kontraste grafikoa enborren batez besteko pisuak konparatzeko, nitrogeno jaso ez zuten eta jaso zutenen artean (jo dezagun banaketa normalak zirela).

Nitrogeno barik	*0.32*	*0.53*	*0.28*	*0.37*	*0.47*	*0.43*	*0.36*	*0.42*	*0.38*	*0.43*
Nitrogenoarekin	*0.26*	*0.43*	*0.47*	*0.49*	*0.52*	*0.75*	*0.79*	*0.86*	*0.62*	*0.46*

• [Binakako datuak] Bi aldagaiak bi zutaberen bidez adierazita:

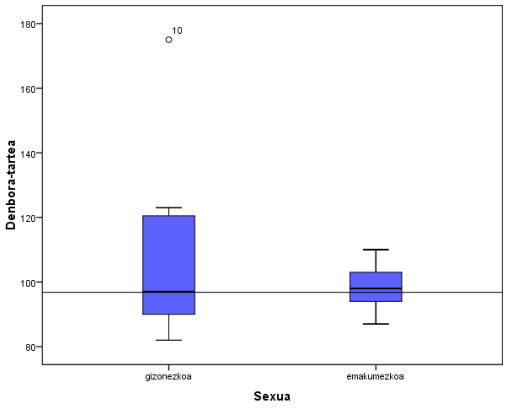
`Gráficos` → `CDA` → `Diagramas de caja` → `Simple[Resúmenes para distintas variables]` → `Definir` → `Las cajas representan` [X_1 `aldagaia,` X_2 `aldagaia]` → `Aceptar`

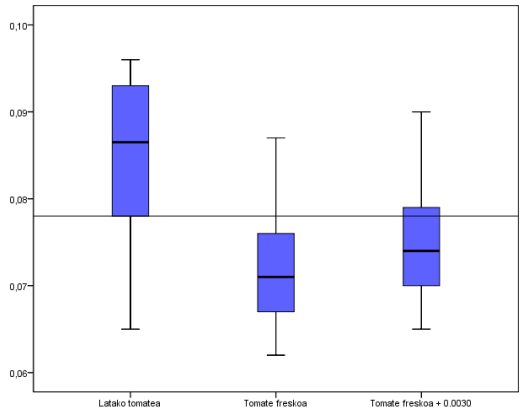
• [Populazio askeak] Zutabe batean bi aldagaiak eta beste zutabe batean sailkatzeko aldagai kualitatiboa:

`Gráficos` → `CDA` → `Diagramas de caja` → `Simple[Resúmenes para grupos de casos]` → `Definir` → `Variable` [X `aldagai kuantitatiboa]. Eje de categorías [aldagai kualitatiboa]` → `Aceptar`

2. laginaren 1. koartila 1. laginaren mediana baino handiagoa denez, batezbestekoen diferentzia estatistikoki nabarmena dela ondoriozta daiteke. $q_2^{Nbarik} < q_1^N \Rightarrow \mu^{Nbarik} < \mu^N$, hau da, nitrogenoa jaso zuten zuhaitzek nitrogenoa jaso ez zutenek baino pisu handiagoa zuten.

• Egin itzazu 5.11 eta 5.13 ariketak grafikoki.





5.11 ariketaren grafikoari begiratuz, ezin da onartu $\mu_1 < \mu_2$, ezin da onartu $\mu_1 > \mu_2$ eta ezin da errefusatu $\mu_1 = \mu_2$.

5.13 ariketaren grafikoari begiratuz, ezin da errefusatu $\mu_1 < \mu_2$ eta ezin da errefusatu $\mu_1' < \mu_2$, non $\mu_1' = \mu_1 + 0{,}0030$ den.

Oharra: Bi aldiz grafikoa sakatu ondoren, marra horizontal bat irudikatzeko aukera dago. Bestalde,

`Analizar` → `Estadísticos descriptivos` → `Explorar` → `Dependientes` [X `aldagaia]. Factores[aldagai kualitatiboa] Estadísticos[percentiles]`

Horrela, q_1 eta q_2 zenbakizko balioak aurki daitezke.

Autoebaluaziorako ariketak

Ireki ezazu `sales.sav (sales.sav)` fitxategia.

1. a) Batez besteko sarrera [beneficio(ingreso)] 2500 dolar delako hipotesia onar daiteke? Eta 2500 dolar baino handiagoa edo txikiagoa dela? b) Batez besteko sarrera 2400 dolar delako hipotesia onar daiteke? Eta 2400 dolar baino handiagoa dela?

2. Konpara ezazu (h-k baten bidez eta grafikoki) batez besteko sarreren berdintasuna, arreta-denbora [atencion] minutu bat baino gutxiago eta 4 minutu baino gehiago den kasuetan.

3. Iparraldekoen eta hegoaldekoen batez besteko sarrera berdina delako hipotesia onar daiteke?

Ireki ezazu `1991 U.S. General Social Survey.sav (Encuesta general USA 1991.sav)` fitxategia.

4. a) Neba-arreben eta seme-alaben batez besteko kopuruak berdinak direlako hipotesia onar daiteke? Eta bata bestea baino handiagoa dela? b) Neba-arreben batezbestekoa seme-alabena baino zehazki 2 pertsona gehiago delako hipotesia onar daiteke?

Ireki ezazu `Cars.sav (Coches.sav)` fitxategia.

5. a) Iragazkia [filter] hautatua eta ez hautatua izateko probabilitatea berdina dela onar daiteke? b) Eta iragazkia ez hautatua izateko probabilitatea %30 baino txikiagoa dela? c) Eta iragazkia ez hautatua izateko probabilitatea %20 baino handiagoa dela?

6. AEBn eta Europan saldutako kotxeen artean, iragazkia hautatua izateko probabilitatea berdina dela onar daiteke? Zein herrialdetan da gertagarriagoa?

Emaitzak:

1. a) $H_0 : \mu = 2500$ kontrastea egitean $t_p = 0{,}643$ eta $p = 0{,}520 > 0{,}05 = \alpha$. Beraz, $\mu = 2500$ onar daiteke %5eko esangura-mailarekin, eta zenbatespen-tarteari begiratuz, $I = 2500 + (-34,\ 67{,}15)$ ezin dugu ondorioztatu $\mu > 2500$ edo $\mu < 2500$ denik.
   b) $H_0 : \mu = 2400$ kontrastea egitean $t_p = 4{,}521$ eta $p = 0{,}000 < 0{,}05 = \alpha$. Beraz, $\mu = 2400$ ezin da onartu %5eko esangura-mailarekin. Eta zenbatespen-tarteari begiratuz, $I = 2400 + (66,\ 167{,}15)$, $\mu > 2400$ ondorioztatu ahal dugu.

2. $H_0 : \mu_1 = \mu_2$ kontrastea egitean $t_p = 10{,}886$ eta $p = 0{,}000 < 0{,}05 = \alpha$. Beraz, $\mu_1 \neq \mu_2$ onar daiteke %5eko esangura-mailarekin. Zenbatespen-tarteari begiratuz, $I = (618{,}854,\ 891{,}109)$, beraz $\mu_1 - \mu_2 > 0$, hau da, $\mu_1 > \mu_2$. Grafikoki, gauza bera ikusten dugu. Argiago ikusteko, grafikoan marra horizontala irudika daiteke edo q_1 eta q_2 balioak kalkula daitezke; izan ere, $q_1^{<1min} = 2252{,}87 > q_2^{>4min} = 2206{,}92 \Rightarrow \mu_1^{<1min} > \mu_2^{>4min}$.

3. $H_0 : \mu_1 = \mu_2$ kontrastea egitean $t_p = -0{,}952$ eta $p = 0{,}341 > 0{,}05 = \alpha$. Beraz, $\mu_1 = \mu_2$ onar daiteke %5eko esangura-mailarekin.

4. a) $H_0 : \mu_1(neba - arreba) = \mu_2(seme - alaba)$ kontrastea egitean $t_p = 24{,}464$ eta $p = 0{,}000 < 0{,}05 = \alpha$. Beraz, $\mu_1 \neq \mu_2$ onar daiteke %5eko esangura-mailarekin. Est-tarteari begiratuz, $I^{0,95}_{\mu_1-\mu_2} = (1{,}870,\ 2{,}196)$, beraz $\mu_1 - \mu_2 > 0$, hau da, $\mu_1 > \mu_2$.
b) $H_0 : \mu_1 = \mu_2 + 2$ kontrastea egitean $t_p = 0{,}393$ eta $p = 0{,}694 > 0{,}05 = \alpha$. Beraz, $\mu_1 = \mu_2 + 2$ onar daiteke %5eko esangura-mailarekin. Zenbatespen-tarteari begiratuz, ezin da ondorioztatu $\mu_1 - \mu_2 > 2$ edo $\mu_1 - \mu_2 < 2$ denik.

5. a) Izan bedi $p =$ "iragazkia ez hautatua" izateko probabilitatea (1. behaketa 0 delako). $H_0 : p = 0{,}5$ kontrastea egitean $p - balioa = 0{,}000 < 0{,}05 = \alpha$. Beraz, $p \neq 0{,}5$ onar daiteke %5eko esangura-mailarekin. b) $H_0 : p \geq 0{,}3$ kontrastea egitean $p - balioa = 0{,}096 > 0{,}05 = \alpha$. Beraz, $p < 0{,}3$ ezin da onartu %5eko esangura-mailarekin. c) $H_0 : p \leq 0{,}2$ kontrastea egitean $p - balioa = 0{,}001 < 0{,}05 = \alpha$. Beraz, $p > 0{,}2$ onar daiteke %5eko esangura-mailarekin.

6. Izan bedi $p_1 =$ "AEBko kotxeen artean iragazkia hautatua" izateko proportzioa eta $p_2 =$ "Europako kotxeen artean iragazkia hautatua" izatekoa. $H_0 : p_1 = p_2$ kontrastea egitean $t_p = -13{,}590$ eta $p = 0{,}000 < 0{,}05 = \alpha$. Beraz, $p_1 \neq p_2$ onar daiteke %5eko esangura-mailarekin. Zenbatespen-tarteari begiratuz, $I = (-0{,}484,\ -0{,}362)$, beraz $p_1 - p_2 < 0$, hau da, $p_1 < p_2$, Europan iragazkia hautatua izatea AEBn baino gertagarriagoa da.

6. SPSS PRAKTIKA

Hipotesi-kontraste ez-parametrikoak

Helburua

Praktika honetan, hipotesi-kontraste ez-parametriko mota batzuk kalkulatzeko bidea aurkezten da: (1) doikuntza-egokitasunerako probak, (2) independentzia- eta homogenotasun-probak, (3) zorizkotasun-probak eta (4) populazioak konparatzeko probak.

6.1. Doikuntza-egokitasunerako kontrastea

• $H_0 : (p_1, p_2, \ldots, p_k) = (p_1^0, p_2^0, \ldots, p_k^0)$ hipotesi nulua kontrastatzeko proba.

Egin dezagun 6.2 ariketa. Hipotesi nulua $H_0 : (p_1, p_2, p_3, p_4) = (0{,}16,\ 0{,}24,\ 0{,}20,\ 0{,}40)$ da. Zutabe batean lorearen kolorearen aldagai kualitatiboa eta bestean behatutako maiztasuna adieraziko ditugu. Programari ohartarazi zenbat aldiz agertzen den behaketa bakoitza (`Datos` → `Ponderar casos`)

`Analizar` → `Pruebas no paramétricas` → `Chi cuadrado` → `Contrastar variables [kolorea]. Valores esperados [0.16, 0.24, 0.20, 0.40]` → `Aceptar`

Lorearen kolorea

	N observado	N esperado	Residual
gorria	70	80,0	-10,0
arrosa	126	120,0	6,0
horia	96	100,0	-4,0
zuria	208	200,0	8,0
Total	500		

Estadísticos de contraste

	Lorearen kolorea
Chi-cuadrado[a]	2,030
gl	3
Sig. asintót.	,566

$Chi-cuadrado = 2{,}030$ χ^2_p estatistikoa da.

$Sig.asintot. = 0{,}566$ $p-balioa$ da.

Interpretazioa: $p > \alpha = 0{,}05$ denez, H_0 onartzen da.

• Kolmogorov-Smirnov-en kontrastea. $H_0 : X = \mathcal{N}(\hat{\mu}, \hat{\sigma})$ (normal) edo $H_0 : X = \mathcal{P}(\hat{\lambda})$ (poisson) edo $H_0 : X = \mathcal{E}(\hat{\lambda})$ (esponentziala) edo $H_0 : X = \mathcal{U}[a, b]$ (uniforme) hipotesi nulua kontrastatzeko proba. Proba honekin, ez da beharrezkoa klaseen bilketa itxarondako maiztasunak txikiak direnean.

Egin dezagun 6.4 ariketa. Hipotesi nulua $H_0 : X = \mathcal{P}(\hat{\lambda})$ da.

Organismo-kopurua	0	1	2	3	4	5	6	7
Lagin-kopurua	15	30	25	20	5	4	1	0

`Analizar` → `Pruebas no paramétricas` → `K-S de 1 muestra` → `Contrastar variables [aldagaia]. Distribucion de contraste [Poisson]` → `Aceptar`

Prueba de Kolmogorov-Smirnov para una muestra

		Organismo-kopurua
N		100
Parámetro de Poisson[a,b]	Media	1,86
Diferencias más extremas	Absoluta	,019
	Positiva	,019
	Negativa	-,015
Z de Kolmogorov-Smirnov		,185
Sig. asintót. (bilateral)		1,000

a. La distribución de contraste es la de Poisson.

b. Se han calculado a partir de los datos.

$Sig.asintot.(bilateral) = 1{,}000$ $p-balioa$ da.

Interpretazioa: $p > \alpha = 0{,}05$ denez, H_0 onartzen da, hots, $X = \mathcal{P}(1{,}86)$.

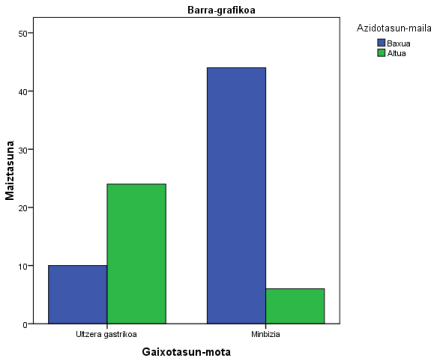
6.2. Independentzia-kontrastea eta homogenotasun-kontrastea

• Egin dezagun 6.6 adibidearen 2x2 kontingentzia-taula. Hipotesi nulua H_0 : azidotasun-maila gaixotasunarekiko askea da. Hiru aldagai osatu behar ditugu, bi kualitatibo (azidotasun-maila eta gaixotasun-mota) eta kuantitatibo bat (maiztasuna adierazteko). Programari ohartarazi zenbat aldiz agertzen den behaketa bakoitza (`Datos` → `Ponderar casos`)

`Analizar` → `Estadísticos descriptivos` → `Tabla de contingencia` → `Filas [gaixotasun-mota]. Columnas [azidotasun-maila]. Estadísticos [chi cuadrado]. Casillas [Frecuencias. Porcentajes]. Mostrar los gráficos de barras agrupados.` → `Aceptar`

Tabla de contingencia Gaixotasun-mota * Azidotasun-maila

			Azidotasun-maila		
			baxua	altua	Total
Gaixotasun-mota	ultzera gastrikoa	Recuento	10	44	54
		Frecuencia esperada	21,9	32,1	54,0
		% de Gaixotasun-mota	18,5%	81,5%	100,0%
		% de Azidotasun-maila	29,4%	88,0%	64,3%
		% del total	11,9%	52,4%	64,3%
	minbizia	Recuento	24	6	30
		Frecuencia esperada	12,1	17,9	30,0
		% de Gaixotasun-mota	80,0%	20,0%	100,0%
		% de Azidotasun-maila	70,6%	12,0%	35,7%
		% del total	28,6%	7,1%	35,7%
Total		Recuento	34	50	84
		Frecuencia esperada	34,0	50,0	84,0
		% de Gaixotasun-mota	40,5%	59,5%	100,0%
		% de Azidotasun-maila	100,0%	100,0%	100,0%
		% del total	40,5%	59,5%	100,0%



Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	30,258[b]	1	,000		
Corrección por continuidad[a]	27,760	1	,000		
Razón de verosimilitud	31,608	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	29,897	1	,000		
N de casos válidos	84				

a. Calculado sólo para una tabla de 2x2.

Yates-en zuzenketa aplikatuz, $Chi-cuadrado = 27{,}76$ χ^2_p estatistikoa da.

$Sig.asintot.(bilateral) = 0{,}000$ $p-balioa$ da.

Interpretazioa: $p < \alpha = 0{,}05$ denez, H_0 errefusatzen da, hots, menpekoak dira.

• Egin dezagun 6.7 ariketa (3x3 kontingentzia-taula). Hipotesi nulua H_0 : erantzuna eta probintzia askeak izatea da. Hiru aldagai osatu behar ditugu, bi kualitatibo (erantzuna eta probintzia) eta kuantitatibo bat (maiztasuna adierazteko). Programari ohartarazi zenbat aldiz agertzen den behaketa bakoitza (`Datos` → `Ponderar casos`)

`Analizar` → `Estadísticos descriptivos` → `Tabla de contingencia` → `Filas [erantzuna]. Columnas [probintzia]. Estadísticos [chi cuadrado]` → `Aceptar`

Tabla de contingencia erantzuna * probintzia

Recuento

		probintzia			
		Araba	Bizkaia	Gipuzkoa	Total
erantzuna	alde	11	13	9	33
	kontra	32	28	27	87
	abstentzioa	7	9	14	30
Total		50	50	50	150

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3,810[a]	4	,432
Razón de verosimilitud	3,739	4	,442
Asociación lineal por lineal	1,918	1	,166
N de casos válidos	150		

$Chi - cuadrado = 3{,}810$ χ^2_p estatistikoa da.

$Sig.asintot.(bilateral) = 0{,}432$ $p - balioa$ da.

Interpretazioa: $p > \alpha = 0{,}05$ denez, H_0 onartzen da, hots, askeak dira.

6.3. Zorizkotasun-kontrastea

Bolada-testa egiteko,

`Analizar` → `Pruebas no paramétricas` → `Rachas` → `Lista contrastar variables [aldagaia]. Punto de corte[?]` → `Aceptar`

• Aldagai kualitatibo dikotomikoa den kasuan, `punto de corte: media` aukera daiteke. Egin dezagun 6.11 ariketa. H_0 : gazteak eta helduak zoriz hautatuak izan dira. Horrela, bolada-kopurua $\hat{R} = 29$, estatistikoa $z_p = -0{,}513$ eta p-balioa $p = 0{,}608 > \alpha$, beraz ezin dugu hipotesi nulua errefusatu, zoriz hautatu dira.

• Aldagai kuantitatiboa den kasuan, `punto de corte: mediana` izan daiteke. Egin dezagun 6.12 ariketa. H_0 : zorizkotasuna. Horrela, mediana $Me = 93{,}75$, bolada-kopurua $\hat{R} = 28$, estatistikoa $z_p = 0{,}572$ eta p-balioa $p = 0{,}568 > \alpha$; beraz, ezin dugu hipotesi nulua errefusatu.

6.4. Populazioak konparatzeko kontrasteak

6.4.1. Bi lagin askeren konparaketa

Mann-Whitney-ren testa egiteko,

`Analizar` → `Pruebas no paramétricas` → `2 muestras independientes` → `Lista contrastar variables [aldagaia kuantitatiboa]. Variable de agrupación [aldagai kualitatiboa]. U de Mann-Whitney` → `Aceptar`

• Egin dezagun 6.13 ariketa. H_0 : ikasle berrien eta errepikatzaileen batez besteko puntuazioak berdinak dira. Horrela, heinen baturak $\hat{R}_1 = 248$ eta $\hat{R}_2 = 217$, estatistikoak $\hat{U}_1 = 97$ eta $z_p = -0{,}643$ eta p-balioa $p = 0{,}520 > \alpha$; beraz, ezin dugu berdintasuna errefusatu.

6.4.2. Bi lagin askeren baino gehiagoren konparaketa

Kruskal-Wallis-en testa egiteko,

`Analizar` → `Pruebas no paramétricas` → `k muestras independientes` → `Lista contrastar variables [aldagaia kuantitatiboa]. Variable de agrupación [aldagai kualitatiboa]. H de Kruskal-Wallis` → `Aceptar`

• Egin dezagun 6.14 ariketa. H_0 : hiru metodoekin batez besteko notak berdinak dira. Horrela, batez besteko heinak $\frac{\hat{R}_1}{n_1} = 8{,}40$, $\frac{\hat{R}_2}{n_2} = 10{,}17$ eta $\frac{\hat{R}_3}{n_3} = 11{,}89$, estatistikoa $K = 1{,}145$ eta p-balioa $p = 0{,}5664 > \alpha$; beraz, ezin dugu berdintasuna errefusatu.

6.4.3. Binakako datuen bi laginen konparaketa

`Analizar` → `Pruebas no paramétricas` → `2 muestras relacionadas` → `Contrastar pares [`X `eta` Y `aldagaiak]. Tipo de prueba[?]` → `Aceptar`

Bi aldagai kuantitatibo jarraitu konparatzeko **zeinudun heinen Wilcoxon-en testa** egiten da, hots, `Tipo de prueba: Wilcoxon`.

• Egin dezagun 6.15 ariketa. H_0 : iragazkinik gabeko batez besteko kontsumoa ez da iragazkinekikoa baino handiagoa. Horrela, heinen baturak $T^+ = 327$ eta $T^- = 108$, estatistikoa $W = -2{,}373$ eta $p = 0{,}018 < \alpha$; beraz, ezin dugu onartu hipotesi nulua, hots, ezin dugu baieztapena errefusatu.

Bi aldagai kualitatibo dikotomiko konparatzeko `Tipo de prueba: Mc Nemar`.

6.4.4. Binakako datuen bi laginen baino gehiagoren konparaketa

`Analizar` → `Pruebas no paramétricas` → `k muestras relacionadas` → `Variables de contraste` [$X_1,\ X_2, \ldots, X_k$ `aldagaiak`]. `Tipo de prueba[?]` → `Aceptar`

Bi aldagai kuantitatibo jarraitu konparatzeko, `Tipo de prueba: Friedmann`.

Bi aldagai kualitatibo konparatzeko, `Tipo de prueba: W de Kendall` (z.a. ordinalak) edo `Q de Cochran` (z.a. dikotomikoak).

Autoebaluaziorako ariketak

• Ireki ezazu `Cars.sav (Coches.sav)` fitxategia.
a) Onar daiteke %5eko esangura-mailarekin iragazkia [filter] aldagaia kontsideratzean hautatua eta hautatua ez izatearen proportzioak berdinak direla?
b) Eta onar daiteke hipotesi nulua $H_0 : (p_1, p_2) = (\frac{1}{4}, \frac{3}{4})$, p_1 = hautatua ez izatearen eta p_2 = hautatua izatearen proportzioak direlarik.
c) Onar daiteke %5eko esangura-mailarekin azelerazioa banaketa normalari darraiola?

(Em.: a) Ez. $\chi^2_p = 85{,}065$ eta $p = 0{,}000 < \alpha = 0{,}05 \Rightarrow H_0$ errefusatzen da.
b) Bai. $\chi^2_p = 0{,}754$ eta $p = 0{,}385 > \alpha = 0{,}05 \Rightarrow H_0$ ez da errefusatzen.
c) Bai. $p = 0{,}326 > \alpha = 0{,}05 \Rightarrow H_0$ onartzen da $\Rightarrow \mathcal{N}(15{,}50,\ 2{,}821)$ banaketari darraio.)

• Ireki ezazu `GSS93 subset.sav (GSS93 reducido.sav)` fitxategia.
a) Onar daiteke %5eko esangura-mailarekin joera politikoa [política] eta sexua aldagaiak askeak direla?
b) Zein da emakumezko liberalen behatutako eta itxarondako maiztasunak? Zenbat pertsona dira kontserbadoreak?
c) Emakumezkoen artean, zein da moderatuak direnen ehunekoa? Eta gizonezkoen artean?
d) Moderatuak direnen artean, zeintzuk dira emakumezkoen eta gizonezkoen portzentajeak?

(Em.: a) Bai. $\chi^2_p = 7{,}215$ eta $p = 0{,}125 > \alpha = 0{,}05 \Rightarrow H_0$ ez da errefusatzen $\Rightarrow$ askeak dira.
b) 110 eta 109.5; 282. c) %38.1 eta %34.5. d) %59.2 eta %40.8)

• Egin itzazu 5.11, 5.13, 8.1 eta 8.6 ariketak.

• Ireki ezazu `survey_sample.sav` fitxategia. Izan bitez X_1 = inkestatutakoak betetako hezkuntzako urte kopuru handiena [educ], X_2 = bere aitak betetakoa [paeduc], X_3 = bere amak betetakoa [maeduc] eta X_4 = bere bikotekideak betetakoa [espeduc] z.a.. Friedmann-en testaren bidez, azter itzazu hauek: a) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ eta b) $H_0 : \mu_2 = \mu_3$. c) Cochran-en testaren bidez, ondoriozta daiteke berriak jasotzeko 5 iturriak (egunkariak, aldizkariak, telebista, irratia eta Internet) berdinak direla?

Emaitzak:

5.11 H_0 : emakumezkoen eta gizonezkoen batez besteko denbora-tarteak berdinak dira. Horrela, estatistikoak $U_1 = 16$ eta $z_p = -0{,}244$ eta p-balioa $p = 0{,}808 > \alpha$; beraz, ezin dugu berdintasuna errefusatu.

5.13 H_0 : tomate fresko eta ontziraturikoen batez besteko kobre-kopurua berdina da. Horrela, estatistikoa $W = -2{,}705$ eta p-balioa $p = 0{,}007 < \alpha$; beraz, ezin dugu onartu hipotesi nulua.

8.1 H_0 : batez besteko heriotza-tasa berdina da urteko sasoi guztietan. Estatistikoa $K = 20{,}205$ eta p-balioa $p < \alpha$; beraz, ezin dugu berdintasuna onartu.

8.6 H_0 : batez besteko altuerak berdinak dira espezieekiko. Estatistikoa $K = 15{,}971$ eta p-balioa $p = 0{,}003 < \alpha$; beraz, ezin dugu berdintasuna onartu.

sav a) n=957, p-balioa $p = 0{,}000 < \alpha$; beraz, ezin dugu onartu berdintasuna. b) n=1907, p-balioa $p = 0{,}977 > \alpha$; beraz, gurasoen artean hezkuntza berdina dela ondorioztatzen da. c) Ez, p-balioa $p = 0{,}000 < \alpha$ baita.

7. SPSS PRAKTIKA

Erregresioa

Helburua

Praktika honen helburua Erregresio Lineal eta Anizkoitzean aplikatzen diren teknikak adieraztea da.

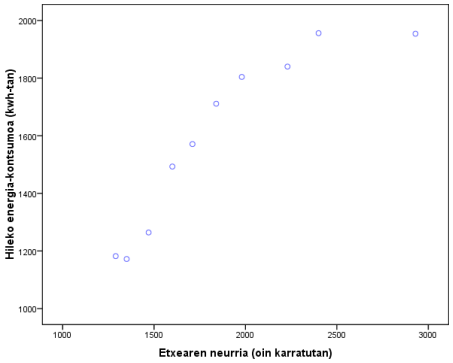
- Lehenik, populazio-eredua proposatzeko grafikoak eta teknikak deskribatzen dira.
- Bigarrenik, parametroak zenbatesteko bidea adierazten da.
- Hirugarrenik, erregresio-ereduaren erabilgarritasuna zehazten da, bai adierazgarritasun orokorra planteatuz, bai koaldagai bakoitzaren garrantzia aztertuz.
- Aldagaien arteko korrelazio lineal bakuna, anizkoitza eta partziala ikertzen dira.
- Hipotesi eta hondarren azterketa komentatzen da.
- Iragarpenak egiteko prozedura azaltzen da.
- Azkenik, bi eranskin daude; lehenengoan, eredua eraikitzeko urratsak agertzen dira.
- Bigarren eranskinean *EXCEL* programarekin egin daitekeena aipatzen da.

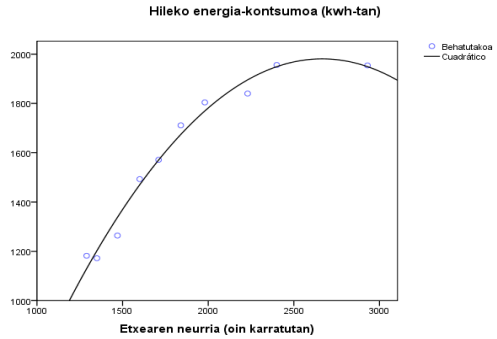
Bereziki, ariketa-zerrendatik 7.5 ariketa garatuko dugu.

7.1. Populazio-eredua proposatzea: $\mathbf{Y = f(X_1, X_2, \ldots, X_k) + \epsilon}$

• Baldin $k = 1$ bada, hodei-puntua egin dezakegu (bestela, k grafikoak, X_j vs Y):

`Gráficos` → `CDA` → `Dispersión/Puntos` → `Dispersión Simple` → `Eje Y, Eje X`





• Erregresio-mota batzuk proposatzen ditugu:

`Analizar` → `Regresión` → `Estimación curvilínea` → `Dependiente[Y]. Variable[X]. Modelo`

Lineal	**Lineala**	$\mathbf{\hat{Y} = a + bX}$		
Logarítmico	Logaritmikoa	$\hat{Y} = a + b\ln(X)$		
Inverso	**Hiperbolikoa**	$\mathbf{\hat{Y} = a + \frac{b}{X}}$		
Cuadrático	**Koadratikoa**	$\mathbf{\hat{Y} = a + b_1X + b_2X^2}$		
Cúbico	Kubikoa	$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3$		
Potencia	**Potentziala**	$\mathbf{\hat{Y} = aX^b}$	$\Leftrightarrow$	$\ln(\mathbf{\hat{Y}}) = \ln(\mathbf{a}) + \mathbf{b}\ln(\mathbf{X})$
Compuesto	Konposatua	$\hat{Y} = ab^X$	$\Leftrightarrow$	$\ln(\hat{Y}) = \ln(a) + \ln(b)X$
S	S	$\hat{Y} = e^{a+\frac{b}{X}}$	$\Leftrightarrow$	$\ln(\hat{Y}) = a + \frac{b}{X}$
Logístico	Logistikoa			
Crecimiento	Hazkundea	$\hat{Y} = e^{a+bX}$	$\Leftrightarrow$	$\ln(\hat{Y}) = a + bX$
Exponencial	**Esponentziala**	$\mathbf{\hat{Y} = ae^{bX}}$	$\Leftrightarrow$	$\ln(\mathbf{\hat{Y}}) = \ln(\mathbf{a}) + \mathbf{bX}$

Bi aldiz sakatzen badugu grafikoa, hau egin dezakegu:

`Elementos` → `Linea de ajuste total` → `Cuadrático` → `Intervalos de confianza`

r^2 (R cuadrado) balioaren arabera eredua hautatzen dugu.

• Egotekotan, aukeratutako ereduaren aldagai berriak osatzen ditugu. Adibide honetan $Y = A + BX + CX^2 + \epsilon$ eredu koadratikoa proposatzen dugunez, X^2 aldagai berria osatu beharko dugu: `Transformar` → `Calcular`

`Analizar` → `Regresión` → `Lineal` → `Dependiente[`Y`]. Independientes[`X, X^2`]` → `Estadísticos[Estimaciones. Intervalos de confianza. Ajuste del modelo. Correlación parcial. Residuos: Durbin-Watson. Diágnosticos por caso: todos los`

`casos]. Gráficos[Y:*ZRESID X:*ZPRED. Histograma. Gráfico de probabilidad normal]. Guardar[Residuos tipificados]`

Aldagai kualitatibo baten arabera aztertzeko: `Variable de seleccion` aukeran aldagaia sartu ondoren, `regla`-ren bidez ikertzeko maila(k) zehaztu.

7.2. Minimo karratuen bidezko metodoa erabiliz parametroak zenbatestea: $\hat{\mathbf{Y}} = \mathbf{a} + \mathbf{b_1X_1} + \mathbf{b_2X_2} + \ldots + \mathbf{b_kX_k}$

Variable dependiente: Hileko energia-kontsumoa (kwh-tan)

	Resumen del modelo					Estimaciones de los parámetros		
Ecuación	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2
Cuadrático	.982	189.710	2	7	.000	-1216.144	2.399	.000

La variable independiente esEtxearen neurria (oin karratutan).

$$\hat{Y} = a + b_1X_1 + b_2X_2 = -1216{,}143887 + 2{,}398930X - 0{,}000450X^2$$

Hiru dezimal baino gehiago ikusteko: taula bi aldiz sakatu eta zenbakia bi aldiz sakatu.

7.3. Ereduaren erabilgarritasuna zehaztea

7.3.1. Doikuntza-egokitasuna

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	.991[a]	.982	.977	46.801	2.079

a. Variables predictoras: (Constante), Etxearen neurriaren karratua, Etxearen neurria (oin karratutan)

b. Variable dependiente: Hileko energia-kontsumoa (kwh-tan)

Korrelazio-koefizientea $r = 0{,}991$, determinazio-koefizientea $r^2 = 0{,}982$, zuzendutako determinazio-koefizientea $\overline{r}^2 = 0{,}977$, zenbatespen-errore estandarra $s = 46{,}801$.

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	831069,546	2	415534,773	189,710	,000[a]
	Residual	15332,554	7	2190,365		
	Total	846402,100	9			

a. Variables predictoras: (Constante), Etxearen neurriaren karratua, Etxearen neurria (oin karratutan)

b. Variable dependiente: Hileko energia-kontsumoa (kwh-tan)

Doikuntza-egokitasun globala aztertzeko estatistikoa $F = 189{,}710$ da eta dagokion $p - balioa = 0{,}000 < 0{,}05$ denez, doikuntza koadratikoa egokia da %5eko esangura-mailarekin.

7.3.2. B_j aldaparekiko inferentzia

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%		Correlaciones		
		B	Error típ.	Beta			Límite inferior	Límite superior	Orden cero	Parcial	Semiparcial
1	(Constante)	-1216.144	242.806		-5.009	.002	-1790.290	-641.998			
	Etxearen neurria (oin karratutan)	2.399	.246	4.049	9.758	.000	1.818	2.980	.912	.965	.496
	Etxearen neurriaren karratua	.000	.000	-3.161	-7.618	.000	-.001	.000	.858	-.945	-.388

a Variable dependiente: Hileko energia-kontsumoa (kwh-tan)

Taulari begiratuz hiru koefizienteak adierazgarriak dira, eta bi koefiziente estandarizatuei begiratuz, $Y =' energia - kontsumoa'$ azaltzeko $X =' etxearen\ neurria'$ aldagaia garrantzitsuena da.

`Matriz de covarianzas` erregresio-koefizienteen bariantza eta kobariantza-matrizea da. Izan ere, $Var(b_1) = S^2_{b_1} = 3{,}4910^{-9}$, $Var(b_2) = S^2_{b_2} = 0{,}060$ eta $Cov(b_1, b_2) = S_{b_1 b_2} = -1{,}4410^{-5}$. Gogora ezazu, $\{H_0 : B_j = 0\}$ hipotesi-kontrasteetan, estatistikoa $t_p = b_j / S_{b_j}$ eta konfiantza-tarteetan $I^{1-\alpha}_{B_j} = (b_j \mp t_{\alpha/2;n-k-1} S_{b_j})$. Kontuan hartu S_{b_j} balioak ere `Error típ.` izeneko 2. zutabean emanda datozela.

7.4. Korrelazioa

Aurreko grafikoan $r_{y1} = 0{,}912$ eta $r_{y2} = 0{,}858$ korrelazioak zein $r_{y1\cdot 2} = 0{,}965$ eta $r_{y2\cdot 1} = -0{,}945$ korrelazio partzialak agertzen dira. Y eta X_1 aldagaien arteko erlazio lineala handia eta gorakorra da; bestalde, Y eta X_2 aldagaien arteko erlazioa handia eta negatiboa da, r_{y2} gezurrezko koerlazioa izanik. Bereziki, korrelazio bakun guztiak ($r_{y1} = 0{,}912$, $r_{y2} = 0{,}858$ eta $r_{12} = 0{,}992$) aurkitzeko eta adierazgarritasuna aztertzeko:

`Analizar` → `Correlaciones` → `Bivariadas`

eta korrelazio partzialak lortzeko, esate baterako $r_{12\cdot y} = 0{,}997$

`Analizar` → `Correlaciones` → `Parciales` → [X,X^2] `controlando` [Y]

$X_1 = X$ eta $X_2 = X^2$ aldagaien arteko erlazioa handiegia da, aipatzeko arazoa dugu, aldez aurretik espero genuena X_2-ren definizioagatik.

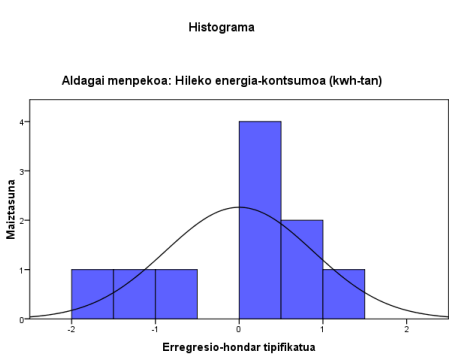
7.5. Hondarren azterketa

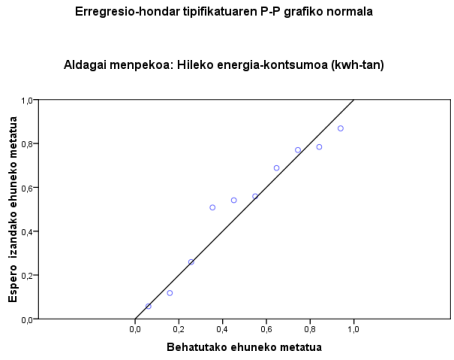
• Grafikoki, $X[X_1]$ *vs* $e^{tip}[ZRE]$ eta $X^2[X_2]$ *vs* $e^{tip}[ZRE]$ puntu-hodeiak egitean (`Graficos` → `CDA` → `Dispersión`) zorizkoak direnez, **zehaztatze-akatsik ez** dagoela ondorioztatzen dugu.

• Aurreko grafikoak uhin-itxurakoak ez direnez eta Durbin-Watson koefizientea (2.079) 2tik hurbil dagoenez hondarren artean **autokorrelaziorik ez** dagoela ondorioztatzen dugu (0tik edo 4tik hurbil egonez gero autokorrelazio negatiboa edo positiboa dagoela esaten da, hurrenez hurren)

• $\hat{Y}^{tip}[ZPRED]$ *vs* $e^{tip}[ZRESID]$ grafikoa $e^{tip} = 0$ ardatzean zentratutako banda horizontal batean dagoenez (ez da elipse edo triangelu itxurakoa, besteak beste), hondarren **homozedastizitatea** onartzen dugu.

• e hondar-aldagaiaren histogramaren bidez, tartetik kanpoko datuak eta normaltasunik eza aurki ditzakegu. Alde batetik, e_i balio guztiak $(-3,\ 3)$ tartean badaude, ez dago tartetik kanpoko daturik. Beste aldetik, ondoan dagoen grafikoaren puntuak 1. koadrantearen erdikaritik hurbil badaude, normaltasuna onartzen da. Espero duguna $e \approx \mathcal{N}(0, S)$, hots, $e^{tip} = e/S \approx \mathcal{N}(0, 1)$ izatea da.





Kolmogorov-Smirnoven kontrastea e^{tip} aldagaian aplikatuz, $p-balioa = 0{,}775$ denez, ezin da hondarren normaltasuna errefusatu %95eko konfiantza-mailarekin. Beraz, **normaltasuna** eta **tartetik kanpoko daturik ez** dagoela onartzen dugu.

• **Multikolinealtasuna** dagoelako susmoa dugu, X eta X^2 aldagaien arteko korrelazioa adierazgarria delako (izan ere, $p - balioa = 0{,}000$ baita). Sakonago aztertzeko, `Diagnosticos de colinealidad` direlakoak aztertu beharko genituzke.

7.6. Doikuntza egokia bada, Y-ren iragarpenak egiteko erabiltzea

Zein da 5. behaketaren iragarpena ereduaren arabera? 1570.06 hondar txikieneko behaketa da.

Zein da 1500 oin karratu dituen etxe baten itxarondako banakako energia-kontsumoa?

`Editor de datos` pantailan $x = 1500$ datu berria gehitzen diegu aldagaia independenteei.

`Analizar` → `Regresión` → `Lineal` → `Guardar [Valores Pronosticados:No tipificados. Intervalos de pronostico:Individuos]`

Editor de datos PRE aldagaian $\hat{y} = 1369{,}6609$ kwh iragarpena agertzen da, $LICI$ eta $UICI$ aldagaietan agertzen dira dagokion konfiantza-mailaren behe-muturrak (1250.3172) eta goi-muturrak (1489.0046), hau da, %95eko konfiantza-tartea $I^{0,95}_{y_0|x_0} = (LICI,\ UICI) = (1250{,}3172,\ 1489{,}0046)$ da.

Batezbestekoaren konfiantza-tartea lortzeko: `Guardar [ Intervalos de pronostico: Media]`; horrela, $I^{0,95}_{\mu_y|x_0} = (LMCI,\ UMCI)$

Autoebaluaziorako ariketak

- 7.1 ariketaren datuak erabiliz, erantzun galdera hauei:

1. Plantea itzazu eredu lineala, hiperbolikoa, koadratikoa, potentziala eta esponentziala. %5eko esangura mailarekin zeintzuk dira erabilgarriak diren ereduak?
2. Ordena itzazu erabilgarriak direnak doikuntza-egokitasunaren arabera, 1.a egokiena izanik.
3. Idatz itzazu erregresio-koefizienteak eredu linealean eta esponentzialean.
4. Eredu linealean, zenbat da bariantza-azaldua eta hondar-bariantza?
5. Kontsidera ezazu $Y = Ae^{BX} + \epsilon$ eredu esponentziala. Azter ezazu doikuntza-egokitasuna.
6. Zein da B-ren %95eko konfiantza-tartea? Zer ondoriozta daiteke?
7. Zein da A-rena? Ondoriozta daiteke $A = 1$ dela hipotesi-kontraste baten bidez?
8. Ereduaren arabera, zein behaketari dagokio ondoen dagoen iragarritako balioa? Zenbat da?
9. Ereduaren arabera, zein da 4. behaketaren iragarritako balioaren %95eko konfiantza-tartea?

1. Lineala, potentziala eta esponentziala.

2. Esponentziala-lineala-potentziala
3. $\hat{Y} = -0{,}1 + 0{,}7X$ eta $\hat{Y} = 0{,}616e^{0{,}347X}$
4. $S_e^2 = 1{,}1$ eta $S_{\hat{y}}^2 = 4{,}9$
5. Ereduaren erabilgarritasuna aztertzeko estatistikoa: $F = 25$ eta beraren p-balioa $p = 0{,}015 < 0{,}05$ denez, eredua erabilgarria da %5eko esangura mailarekin. Gainera eredua oso egokia da korrelazio-koefizienteak (balio absolutuan) 1etik hurbil daudelako, $r = 0{,}945, r^2 = 0{,}893, \overline{r} = 0{,}857$. Bestalde, zenbatespen-errore estandarra $S = 0{,}219$
6. $I_B^{0,95} = (0{,}126, 0{,}567)$, 0 barne ez dagoenez presioa (X) adierazgarria da konpresioa (Y) azaltzeko. Unitate bat gehitzean presioan, konpresioa 0.126 eta 0.567 unitate artean aldatzea espero dugu.
7. $a' = \ln a = -0{,}485 \Rightarrow a = e^{a'} = e^{-0{,}485} = 0{,}616$ eta $I_{A'}^{0,95} = (-1{,}217, 0{,}246) \Rightarrow I_A^{0,95} = (e^{-1{,}217}, e^{0{,}246}) = (0{,}296, 1{,}279)$. Bai, $p = 0{,}125 > 0{,}05 \Rightarrow a' = 0 \Leftrightarrow a = e^0 = 1$
8. Lehenengoa, $\hat{y'}_1 = -0{,}1386 \Rightarrow \hat{y}_1 = 0{,}8706$
9. $(0{,}1057,\ 1{,}6964) \Rightarrow (1{,}1115,\ 5{,}4545)$

• Egin itzazu 7.8 eta 7.9 ariketak (Oharra 9. d) ANOVA taulari begiratuz, $S_{\hat{y}}^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{n} = \frac{484206{,}3492}{5} = 96841{,}2698$, $S_e^2 = \frac{\sum e_i^2}{n} = \frac{7793{,}6508}{5} = 1558{,}7302$ eta $S_y^2 = \frac{\sum(y_i - \bar{y})^2}{n} = \frac{492000}{5} = 98400$)

• [Aukerakoa] Egin itzazu 7.4 eta 7.5 ariketak aldagai kualitatibo bat erabiliz (adibidez, lurraldea).

A eranskina: Eredu anizkoitzaren eraikuntza urratsez urrats

Ireki ezazu `Cars.sav (Coches.sav)` fitxategia

Demagun kontsumoaren erregresio-eredua egin nahi dugula, hau da, $Y =$**Kontsumoa**. Aldagai aske kuantitatibo guztiak hauek dira: $X_1 = Cilindrada$, $X_2 = Potencia$, $X_3 = Peso$, $X_4 = Aceleracion$. Zeintzuk sartuko ditugu erregresio-ereduan?

`Analizar` → `Regresión` → `Lineal` → `Dependiente[Consumo]. Independientes[Cilindrada,Potencia, Peso, Aceleración]` → `Método[Pasos sucesivos]. Estadísticos[Estimaciones. Intervalos de confianza. Matriz de covarianzas. Correlación parcial. Residuos: Durbin-Watson. Diágnosticos por caso: todos los casos]. Gráficos[Y:*ZRESID X:*ZPRED. Histograma. Gráfico de probabilidad normal]. Guardar[Residuos tipificados]`

Pearson-en korrelazioak ikustean kontsumoarekiko aldagai aske guztien menpekotasun lineala adierazgarria da. Korrelazio altuena duena, **Pisua**, sartzen da ereduan lehenengo urratsean. Izan ere, $r_{Kon\ Pis} = r_{y3} = 0{,}837$ handiena da.

Bigarren urratsean, `Variables excluidas` taulari begiratuz, datu hauek agertzen dira. `Beta dentro:` koefiziente estandarizatuak (aldagai askeak ereduan sartzekotan), estatistikoa eta p-balioa; `correlación parcial:` aldagai askea eta kontsumoaren arteko korrelazio partziala (sartuta dagoen aldagaiaren eragina eliminatuz) eta `tolerancia:` aldagaiaren bariantzaren proportzioa (sartutako aldagaiaren bariantzagatik ez azaldua), zenbat eta 0tik (1etik) hurbilago, orduan eta sartzeko beharra gutxiago (handiago). p-balioa txikiena (eta 0.05 baino txikiagoa) duen aldagaia ereduan sartuko den hurrengoa izango da. Bat baino gehiago denean, korrelazio partziala altuena duena, kasu honetan, **Potentzia**. Izan ere, $r_{Kont\ Pot\cdot Pis} = r_{y2\cdot 3} = 0{,}422$ handiena da. Bere tolerancia$= 1 - r^2_{Pot\ Pis} = 1 - r^2_{23} = 1 - 0{,}857^2 = 0{,}265$ da.

Hirugarren urratsean gelditu da metodoa, sartuta ez dauden beste aldagai guztien p-balioak 0.05 baino handiagoak direlako. Beraz, ez dira sartuko ereduan.

Azkenik, eredua ondokoa da:

$$Kontsumoa = 0{,}196 + 0{,}006\ Pisua + 0{,}046\ Potentzia$$

Kontsumo-aldagaia Pisua eta Potentziaren bidez azal daiteke %75ean ($r = 0{,}868$, $r^2 = 0{,}753$ eta $\overline{r}^2 = 0{,}752$). Eta eredua erabilgarritzat jotzen dugu, $F = 593{,}635$, p-balioa$= 0{,}000$ delarik.

B. eranskina: Zer egin daiteke *EXCEL* programarekin?

• $Y = f(X)$ eredua aztertzeko: a) X vs Y hodei-puntua egiteko {`Asistente para gráficos` → `Dispersión` → `X vs Y`}; b) Erregresio-koefizienteak eta r^2 lortzeko (esate baterako, $Y = a + bX + cX^2$) {`Grafico` → `Agregar linea de tendencia` → `Tipo de tendencia o regresión [Polinomial orden 2]. Opciones[Presentar ecuacion y` r^2`]`}

• $Y = f(X_1, X_2, \ldots, X_k)$ eredua sakontasun handiagoz aztertzeko: {`Herramientas` → `Análisis de datos` → `Regresión`}

`Microsoft Office Excel 2007` erabiltzen baduzu: `Office` → `Opciones de Excel`→ `Complementos` → `Complementos de Excel. Ir.` → `Herramientas para análisis. Herramientas para análisis - VBA. Aceptar`. Behin instalazioa eginda, `Datos` → `Análisis de Datos` → `Regresión`

• Korrelazio bakuna: {`Herramientas` → `Análisis de datos` → `Coeficientes de correlación de Pearson`}

8. SPSS PRAKTIKA

Bariantza-analisia

Helburua

Praktika honen jomuga ANOVA edo bariantza-analisia aztertzea da. Faktore bakarra zein faktore biko analisia kontsideratuko da populazio-batezbestekoak konparatzeko asmoz. Diferentziak adierazgarriak diren kasuetan, batezbestekoen binakako berdintasuna adierazten da.

8.1. Faktore bakarreko bariantza-analisia

• Egin dezagun 8.6 ariketa. Hipotesi nulua $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ da, altuera espezieekiko independentea izatea da. Bi aldagai osatu behar ditugu, kualitatibo bat (espeziea: 1=pinea, 2=pinaster, 3=silvestris, 4=laricio, 5=halopensis) eta kuantitatibo bat (altuera).

`Analizar` → `Comparar medias` → `ANOVA de un factor` →

- **Post-hoc:** `Asumiendo varianzas iguales: Scheffé, Tukey. No asumiendo varianzas iguales: T2 de Tamhane.`
- **Opciones:** `Estadísticos.Descriptivos. Efectos aleatorios y fijos. Prueba de homogeneidad de la varianza. Brown-Forsythe. Welch. Grafico de medias.`

`Prueba de homogeneidad de la varianza` izeneko taulan ikusten dugunez, $p = 0{,}129 \Rightarrow H_0$ onartzen dugu, hau da, bariantzen berdintasunaren hipotesia betetzen da.

`ANOVA` taulari begiratuz, estatistikoa $F = 6{,}332$ da eta beraren p-balioa $p = 0{,}001$. Beraz, faktoreak eragina duela onartzen badugu, ezin da ondorioztatu batezbesteko guztiak berdinak direnik.

Batezbestekoen binakako berdintasuna aztertzean, lagin-tamainak diferenteak direnez, **Scheffe**-ren arabera %5eko esangura-mailarekin bi batezbestekoen bi diferentzia adierazgarriak ikusten dira: $\mu_{pinaster} > \mu_{pinea}$ eta $\mu_{silvestris} > \mu_{pinea}$. Bestalde, **hiru talde homogeneo** osatzen dira: lehenengoan halopensis dago eta hirugarrenean silvestris espeziea; pinea eta laricio 1. edo 2.ean egon daitezke eta laricio eta pinaster 2.ean edo 3.ean. Proba biak (binakako konparaketa anizkoitzak eta talde homogenoak osatzea) ez dira metodo baliokideak. Lagin-tamaina antzekoak badira, **Tukey**-ren metodoa erabiliko genuke.

(I) espezie	(J) espezie	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
pinea	pinaster	-1,31267	,35715	,024	-2,4991	-,1263
	silvestris	-1,43000*	,39991	,030	-2,7564	-,1036
	laricio	-,49467	,41880	,842	-1,8859	,8965
	halopensis	,16667	,48905	,998	-1,4579	1,7912
pinaster	pinea	1,31267	,35715	,024	,1263	2,4991
	silvestris	-,11733	,35715	,998	-1,3037	1,0691
	laricio	,81800	,37882	,350	-,4404	2,0764
	halopensis	1,47993	,45528	,058	-,0330	2,9917
silvestris	pinea	1,43000	,39991	,030	,1036	2,7564
	pinaster	,11733	,35715	,998	-1,0691	1,3037
	laricio	,93533	,41880	,317	-,4559	2,3265
	halopensis	1,59667	,48905	,056	-,0279	3,2212
laricio	pinea	,49467	,41880	,842	-,8965	1,8859
	pinaster	-,81800	,37882	,350	-2,0764	,4404
	silvestris	-,93533	,41880	,317	-2,3265	,4559
	halopensis	,66133	,50509	,787	-1,0165	2,3392
halopensis	pinea	-,16667	,48905	,998	-1,7912	1,4579
	pinaster	-1,47993	,45528	,058	-2,9917	,0330
	silvestris	-1,59667	,48905	,056	-3,2212	,0279
	laricio	-,66133	,50509	,787	-2,3392	1,0165

*. La diferencia de medias es significativa al nivel 0.05.

espezie	N	Subconjunto para alfa = 0.05		
		1	2	3
halopensis	3	6,9267		
pinea	6	7,0933	7,0933	
laricio	5	7,5880	7,5880	7,5880
pinaster	10		8,4060	8,4060
silvestris	6			8,5233
Sig.		,672	,084	,343

Se muestran las medias para los grupos en los subconjuntos homogéneos.

a. Usa el tamaño muestral de la media armónica = 5,172.

b. Los tamaños de los grupos no son iguales. Se utilizará la media armónica de los tamaños de los grupos. Los niveles de error de tipo I no están garantizados.

• Egin dezagun 8.1 ariketa. Hipotesi nulua $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ da, heriotza-tasa urtesasoiarekiko independentea izatea da. Bi aldagaiak osatu ondoren (urtesasoia: 1=udaberria, 2=uda, 3=udazkena, 4=negua) ondokoa egiaztatzen dugu.

(I) urtesasoia	(J) urtesasoia	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
udaberria	uda	.56667*	.09718	.001	.2453	.8880
	udazkena	-.43333	.15741	.166	-1.0137	.1470
	negua	-.96667*	.20467	.020	-1.7573	-.1761
uda	udaberria	-.56667*	.09718	.001	-.8880	-.2453
	udazkena	-1.00000*	.16432	.002	-1.5794	-.4206
	negua	-1.53333*	.21003	.001	-2.3153	-.7513
udazkena	udaberria	.43333	.15741	.166	-.1470	1.0137
	uda	1.00000*	.16432	.002	.4206	1.5794
	negua	-.53333	.24381	.291	-1.3449	.2782
negua	udaberria	.96667*	.20467	.020	.1761	1.7573
	uda	1.53333*	.21003	.001	.7513	2.3153
	udazkena	.53333	.24381	.291	-.2782	1.3449

*. La diferencia de medias es significativa al nivel .05.

Bariantzak desberdinak direnez, ANOVA ikerketa ez da egokiena. Kasu honetan, **Brown-Forsythe**-ren edo **Welch**-en estatistikoei begiratuz ondorioak aterako ditugu: faktoreak berriro eragina du. Eta binakako azterketa egiteko `Comparaciones múltiples post-hoc` atalean aukera batzuk daude. Esate baterako, **Tamhane**-ren arabera, %5eko esangura-mailarekin ondoko batezbestekoen diferentziak adierazgarriak dira: $\mu_1 \neq \mu_2$, $\mu_1 \neq \mu_4$, $\mu_2 \neq \mu_3$ eta $\mu_2 \neq \mu_4$. Heriotza-tasa baxuena udan gertatzen da; bigarrenik, udaberrian edo udazkenean; eta heriotza-tasa altuena udazkenean edo neguan gertatzen da.

8.2. Faktore biko bariantza-analisia (n > 1)

• Egin dezagun 8.12 ariketa. Hipotesi nulua $H_0 : \mu_{11} = \mu_{12} = \mu_{13} = \mu_{21} = \mu_{22} = \mu_{23}$ da. Hiru aldagai osatu behar ditugu, bi kualitatibo: sexua (1=emakumezkoa, 2=gizonezkoa) eta tratamendu-mota (1=marihuana, 2=alkohol, 3=droga gabekoa) eta kuantitatibo bat, puntuazioa.

`Analizar` → `Modelo Lineal General` → `Univariante` → `Dependiente[Puntuazioa].`
`Factores fijos [Sexua. Tratamendu-mota]` →

- **Gráficos**: `Eje horizontal[A]. Líneas distintas [B]. Añadir`

  `Eje horizontal[B]. Líneas distintas [A]. Añadir`
- **Post-hoc:** `Scheffé. Tukey. Tamhane`
- **Opciones**: `Medias marginales: GLOBAL, A, B, A*B. Comparar los efectos principales. Descriptivos. Prueba de homogeneidad de varianza`

$p = 0{,}145$ denez, bariantzen berdintasuna onartzen dugu.

Tratamendu-mota faktorearekiko estatistikoa $F = 11{,}170$ eta p-balioa $p = 0{,}002$ direnez, faktore honek eragina du puntuazioetan. Hala ere, sexuak ez du eraginik emaitzetan ($p = 0{,}882$), ezta interakzioak ere ($p = 0{,}324$).

Eragina duen faktorea aztertzen dugu. Tukey-ren arabera, tratamendu-mota faktorean %5eko esangura-mailarekin ondoko batezbestekoen diferentziak adierazgarriak dira: $\mu_1 \neq \mu_2$ eta $\mu_2 \neq \mu_3$. Bi multzo homogenoak ditugu: alde batetik, alkohola eta, beste aldetik, marihuana eta droga gabekoa. Puntuazio txarrena alkoholaren eraginpean gertatzen da.

8.3. Faktore biko bariantza-analisia (n = 1)

• Egin dezagun 8.9 ariketa. Hipotesi nulua $H_0 : \mu_{11} = \ldots = \mu_{44}$ da. Hiru aldagai osatu behar ditugu, bi kualitatibo (langilea eta makina) eta kuantitatibo bat (unitate-kopurua).

`Analizar` → `Modelo Lineal General` → `Univariante` → `Dependiente[Unitate-kopurua]. Factores fijos [Langilea. Makina.]` →

- **Modelo**: `Personalizado. Efectos principales.` Askatu: `Incluir la intersección en el modelo`
- **Gráficos**: `Eje horizontal[A]. Líneas distintas [B]. Añadir`

  `Eje horizontal[B]. Líneas distintas [A]. Añadir`
- **Post-hoc:** `Scheffé. Tukey. Tamhane`
- **Opciones**: `Medias marginales: GLOBAL, A, B. Comparar los efectos principales. Descriptivos. Prueba de homogeneidad de varianza.`

Langileen faktorearekiko estatistikoa $F = 0{,}690$ eta p-balioa $p = 0{,}581$ denez, faktore horrek ez du eraginik errendimenduan. Hala ere, makinaren arabera errendimendua aldatzen da ($F = 11{,}859$ eta $p = 0{,}002$).

Eragina duen faktorea aztertzen dugu. Tukey-ren arabera, makina-mota faktorean %5eko esanguramailarekin 1. makinaren etekina beste gainontzekoekin konparatuz desberdina dela ondoriozta daiteke. Bi multzo homogenoak ditugu: alde batetik 1. makina, etekin handiena ematen duena, eta bestalde, 2., 3 . eta 4. makinena.

Autoebaluaziorako ariketak

• Egin ezazu 8.11 ariketa.

a) Bariantzen berdintasuna onartzen da?

b) Aipa itzazu barren bolumena azaltzeko adierazgarriak diren faktoreak.

c) Lekuari dagokionez, zer ondoriozta daiteke?

d) Gari-motari dagokionez, zer ondoriozta daiteke?

Sol.: a) Bai, $p = 0{,}578$ delako. b) Lekuaren, gari-motaren eta interakzioaren eraginekiko estatistikoak $F' = 164{,}367$, $F'' = 78{,}247$ eta $F''' = 8{,}369$ dira, non p-balioak $p = 0{,}000$, $p = 0{,}000$ eta $p = 0{,}001$ diren. Beraz, hirurek eragina dute barren bolumena azaltzeko. c) Binakako batezbestekoak konparatzean, hiru mailak desberdinak direla ikusten da. Barren bolumen txikiena 2. lekuan eta handiena 3.ean gertatzen dira. d) Hiru mailak ere desberdinak dira, 2. gari-motarekin barren bolumen txikiena lortzen da eta handiena 3.arekin.

• Egin ezazu 8.10 ariketa.

a) Aipa itzazu etekina azaltzeko adierazgarriak diren faktoreak.

b) Lur-zatiari dagokionez, zer ondoriozta daiteke?

c) Ongarri-motari dagokionez, zer ondoriozta daiteke?

Sol.: a) Lur-zatiak ez du eraginik ($p = 0{,}640$), baina ongarri-motak eragina du ($p = 0{,}000$) etekina azaltzeko. b) Binakako batezbestekoak konparatzean, bost mailak berdinak direla ikusten da. Izan ere, etekina azaltzeko lur-zatiak ez du eraginik. c) Ongarri-mota faktorean %5eko esangura-mailarekin 2. ongarriaren etekina beste gainontzekoekin konparatuz desberdina dela ondoriozta daiteke. Bi multzo homogenoak ditugu: alde batetik 2. ongarria, etekin handiena ematen duena eta bestalde, 1., 3 . eta 4. ongarriena.

A. eranskina: Zer egin daiteke *EXCEL* programarekin?

• Faktore bakarreko bariantza-analisia

Idatz ezazu aldagaia k (tratamendu-kopurua) zutabeetan (edo k errenkadetan).

`Herramientas` → `Análisis de datos` → `Análisis de varianza de un factor` → `Agrupado por columnas (filas)`

• Faktore biko bariantza-analisia ($n = 1$)

Idatz itzazu datuak ariketa-zerrendan agertzen diren moduan.

`Herramientas` → `Análisis de datos` → `Análisis de varianza de dos factores con una sola muestra por grupo`

• Faktore biko bariantza-analisia ($n > 1$)

Idatz itzazu datuak ariketa-zerrendan agertzen diren moduan.

`Herramientas` → `Análisis de datos` → `Análisis de varianza de dos factores con varias muestras por grupo` → `Filas por muestra[`n `(lagin bakoitzaren tamaina)]`

9. SPSS PRAKTIKA

Kalitatearen kontrol estatistikoa

Helburua

Praktika honen xede nagusia zea da: kalitatearen kontrol estatistikoan azaltzen diren oinarrizko kontrol-grafikoak eraikitzea. Alde batetik, aldagaien grafikoak eta atributuen grafikoak agertzen dira; bestalde, Paretoren diagrama dugu.

9.1. Aldagaien grafikoak

9.1.1. Batez besteko balioak eta heinak ($\overline{x}$ eta R kontrol-grafikoak)

Egin dezagun 9.1. ariketa. Datuak sartzeko eraren arabera bi bide daude kontrol-grafikoa egiteko. 1. bidea ondokoa da: zutabe batean `pisua` aldagaia osatzen dugu ($k \cdot n = 20 \cdot 4 = 80$ datuekin) eta beste zutabe batean `lagina` izenekoa, behaketa bakoitzari dagokion lagina adieraziz (hots, $\{1, 1, 1, 1, 2, 2, 2, 2, ..., 20, 20, 20, 20\}$). Orduan,

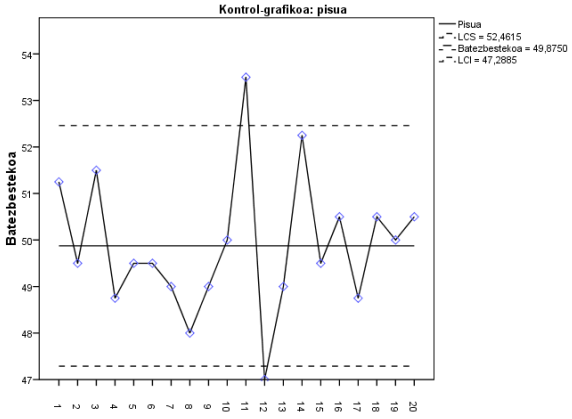
`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de variables. X-barra, R, s. Organización de los datos: los casos son unidades.`→ `Medida del proceso [pisua]. Subgrupos definidos por [lagina]. Gráficos: X-barra con rango. Mostrar gráfico R.`

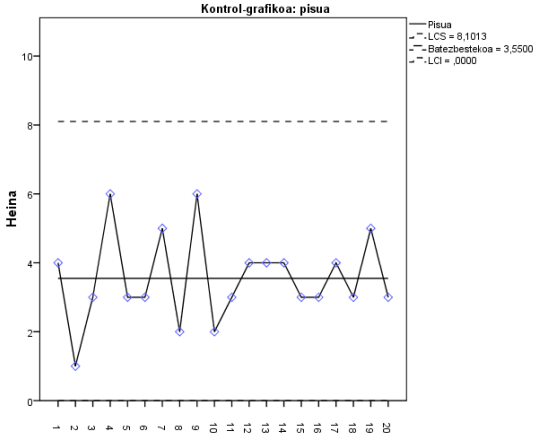
2. bidea erabiltzekotan, $n = 4$ tamainako $k = 20$ zutabe osatzen ditugu (`lagin1,lagin2,...,lagin20` izendatuz) azpilagin bakoitzaren datuekin, eta beste zutabe batean `behaketa` izenekoa: behaketa bakoitzaren identifikazioa idazten dugu, adibidez, $\{1, 2, 3, 4\}$. Orduan,

`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de variables. X-barra, R, s. Organización de los datos: los casos son subgrupos.`→ `Muestras`

`[lagin1,...,lagin20]. Subgrupos etiquetados mediante [behaketa]. Gráficos: X-barra con rango. Mostrar gráfico R.`

Edozein bideri jarraituz, kontrol-grafiko bi hauek lortuko ditugu:





11. eta 12. behaketak kontrol-mugetatik kanpo daudenez, prozesua kontroletik kanpo dago. Prozesuan arazoak egon daitezke, eta batezbesteko arraroen kausa aztertu beharko dira.

9.1.2. Batez besteko balioak eta desbideratze estandarrak ($\overline{x}$ eta s kontrol-grafikoak)

Kasu honetan, analogoki definitu ahal ditugu datuak sartzeko bi bideak eta horren arabera `Organización de los casos` zehaztu ahal dugu. Orduan,

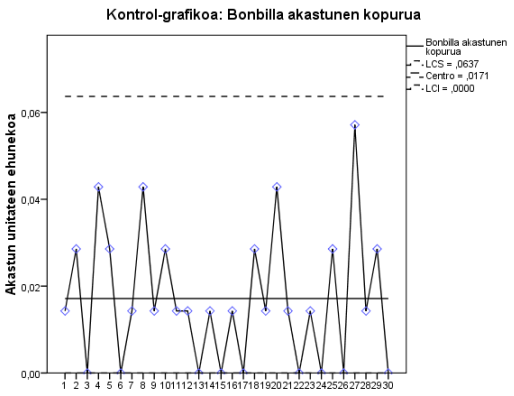
`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de variables. X-barra, R, s. Organización de los datos.` → `Medida del proceso [pisua]. Subgrupos definidos por [lagina]. Gráficos: X-barra con desviación típica. Mostrar gráfico s.`

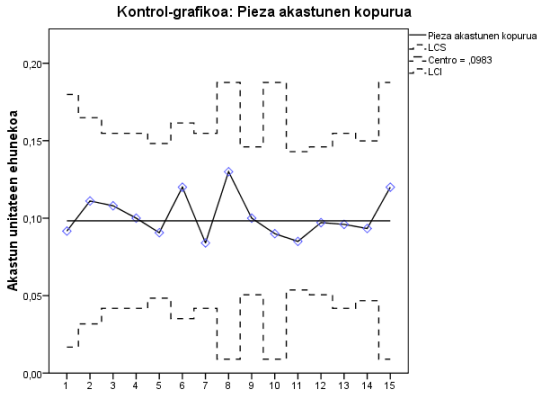
9.2. Atributuen grafikoak

9.2.1. Akastun unitateen ehunekoa (p kontrol-grafikoa)

Egin dezagun 9.2. ariketa. Osatuko dugu $k = 30$ tamainako `akatsak` izeneko aldagaia, azpilagin bakoitzaren akastun bonbillen kopurua adieraziz.

`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de atributos: p,np. Organización de los datos: los casos son subgrupos.` → `Número de disconformes: [akatsak]. Tamaño de la muestra. Constante: 70. Diseño: p (proporción disconforme)`



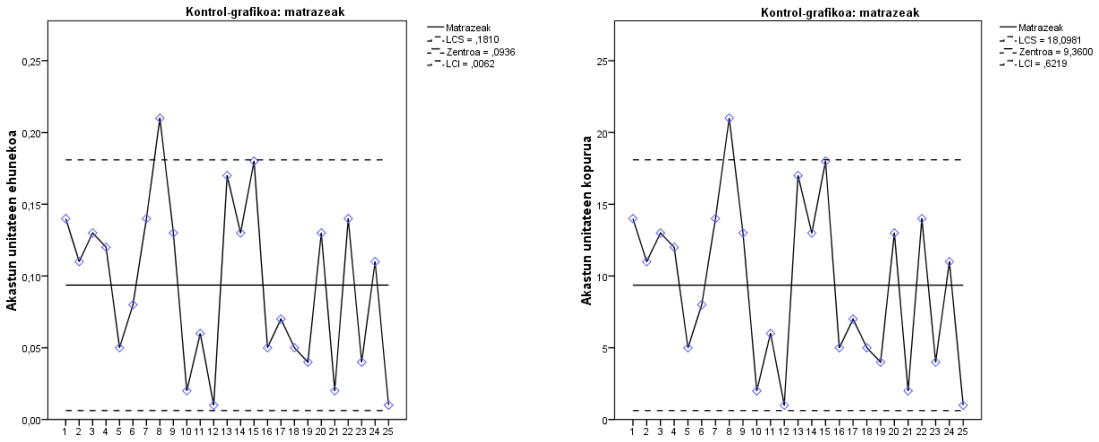


Egin dezagun 9.3. ariketa. Osatuko ditugu bi aldagai: lehenengoa $k = 15$ tamainako `akatsak` izeneko aldagaia, azpilagin bakoitzaren akastun piezen kopurua adieraziz, eta bigarrena `tamainak`, azpilagin bakoitzaren tamaina, hots, aztertutako pieza kopurua.

`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de atributos: p,np. Organización de los datos: los casos son subgrupos.` → `Número de disconformes: [akatsak]. Tamaño de la muestra. Variable: tamainak.`

9.2.2. Akastun unitateen kopurua (*np* kontrol-grafikoa)

Egin dezagun 9.4. ariketa. Akastun matrazeen aldagaia osatu eta gero, p grafikoa egitean aurreko urratsei jarraitu eta *np* kasuan honako hau aldatu: `Diseño: np (número de disconformes)`. Bi kasuetan, `Tamaño de la muestra. Constante: 100`



9.2.3. Batez besteko akatsen kopurua unitateko (u kontrol-grafikoa)

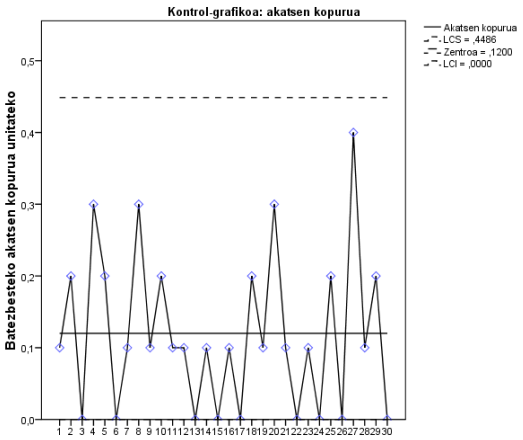
Egin dezagun 9.5 ariketa. Osatuko dugu akatsen aldagai bakarra.

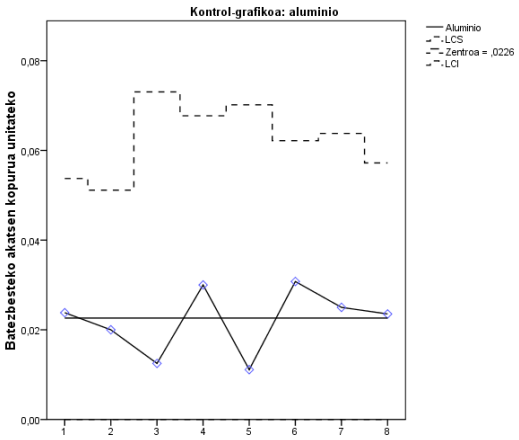
`Analizar` → `Control de calidad` → `Gráficos de control` → `Gráficos de atributos: c, u. Organización de los datos: los casos son subgrupos.` → `Número de disconformidades: [akatsak]. Tamaño de la muestra. Constante: 10. Diseño: u (Disconformidades por unidad)`

Unitateko batez besteko akatsen kopurua 0.12 da, eta behaketa guztiak goi-muga baino baxuagoak direnez, prozesua kontrolaren menpe dago.

Egin dezagun 9.6 ariketa. Osatuko ditugu bi aldagai: lehenengoa $k = 8$ tamainako `akatsak` izeneko aldagaia akatsen kopuruak adieraziz, eta bigarrena `tamainak` deritzona, behaketa bakoitzean aztertutako azalera. SPSSren arabera tamainek zenbaki arruntak izan behar dutenez, adieraz dezakegu tamaina dezimetro karratutan `Transformar` → `Calcular variable` erabiliz. Eta kasu horretan, aurreko bideari jarraituz baina: `Tamaño de la muestra. Variable: [tamainak].`

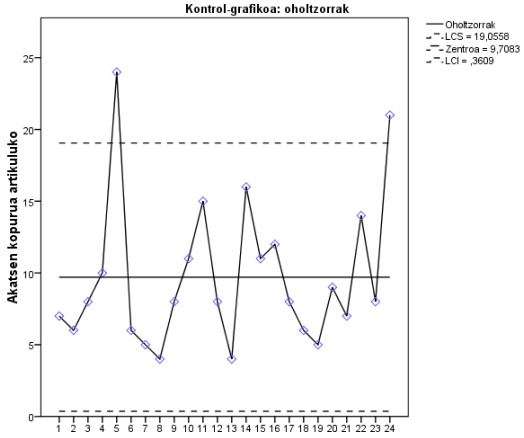
Grafikoan dezimetro karratuko akatsen kopurua adierazten denez, OY eskala bider 100 eginez gero, balioak metro karratuko akatsen kopuruak izango lirateke. Adibidez, marra zentralean $\overline{u} = 0{,}0226$ akats dezimetro karratuko, hots, $\overline{u} = 2{,}26$ akats metro karratuko. Behaketa guztiak goi-muga baino baxuagoak direnez, prozesua kontrolaren menpe dago.





9.2.4. Akatsen kopurua artikuluko (c kontrol-grafikoa)

Egin dezagun 9.7 ariketa. Akatsen aldagaia osatu eta gero, u grafikoa egitean aurreko urratsei jarraitu eta c kasuan honako hau aldatu: `Constante: 1. Diseño: c (número de disconformidades)`.

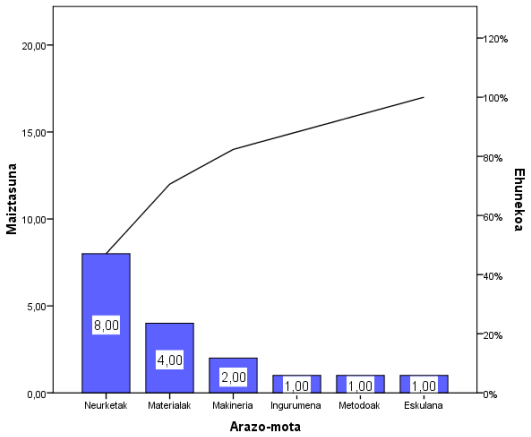


Marra zentralean ikusten denez, akatsen kopurua oholtzorreko 9.7083 da. Bi behaketa 19.0558 goimuga baino handiagoak dira; izan ere, 5. eta 24. behaketak (akatsak 21 eta 24 izanik, hurrenez hurren). Beraz, prozesua kontroletik kanpo dago.

9.3. Paretoren diagrama

Egin dezagun 9.8 ariketa. Horretarako osatu behar ditugu arazoen aldagai kualitatiboa `[arazo-mota]` eta maiztasun absolutuen aldagaia `[maiztasuna]` eta `Datos` → `Ponderar casos` → `Ponderar casos mediante: maiztasuna` egin.

`Analizar` → `Control de calidad` → `Gráficos de Pareto. Valores individuales de los casos` → `Valores[maiztasuna]. Etiqueta de categorías: variable[arazo-mota]`


20,00
15,00
10,00
5,00
0,00
Maiztasuna
120%
100%
80%
60%
40%
20%
0%
Ehunekoa
8,00
4,00
2,00
1,00
1,00
1,00
Neurketak
Materialak
Makineria
Ingurumena
Metodoak
Eskulana
Arazo-mota

Eranskina: ariketak

Atal honetan, SPSSko praktikak lantzeko erabilitako eta proposatutako ariketak kontsulta daitezke, batzuk bibliografiako liburuetatik aterata daude.

1. Probabilitatea
2. Zorizko aldagai diskretu eta jarraituak

   2.1 Aseguru-etxe bateko langile batek bost adin berdineko gizabanakori aseguru-poliza saldu die. Taula aktuarialen arabera, adin horretako gizabanako batek 30 urte baino gehiago bizitzeko probabilitatea 3/5ekoa da. Kalkula itzazu:

   a) 30 urte barru, gutxienez 3 pertsona bizirik egoteko probabilitatea.

   b) 30 urte barru, gehienez 2 pertsona bizirik egoteko probabilitatea.

   (Em.: a) 0.6826, b) 0.3174)

   2.2 Gasolina-zerbitzugune batera iristen diren automobil-kopurua orduko 204koa da, batez beste. Baldin zerbitzugune horrek gehienez minutuko 10 automobil zerbitza baditzake, kalkula ezazu minutu zehatz batean zerbitza daitezkeenak baino automobil gehiago iristeko probabilitatea.

   (Em.: 0.0007)

   2.3 Populazio batean 12000 euro baino gehiago kobratzen duen gizabanako-kopurua %0.004koa da. Kalkula ezazu, aztertutako 5000 gizabanakoren artean gehienez 2 pertsona aipatutako kantitatea kobratzeko probabilitatea, kontsultatutako guztiek erantzuten dutela jorik.

   (Em.: 0.9988)

   2.4 Fabrikatze-prozesu batean eguneroko akastun unitate-kopurua 10 parametroko poisson banaketari darraiola ezaguna da. Kalkula ezazu 150 egunetan akastun unitate-kopurua 1480 baino handiagoa izateko probabilitatea eta 1480 eta 1520 artean egotekoa. Eta akastun unitate-kopurua 10eko itxaropena eta bariantza dauzkan zorizko aldagaia baldin bada?

   (Em.: 0.6972, 0.3944; 0.6972, 0.3944)

   2.5 Izan bedi $X : \mathcal{N}(1500,\ 38{,}7)$ zorizko aldagaia. Zein da aldagaiaren balioa

   a) non banaketa-funtzioa 0.5 den? Balio horri mediana deitzen zaio.

   b) non banaketa-funtzioa 0.25 den? Balio horri 1. koartila deitzen zaio.

   (Em.: a) 1500, b) 1474)

3. Estatistika deskribatzailea

3.9 Ondoko taulan, 35 egunean zehar lurralde batean botatzen diren zuhaitzen kopuruak adierazten dira. Kalkula itzazu banaketa honentzat ezagutzen dituzun a) joera zentraleko estatistikoak, b) sakabanapen-estatistikoak eta c) forma-estatistikoak.

200	300	100	100	0	200	0	0	200	0
0	400	0	0	400	0	1000	300	0	0
0	0	0	0	100	0	300	0	200	200
0	0	100	300	0					

(*Em.* : *a*)$\overline{x} = 125{,}71$ *zuhaitz*, $Me = 0$ *zuhaitz*, $Mo = 0$ *zuhaitz*; *b*) $R = 1000$ *zuhaitz*, $RI = 200$ *zuhaitz*, $S_n^2 = 39053{,}06$ *zuhaitz*^{2}, $S_n = 197{,}62$ *zuhaitz*, $CV = 157{,}20\,\%$, *c*) $\nu = 0{,}64$, $g_3 = 2{,}57$, $g_4 = 8{,}32$)

4. Konfiantza tartezko zenbatespena

4.1 Hiri txiki batean ur-erabilerari buruzko ikerkuntza batean, 25 etxebizitzatako zorizko lagina ateratzen da. Banaketa normalari darraion X aldagaia, *asteko erabilitako ur litro-kopurua*, aztertuko da. Zoriz aukeraturiko aste batean, ondoko balioak lortu ziren:

175	185	186	118	158	150	190	178	137	175
180	200	189	200	180	172	145	192	191	181
183	169	172	178	210					

Datu horientzat, $\sum_{i=1}^{25} x_i = 4394$ eta $\sum_{i=1}^{25} x_i^2 = 782666$. a) Egiazta itzazu balio horiek. b) Informazio hau erabiliz, zenbatets itzazu μ, σ^2 eta σ. c) Lor ezazu μ-rako %90eko konfiantza-tartea. d) Hiriko ur-depositua asteko 160 litroko batez besteko kontsumoa baimentzeko aski handia da. Uste al duzu hirian ur-falta arazoren bat egon daitekeenik? Azal ezazu erantzuna lortutako konfiantza-tartean oinarriturik.

(*Em.: b)* $\hat{\mu} = 175{,}76$ *ur-litro/aste*, $\hat{\sigma}^2 = 432{,}36$ *(ur-litro/aste)*^{2} *eta* $\hat{\sigma} = 20{,}79$ *ur-litro/aste. c)* $I_\mu^{0,90} = (168{,}65,\ 182{,}87)$. *d) Ur-falta dago.)*

4.14 Ondoko taulan, enpresa bateko langileen denbora-tarteak (segundotan) ekintza bat egiteko adierazten dira. Demagun populazio normalak direla.

Emakumezkoa	103	94	110	87	98		
Gizonezkoa	97	82	123	92	175	88	118

a) Kalkula ezazu batezbestekoaren puntu-zenbatespena eta %90 mailako KT.

b) Kalkula itzazu gizonezkoen eta emakumezkoenak.

c) Demagun bariantza berdineko populazioak direla. Kalkula ezazu populazioaren batezbestekoen arteko diferentziarako %90 mailako KT.

d) Kalkula ezazu %90eko σ_1^2/σ_2^2-ren KT, non σ_1^2, σ_2^2 bi populazioen bariantzak diren. Zer esan daiteke aurreko atalean egin dugun suposizioari buruz?

e) Kalkula ezazu berriro populazioaren batezbestekoen arteko diferentziarako %90 mailako KT. Zein ondorio atera dezakegu?

(*Em.: a)* $\hat{\mu} = 105{,}58$, $I_\mu^{0,90} = (92{,}54,\ 118{,}63)$, *b)* $\hat{\mu_1} = 110{,}71$, $\hat{\mu_2} = 98{,}40$, $I_{\mu_1}^{0,90} = (87{,}08,\ 134{,}35)$, $I_{\mu_2}^{0,90} = (90{,}07,\ 106{,}73)$ *c)* $I_{\mu_1-\mu_2}^{0,90} = (-39{,}4032,\ 14{,}7832)$, *d)* $I_{\sigma_1^2/\sigma_2^2}^{0,90} =$

$(0{,}0163,\ 0{,}4537) \Rightarrow 1 \notin I \Rightarrow \sigma_1^2 \neq \sigma_2^2$, *e)* $I^{0,90}_{\mu_1-\mu_2} = (-36{,}5220,\ 11{,}8980) \Rightarrow 0 \in I \Rightarrow$ *ezin da ondorioztatu bata bestea baino luzeagoa denik: berdintasuna ez da baztertzen)*

4.15 Espektofotometriaren bidez, tomate freskoen eta ontziraturikoen tomateen nahitaezko elementuak ikertu dira. Horretarako, kobre-kopurua konparatu da tomate freskoetan eta tomate berberetan, haiek ontziratu ondoren. Datuak ondokoak dira:

	1	2	3	4	5	6	7	8	9	10
Freskoa	0.066	0.079	0.069	0.076	0.071	0.087	0.071	0.073	0.067	0.062
Latakoa	0.085	0.088	0.091	0.096	0.093	0.095	0.079	0.078	0.065	0.068

Kalkula ezazu populazioaren batezbestekoen arteko diferentziarako %98 mailako KT. Zein ondorio atera dezakegu? Demagun populazioak normalak direla.
(Em.: $I^{0,98}_{\mu_1-\mu_2} = (-0{,}0192,\ -0{,}0042) \Rightarrow \mu_1 < \mu_2$*)*

4.17 Suzirien aireratze-instalazio berri bat ikertzen ari da. Dagoen sisteman jaurtiketa arrakastatsuen probabilitatea $p = 0{,}8$ da. Sistema berriarekin 40 jaurtiketa esperimental egiten dira, horietatik 34 arrakastatsuak. Kalkula ezazu %95eko p-ren konfiantza-tartea. Ondoriozta daiteke sistema berria hobea dela?
(Em.: $I^{0,95}_{p} = (0{,}7393,\ 0{,}9607) \Rightarrow Ez$*)*

4.18 Artikulu mota bat saltzen duen enpresa batek A markakoak B markakoak baino gehiago saltzen direla baieztatzen du, diferentzia %8koa dela. 200 bezeroren artean 42k A markako artikulua nahiago dute, eta 150 bezeroen artean 18k B markakoa. Kalkula ezazu %94ko konfiantza-tartea bi marken proportzioen arteko diferentziarako. Erabaki ezazu ea %8ko diferentziaren baieztapena baliozkotzat har daitekeenetz.
(Em.: $I^{0,94}_{p_1-p_2} = (0{,}0164,\ 0{,}1636) \Rightarrow Bai$*)*

5. Hipotesi-kontraste parametrikoak

5.7 (Zenbatespena: 4. ariketa) Hiri txiki batean, banaketa normalari darraion X aldagaia, *asteko erabilitako ur litroen kopurua*, aztertu nahi da. Horretarako, 25 etxebizitzako zorizko lagin bat hautatzen da, ondoko emaitzak lorturik: $\overline{x} = 175{,}76$ ur-litro/aste eta $s_{n-1} = 20{,}79$ ur-litro/aste. Hiriko ur-depositua, asteko 160 litroko batez besteko kontsumoa baimentzeko aski handia bada, uste al duzu hirian ur-falta arazoren bat egon daitekeenik? Plantea eta ebatz ezazu hipotesi-kontrastea 0.10eko esangura-mailarekin.
(Em.: $H_0:\ \mu \leq 160$, $t_p = 3{,}7903$ *edo* $p < 0{,}0005 \Rightarrow H_0$ *errefusatu* $\Rightarrow \mu > 160 \Rightarrow$ *Bai, ur-falta egongo da.)*

5.11 (Zenbatespena: 4. ariketa) Enpresa batean ekintza bat egiteko langileek behar duten denbora-tartean (segundotan) sexuak izaniko eragina ikertu nahi dute. Lagina hautatu ondoren, ondoko datuak lortu ziren:

Denbora-tartea	n_i	$\overline{x}_i$	s_i
Emakumezkoa	5	98.40	8.735
Gizonezkoa	7	110.71	32.185

Demagun populazio normalak direla, 0.1eko esangura-mailarekin langileen sexuak ekintza egiteko eraginik ez duela dela onar daiteke?
(Em.: $F_p = 0{,}0737$ *edo* $p < 0{,}05 \Rightarrow \sigma_1^2 \neq \sigma_2^2$; $t_p = -0{,}9635$ *edo* $p > 0{,}20 \Rightarrow H_0$ *onartu* $\Rightarrow \mu_1 = \mu_2 \Rightarrow$ *Bai)*

5.13 (Zenbatespena: 4. ariketa) Espektofotometriaren bidez, tomate fresko eta ontziraturiko tomateen nahitaezko elementuak ikertu dira. Horretarako, kobre-kopurua konparatu da tomate freskoetan eta tomate berberetan, haiek ontziratu ondoren. Datuak aurreko kapituluarenak direla. Demagun populazio normalak direla, 0.02ko esangura-mailarekin zeintzuk onar daitezke?:

a) tomate ontziraturikoen eta freskoen batez besteko kobre-kopurua berdina da,

b) ontziraturikoen batez besteko kobre-kopurua freskoena baino altuagoa da,

c) ontziraturikoen batez besteko kobre-kopurua freskoena baino gutxienez 0.0030 altuagoa da.

(Em.: a) $t_p = -4{,}4077$ *edo* $p < 0{,}002 \Rightarrow H_0$ *errefusatu* $\Rightarrow \mu_1 \neq \mu_2 \Rightarrow$ *Ez,*
b) $t_p = -4{,}4077$ *edo* $p < 0{,}001 \Rightarrow H_0$ *errefusatu* $\Rightarrow \mu_2 > \mu_1 \Rightarrow$ *Bai,*
c) $t_p = -3{,}2776$ *edo* $p > 0{,}995 \Rightarrow H_0$ *onartu* $\Rightarrow \mu_2 \geq \mu_1 + 0{,}0030 \Rightarrow$ *Bai)*

5.17 (Zenbatespena: 4. ariketa) Suzirien aireratze-instalazio berri bat ikertzen ari da. Dagoen sisteman jaurtiketa arrakastatsuen probabilitatea $p = 0{,}8$ da. Sistema berriarekin 40 jaurtiketa esperimental egiten dira, horietatik 34 arrakastatsuak. 0.05eko esangura-mailarekin zer ondoriozta daiteke: a) lehenengo baieztapena, b) sistema berria hobea dela, c) sistema berria txarragoa dela. d) Konfiantza-tartea kalkulatuz azal itzazu lortutako emaitzak.

(Em.: a) $z_p = 0{,}7906$ *edo* $p = 0{,}4296 \Rightarrow H_0$ *onartu* $\Rightarrow p = 0{,}8 \Rightarrow$ *Bai , b)* $z_p = 0{,}7906$ *edo* $p = 0{,}2148 \Rightarrow H_0$ *onartu* $\Rightarrow p \leq 0{,}8 \Rightarrow$ *Ez, c)* $z_p = 0{,}7906$ *edo* $0{,}7852 \Rightarrow H_0$ *onartu* $\Rightarrow p \geq 0{,}8 \Rightarrow$ *Ez. d)* $I_p^{0{,}95} = (0{,}7393,\ 0{,}9607) \Rightarrow$ *Sistema berria berdina da.)*

5.20 (Zenbatespena: 4. ariketa) Artikulu mota bat saltzen duen enpresa batek A markakoak B markakoak baino gehiago saltzen direla baieztatzen du, diferentzia %8koa dela. 200 bezeroren artean 42k A markako artikulua nahiago dute, eta 150 bezeroen artean 18k B markakoa. Erabaki ezazu ea %8ko diferentziaren baieztapena baliozkotzat har daitekeen 0.06ko esangura-mailarekin.

(Em.: $z_p = 0{,}2554$ *edo* $p = 0{,}7948 \Rightarrow H_0$ *onartu* $\Rightarrow p_1 - p_2 = 0{,}08 \Rightarrow$ *Bai)*

6. Hipotesi-kontraste ez-parametrikoak

6.2 Mendel-en legeak esperimentalki egiaztatu nahi ditugu. Horretarako, 500 landare gurutzatu ziren, eta teoriaren arabera, lore gorri, arrosa, hori eta zuriko landare-kopuruek, hurrenez hurren, 8, 12, 10 eta 20 zenbakiekin proportzionalak izan beharko lukete. Lorturiko datuak 70, 126, 96 eta 208 izan ziren, hurrenez hurren.

(Em.: $\chi^2_p = 2{,}03 \leq \chi^2_{0{,}05;3} = 7{,}815 \Rightarrow$ *%5eko esangura-mailarekin ezin dugu* H_0 *errefusatu. Mendel-en legea esperimentalki egiaztatu da, zeren lore gorri, arrosa, hori eta zuriko landare-kopuruak, hurrenez hurren, 8, 12, 10 eta 20 zenbakiekin proportzionalak baitira.)*

6.4 Ondoko taulan urmael batetik ateratako 100 laginen organismo-kopurua adierazten da. Froga ezazu datu horiek Poisson banaketa batetik aterata daudela.

Organismo-kopurua	0	1	2	3	4	5	6	7
Lagin-kopurua	15	30	25	20	5	4	1	0

(Em.: $\chi^2_p = 1{,}15 \leq \chi^2_{0{,}05;3} = 7{,}815 \Rightarrow$ *%5eko esangura-mailarekin ezin dugu* H_0 *errefusatu. Beraz, banaketa* $\mathcal{P}(1{,}86)$*.)*

6.7 EAEn egindako azterketa soziologiko batean galdera baten erantzunak probintziaka banatzen direnetz ikertu nahi dute. Taulan agertzen den maiztasun-banaketa lortu zen. Erantzuna probintziaren menpe dagoela onar daiteke?

	A	B	G
Alde	11	13	9
Kontra	32	28	27
Abstentzioa	7	9	14

(Em.: $\chi^2_p = 3{,}81 < \chi^2_{0{,}05;4} = 9{,}488 \Rightarrow$ *%5eko esangura-mailarekin onartzen dugu erantzuna probintziaren menpe ez dagoela.)*

6.11 Inkesta batean 60 pertsona hautatu dira, (1) gazteak eta (2) helduak honela ordenatuta agertuz. Ondoriozta daiteke zoriz hautatu direla?:

H G H H H G G G H H G H H H H G G G G H G H G G H H H G G H
G G G G H H G H H G G G G H H G H H H H G G H H G H G G H H

(Em.: $n_1 = 29$, $n_2 = 31$, $\hat{R} = 29$ *denez,* $z_p = -0{,}5127 \in S_0$ *edo* $p > 0{,}05 \Rightarrow$ *%5eko esangura-mailarekin ezin dugu zorizkotasuna errefusatu.)*

6.12 Kalitate-ikuskatzaile batek, langileek egiten duten piezen kopurua ikertzeko asmoz, 50 egunean zehar langile batek egindako kopurua jaso du (10 unitatetan neurtua), ondoko emaitzak lorturik:

100 110 80 75 130 95 105 125 140 85 115 120 150 60 77,5 92 112

83 136 65 72,5 89 160 90 114 155 55 124 92,5 50 115 120 150 60

77,5 92 112 83 136 65 72,5 89 160 90 114 155 55 124 92,5 50

Ondoriozta daiteke zoriz aukeratu dela lagina?

(Em.: $n_1 = 25$, $n_2 = 25$, $\hat{R} = 28$ *eta* $Me = 93{,}75$ *denez,* $z_p = 0{,}5715 \in S_0$ *edo* $p > 0{,}05 \Rightarrow$ *%5eko esangura-mailarekin ezin dugu zorizkotasuna errefusatu.)*

6.13 Klase batean adimen-proba bat egin da (1) ikasle berrien eta (2) errepikatzaileen artean emaitzak berdin banatzen diren ala ez aztertzeko asmoz. 30 ikasleren notak aztertu dira, puntuazioa maximoa 100 dela. Zer ondoriozta daiteke?

B	92	12	83	36	65	72,5	89	60	90	14	55	55	24	92,5	50
E	100	10	80	75	30	95	5	25	40	85	15	20	50	60	77,5

(Em.: $n_1 = 15, n_2 = 15, R_1 = 248, R_2 = 217$ *eta* $U_1 = 97$ *denez,* $z_p = -0{,}6429 \in S_0$ *edo* $p > 0{,}05 \Rightarrow$ *%5eko esangura-mailarekin ezin dugu berdintasuna errefusatu.)*

6.14 Pedagogo batek konparatu nahi ditu irakasteko ondoko hiru metodook: (1) online, (2) erdipresentziazkoa eta (3) presentziazkoa. Horretarako ikasturte batean zehar 3 metodoekin irakasten den ikasgai bat hautatzen du eta metodo bakoitzarekin zorizko lagin bakun bat aukeratzen da, nota finalak honako hauek izanik:

Online	78	80	65	57	89				
Erdipresentziazkoa	74	88	82	93	55	70			
Presentziazkoa	68	83	50	91	84	77	94	81	92

Zer ondoriozta daiteke %90eko konfiantza-mailarekin?

(Em.: $n_1 = 5, n_2 = 6, n_3 = 9, R_1 = 42, R_2 = 61, R_3 = 107$ *eta* $K = 1{,}1451$ *denez,* $K \in S_0 = (0,\ 4{,}605)$ *edo* $p > 0{,}10 \Rightarrow$ *%10eko esangura-mailarekin ezin dugu berdintasuna errefusatu.)*

6.15 Enpresa baten arabera beraiek ekoitzitako iragazkia autoen karburagailu hasieran kokatuta erregai-kontsumoa murrizteko baliagarria da. Informazioa kontrastatzeko 30 auto hautatu ziren eta bakoitzaren kontsumoa neurtu zen iragaki barik (ez) eta iragazkiarekin (bai). Behatutako datuak (l/100 km-tan neurtuak) ondoko taulan agertzen direnak badira, zer ondoriozta daiteke?

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ez	6,8	7,0	7,2	9,0	9,1	10,0	9,2	8,5	8,0	8,9	9,3	10,1	6,5	7,8	6,9
Bai	6,4	6,5	7,3	8,8	8,8	9,0	9,4	8,1	7,5	8,9	9,2	10,5	6,4	8,0	6,5

	16	17	18	19	20	21	22	23	24	25	26	27	38	29	30
Ez	7,4	8,7	9,3	8,2	8,0	7,0	9,3	7,0	6,9	10,0	9,4	8,0	7,8	9,0	9,5
Bai	7,1	9,0	9,7	8,0	7,4	6,7	9,9	6,6	7,0	9,0	8,6	7,1	8,8	8,3	8,2

(Em.: $n = 29$, $T^+ = 327$ eta $W = 2{,}3677$ denez, $W \notin S_0 = (-\infty,\ 1{,}645)$ edo $p < 0{,}05 \Rightarrow$ %5eko esangura-mailarekin ezin dugu onartu hipotesi nulua, hots, ezin dugu baieztapena errefusatu) .

7. Erregresioa

7.1 Material isolatzaile berri baten asmatzaileak presio batzuen eraginpean egongo den bi hazbetako lodieradun ale baten konpresioa neurtu nahi du. Horretarako 5 ale kontsideratzen dira. Presioak (hazbete koadroko 10 libratan neurtuta) eta lortutako konpresioak (0.1 hazbetetan) ondoko taulan agertzen dira:

Ale	1	2	3	4	5
Presioa	1	2	3	4	5
Konpresioa	1	1	2	2	4

a) Egin ezazu puntu-hodeia. b) Ebatzi presioarekiko konpresioaren karratu txikienen erregresio-zuzena. c) Egiazta ezazu hondarren batezbestekoa nulua dela ($\sum e_i = 0$). Kalkula ezazu $\hat{\sigma}^2$. d) Ondoriozta daiteke presioa adierazgarria dela %5eko esangura-mailarekin?, e) Kalkula itzazu korrelazio koefizientea eta determinazio-koefizientea.

(Em.: b) $a = -0{,}1$, $b = 0{,}7$, c) $S^2 = 0{,}3667$, d) $3{,}656 \notin (-3{,}182, 3{,}182) \Rightarrow B \neq 0 \Rightarrow$ bai, e) $r = 0{,}904, r^2 = 0{,}817$)

7.4 Elektrizitate-konpainia batean, X etxearen neurriaren (oin karratutan) eta eredu hiperbolikoaren araberako Y etxebizitzaren hileko energia-kontsumoa (kwh-tan) iragarri nahi da. a) Kalkula itzazu parametroen karratu txikienen bidezko zenbatespenak. b) Kalkula ezazu zenbatespenaren errore-estandarra. c) Zein da korrelazio-koefizientea? d)Zein da 1500 oin karratu duen etxe baten itxarondako energia-kontsumoa?

Neurria	1290	1350	1470	1600	1710	1840	1980	2230	2400	2930
Kontsumoa	1182	1172	1264	1493	1571	1711	1804	1840	1956	1954

(Em.: a) $\hat{Y} = 2786{,}87 - 2105963{,}35\frac{1}{X}$, b) $S = 69{,}9154$, c) $r = -0{,}9766$, d) $\hat{y} = 1382{,}89$)

7.5 Aurreko adibidean, $Y = A + BX + CX^2 + \epsilon$ ereduaren araberako etxebizitzaren hileko energia-kontsumoa (kwh-tan) iragarri nahi da. a) Kalkula itzazu A, B eta C parametroen karratu txikienen bidezko zenbatespenak. b) Kalkula ezazu zenbatespenaren errore-estandarra. c) Kalkula ezazu $\hat{B}$-ren zenbatetsitako errore estandarra. d) Y iragartzeko X^2 aldagaiak informazio ematen du? e) Zein da aldagai garrantzitsuena? f)

Doikuntza-egokitasun globala aztertu. g) Zein da 1500 oin karratu duen etxe baten itxarondako energia-kontsumoa? h) Kalkula eta konpara itzazu aldagaien arteko korrelazio eta korrelazio partzialeko koefizienteak. (Oharrak: kontsidera ezazu $\hat{Y} = a+b_1X_1+b_2X_2$ motako erregresio lineal anizkoitza non $X_1 = X$ eta $X_2 = X^2$ diren. Kalkulu-orrien programa erabil dezakezue, esate baterako, EXCEL.)

(Em.: a) $\hat{Y} = -1216{,}1449 + 2{,}3989X - 0{,}00045X^2$*, b)* $S = 46{,}801$*, c)* $0{,}2458$ *, d) Bai,* $t_p = -7{,}618 \notin (-2{,}365, 2{,}365)$*, e)* $|\beta_1| = |4{,}049| > |\beta_2| = |-3{,}161|$*, f)* $r = 0{,}991, r^2 = 0{,}982, \overline{r}^2 = 0{,}977, F_p = 189{,}710 > F_{0{,}05;2,7}$*, g)* $\hat{y} = 1369{,}6609$*, h)* $r_{y1} = 0{,}9120, r_{y2} = 0{,}8576, r_{12} = 0{,}9925$ *eta* $r_{y1\cdot 2} = 0{,}9652, r_{y2\cdot 1} = -0{,}9446, r_{12\cdot y} = 0{,}9969$*)*

7.8 Hurrengo taulak igeriketa-talde baten hamar kideen altuera cm-tan, pisua kg-tan eta adina urtetan adierazten ditu:

Altuera	144	132	160	147	146	148	136	152	140	141
Pisua	38	32	46	39	38	46	28	45	36	32
Adina	9	6	12	10	9	10	6	11	8	7

a) Kalkula ezazu altueraren erregresio anizkoitzaren planoa, pisuarekiko eta adinarekiko. Urte bat gehiago izateagatik zein da itxarondako altueraren gehikuntza? Eta kilogramo bat gehiago pisatzeagatik? b) Zenbatets ezazu 10 urteko eta 48 kg-ko igerilari baten altuera. c) Kalkula ezazu altuera eta adinarekiko pisuaren erregresio anizkoitzaren planoa. d) Zein da eredu egokiena?

(Em.: a) $\hat{Altuera} = 116{,}322 - 0{,}401\ Pisua + 4{,}943 Adina$*, b)* $146{,}504$ *cm, c)* $\hat{Pisua} = 54{,}232 - 0{,}375\ Altuera + 4{,}317 Adina$*, d) lehenengoa (*$r^2_{alt} = 0{,}940 > 0{,}909 = r^2_{pis}$*)*

7.9 Bost familiari dagozkien hurrengo datuak dauzkagu, eurotan emanda:

Familia	Aurrezpena S	Diru-sarrera Y	Kapitala W
A	600	8000	12000
B	1200	11000	6000
C	1000	9000	6000
D	700	6000	3000
E	300	6000	18000

a) Kalkulatu S-ren erregresio-ekuazioa, Y eta W-rekiko. b) Kalkula itzazu aurrezpenaren korrelazio koefiziente bakunak eta partzialak. c) Zein da eragin handiagoa duen aldagaia? d) Kalkulatu hondar-bariantza eta bariantza azaldua. Zeintzuk dira bariantza totaletik bariantza azalduaren eta azaldu gabekoaren portzentajeak? Zein da zuzendutako determinazio-koefiziente anizkoitza? e) Zein aurrezpena aurreikusiko zenuke 5000 euroko kapitala eta 8000 euroko diru-sarrera dauzkan familia batentzat? (Oharra: kalkuluak errazteko aldagaien unitateak mila eurotan adieraz itzazu.)

(Em.: a) $\hat{S} = 103{,}6508 + 0{,}1151Y - 0{,}0294W$*, b)* $r_{sy} = 0{,}8737$*,* $r_{sw} = -0{,}7485$*,* $r_{sy\cdot w} = 0{,}9818$ *eta* $r_{sw\cdot y} = -0{,}9660$*, c)* $|\beta_y| = |0{,}696| > |\beta_w| = |-0{,}502|$*, d)* $s^2_e = 1558{,}7302$*,* $s^2_{\hat{y}} = 96841{,}2698$*,* $\%98{,}42$ *eta* $\%1{,}58$ *eta* $\overline{r}^2 = 0{,}9683$*, e)* $877{,}4603$ *euro)*

8. Bariantza-analisia

8.1 1964-1969 urteetan jaso ziren ondoko datuak. Determina ezazu ea urte-sasoien arabera heriotza-indizeen desberdintasuna adierazgarria den ala ez, 0.05eko esangura-mailarekin. Urteka heriotza-indizeak independenteak direla jotzen da.

Urte-sasoia	Behaketak
Udaberria	9, 9.3, 9.3, 9.2, 9.4, 9.1
Uda	8.8, 8.7, 8.8, 8.6, 8.7, 8.3
Udazkena	9.4, 9.4, 10.3, 9.8, 9.4, 9.6
Negua	10.6, 9.8, 10.9, 10.2, 9.7, 9.9

Ondorioak atera ahal izateko, zeintzuk dira jo behar diren baldintzak?

(Em.: $F = 24{,}58 > F_{0{,}05;3,20} = 3{,}10 \Rightarrow$ *%5eko esangura-mailarekin urte-sasoien arabera heriotza-indizeen desberdintasuna adierazgarria da.)*

8.6 Pinu-multzo baten altuera (metrotan) eta espeziea ezagutuz, altueraren arabera pinu guztiak berdinak direla kontsidera daiteke?

Pinea	8.52	Pinaster	8.52	Pinea	8.13
Pinaster	6.45	Pinea	6.43	Halapensis	7.17
Silvestris	7.41	Pinea	6.21	Pinaster	8.40
Pinea	7.15	Halapensis	7.07	Silvestris	8.87
Pinaster	8.73	Pinaster	8.83	Pinea	6.12
Laricio	7.55	Pinaster	8.53	Pinaster	8.91
Halapensis	6.54	Laricio	7.84	Silvestris	8.81
Laricio	7.74	Silvestris	8.59	Laricio	7.40
Silvestris	8.65	Laricio	7.41	Pinaster	8.19
Silvestris	8.81	Pinaster	8.94	Pinaster	8.56

(Em.: $F = 6{,}33 > F_{0{,}05;4,25} = 2{,}76 \Rightarrow$ *ezin dugu onartu pinu guztiak berdinak direnik).*

8.9 Termometroak ekoizten dituen lantegi batean, marka desberdinetako lau makina erabiltzen dira, non makina bakoitza lau langilek erabiltzen duten. Egunero langileko eta makinako ekoitzitako unitate-kopurua ondokoa da:

Unitate-kopurua	Makina			
Langilea	A	B	C	D
1	14	9	7	8
2	12	11	10	9
3	16	8	8	11
4	14	8	6	10

Egin ezazu dagokion bariantza-analisia, suposaketak eta ondorioak zehaztuz.

(Em.: $F'(L) = 0{,}69 < F_{0{,}05;3,9} = 3{,}86 \Rightarrow$ *langileak ez du eraginik unitate-kopuruan.* $F''(M) = 11{,}86 > F_{0{,}05;3,9} = 3{,}86 \Rightarrow$ *makina-motaren arabera unitate-kopurua aldatzen da. %5eko esangura-mailarekin unitate-kopuruak desberdinak dira).*

8.10 Patata ekoizpenarekiko lau ongarrien arteko desberdintasunak ikertzeko asmoz, 5 finka kontsideratu ziren. Finka bakoitza tamaina eta mota berberako lau lurzatitan azpizatitu zen. Zoriz, finka bakoitzeko lurzati bakoitzean banatu ziren ongarriak. Ongarrien arabera edo finken arabera, desberdintasunak adierazgarriak direla esan daiteke? Etekinen (tonatan) datuak ondokoak izan ziren:

Etekina	Ongarria			
Lurzatia	1	2	3	4
1	2.1	2.2	1.8	2.1
2	2.2	2.6	1.9	2.0
3	1.8	2.7	1.6	2.2
4	2.0	2.5	2.0	2.4
5	1.9	2.8	1.9	2.1

(Em.: $F'(L) = 0{,}65 < F_{0{,}05;4,12} = 3{,}26 \Rightarrow$ *lurzatien arabera desberdintasunak ez dira adierazgarriak.* $F''(O) = 14{,}04 > F_{0{,}05;3,12} = 3{,}49 \Rightarrow$ *ongarrien arabera desberdintasunak adierazgarriak dira. %5eko esangura-mailarekin desberdintasunak adierazgarriak dira.)*

8.11 Hiru gari mota hiru leku desberdinetan landatu dira. Hazi eta gero, mota bakoitzaren lagin bat hartzen da (guztira 9), irina lortzeko. Lagin bakoitzeko irinarekin, hiruna ogi-barra egiten dira. Hurrengo taulan, lortutako 27 barren bolumenak (unitate estandarretan) agertzen dira. %1eko esangura-mailarekin, kontrasta ezazu ea gari-motak edo lekuak eraginik daukaten barren batez besteko bolumenean. Bi faktoreen arteko interakzioarik dagoela esan daiteke?.

Bolumena	Gari mota		
Lekua	1	2	3
1	15.2 13.8 14.3	13.4 16.5 15.2	20.4 18.2 16.3
2	7.6 4.8 3.9	4.8 2.7 3.9	12.2 11.8 13.4
3	19.2 17.5 18.4	11.3 13.4 12.6	22.3 25.1 24.2

(Em.: $F'(L) = 164{,}37 > F_{0{,}01;2,18} = 6{,}01 \Rightarrow$ *lekuak eragina du barren bolumena azaltzeko.* $F''(G) = 78{,}25 > F_{0{,}01;2,18} = 6{,}01 \Rightarrow$ *gari-motak ere bai.* $F''' = 8{,}37 > F_{0{,}01;4,18} = 4{,}58 \Rightarrow$ *lurzati eta gari-motaren artean interazkioa dago. %1eko esangura-mailarekin desberdintasunak adierazgarriak dira.)*

8.12 Zirkulazio-istripuetan droga batzuek izaniko eragina ikertu nahi da. Horretarako, gidatzeko simulagailu batean jarri ziren hiru emakume eta hiru gizon bakoitza tratamendu baten menpe: marihuanaren eraginpean, alkoholaren eraginpean eta drogarik gabe. Simulagailuak 0 eta 35 puntu artean ematen ditu, puntuazio altuenak gidatzeko egoera onenekin erlazionatuta daudela. Ondoko datuak lortu ziren:

Pertsona	Marihuana	Alkohola	Droga barik
E	19	8	21
E	18	10	31
E	25	10	26
G	20	18	28
G	17	7	14
G	21	16	24

Sexuaren arabera, hiru drogen eragina berdina dela ondoriozta daiteke? Zer esan daiteke sexuaren eta hartutako drogaren arteko interakzioari buruz?

(Em.: $F'(S) = 0{,}02 < F_{0{,}05;1,12} = 4{,}75 \Rightarrow$ *sexuaren arabera ez dago desberdintasun adierazgarririk.* $F''(T) = 11{,}17 > F_{0{,}05;2,12} = 3{,}89 \Rightarrow$ *droga-motak eragina du puntuazioan.*

$F''' = 1{,}24 < F_{0{,}05;2,12} = 3{,}89 \Rightarrow$ sexuaren eta droga-motaren arteko interakziorik ez dago. %5eko esangura-mailarekin desberdintasunak adierazgarriak dira.)

9. Kalitatearen Kontrol Estatistikoa

9.1 [*SPSS*] Zementu-fabrika batean, 9 ordu eta erditan zehar, erdi ordutan behin 4 zementuzaku hautatu dira zoriz, haien pisuak honako hauek izanik:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
7	$7^{\frac{1}{2}}$	8	$8^{\frac{1}{2}}$	9	$9^{\frac{1}{2}}$	10	$10^{\frac{1}{2}}$	11	$11^{\frac{1}{2}}$	12	$12^{\frac{1}{2}}$	13	$13^{\frac{1}{2}}$	14	$14^{\frac{1}{2}}$	15	$15^{\frac{1}{2}}$	16	$16^{\frac{1}{2}}$
53	50	52	46	51	48	51	48	47	49	54	49	50	53	50	51	47	52	50	49
51	49	50	47	48	49	46	47	52	50	55	45	48	54	49	50	51	51	53	50
52	50	53	52	49	51	49	48	46	50	53	46	51	52	48	49	48	50	48	51
49	49	51	50	50	50	50	49	51	51	52	48	47	50	51	52	49	49	49	52

a) Egin ezazu batezbestekoen eta heinen grafikoa. b) Egin ezazu desbiderapenen eta heinen grafikoa. Zer ondoriozta daiteke?

(Em.: a) $\overline{x} = 49{,}8750,\ LCI = 47{,}2885,\ LCS = 52{,}4615$ eta $\overline{x} = 3{,}55,\ LCI = 0,\ LCS = 8{,}1013$; b) $\overline{x} = 49{,}8750,\ LCI = 47{,}3245,\ LCS = 52{,}4255$ eta $\overline{s} = 1{,}5667,\ LCI = 0,\ LCS = 3{,}55$; c) Aldakortasuna kontrol-mugen artean dago (hots, batezbestekoen inguruan homogenotasun handia dago) baina 11. eta 12. behaketak kontrol-mugetatik kanpo daude).

9.2 [*SPSS*] Bonbillak ekoizten dituen enpresa batean, 70 bonbillako 30 zorizko lagin hautatu ondoren, akastun bonbillen kopurua ondokoa izan zen, hurrenez hurren: 1, 2, 0, 3, 2, 0, 1, 3, 1, 2, 1, 1, 0, 1, 0, 1, 0, 2, 1, 3, 1, 0, 1, 0, 2, 0, 4, 1, 2, 0. Zenbatets ezazu ekoitzitako akastun bonbillen ehunekoa eta egin ezazu akastun bonbilla-ehunekoaren kontrol-grafikoa.

(Em.: $\hat{p} = 0{,}0171,\ LCI = 0,\ LCS = 0{,}0637\ LCI \leq p_i \leq LCS,\ \forall i \Rightarrow$ Kontrolaren menpe)

9.3 [*SPSS*] Produzkio-prozesu batean hainbat tamainatako 15 lagin kontsideratu dira akastun piezen kopurua ondokoa izanik:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n_i	120	180	250	250	320	200	250	100	350	100	400	350	250	300	100
x_i	11	20	27	25	29	24	21	13	35	9	34	34	24	28	12

Zenbatets ezazu ekoitzitako akastun piezen ehunekoa eta egin ezazu akastunen ehunekoaren kontrol-grafikoa.

(Em.: $\hat{p} = 0{,}0983,\ LCI = \{0,0161,\ 0,0310,\ 0,0411,\ 0,0411,\ 0,0476,\ 0,0344,\ 0,0411,\ 0,0084,\ 0,0498,\ 0,0084,\ 0,0529,\ 0,0498,\ 0,0411,\ 0,0460,\ 0,0084\},\ LCS = \{0,1785,\ 0,1636,\ 0,1536,\ 0,1536,\ 0,1470,\ 0,1602,\ 0,1536,\ 0,1862,\ 0,1448,\ 0,1862,\ 0,1418,\ 0,1448,\ 0,1536,\ 0,1487,\ 0,1862\}\ LCI_i \leq p_i \leq LCS_i,\ \forall i \Rightarrow$ Kontrolaren menpe)

9.4 [*SPSS*] Produzkio-prozesu batean 100 matrazeko 25 lagin kontsideratu dira akastun matrazeen kopuruak ondokoak izanik: 14, 11, 13, 12, 5, 8, 14, 21, 13, 2 ,6, 1, 17, 13, 18, 5, 7, 5, 4, 13, 2, 14, 4, 11, 1. Zenbatets itzazu ekoitzitako akastun matrazeen ehunekoa eta akastun matrazeen kopuru osoa, eta egin eta interpreta itzazu kontrol-grafikoak.

(Em.: Akastun matrazeen proportzioa: $\hat{p} = 0{,}0936$, $LCI = 0{,}0062$, $LCS = 0{,}1810$ *eta akastun matrazeen kopuru osoa:* $n\hat{p} = 9{,}3600$, $LCI = 0{,}6219$, $LCS = 18{,}0981$. *8. behaketa goiko kontrol-mugatik at dagoenez, prozesua kontroletik kanpo dago.)*

9.5 [*SPSS*] Industria batean artikulu berezi baten ekoizpena kontrolatzeko, bi ordutan behin, 10 unitateko loteak hartzen dira. Baldin 30 lagin zoriz hautatu badira, akatsen kopurua ondokoa izanik: 1, 2, 0, 3, 2, 0, 1, 3, 1, 2, 1, 1, 0, 1, 0, 1, 0, 2, 1, 3, 1, 0, 1, 0, 2, 0, 4, 1, 2, 0. Zer ondoriozta daiteke batez besteko akatsen kopuruaz?

(Em.: $\overline{u} = 0{,}12$, $LCI = 0$, $LCS = 0{,}4486$ $LCI \leq u_i \leq LCS$, $\forall i \Rightarrow$ *Kontrolaren menpe)*

9.6 [*SPSS*] Aluminio anodizatua ekoizten duen enpresa batek prozesua ikuskatzean metro karratuko akatsak zenbatu ditu. Prozesua kontrolpean dagoela ondoriozta daiteke?

Lagina	1	2	3	4	5	6	7	8
Ordua	7	8	9	10	11	12	13	14
Behatutako m^2-en kopurua	2,1	2,5	0,8	1	0,9	1,3	1,2	1,7
Behatutako akatsen kopurua	5	5	1	3	1	4	3	4

(Em.: $\overline{u} = 2{,}26$, $LCI_i = 0$, $LCS_i = \{5,3737,\ 5,1138,\ 7,3042,\ 6,7717,\ 7,0157,\ 6,2172,\ 6,3787,\ 5,7205\}$ $LCI_i \leq u_i \leq LCS_i$, $\forall i \Rightarrow$ *Kontrolaren menpe)*

9.7 [*SPSS*] Egurrezko enpresa batean ekoitzitako oholtzorren akatsen kalitate-kontrola egin nahi da. 24 oholtzorreko lagina hautatu ondoren, pieza bakoitzaren akatsen kopurua ondokoa izan zen: 7, 6, 8, 10, 24, 6, 5, 4, 8, 11, 15, 8, 4, 16, 11, 12, 8, 6, 5, 9, 7, 14, 8, 21. Egin ezazu oholtzorreko akatsen kopururako grafikoa.

(Em.: $\overline{c} = 9{,}7083$, $LCI = 0{,}3609$, $LCS = 19{,}0558$ *5. behaketa (24 akats) eta azkena (21 akats) goi-mugatik kanpo dagoenez, prozesua kontroletik kanpo dago*

9.8 [*SPSS*] Laboreak ekoizten dituen enpresa baten Salmenta Sailean bezeroen kexak aztertzen ari dira. Horretarako, lan-talde batek bukatutako produktuen biltegian arazoen kausa-efektuak ikertu ondoren, arazo nagusiak honela sailkatu dira: (1) makineria (zakubetetzailea, gaizki betetzea, etiketatzailea, paletitzailea), (2) ingurumena (kutsatzea, hautsa), (3) materialak (lehengaiak, gaizki ixtea edo irekitzea), (4) metodoak (zaku eta paleten manipulazioa), (5) neurketak (pisatzea, nahasturaren neurria), (6) eskulana (makineriaren manipulazioa, zaku eta paleten garraioa). Demagun atzemandako asteroko arazoak ondokoak izan direla: 2 makineria motakoak, 1 ingurumen motakoa, 4 materialengatik, 1 metodoagatik, 8 neurketengatik, 1 eskulana motakoa. Egin ezazu Paretoren diagrama. Grafikoari begiratuz, zein izango da lehenengo helburua?

(Em.: Neurketei buruzko alderdi guztiak hobetzea; izan ere, arazoen %47 baita. $H = \{0{,}4706,\ 0{,}7059,\ 0{,}8235,\ 0{,}8824,\ 0{,}9412,\ 1\}$*)*

Bibliografia

[1] J. M. Casas-Sánchez. *Inferencia estadística.* Centro de Estudios Ramón Areces, 1997.

[2] J. L. Devore. *Probabilidad y Estadística para Ingeniería y Ciencias.* International Thomson, 2001.

[3] J. M. Vilar Fernández. *Modelos Estadísticos Aplicados.* Servicio de Publicaciones de la Universidad de la Coruña, 2003.

[4] R.L. Berger G. Casella. *Statistical Inference.* Duxbury Press, 1990.

[5] D. Peña. *Estadística modelos y métodos.* Alianza Editorial, 1987.

[6] C. Pérez. *Técnicas estadísticas con SPSS.* Prentice Hall, 2004.

[7] R. Ardanuy Q. Martín, M. T. Cabero. *Paquetes Estadísticos SPSS 8.0.* Hespérides, 1999.

[8] V.K. Rohatgi. *Statistical inference.* Editorial Wiley, 1984.

[9] V.K. Rohatgi. *An introduction to probability theory and mathematical statistics.* John Wiley and sons, 2000.

[10] S. Ríos. *Ejercicios de Estadística.* Paraninfo, 1989.

[11] L. Ruiz-Maya. *Problemas de Estadística.* Editorial AC, 1989.

[12] M. R. Spiegel. *Estadística.* MacGraw-Hill, 2002.

[13] R.E. Walpole. *Probabilidad y Estadística para Ingeniería y Ciencias.* Pearson Educacion, 2007.

[14] J. N. Millar y J. C. Millar. *Estadística y Quimiometría para Química Analítica.* Prentice Hall, Pearson Educación, S.A. Madrid, 2002.

[15] I. Miller y J.E. Freund. *Probabilidad y Estadística para Ingeniería y Ciencias.* Prentice Hall, 1997.

[16] R.L. Scheaffer y J.T. McClave. *Probabilidad y Estadística para Ingeniería.* Iberoamericana, 1993.

[17] R.E. Walpole y Myers. *Probabilidad y Estadística.* Mc Graw Hill, 1992.

[18] G. Velasco y P.M. Wisniewski. *Probabilidad y estadística para Ingeniería y Ciencias.* Thomson Learning, 2001.

[19] J. Kickinson y S. Chakraborti. *Non parametric statistical inference.* Editorial Dekker Inc, 1992.

[20] W. Mendenhall y T. Sincich. *Probabilidad y Estadística para Ingeniería y Ciencias.* Prentice Hall Hispanoamericana, 1997.