ESTATISTIKA ETA DATUEN ANALISIA

IX: Bi aldagai kuantitatiboen baterako azterketa: korrelazioa

Egilea: Josemari Sarasola

Gizapedia

gizapedia.hirusta.io

9.1 Atributu-aldagai kuantitatibo erlazio estatistikoa

9.1.1 Aldagai independentea: atributua

9.1.1.1 Eta edo korrelazio-ratioa

9.1.2 Aldagai independentea: kuantitatibo diskretua

9.1.3 Aldagai independentea: kuantitatibo jarraitua

9.2 Korrelazioa aztertzeko abiapuntua: puntu-hodeia

9.3 Korrelazio-motak

9.4 Kobariantza

9.5 Pearson-en korrelazio koefiziente lineala

9.6 Korrelazio monotonikoa: Spearman koef.

9.7 Aldagai dikotomiko-aldagai kuantitatibo korrelazioa

9.8 Aldagai dikotomiko-aldagai dikotomiko korrelazioa

9.9 Korrelazio partziala

9.10 Sasiko korrelazioa

9.11 Ariketak

9. gaia: Bi aldagai kuantitatiboen baterako azterketa: korrelazioa

Aurreko ikasgaian bi aldagai kualitatiboen arteko asoziazioa ikasita, ikasgai honetan bi aldagai kuantitatiboen arteko erlazio estatistikoa, **korrelazioa** alegia, ikasi behar dugu. Hori baino lehen, ordea, tarteko egoera bat landuko dugu: atributu baten eta aldagai kuantitatibo baten arteko erlazioa alegia.

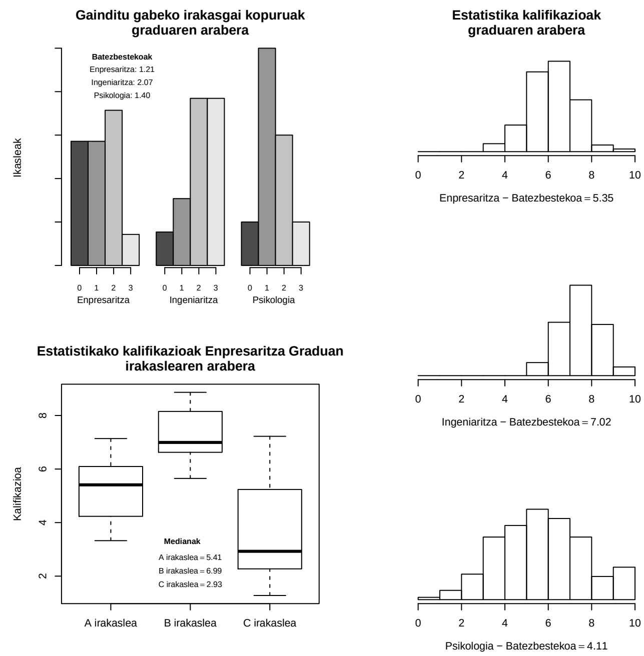
9.1 Atributu-aldagai kuantitatibo erlazio estatistikoa

Aldagai kualitatibo baten eta kuantitatibo baten arteko erlazio estatistikoa aztertzeko lehen pausoa **aldagai independentea eta dependentea** zehaztea da. Ondoren, prozedura orokorra aldagai dependentearen datuak aldagai independentearen kategoria, balio edo balio-tarteen arabera bereizi eta datu-azpimultzo bakoitza bere aldetik aztertu eta azterketa horiek elkarrekin alderatzea da.

9.1.1 Aldagai independentea: atributua

Kasu honetan, atributuaren kategoriaren arabera datuak bereizi eta sortutako datu-azpimultzo kuantitatiboak (beste aldagai kuantitatiboari buruzkoak) banan-banan aztertu behar dira, horretarako diagrama egokiak erabiliz (puntu-diagramak, histogramak, kaxa-diagramak) eta estatistiko deskribatzaileak erabiliz (batezbesteko aritmetikoa, desbideratze estandarra eta abar).

Irudiotan adibide batzuk azaltzen dira:



Irudia 9.1: **Atributua da aldagai independentea**. *(Goian ezkerrean)* Gradua eta gainditu ez diren irakasgaien kopurua jaso dira ikasle batzuengan. Aldagai independentea gradua da, eta horren arabera banatzen dira datuak azterketa egiteko. Batezbestekoei erreparatuz, enpresaritzan gutxiago eta ingeniaritzan gehiago suspenditzen da. *(Behean ezkerrean)* Irakaslea eta estatistikan izandako kalifikazioa jaso dira Enpresaritzako ikasle batzuen artean. Aldagai independentea irakaslea da, irakaslearen arabera izan baitaitezke ezberdinak notak. B irakaslearekin nota handiagoa lortzen da orokorrean. *(Eskuinean)* Ikasle batzuen artean estatistika nota eta gradua jaso dira. Graduaren arabera lortzen da estatistikan nota handiagoa edo txikiagoa, beraz gradua da aldagai independentea. Ingeniaritzan lortzen da orokorrean nota handiena.

9.1.1.1 Eta edo korrelazio-ratioa

Korrelazio-ratioak, eta (η) ere deituak, aldagai kualitatiboak aldagai kuantitatiboak zenbateraino azaldu edo esplikatzen duen neurtzen du. Honela kalkulatzen da:

$$\eta^2 = \frac{\sum_x n_x(\overline{y}_x - \overline{y})^2}{\sum_{x,i}(y_{xi} - \overline{y})^2}$$

non,

- $\overline{y}$ datu guztien batezbestekoa den,
- $\overline{y}_x$ x kategoriako batezbestekoa den,
- y_{xi} datu bakoitza den.

η-ren balio zehatza asoziazio-neurrien balioa bezalatsu interpretatzen da.

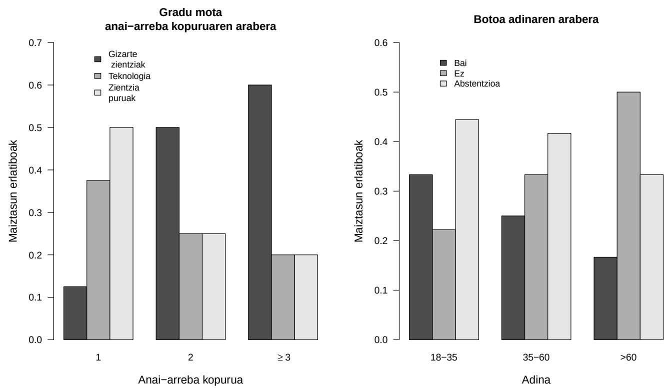
Atributuaren kategoriak bi bakarrik direnean, Cohen-en d erabil daiteke aukeran, baina horretarako gogoratu behar da bi kategorietako datuen bariantzak berdinak izan behar direla.

9.1.2 Aldagai independentea: kuantitatibo diskretua

Adibidez, pertsona batzuengan anai-arreba kopurua eta aukeratutako karrera jaso direnean, aldagai independentea anai-arreba kopurua da. Hala, bi aldagaien arteko erlazio estatistikoa aztertzeko, aukeratutako gradu-motari buruzko datuak anai-arreba kopuruaren arabera sailkatuko ditugu. Ondoren, gradu-motari buruzko datu-azpimultzo bakoitza bere aldetik aztertu (maiztasun-taulak eta barra-diagramak baliatuz esaterako) eta emaitzak (gradu bakoitzaren portzentajeak) elkarrekin alderatuko dira.

9.1.3 Aldagai independentea: kuantitatibo jarraitua

Adibidez, pertsona batzuengan adina eta bozketa batean emandako botoa (bai-ez-abstentzioa) jaso direnean, aldagai independentea adina da. Hala, botoa adinaren arabera aztertuko dugu. Adinak balio ezberdin asko hartzen dituenez, adin-tarteak osatu eta tarte horien arabera botoari buruzko datu-azpimultzoak izango ditugu. Hain zuzen, aldagai kuantitatiboak balio asko hartzen dituenean, beste aldagaiari buruzko datu-azpimultzoak sortzeko modu bakarra aldagai kuantitatiboan tarteak osatzea da. Adibidearekin jarraituz, adin-tarte bakoitzean sektore ekonomikoari buruzko portzentajeak kalkulatu eta beste tarteetakoekin alderatu ditugu. Barra-diagramak ere era daitezke alderatzeko.



Irudia 9.2: **Aldagai independentea kuantitatiboa da**. *(Ezkerrean)* Anai-arreba bakarren kasuan, zientzia eta teknologia dira gehien aukeratzen direnak. Anai-arreba kopuru handiko familietan, berriz, gizarte-zientziak nagusitzen dira. *(Eskuinean)* Ikusten denez, zenbat eta adin handiagoa, baiezko botoa emateko joera handiagoa da.

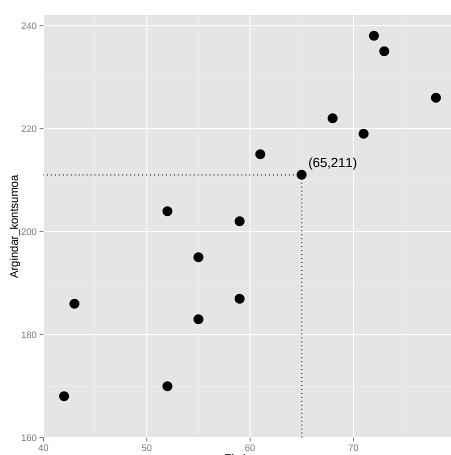
9.2 Korrelazioa aztertzeko abiapuntua: puntu-hodeia

Hemendik aurrera **korrelazioa** soilik, bi aldagai kuantitatiboen arteko erlazio estatistikoa alegia, aztertu behar dugu ikasgaian zehar. Horretarako abiapuntua bi aldagaien datuak lotzen dituen **puntu-hodeia** edo **sakabanatze-diagrama** izenekoa da (ingelesez, *scatter plot*, datu guztiak kartesiar diagrama batean iruditzen dituena besterik ez dena. Puntu-hodeia aztertuz korrelazioaren nondik norakoa azter daiteke.

Adibidea: Lantegi batean argindar-kontsumoa eta ekoizpena jaso dira 15 egunetan zehar. Honakoak dira datuak:

Ekoizpena	65	72	59	68	52	55	71	73
Argindar-kontsumoa	211	238	187	222	204	195	219	235
Ekoizpena	43	59	78	61	52	55	42	
Argindar-kontsumoa	186	202	226	215	170	183	168	

Puntu-hodeia eratu eta interpretatu behar da.

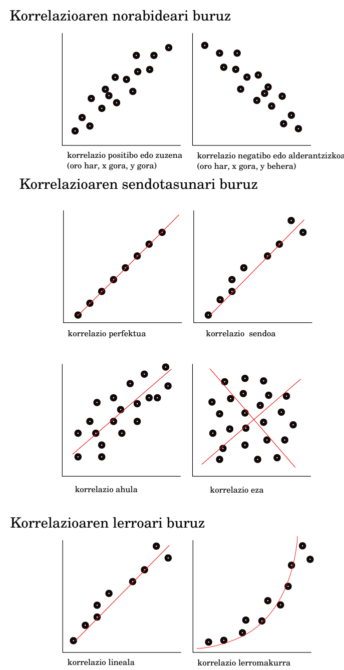


Irudia 9.3: **Puntu-hodeia**. Diagramaren eraketa erakusteko, lehen datu-bikotearen koordenatuak zehaztu dira. Ikusten denez, *oro har*, zenbat eta ekoizpen handiagoa, argindar-kontsumoa ere hainbat eta handiagoa da.

9.3 Korrelazio-motak

Norabidearen arabera,

- **korrelazio positibo** edo **zuzena**, aldagai bat gehitzean, orokorrean beste aldagaia ere gehitu egiten denean;
- **korrelazio negatibo** edo **alderantzizkoa**, aldagai bat gehitzean, orokorrean beste aldagaia murriztu egiten denean.



Sendotasunaren arabera,

- **korrelazio perfektua**, puntuak lerro berean daudenean;
- **korrelazio sendoa**, aldagaien arteko korrelazioa argia eta estua denean;
- **korrelazio ahula**, aldagaien arteko korrelazioa lausoa denean;
- **korrelazio hutsa**, aldagaien artean batere korrelazioa badaezpadakoa denean.

Lotura funtzionalaren arabera,

- **korrelazio lineala**, aldagaien arteko korrelazioa zuzen baten araberakoa denean gutxi gorabehera;
- **korrelazio lerromakurra**, aldagaien arteko korrelazioa kurba baten araberakoa denean gutxi gorabehera.

9.4 Kobariantza

Kobariantza *korrelazio linealaren norabidea* neurtzen duen koefiziente bat da. Nabarmendu behar dira *korrelazio lineala* soilik neurtzen duela, eta horren *norabidea*. Ez du korrelazioaren sendotasunari buruz inongo informaziorik ematen.

Bi formula hauekin kalkula daiteke:

$$s_{xy} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum_i x_i y_i}{n} - \overline{xy}$$

Lehen formula kobariantzaren definizioari eta jatorrizko eraketari dagokio. Bigarrenarekin aiseago burutzen dira kalkuluak.

Kobariantzak edozein balio har dezake, positibo zein negatibo, inongo mugarik gabe. Honela interpretatzen da: kobariantza positiboa bada, korrelazio positibo edo zuzena da; negatiboa bada, korrelazio negatiboa edo alderantzizkoa da. Arestian adierazi bezala, ez du korrelazioaren sendotasunari buruz inongo informaziorik ematen.

Beste alde batetik, kobariantzaren unitateak alagaien unitateen biderkadura da; adibidez, altuerak *cm*-tan eta pisuak *kg*-tan jasota, bi aldagaien arteko kobariantzaren unitateak $cm \times kg$ dira.

9.5 Pearson-en korrelazio koefiziente lineala

Kobariantzak sendotasunari buruzko informazioa ez ematearen arrazoia unitateekin du zerikusirik: adibidez, altueraren (cm) eta pisuaren (kg) arteko kobariantza kalkulatzean, altuera cm-tan eman ordez m-tan emango bagenu kobariantza berria cm kalkulatutakoa zati 100 izango lizateke. Noski, horregatik ezin dugu esan korrelazio txikiagoa denik, unitatea bakarrik aldatu diegunez, datuak berdinak direlako.

Sendotasuna neurtzeko egokia izateko, beraz, unitatea ezabatu beharko diegu datuei, datuak dimentsiogabetu alegia, eta hori jada ikasita daukagun estandarketa delakoaz egiten dugu (gogoratu $z = (x - \overline{x})/s$). Beraz, kobariantza x eta y datuekin kalkulatu ordez, z_x eta z_y datuekin eman beharko da:

$$s_{z_x z_y} = \frac{\sum z_x z_y}{n} - \overline{z_x} \cdot \overline{z_y}$$

Lehenbizi, aldagai estandartuen batezbestekoa 0 dela frogatuko dugu:

$$\overline{z_x} = \frac{\sum \left(\frac{x_i - \overline{x}}{s_x}\right)}{n} = \frac{1}{s_x} \frac{\sum (x_i - \overline{x})}{n} = \frac{1}{s_x} \left(\frac{\sum x_i}{n} - \frac{n\overline{x}}{n} \right) = \frac{1}{s_x} (\overline{x} - \overline{x}) = 0$$

Horrela,

$$s_{z_x z_y} = \frac{\sum z_x z_y}{n} = \frac{\sum \left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right)}{n} = \frac{1}{s_x s_y} \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{s_{xy}}{s_x s_y}$$

Azken adierazpenari Pearson-en korrelazio-koefiziente lineala deritzo, eta honela adierazten da:

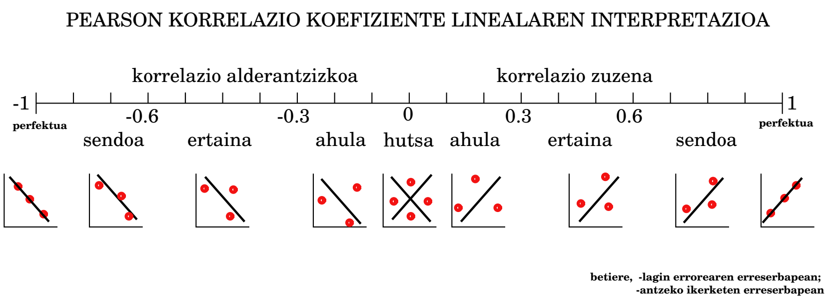
$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Koefizienteak [-1,1] tarteko balioak hartzen ditu, korrelazio lineala soilik neurtzen du (kobariantzak bezalaxe), eta honela interpretatzen da:

- **norabideari buruz**,
  - $r_{xy} > 0$ badugu, korrelazioa positiboa edo zuzena da;
  - $r_{xy} < 0$ badugu, korrelazioa negatiboa edo alderantzizkoa da;
- **sendotasunari buruz**,
  - $|r_{xy}| < 0.3$ badugu, korrelazioa ahula dela esango dugu;
  - $0.3 < |r_{xy}| < 0.6$ badugu, korrelazioa ertaina dela;

- $|r_{xy}| > 0.6$ badugu, korrelazioa sendoa dela;
- $|r_{xy}| = 1$ badugu, korrelazioa perfektoa da (puntu guztiak zuzen batean lerrokatuta daude).

Kasu guztietan, intepretazioa antzeko ikerketen emaitzen eta lagin-errorearen erreserbapean izango da.



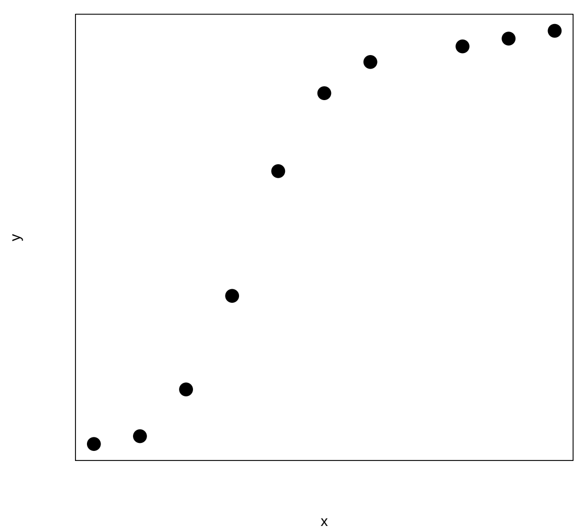
Irudia 9.4: Korrelazioa interpretatzeko erregela.

9.6 Korrelazio monotonikoa: Spearman-en korrelazio-koefizientea

Korrelazio monotonikoak aldagai batek gora egitean besteak orokorrean gora edo behera egiten duen adierazten du, baina erlazio hori lineala edo lerromakurra den kontuan hartu gabe. Pearsonen korrelazio-koefizienteak korrelazio monotoniko linealak bakarrik jasotzen ditu, baina ez lerromakurrak. Spearmanen korrelazio-koefizientea (r_s) mota guztietako korrelazio monotonikoak, linealak nahiz lerromakurrak, jasotzen dituen estatistikoa da, bi aldagaietako elementuen mailen arteko Pearsonen koefizientea besterik ez dena, $(m(x_i), m(y_i))$ bi aldagaietarako ezarritako mailak izanik:

$$r_s = r_{m(x_i),m(y_i)}$$

$[-1, 1]$ tarteko balioak hartzen ditu, eta Pearson koefizientea bezala interpretatzen da.



Irudia 9.5: x aldagaiak gora egiten duenean, y aldagaiak beti egiten du gora. Korrelazioa monotoniko perfektoa da beraz. Pearsonen koefizienteak ez du erlazio perfekto hori jasotzen eta 1etik beherako korrelazioa ematen du (0.933), korrelazio lineala soilik jasotzen duelako. Korrelazio monotonikoa (gora-gora edo gora-behera) besterik jaso nahi ez badugu, Spearman koefizientea erabili behar da, kasu honetan 1 emango duena.

9.7 Aldagai dikotomiko-aldagai kuantitatibo korrelazioa

Pearsonen korrelazio-koefizientea kalkulatu ahal izateko, bi aldagaiak kuantitatiboak izan behar dira. Bi aldagaietako bat kualtitatiboa denean, kalkuluak ezin dira egin noski. Aldagaia *dikotomikoa* denean ordea, bi kategoria bakarrik hartzen dituenean alegia (adibidez, sexuaren kasuan, gizon/emakume; azterketa bat gainditu den, bai/ez), ahal da korrelazio-koefiziente lineala kalkulatu, aldagai dikotomikoaren bi kategoriei 0 eta 1 esleituz, hurrenez hurren. Interpretazioa egiterakoan, tentuz ibili behar da, eta 0/1 balioak behar bezala deskodetu behar dira.

9.7.1 Dikotomiko-kuantitatibo korrelazioaren aplikazioa: item-test korrelazioa

Test bateko item edo galdera baten erantzuna gaizki/ongi edo 0 puntu/1 puntu motakoa denean, galdera horren erantzunaren eta test osoko puntuazioaren arteko korrelazioa kalkulatu daiteke. Gaizki 0 eta ongi 1 dikotomizazioarekin, baiezta daiteke galdera egokiak testeko puntuazio osoarekin korrelazio positibo eta altua dutenak dira, eta galdera desegokiak, testak jasotzen duenarekin zer ikusirik ez dutelako, korrelazio ahula edo negatiboa dutenak.

9.7.2 Cronbach-en alfa

Aurreko aplikazioarekin loturik, test bati barne koherentzia eskatu behar zaio, item edo galdera guztiek gauza bera neurtzea, izan konstruktoa, izan ezagutza edo trebezia. Horretarako, estatistika baliabide zenbait eskaintzen ditu, adibidez item jakin bateko erantzunaren eta test osoko puntuazioaren arteko korrelazio koefizientea eskatzea, aurreko sekzioan ikusi dugun bezala. Baina horrela azkenean galdera bakoitzaren koherentzia maila eskuratzen dugu eta ez test osoko barne koherentziaren neurri bat.

Cronbach-en alfa (Lee Cronbach, 1951) berriz, test bateko barne koherentzia globala neurtzen duen koefizientea da:

$$\alpha = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum_{i=1}^{k} S_i^2}{S_t^2}\right]$$

- S_t^2 izanik test osoko puntuazioa;
- S_i^2 item edo galdera bakoitzaren puntuazioa;

- k izanik item kopurua.

Cronbach-en alfak testeko item edo galderen korrelazio bateratua neurtzen du, eta beraz, test osoko barne koherentziaren neurria litzateke.

Honela interpretatzen da:

- 1 denean, itemak perfektuki daude korrelazionaturik.
- 0 denean, ez dago itemen arteko korrelaziorik.
- Negatiboa denean, itemak modu negatiboan daude korrelazionaturik; eta orduan, testa ez da koherentziaz eratua, inkoherentziaz baizik.

Orokorrean, 0.7-0.9 bitarteko balioak dira egokienak testa koherentea dela baieztatzeko. Hortik gora, testa koherentea litzateke baina galderen artean erredundantzia litzateke.

Interpretatzerakoan, kontuan hartu behar da baita ere item kopuru handiak Cronbach alfa handia izatera eramaten duela orokorrean, eta item kopuru txikiak alfa txikia izatera. Beraz, galdera askoko testetan Cronbachen alfrako balio handiagoak eskatu beharko lirateke.

Koefizientearen balio txikia denean, puntuazio totalarekin korrelazio eskasa duten galderak ezabatu edo besterik gabe galderak gehitzea izan daiteke irtenbidea Cronbachen alfaren balio onargarri batera heltzeko.

Cronbach-en alfa puntuazioa totala batuketaz kalkulatzen duten testetan soilik erabiltzen da.

Azkenik ohartu behar da Cronbach-en alfa ez duela zehazten testak orokorrean barne koherentzia duen, baizik eta testeko erantzunek barne koherentzia duten, erantzun jakin batzuetarako kalkulatzen den koefizientea baita.

KONSTRUKTOA: HEAVY METAL MUSIKAREKIKO ZALETASUNA

Test erredundantea	Test inkoherentea	Test kontraesankorra
• ACDC taldea gogoko al duzu? • Metallica taldea gogoko al duzu? • Iron Maiden taldea gogoko al duzu?	• ACDC taldea gogoko al duzu? • Vox-i inoiz eman al diozu botoa? • Makarroiak gogoko al dituzu?	• ACDC taldea gogoko al duzu? • Gregoriar kantua gogoko al duzu?
$\alpha \sim 1$	$\alpha \sim 0+$	$\alpha < 0$

KONSTRUKTOA: HEAVY METAL MUSIKAREKIKO ZALETASUNA

Test koherentea
• ACDC taldea gogoko al duzu? • Inoiz joan zara heavy metaleko kontzertu batera? • Inoiz erosi duzu heavy metaleko diskoa?
$\alpha \in [0.7, 0.9]$

9.8 Dikotomiko-dikotomiko korrelazioa

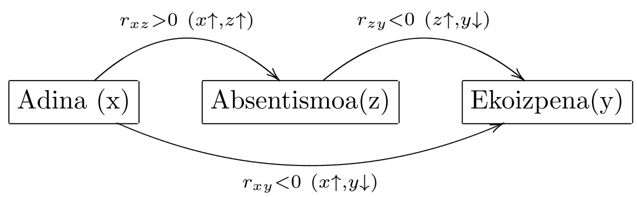
Bi aldagaiak dikotomikoak direnean ere kalkula daiteke korrelazio-koefiziente lineala. Nahikoa da bi aldagai dikotomikoetako bina kategoriei 0/1 balioak esleitzea. Interpretazioa 0/1 balioak deskodetuz egiten da.

9.9 Korrelazio partziala

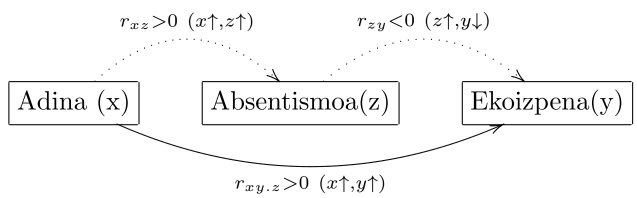
Korrelazio partziala bi aldagai kuantitatiboen arteko korrelazioa kalkulatzeko erabiltzen da, baina zeharkako eragina eragiten duten aldagaien efektua baztertuz.

Demagun lantegi bateko langileen adinak, azken urtean bakoitzak izan dituen absentismo-egunak eta ekoizpena jaso direla. Hipotesi teorikoak dioenez, langilea zenbat eta zaharragoa (adina zenbat eta handiagoa), ekoizpena hainbat eta handiagoa da. Adinaren eta ekoizpenaren arteko korrelazio-koefizientea kalkulaturik, ordea, negatiboa ematen du. Zer dela-eta suertatzen da uste denaren aurkakoa? Tarteko aldagai bat dago, absentismoa alegia, emaitzak distortsionatu egiten dituena. Hain zuzen, adinean gora, absentismoak ere gora egiten du, eta horrek ekoizpena behera dakar. Beraz, absentismoa da, horren tarteko eraginagatik, ekoizpena behera ekarri eta adinean gora ekoizpenak behera egiten duelako itxurazko korrelazioa sorrarazten duena. Absentismoa *aldagai nahaslea* dela esango dugu. Aldagai nahasle horren efektua saihesteko, tarteko aldagaiaren eragina saihestu edo ezabatu behar da, kasu honetan absentismoarema. Hori korrelazio-koefiziente partzialaren bitartez egiten da. Adina x, absentismoa z eta ekoizpena y izanik, honela kalkulatzen da x eta y aldagaien arteko korrelazio-koefiziente partziala, z aldagairen efektua baztertuz:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{1 - r_{xz}^2}\sqrt{1 - r_{yz}^2}}$$



Irudia 9.6: Absentismoaren tarteko eraginagatik (goiko geziak), adinaren eta ekoizpenaren arteko korrelazioa negatibo suertatzen da (beheko gezia), aldez aurretik uste denaren bestera, adinak lanpostuan esperientzia handiagoa dakarrelako.



Irudia 9.7: Oztopoa gainditzeko, absentismoaren eragina baztertu behar da (goiko geziak), korrelazio partzialeko koefiziente baten bitartez (beheko gezia), orokorrean korrelazio positiboa emango duena, uste denarekin bat.

9.10 Sasiko korrelazioa

Sasiko korrelazioa edo **korrelazio faltsua** (ingelesez, *spurious correlation*; gazteleraz, *correlación espúrea*) elkarrekin zerikusi handirik ez duten aldagaiak uztartzean gerta daiteke. Adibidez, biztanleria nahiko konstantea duen herrialde bateko urteko ogi-kontsumoaren eta hurrengo urteko jaiotza-kopuruaren arteko korrelazio-koefizientea kalkulaturik, bi aldagaien arteko korrelazio-koefizientea positibo eta sendoa suertaturik, ezin da ondorioztatu, ogia afrodisiakoa denik edota bikoteen ugalkortasuna gehitzen duenik, bi aldagaiak horrela lotzeko fundamenturik ez dagoelako; bi aldagai horien arteko korrelazioa sasikoa dela esango dugu orduan.

Beraz, aldagaien arteko korrelazioa aztertzerakoan, aldagaien arteko lotura aurrez teorikoki justifikatuta egon behar da, emaitzak adierazgarriak izango badira.

Honen guztiaren ondorio gisa, estatistikan maiz errepikatzen den leloa aipatu behar dugu: *korrelazioak ez du kausazioa inplikatzen* (ingelesez, *correlation doesn't imply causation*).

9.11 Ariketak

1. Ikasle batzuen gela eta kalifikazioak jaso dira:

Gela	A	A	A	A	B	B	B	B	C	C	C
Nota	8	5	6	9	6	6	7	7	5	6	5

   Bi aldagaien arteko erlazio estatistikoa aztertu, eta korrelazio-ratioaren kalkuluarekin eta interpretazioarekin bukatuz.

2. Ikasle zenbaiten gainean, matematika nota, bigarren hizkuntzako nota, 200 m korritzeko behar duten denbora (segundutan) eta egunean zehar telefono mugikorrarekin zenbat denbora ibili diren (minututan):

Matematika	5.2	5.2	8.3	8.9	4.3	7.4	7.8	9.0	5.6
Hizkuntza	3.4	5.4	7.8	9.2	2.1	6.8	8.3	8.7	4.5
Lasterketa	36	77	58	42	91	74	24	35	56
Telefonoa	45	54	12	15	66	28	22	10	39

   Matematika notaren eta beste hiru aldagaien arteko puntu hodeiak marraztu eta interpretatu behar dira.

3. Enpresa zenbaitetan I+G alorrean izandako gastu eta mozkin totalen portzentajea jaso da fakturazioari buruz:

I+G	2	3	1	4	6
Mozkinak	15	20	12	24	22

   (a) Bi aldagaien arteko kobariantza kalkulatu eta interpretatu, formula erosoa nahiz jatorrizkoa baliatuz.

   (b) Bi aldagaien arteko korrelazio-koefiziente lineala kalkulatu eta interpretatu. Esan al daiteke ziurtasun handiz I+G gastuak mozkinak gehitzen dituela?

4. Enpresa batean asteko ekoizpenari eta unitate kostuari buruzko datuak jaso dira aste zenbaitetan zehar:

Astea	1	2	3	4	5	6	7	8	9	10	11	12
Ekoizpena	10	12	15	18	22	27	33	39	42	48	56	68
Unitateko kostua	99	58	48	41	38	36	35	35	34	34	34	34

   (a) Kobariantza kalkulatu eta interpretatu.

   (b) Korrelazio-koefiziente lineala kalkulatu eta interpretatu.

   (c) Bi aldagaien artean nolako erlazio estatistikoa dago: lineala ala lerromakurra? Puntu hodeia marraztu galdera erantzuteko.

   (d) Aldagai batek gora egiten duenean, besteak zenbateraino egiten duen behera neurtu ezazu koefiziente egokia kalkulatuz, eta aurreko emaitzekin alderatu.

5. Ikasle zenbaiten sexua kontuan harturik, froga batean izandako kalifikazioak jaso dira:

Sexua	g	e	g	e	g	e	e
Kalifikazioa	6.7	8.5	6.9	7.2	8.0	8.6	9.2

   Korrelazio koefiziente lineala kalkulatu eta interpretatu. Zenbateraino du eragina sexuak kalifikazioetan?

6. Test bateko puntuazio totala eta galdera jakin bat ongi erantzun duten galdetu zaizkie ikasle zenbaiti (galdera, z: zuzen, o: oker):

Puntuazioa	87	46	65	72	82	61	39	42
Galdera	z	o	z	o	o	z	z	z

   Galdera hori baztertzea komenigarria den azter ezazu, item-test korrelazioaren bitartez.

7. Irakasle batek 5 galderako test bat prestatu du ikasgai baten ulermena ebaluatzeko. Honako hauek dira erantzunak eta puntuazio totalak:

	Ikasleak							
Item zenbakia	A	B	C	D	E	F	G	H
1. itema	1	1	1	1	1	1	1	0
2. itema	1	1	1	1	1	0	1	0
3. itema	1	1	1	1	0	0	1	0
4. itema	1	1	0	1	0	0	1	0
5. itema	1	1	0	0	0	0	0	0
Puntuazio totala	5	5	3	4	2	1	4	0

   Kalkulatu Cronbach-en alpha eta interpreta ezazu

8. Matematikako 10 itemeko test batean, honako hauek dira ikasleek lorturiko puntuazio totalak:

Ikaslea	A	B	C	D	E	F	G	H	I	J
Puntuazioa	9	9	8	7	7	6	6	5	5	4

Item-test korrelazioak eta itemeko bariantzak ere kalkulatu dira (0: itema gaizki erantzun, 1: itema ongi erantzun), lehenengo itemerako ezik:

Item zenbakia	1	2	3	4	5	6	7	8	9	10
$r(item, puntuazioa)$	?	-0.21	0.26	0.53	0.64	0.38	0.65	0.59	0.74	0.66
s_i^2	?	0.24	0.25	0.21	0.21	0.24	0.16	0.25	0.24	0.21

Lehenengo itemean ikasleek izan dituzten puntuazioak hauek dira:

Ikaslea	A	B	C	D	E	F	G	H	I	J
Puntuazioa	0	0	0	1	0	1	0	1	1	1

(a) Kalkulatu Cronbach-en alpha eta interpreta ezazu.

(b) Testaren barne koherentzia gehitzeko, zer egingo zenuke?

9. Botika berri bat asmatu ondoren, urtebetez frogatu zen gaixotasun bat pairatzen zuten pertsonen zenbaiten artean. Beste batzuek ohiko tratamenduarekin jarraitu zuten. Urtebete pasa ondoren, gaixotasun-egoera arindu den jaso zen. Emaitzak honako hauek dira:

Botika berria?	b	b	b	b	b	e	e	e	e	e
Gaixotasuna arindu?	b	e	b	b	b	e	e	e	e	b

Bi aldagaien arteko korrelazio koefiziente lineala baliatuz, erabaki botika eraginkorra izan daitekeen.

10. Test bat gainditu duten ala ez eta testeko galdera jakin bat ongi erantzun duten ala ez jaso da azterketa bat egin duten pertsona zenbaiten artean. Emaitzak hauek dira:

	Galdera ongi erantzun da?		Guztira
Testa gainditu?	Bai	Ez	
Bai	36	16	52
Ez	12	54	66
Guztira	48	70	118

Bi aldagaien arteko korrelazio koefiziente lineala baliatuz, azter ezazu galderaren egokitasuna testaren kalifikazioa zehazterakoan.

11. Administrazio publikoko oposizioetan lehiakideen artean, sexua (g:gizon, 0; e:emakume, 1), maila ekonomikoa (b:baxua, 0; a:altua, 1), lorturiko nota eta oposizioa prestatzeko bi urteko prestakuntza intentsiboa eskaintzen duen akademia batera (bai, 1; ez, 0) joan diren jaso dira eta aldagai guztien arteko korrelazio-koefiziente linealak kalkulatu:

r_{xy}	Sexua	Maila ekon.	Nota	Akademia
Sexua	1	-0.258	-0.060	0
Maila ekon.	-0.258	1	0.669	0.774
Nota	-0.060	0.669	1	0.904
Akademia	0	0.774	0.904	1

   (a) Azaldu aurreko korrelazio-matrizearen ezaugarriak.

   (b) Aldagai horien arteko korrelazioari buruz hipotesiak planteatu, eta hasieran batean itxuraz betetzen ez diren kasuetan, korrelazio partzialaren kontzeptua baliatu hipotesia egiaztatzeko.

12. 25-35 urte bitarteko gazte batzuen errentak eta non bizi diren, alokatutako etxebizitza batean (A), etxebizitza propio batean (P) edo gurasoen etxean (G) jaso dira. Hona datuak:

Errenta	16	17	19	21	25	28	30	32	34	36	39
	40	41	43								
Etxea	A	G	G	G	A	A	A	G	P	P	P
	A	P	P								

   (a) Errenta eta etxebizitzaren arteko korrelazioa esploratu ezazu, datuak laburbilduz eta diagrama egoki batez.

   (b) Korrelazioaren sendotasuna aztertu, errentaren eta etxebizitza kategoria bakoitzeko maiztasun erlatiboaren arteko korrelazio-koefiziente lineala kalkulatuz.

   (c) Korrelazioaren sendotasuna aztertu, eta (η) kalkulatuz.

Ebazpenak

(1) ariketa

1. Aldagai dependentea eta independentea bereizi:

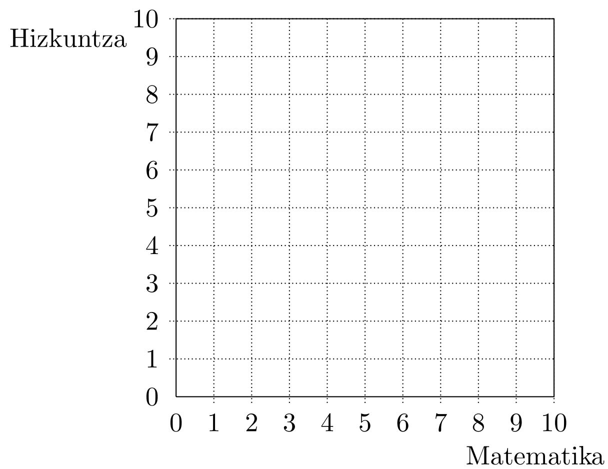
2. Datuak bereizi aldagai independentearen arabera:

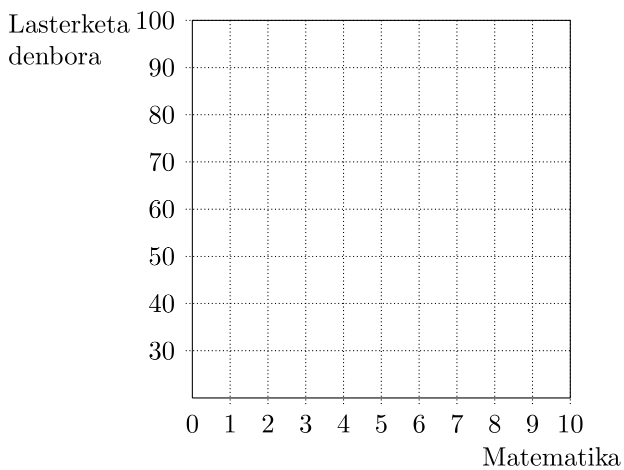
3. Datu-azpimultzoak aztertu bakoitza bere aldetik:

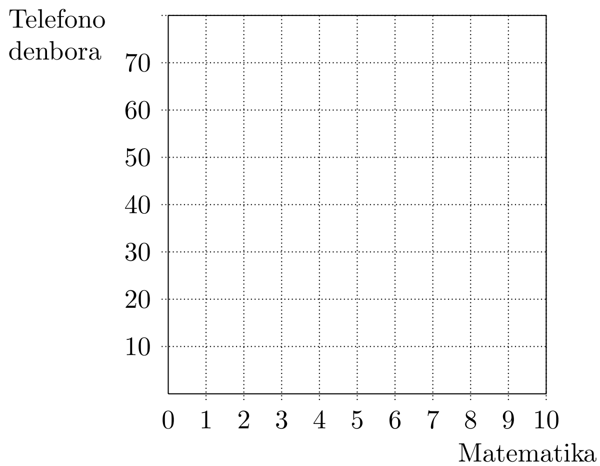
4. Norabidea aztertu:

5. Sendotasuna aztertu (eta kalkulatu):

(2) ariketa


Hizkuntza
10
9
8
7
6
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9 10
Matematika


Lasterketa denbora
100
90
80
70
60
50
40
30
0 1 2 3 4 5 6 7 8 9 10
Matematika



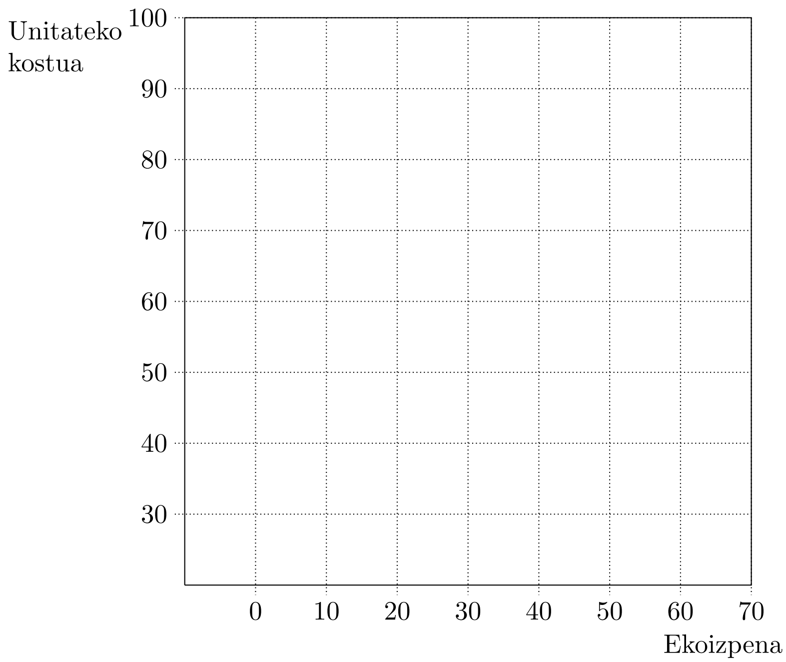
(3) ariketa

x_i (I+G)	y_i (Mozkinak)	$x_i y_i$	x_i^2	y_i^2
2	15			
3	20			
1	12			
4	24			
6	22			

x_i	y_i	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$(x_i - \overline{x}) \times (y_i - \overline{y})$
2	15			
3	20			
1	12			
4	24			
6	22			

(4) ariketa

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
10	99			
12	58			
15	48			
18	41			
22	38			
27	36			
33	35			
39	35			
42	34			
48	34			
56	34			
68	34			


Unitateko kostua
100
90
80
70
60
50
40
30
0
10
20
30
40
50
60
70
Ekoizpena

(4d)

x_i	y_i	$m(x_i)$	$m(y_i)$	$m(x_i)m(y_i)$	$m(x_i)^2$	$m(y_i)^2$
10	99					
12	58					
15	48					
18	41					
22	38					
27	36					
33	35					
39	35					
42	34					
48	34					
56	34					
68	34					
		78	78		650	644.5

(5) ariketa

Kodifikazioa:

x_i	y_i			
	6.7			
	8.5			
	6.9			
	7.2			
	8.0			
	8.6			
	9.2			

(6) ariketa

Kodifikazioa:

x_i	y_i			

(7) ariketa

x_1	x_2	x_3	x_4	x_5	x_T	x_1^2	x_2^2	x_3^2	x_4^2	x_5^2	x_T^2

$$s_x^2 = \frac{\sum x^2}{n} - \overline{x}^2$$

- $s_1^2 =$
- $s_2^2 =$
- $s_3^2 =$
- $s_4^2 =$
- $s_5^2 =$
- $s_T^2 =$

$$\alpha = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum_{i=1}^{k} S_i^2}{S_t^2}\right] =$$

(8) ariketa

(a)

x_1	x_1^2	x_T	x_T^2

$$\alpha = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum_{i=1}^{k} S_i^2}{S_t^2}\right] =$$

(b)

x_1	x_T	$x_1 x_T$	x_1^2
0	9		
0	9		
0	8		
1	7		
0	7		
1	6		
0	6		
1	5		
1	5		
1	4		

x_T	x_T^2
9	
9	
8	
7	
7	
6	
6	
5	
5	
4	

$$\alpha = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum_{i=1}^{k} S_i^2}{S_t^2}\right] =$$

(9) ariketa

Kodifikazioa:

x_i	y_i			

(10) ariketa

Kodifikazioa:

x_i	y_i				

(11) ariketa

(a)

Ezaugarriak:

-
-

(b)

HIPOTESIA	DATUEK ERAKUTSIA	INTERP. ETA ZER EGIN

(12) ariketa

(a) Aldagai bat kuantitatiboa denez, eta bestea kualitatiboa eta ez dikotomikoa, korrelazioa aldagai independentea eta dependentea bereiziz egiten da, datuak aldagai independentearen arabera bereiziz, eta aldagai dependentearen datu azpimultzoak bakoitza bere aldetik aztertuz.
15-25: ....
25-35: ....
35-45: ....

Etxe-erregimena	% (15-25)	% (25-35)	% (35-45)
Gurasoen etxean			
Alokairuan			
Etxe propioan			

....
....
....
....
....
....

(b)

Errenta (x)	% (gurasoen etxean) (y)	xy	x^2	y^2
$15-25 \rightarrow 20$			400	
$25-35 \rightarrow 30$			900	
$35-45 \rightarrow 40$			1600	
90			2900	

$$\overline{x} = \frac{90}{3} = 30 \;\; ; \;\; \overline{y} = \frac{\quad}{3} =$$

$$s_{xy} = \frac{\sum xy}{n} - \overline{x}\overline{y} =$$

$$s_x = \sqrt{\frac{2900}{3} - 30^2} = 8.16; \quad s_y = \sqrt{\frac{\qquad}{\qquad} - \qquad} =$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\qquad}{\qquad} =$$

Errenta (x)	% (alokairuan) (y)	xy	x^2	y^2
$15-25 \to 20$			400	
$25-35 \to 30$			900	
$35-45 \to 40$			1600	
90			2900	

$$\overline{x} = \frac{90}{3} = 30 \;\; ; \;\; \overline{y} = \frac{}{3} =$$

$$s_{xy} =$$

$$s_x = 8.16; \quad s_y = \sqrt{\frac{}{} - \quad} =$$

$$r_{xy} = \frac{}{} =$$

Errenta (x)	% (etxe propioa) (y)	xy	x^2	y^2
$15-25 \to 20$			400	
$25-35 \to 30$			900	
$35-45 \to 40$			1600	
90			2900	

$$\overline{x} = \frac{90}{3} = 30 \;\; ; \;\; \overline{y} = \frac{}{3} =$$

$$s_{xy} =$$

$$s_x = 8.16; \quad s_y = \sqrt{\frac{}{} - \quad} =$$

$$r_{xy} = \frac{}{} =$$

Laburpen-taula:

Erregimena	$r_{errenta, \% \ erregimena}$
Gurasoenean	-0.96
Alokairuan	-0.11
Etxe propioan	0.96

Esplorazioko emaitzak berretsita geratzen dira: zenbat eta errenta handiagoa, gurasoen etxean bizi direnak urriago ($r_{xy} < 0$) eta etxe propioa dutenak ugariago ($r_{xy} < 0$). Emaitza hauekin gainera, korrelazio horiek sendoak direla dakigu: ($|r_{xy}| = 0.96 > 0.6$). Alokairuan bizi direnekiko korrelazioa berriz, oso ahula da ($|r_{xy}| = 0.11 < 0.3$) eta beraz, ez dago loturarik errentaren eta alokairuan bizi direnen kopuruaren artean.

(c)

Etha kalkulatzeko, datuak aldagai kualitatiboaren, eta beraz etxe erregimenaren arabera bereizi behar dira:

Gurasoen etxean: ....
Alokairuan: ....
Etxe propioan: ....

Erregimena	n	$\overline{x}$
Gurasoenean		
Alokairuan		
Etxe propioan		

$$\eta = \frac{\qquad\qquad\qquad\qquad\qquad\qquad}{} = \frac{}{1123} =$$

....
....
....