

# J3 Jarduera

## Datuen kudeaketa.

Pilar González eta Susan Orbe

Ekonomia Aplikatua III (Ekonometria eta Estatistika) Saila

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

## J3.1. Eskuzko barnerapena.

### Adierazburua.

Enpresa familiar batek aztertu nahi du, hondartzan alokatzen dituzten eguzkitako kopurua ( $A$ , asteroko unitatetan) eta astean konstante mantentzen den eguneko alokatze prezioaren arteko erlazioa. Enpresak, aurreko epealdietako datuak, maiatzaren lehen astetik iraileko azken aurreko astera arteko datuak, gordeta ditu:

$t$	1	2	3	4	5	6	7	8	9	10	11
$A$	347	251	339	310	305	317	364	403	323	322	401
$P$	4,5	4,5	4,5	5	5	6	6	7	7	8	9
$t$	12	13	14	15	16	17	18	19	20	21	
$A$	394	424	442	462	491	437	422	433	393	429	
$P$	9	10	10	12	15	15	12	10	10	9	

## J3.1. Eskuzko barnerapena.

### Zereginak.

- a. Barneratu itzazu datuak eskuz Gretl programan.
- b. Barneratu itzazu aldagaien ezaugarriak.
- c. Barneratu ezazu datu-multzoaren informazioa.
- d. Gorde itzazu datu horiek Gretl formatuko datu-fitxategi batean eguzkitakoak.gdt izenarekin.

## J3.1 Datu-fitxategiak eraikitzen: Excel formatua irakurtzen.

### Adierazburua.

Euskal Autonomia Erkidegoko (EAE) azken urteetako autoen salmentaren eboluzioa aztertu nahi da. Horretarako aldagai hauen datuak eskuragarri daude:

*MATEAE*: EAEn matrikulatuta dauden auto kopurua.

*MATSP*: Espainiako estatuan matrikulatuta dauden auto kopurua.

*P*: Brent petrolio kupelaren prezioa dolarretan.

Datu horiek `matrikulazioak.xls` datu-fitxategian aurkitzen dira, Matrikulazio serieak deituriko orrian.

## J3.1. Autoen matrikulazioa.

### Zereginak.

- a. Inporta itzazu datuak Gretl programara.
- b. Barneratu itzazu aldagaien ezaugarriak.
- c. Gorde itzazu datuak Gretl formatuko datu-fitxategi batean `matrikulazioak.gdt` izenarekin.
- d. Ondoren, datuak denborazko serieak direla konturatzean (2004ko urtarriletik 2012ko abendura), informazio hori barneratu nahi da. Aldatu ezazu `matrikulazioak.gdt` datu-fitxategiaren egitura.

## J3.1 Datu-fitxategiak eraikitzen: lagindegiko fitxategia.

### Zereginak.

Ireki ezazu Wooldridge (2001) liburuko karpetan dagoen `wage1.gdt` datu-fitxategia.

- Gorde ezazu fitxategi hori zure karpetaaren batean.
- Interesatzen zaizkigun aldagaiak `wage`, `educ`, `exper`, `tenure`, `nonwhite`, `female`, `married`, `numdep` eta `smsa` dira. Ezabatu itzazu gainerako aldagaiak.
- Aldatu itzazu aldagaien izenak eta deskriapenak euskeraz agertu daitezcan.

*wage* - soldata: batezbesteko soldata, orduko dolarretan.

*educ* - hezk: hezkuntza urteak.

*exper* - esper: esperientzia urteak.

*nonwhite* - arraza: = 1 zuria ez bada.

*female* - emakumea: = 1 emakumea bada.

*married* - ezkondua: = 1 ezkondua bada.

*numdep* - mendekoak: dependenteen kopurua.

*smsa* - hiria: = 1 hirian bizi bada.

*tenure* - antzintasuna: enpresa berdinean lan egindako urte kopurua.

- Gorde ezazu datu-fitxategia `soldata.gdt` izenarekin.



## J3.1 Datu-fitxategiak eraikitzen: txt formatua irakurtzen.

### Adierazburua.

Euskal Herriko landa-etxei buruzko informazio ugari aurkitu daiteke nekatur.net web orrian. NBizkaia.txt datu-fitxategiak honako aldagai hauen informazioa eskaintzan du:

Gela bikoitzen gaueko batezbesteko prezioa (PR, eurotan).

Gosariaren prezioa (GP, pertsonako eurotan).

Logela kopurua (L).

Lokalizazioa (LOK).

Parke natural baterako hurbiltasuna (PN).

Laku edo urtegi baterako hurbiltasuna (LK).

Hondartza baterako hurbiltasuna (HD).

Wifirako sarbidea (WIFI).

Kalitateko Q izatea (Q).

Kalitate turistikoaren konpromezua izatea (KT).

## J3.1 Bizkaiko landa-etxeak.

### *NBizkaia.txt* datu-fitxategiko kodifikazioa.

NBizkaia.txt fitxategiko LOK, PN, LK, HD, WIFI, Q eta KT aldagaien kodifikazioak hauek dira:

- LOK:                    A, hiritik kanpo badago.  
                          C, hiritik hurbil badago.
- PN, LK eta HD:       V, 20 km baino hurrunago badago.  
                          R, km 1 baino hurbilago badago.  
                          N, zerbitzua ez dadu.  
                          P, zerbitzu propioa bada.
- WIFI:                   G, sarbidea dohakoia bada.  
                          N, sarbiderik ez badu.  
                          S, sarbidea ordaindu behar bada.
- Q eta KT:             N, ez baldin badute  
                          S baldin badute.

## J3.1 Bizkaiko landa-etxeak.

### Zereginak.

- a. Inporta ezazu NBizkaia.txt datu-fitxategia. Izan ezazu kontuan NBizkaia.txt fitxategiaren formatua testuarena dela, datuak zeharkakoak direla, aldagaiak tabuladorekin banatuta daudela eta hamarrekoak adierazteko ikurra koma (,) dela.
- b. Begira ezazu Gretl programak ezarritako kodifikazioa aldagai kualitatiboentzat.
- c. Jarri itzazu aldagaien ezaugarriak eta egiazta ezazu aldagai kualitatiboak diskretu bezala agertzen direla.
- d. Gorde ezazu datu-fitxategia NekaturBizkaia.gdt bezala.

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

## J3.2 Jarduera. Datuen eraldakuntza.

### Behaketak gehitu.

Enpresa familiarrak bere fitxategietan begiratzean aurreko denboraldiko irailaren azken asteko behaketak aurkitzen ditu:

$t$		22
$A$		379
$P$		9

Sartu itzazu behaketa berriak eskuz eguzkitakoak.gdt fitxategira.

## J3.2. Eguzkitakoen alokairua.

### Erantsi aldagaiak.

Enpresa familiarraren ahulkulari batek, alokatutako eguzkitakoak eguraldiaren ezaugarriaren mendekoa izan daitekela uste du. Ondorioz, honako aldagai hauek barneratzea erabakitzen du:

$T$ : asteko batezbesteko tenperatura gradu zentigradutan neurtua.

$H$ : astean zehar haizea egon bada ala ez.

Barneratu itzazu eguzkitakoak.gdt fitxategian taulan ematen diren aldagaien balioak.  $H$ , aldagaiari dagokionez eman iezaiozu 1 balioa haizedun astea izan bada eta 2 balioa haize gabeko astea izan bada.

$t$	1	2	3	4	5	6	7	8	9	10	11
$T$	26	23	25	23	23	25	28	33	28	30	33
$H$	bai	bai	ez	bai	bai	ez	ez	ez	ez	bai	bai

$t$	12	13	14	15	16	17	18	19	20	21	22
$T$	32	34	35	36	40	36	34	36	33	35	32
$H$	bai	ez	ez	ez	ez	ez	bai	ez	ez	ez	bai

## J3.2. Autoen matrikulazioa.

### Aldagai berriak exceletik inportatu.

Euskal Autonomia Erkidegoan (EAE) azken urteetako auto matrikulatuen kopuruan egoera ekonomikoaren joerak eragina izan duela pentsatzen da. Aldagai hori neurtzeko Industriako Produkzio Indizea (IPI) erabiliko da proxy-aldagaia bezala.

Matrikulazioak.xls datu-fitxategiko, IPI izeneko orrian, dagoen informazioa hau da:

*IPISP*: Industriarako Produkzio Indizea Espainiako estatuarentzat (datu gordinak).

*IPISPUA*: IPI-ren urteroko aldakuntza tasa espainiako estatuan.

*PIEAE*: EAEko Industriarako Produkzio Indizea urtaroko joera kenduta.

*PIEAEUA*: EAEko IPIren urteroko aldakuntza tasa.

## J3.2. Autoen matrikulazioa.

### Zereginak.

- a. Inporta itzazu datu horiek Gretlera.
- b. Barneratu itzazu aldagaien ezaugarriak.
- c. Gorde itzazu datuak Gretl formatuko datu-fitxategian eta izendatu ezazu  
`matrikulazioak.gdt`  
bezala datu-fitxategi berria.



## J3.2. Datuen simulazioa.

### Adierazburua.

Ireki ezazu *Datu-multzo berria* honako ezaugarri hauekin:

- Zeharkako datuak.
- Lagin-tamaina: 1000.

a. Simula itzazu aldagai hauek:

$u$ , zero batezbesteko eta 9 bariantzako banaketa normal batetik.

$X_2$ , 6 batezbesteko eta 25 bariantzako banaketa normal batetik.

$X_3$ , 16 eta 65 balioen tarteko banaketa uniforma batetik.

b. Simula ezazu aldagai hau:

$$Y = 3 + 0,5 X_2 - 2,4 X_3 + u$$

c. Gorde itzazu datu hauek `simulatuak.gdt` deituriko datu-fitxategian.

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.**
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

## J3.3. Analisi deskribatzailea.

### Bizkaiko landa-etxeak.

- a. Lor itzazu NekaturBizkaia.gdt datu-fitxategiko aldagai kuantitatiboen estatistiko nagusiak.

Zein da logelaren batezbesteko prezioaren ibiltartea?, eta logelen kopuruaren aldagaiarena? Zein dira gosariaren pertsonako prezio minimoa eta maximoa?

- b. Lor ezazu aldagai kuantitatiboen korrelazio matrizea. Iruzkindu itzazu emaitzak.

### Soldatak.

- a. Lor itzazu soldatak.gdt fitxategiko soldata, hezkuntza, esperientzia eta aintzintasuna aldagaien estatistiko nagusiak. Iruzkindu itzazu emaitzak.
- b. Lor ezazu aurreko aldagaien korrelazio matrizea. Iruzkindu itzazu emaitzak.

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

## J3.4. Denborazko serien grafikoa.

### Alokatutako eguzkitakoak.

Ireki ezazu `eguzkitakoak.gdt` datu-fitxategia eta irudikatu ezazu alokatutako eguzkitakoak, prezioa eta tenperatura aldagaien denborazko serien grafikoa.

- a. Irudikatu itzazu aldagaien serieak batera grafiko batean.
- b. Iruzkindu itzazu emaitzak:
  - Serietan joeraren bat ikusten duzu?
  - Serietan epealdiko joeraren bat ikusten duzu?

## J3.4. Denborazko serien grafikoa.

### Matrikulatutako autoak.

Ireki ezazu matrikulazioak.gdt datu-fitxategia:

- a. Adierazi itzazu grafikoki EAEn eta Espainian matrikulatutako autoen eta petrolioaren prezioaren denborazko serieak:
  - Grafiko berean.
  - Banakako grafikoetan.
  
- b. Irudikatu itzazu grafiko berean EAEko eta Espainiako IPI serieak eta urteroko aldakuntza serieak.
  
- c. Iruzkindu itzazu emaitzak:  
Serietan joeraren bat ikusten duzu?  
Serietan urtaroko joeraren bat ikusten duzu?

## J3.4. Grafiko bibarianteak.

### Alokatutako eguzkitakoak.

Ireki ezazu eguzkitakoak.gdt datu-fitxategia eta adierazi ezazu grafikoki honako aldagai hauen arteko erlazioak:

- Alokatutako eguzkitakoak prezioarekiko.
- Alokatutako eguzkitakoak tenperaturarekiko.
- Alokatutako eguzkitakoak haizearekiko.
- Iruzkindu itzazu emaitzak.

## J3.4. Grafiko bibarianteak.

### Matrikulatutako autoak.

Ireki ezazu matrikulazioak.gdt datu-fitxategia eta irudikatu itzazu honako aldagai hauen arteko erlazioak:

- EAEn matrikulatutako autoak petrolioaren prezioarekiko.
- EAEn matrikulatutako autoak EAeko IPIrekiko.
- EAEn matrikulatutako autoak IPIren urteroko aldakuntzarekiko.
- Iruzkindu itzazu emaitzak.



## J3.4. Grafiko bibarianteak.

### Soldatak.

Ireki ezazu soldatak.gdt datu-fitxategia eta irudikatu itzazu honako aldagai hauen erlazioak:

- a. Soldatak hezkuntzarekiko.
- b. Soldatak esperientziarekiko.
- c. Soldatak aintzinasunarekiko.
- d. Esperientzia aintzinasunarekiko.
- d. Esperientzia hezkuntzarekiko.
- e. Hezkuntza aintzinasunarekiko.
- f. Iruzkindu itzazu emaitzak.

## J3.4. Grafiko tribarianteak.

### Alokatutako eguzkitakoak.

Ireki ezazu `eguzkitakoak.gdt` datu-fitxategia eta irudikatu ezazu alokatutako eguzkitakoak, prezioarekin eta tenperaturarekin duen erlazioa.

### Matrikulatutako autoak.

Ireki ezazu `matrikulazioak.gdt` datu-fitxategia eta irudikatu ezazu honako aldagai hauen erlazioak:

- Espanian matrikulatutako autoak petrolioaren prezioarekiko eta Espainiako IPIrekiko.
- Espanian matrikulatutako autoak petrolioaren prezioarekiko eta Espainiako IPIren urteroko aldakuntzarekiko.
- Iruzkindu itzazu emaitzak.

## J3.4. Grafiko tribarianteak.

### Soldatak.

Ireki ezazu soldatak.gdt datu-fitxategia eta irudikatu itzazu honako aldagai hauen arteko erlazioak:

- Soldatak, hezkuntza eta esperientzia aldagaiekiko.
- Soldatak, esperientzia eta aintzinasuna aldagaiekiko.
- Soldatak, aintzinasuna eta hezkuntza aldagaiekiko.
- Esperientzia, hezkuntza eta aintzinasuna aldagaiekiko.
- Iruzkindu itzazu emaitzak.

- 1 J3.1 Jarduera. Datu-fitxategiak eraikitzen.
- 2 J3.2 Jarduera. Datuen eraldakuntza.
- 3 J3.3 Jarduera. Analisi deskribatzailea.
- 4 J3.4 Jarduera. Analisi grafikoa.
- 5 J3.5 Jarduera. Aldagai bakarraren analisia.

## J3.5 Jarduera. Aldagai bakarraren analisia: banaketa

### Alokatutako eguzkitakoak.

Ireki ezazu eguzkitakoak.gdt datu-fitxategia eta azter itzazu alokatutako eguzkitakoak, prezioa eta tenperatura aldagaien banaketak.

- Kontrasta ezazu hiru aldagaien normaltasuna.
- Lor itzazu alokatutako eguzkitakoak, prezioa eta tenperatura aldagaien maiztasun banaketak normaltasun eta gamma kontrasteen informazioarekin.
- Iruzkindu itzazu emaitzak.

Aldagaien bat banaketa normal batetik eratorria da?

Aldagaien bat gamma banaketa batetik eratorria da?

Zer ezaugarri ikus daiteke lortutako maiztasun banaketetan?

## J3.5 Jarduera. Aldagai bakarraren analisia: banaketa

### Matrikulatutako autoak.

Ireki ezazu matrikulazioak.gdt datu-fitxategia.

- Estima itzazu EAEn eta Espainian matrikulatutako auto aldagaien dentsitate funtzioak kernel gaussianoa erabiliz.
- Lor itzazu aldagai horien maiztasun banaketak normalaren kontrastearen informazioarekin
- Lor itzazu aldagai horien Q-Q grafikoak.
- Interpreta itzazu emaitzak.

Lortutako dentsitate funtzioen ezaugarriak antzekoak dira?, zertan bereizten dira?

Zer ezaugarri bereizten dituzu lortutako maiztasun banaketetan?

Maiztasun grafikoek eta Q-Q grafikoek ematen duten informazioa berdina da?

## J3.5 Jarduera. Aldagai bakarraren analisia: banaketa

### Bizkaiko landa-etxeak.

Ireki ezazu NekaturBizkaia.gdt datu-fitxategia.

- a. Aztertu ezazu logela kopurua aldagaiaren banaketa.
- b. Lor ezazu batezbesteko prezioa aldagaiaren maiztasun banaketa kontrasteen informazioarekin.
- c. Lor ezazu aldagai kualitatiboen maiztasun banaketak.
  - c.1. Zenbat landa-etxe daude hondartzatik kilometro bat baino gutxiagora?
  - c.2. Zenbat landa-etxe daude laku batetik 20 kilometro baino gehiagora?
  - c.3. Zenbat landa-etxe daude parke natural batean kokatuta?
  - c.4. Zenbat etxe daude hondartzatik kilometro bat baino gutxiagora?, eta parke natural batetik kilometro bat baino gutxiagora? Zenbat etxe daude hondartzatik 20 kilometro baino gehiagora?, eta parke natural batetik 20 kilometro baino gehiagora?
  - c.5. Landa-etxeen zer portzentai dago hiriko zentrotik kanpo kokatuta?
  - c.6. Zenbat landa-etxeek dute interneterako konexioa?, horietatik zenbatek dute dohako konexioa?
  - c.7. Landa-etxeen zer portzentaiak ez du kalitateko Q? Zenbatek ez dute kalitate turistikoaren konpromezua?

### Bizkaiko landa-etxeak.

- d. Lor itzazu batezbesteko prezioaren kaxa grafikoak, aldagai bereizle bezala eskuragarri dauden aldagai diskretuak erabiliz.
- e. Iruzkindu itzazu emaitzak.

Zer esan daiteke batezbesteko prezio aldagaiaren banaketari buruz?

Logela kopurua aldagaia banaketa zehatz batetik eratorria delako lagin-ebidentziarik dago?

Zer informazio ematen dute LOK eta WIFI aldagaien maiztasun banaketek?

Zein dira batezbesteko prezio aldagaian gehien eragiten duten aldagai diskretuak?



## J3.5 Jarduera. Aldagai bakarraren analisia: kaxa grafikoak.

### Oilaskoaren kontsumoa.

Ireki ezazu oilasko.gdt datu-fitxategia aldagaien kaxa grafikoak aztertzeko.

- Irudikatu ezazu oilasko kontsumoaren kaxa grafiko sinplea eta  $X_6$  aldagai bereizlea erabiliz lortzen dena.
- Irudikatu ezazu errenta aldagaiaren kaxa grafikoa.
- Irudikatu ezazu prezio aldagai guztien kaxa grafikoak.
- Irudikatu ezazu prezio aldagai guztien kaxa grafikoak  $X_6$  aldagai bereizlea erabiliz.
- Iruzkindu eta interpreta itzazu emaitzak.