6. gaiari buruzko autoebaluazioaren ebazpena

1. Birus baten bizitza aztertzeko asmoz, laborategi batean 300 ale aztertu ziren, bizitza aldia (ordutan) ondoko moduan banatuta egonik:

Bizitza-aldia	1.0	3.0	4.5	5.5	7.0	9.0
Portzentajea	%8	%25	%20	%20	%20	%7

a) Eraiki ezazu maiztasun-taula.

b) Kalkula itzazu lortatutako bizitza-aldien joera-zentraleko estatistikoak eta CV aldakuntza-koefizientea.

c) 15 ale *bizitza luzekoak* badira, zein izan da bizitza-aldi horretarako ordu-kopuru minimoa?

d) Irudika ezazu kutxa-diagrama. Zer esan daiteke outlierrei buruz?

Ebazpena:

a) Izan bedi $X =$ '*birusaren bizitza-aldia (ordutan)*' zorizko aldagaia. Emandako datuak x_i balioak eta $100 \cdot h_i$ portzentajeak direnez eta kontuan hartuta laginaren tamaina $n = 300$ alekoa dela, maiztasun absolutuen zutabea eraiki dezakegu, izan ere, $f_i = n \cdot h_i$ eta horrela, maiztasun-taula ondokoa da:

Taula 0.i. Maiztasun-taula.

x_i	f_i	F_i	h_i	H_i	$x_i f_i$	$x_i^2 f_i$
1.0	24	24	0.08	0.08	24	24
3.0	75	99	0.25	0.33	225	675
4.5	60	159	0.20	0.53	270	1215
5.5	60	219	0.20	0.73	330	1815
7.0	60	279	0.20	0.93	420	2940
9.0	21	300	0.07	1.00	189	1701
$\sum$	300		1		1458	8370

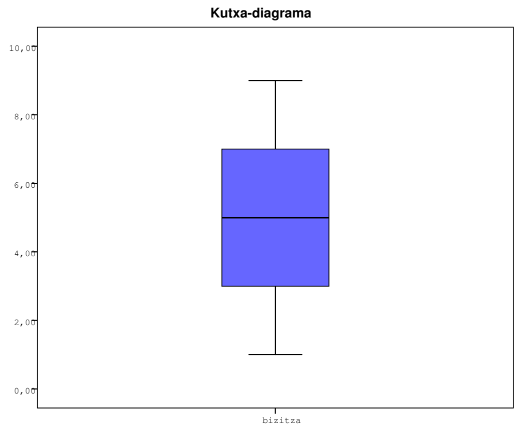
Taula 0.i : Maiztasun-taula.

b) Joera-zentraleko estatistikoak: batezbestekoa, mediana eta moda dira. Batezbestekoa kalkulatuko dugu:
$\overline{x} = \frac{\sum_{i=1}^{300} x_i}{300} = \frac{\sum_{i=1}^{6} x_i f_i}{300} = \frac{1458}{300} = 4.86$ ordu.
Mediana kalkulatzeko: $H_2 < 0.50 < H_3 \Rightarrow Me = x_3 = 4.5$ ordu.
Azkenik, moda: $max f_i = f_2 = 75 \Rightarrow Mo = 3.0$ ordu.
Gainera, $CV = \frac{s_n}{\overline{x}} \cdot 100$, hau da, s_n desbideratze estandarra behar dugu, $s_n =$

$\sqrt{\frac{\sum_{i=1}^{300} x_i^2}{300} - (\overline{x})^2} = \sqrt{\frac{\sum_{i=1}^{6} x_i^2 f_i}{300} - (\overline{x})^2} = \sqrt{\frac{8370}{300} - 4.86^2} = 2.0689$ ordu. Hortaz, $CV = \frac{2.0689}{4.86} \cdot 100 = \%42.5702$ aldakuntza-koefizientea da.

c) 15 ale %5 portzentajeari dagokionez (hots, $\frac{15}{300} = 0.05$) eta bizitza luzeenari buruz galdetzen dutenez, 95. pertzentila kalkulatu behar dugu, $H_5 < 0.95 < H_6$ denez, $p_{95} = 9.0$ ordu *bizitza-aldi luzeen* minimoa da.

d) Kutxa-diagrama eraikitzeko p_{25}, p_{50} eta p_{75} kalkulatu behar ditugu. Maiztasun-taulari begiratuz, argi dago $p_{25} = 3.0$, $p_{50} = Me = 4.5$ eta $p_{75} = 7.0$. Hortaz, barne hesiak: $m_1 = 3.0 - 1.5(7.0 - 3.0) = 3.0 - 6.0 = -3.0$ eta $m_3 = 7.0 + 1.5(7.0 - 3.0) = 7.0 + 6.0 = 13.0$ eta kanpo hesiak: $M_1 = 3.0 - 3.0(7.0 - 3.0) = 3.0 - 12.0 = -9.0$ eta $M_3 = 7.0 + 3.0(7.0 - 3.0) = 7.0 + 12.0 = 19.0$. Beraz, ez dago balio arrarorik edo outlierrik.



2. Ondoko zurtoin eta hosto grafikoak, 0 eta 100 artean balioak hartzen dituen test baten puntuazioa adierazten du.

3 | 1
4 | 112
5 | 0035
6 | 0000144
7 |
8 | 1149
9 | 3

a) Eraiki ezazu maiztasun-taula eta irudika ezazu barra grafikoa.

b) Kalkula itzazu joera zentraleko estatistikoak. Banaketaren balio maximoa 5 puntu gehiago izatekotan, zelan aldatuko zen mediana? Azal ezazu erantzuna.

c) Kalkula itzazu desbideratze estandarra eta aldakuntza-koefizientea.

d) Nota handienen %15-en artean zein da puntuzioa minimoa?

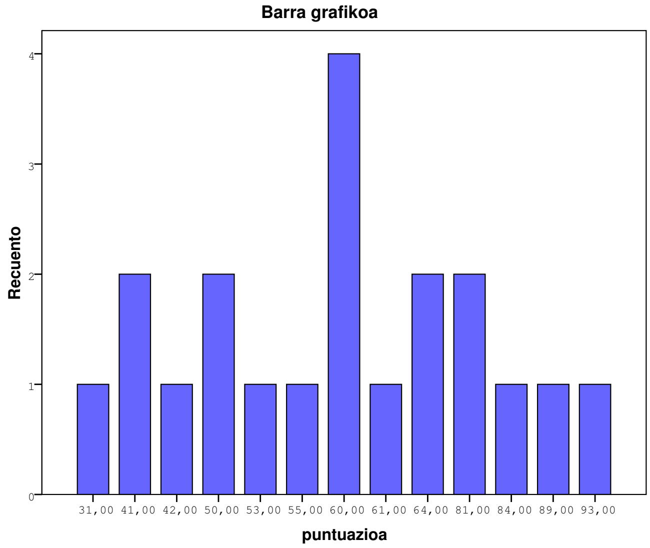
Ebazpena:

a) Izan bedi $X =$ '*testaren puntuazioa*' zorizko aldagaia. Emandako zurtoin eta hosto grafikoan banaketaren itxura ikusteaz aparte, balio guztien zerrenda ordenatua dugu. Izan ere, x_i balioak eta f_i maiztasun absolutuak agertzen dira, horrela, maiztasun-taula eraiki dezakegu ondoko taulan adierazten den moduan.

Taula 0.ii. Maiztasun-taula.

x_i	f_i	F_i	h_i	H_i	$x_i f_i$	$x_i^2 f_i$
31	1	1	0.05	0.05	31	961
41	2	3	0.10	0.15	82	3362
42	1	4	0.05	0.20	42	1764
50	2	6	0.10	0.30	100	5000
53	1	7	0.05	0.35	53	2809
55	1	8	0.05	0.40	55	3025
60	4	12	0.20	0.60	240	14400
61	1	13	0.05	0.65	61	3721
64	2	15	0.10	0.75	128	8192
81	2	17	0.10	0.85	162	13122
84	1	18	0.05	0.90	84	7056
89	1	19	0.05	0.95	89	7921
93	1	20	0.05	1.00	93	8649
$\sum$	20		1		1220	79982

Taula 0.ii : Maiztasun-taula. Barra grafikoa x_i versus f_i (edo h_i) adierazten duen ondoko grafikoa da.



0.1. irudia. 2. a) atalaren barra-grafikoa.
0.1 irudia: 2. a) atalaren barra-grafikoa

b) Joera-zentraleko estatistikoak: batezbestekoa, mediana eta moda dira. Batezbestekoa kalkulatzeko: $\overline{x} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{\sum_{i=1}^{13} x_i f_i}{20} = \frac{1220}{20} = 61$ puntu. Mediana kalkulatzeko: $H_6 < 0.50 < H_7 \Rightarrow Me = x_7 = 60$ puntu. Azkenik, moda: $max f_i = f_7 = 4 \Rightarrow Mo = 60$ puntu. Banaketaren balio maximoa 5 puntu gehiago izatekotan, mediana ez zen aldatuko, berriro $H_6 < 0.50 < H_7 \Rightarrow Me = x_7 = 60$ puntu izango genukeelako.

c) $s_n = \sqrt{\frac{\sum_{i=1}^{20} x_i^2}{20} - (\overline{x})^2} = \sqrt{\frac{\sum_{i=1}^{13} x_i^2 f_i}{20} - (\overline{x})^2} = \sqrt{\frac{79982}{20} - 61^2} = 16.6763$ puntutako desbideratze estandarra denez, $CV = \frac{s_n}{\overline{x}} \cdot 100 = \frac{16.6763}{61} \cdot 100 = \%27.3382$ aldakuntza-koefizientea da.

d) 85. pertzentila kalkulatu behar dugu, $H_{10} = 0.85$ denez, $p_{85} = \frac{x_{10}+x_{11}}{2} = \frac{81+84}{2} = 82.5$ puntu da puntuazioa minimoa.

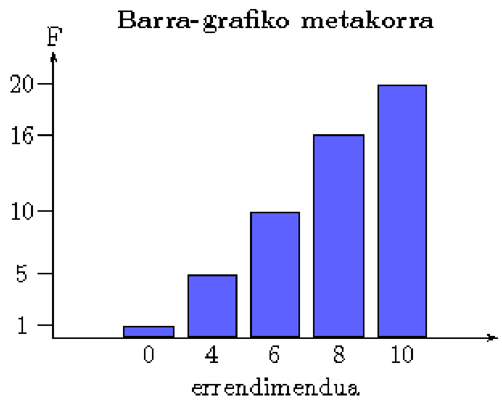
3. Proba batzuen bidez 20 ikasleren adimen-koefizienteak neurtu dira, lortutako emaitzak 76, 88, 96, 100, 108 eta 124 dira eta balio bakoitzari dagokion maiztasun metatu erlatiboa 0.15, 0.15, 0.55, 0.80, 0.95 eta 1.00 dira, hurrenez hurren.

a) Eraiki ezazu adimen-koefizienteen maiztasun-taula estatistikoa.

b) Zein da maiztasun handieneko adimen-koefizientea?

c) Zein pertzentilari dagokio 98 adimen-koefizientea?

d) Ikasleren talde berberari errendimendu-proba bat egiten zaio, emaitzak ondoko grafiko metakorran adierazirik. Eraiki ezazu errendimendu-proben maiztasun-taula estatistikoa eta kalkula ezazu mediana.



0.2. irudia. 3. d) atalaren barra-grafiko metakorra.

0.2 irudia: 3. d) atalaren barra-grafiko metakorra

e) Zeintzuk dira datu sakabanatuenak, adimen-koefizienteak edo errendimendu-puntuazioak?

Ebazpena:

a) Izan bedi $X =$ '*adimen-koefizientea*' izeneko z.a. Emandako datuak x_i balioak eta H_i maiztasun metatu erlatiboak direnez eta kontuan hartuta laginaren tamaina $n = 20$ ikaslerena dela, maiztasun erlatiboen eta maiztasun metatuen zutabeak eraiki ditzakegu, izan ere, $F_i = n \cdot H_i$ eta horrela, maiztasun-taula ondokoa da:

Taula 0.iii. Maiztasun-taula.

x_i	f_i	F_i	h_i	H_i	$x_i f_i$	$x_i^2 f_i$
76	3	3	0.15	0.15	228	17328
88	0	3	0.00	0.15	0	0
96	8	11	0.40	0.55	768	73728
100	5	16	0.25	0.80	500	50000
108	3	19	0.15	0.95	324	34992
124	1	20	0.05	1.00	124	15376
$\sum$	20		1		1944	191424

Taula 0.iii : Maiztasun-taula.

b) Maiztasun handiena $f_3 = 8$ da eta dagokion adimen-koefizientea $x_3 = 96$ da.

c) Baldin $p_j = 98$ bada, balioen zutabeari begiratuz 98 balioa aurkitzen ez dugunez, $p_j = \frac{96+100}{2} = \frac{x_3+x_4}{2}$, beraz $H_3 = 0.40 = j/100$. pertzentilari dagokio, hau da, p_{40} edo %40 pertzentilari dagokio 98 adimen-koefizientea.

d) Izan bedi $Y =$ '*errendimendu-puntuazioa*' izeneko zorizko aldagaia. Maiztasun-taula estatistikoa grafiko metakorraren datuak erabiliz eraikiko dugu. Izan ere, grafiko

Taula 0.iv. Maiztasun-taula.

y_i	f_i	F_i	h_i	H_i	$y_i f_i$	$y_i^2 f_i$
0	1	1	0.05	0.05	0	0
4	4	5	0.20	0.25	16	64
6	5	10	0.25	0.50	30	180
8	6	16	0.30	0.80	48	384
10	4	20	0.20	1.00	40	400
$\sum$	20		1		134	1028

horren OX ardatzean aldagaiaren balioak agertzen dira, y_i, eta OY ardatzean, berriz, maiztasun absolutuak, F_i. Horrela,
Taula 0.iv : Maiztasun-taula. Azkenik, mediana kalkulatu dugu, $H_3 = 0.50$ denez, $Me = \frac{x_3+x_4}{2} = \frac{6+8}{2} = 7$ puntu. Grafikoan ikusten denez $Me = 7$ puntuazioaren ezkerraldean behaketen %50 geratzen da, hau da, 10 ikasle.

e) Aldagai desberdinak direnez, sakabanatuena aurkitzeko aldakuntza-koefizienteen konparaketaz baliatuko gara. Horretarako,
$\overline{x} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{\sum_{i=1}^{6} x_i f_i}{20} = \frac{1944}{20} = 97.2,$
$s_{n,x}^2 = \frac{\sum_{i=1}^{6} x_i^2 f_i}{20} - (\overline{x})^2 = \frac{191424}{20} - 97.2^2 = 123.36 \Rightarrow$
$s_{n,x} = 11.1068$ eta
$\overline{y} = \frac{\sum_{i=1}^{5} y_i f_i}{20} = \frac{134}{20} = 6.7,$
$s_{n,y}^2 = \frac{\sum_{i=1}^{5} y_i^2 f_i}{20} - (\overline{y})^2 = \frac{1028}{20} - 6.7^2 = 6.51 \Rightarrow$
$s_{n,y} = 2.5515$. Beraz,
$CV_X = \frac{s_{n,x}}{\overline{x}} \cdot 100 = 11.4267\% < CV_y = \frac{s_{n,y}}{\overline{y}} \cdot 100 = 38.0816\%$
denez, errendimendu-puntuazioa adimen-koefizientea baino sakabanatuagoa da.

4. Populazio batean ur-arazoa aztertu nahi dute. Horretarako 11 egunetan jasotako euria neurtu zen (l/m^2-tan), hurrengo emaitzak lorturik:

Euria	10	12	14	16	18	20	22
Egun-kopurua	1	2	1	3	1	2	1

a) Irudikatu zurtoin eta hosto grafikoa. Kalkula ezazu ezagutzen dituzun joera-zentraleko estatistiko guztiak. Zer esan daiteke banaketaren itxurari buruz?

b) Zein balioren artean egongo da egun euritsuenetariko %20a?

c) Estima ezazu populazio-batezbestekoa eta populazio-bariantza, puntu-estimazioak erabiliz.

d) Demagun jasotako ur kantitatea banaketa normalari darraiola. Laginean oinarrituta, ondoriozta ezazu ea %95-eko konfiantza-mailaz populazio-batezbestekoa 15 l/m^2 izatea posiblea den.

e) Laginean oinarrituta, ondoriozta ezazu ea %95-eko konfiantza-mailaz populazio-desbidazio estandarra 5 $\mathrm{l/m^2}$ baino txikiagoa izatea posiblea den.

f) Zein da populazio-batezbestekoa lagin-batezbestekotik goitik edo behetik gehienez 0.05 $\mathrm{l/m^2}$-tan desberdina izateko probabilitatea?

g) Benetako batezbestekoan zentratutako zein tartetan egongo da lagin-batezbestekoa, aldien %99etan?

Ebazpena:

a) Izan bedi $X =$ '*jasotako ur kantitatea*' zorizko aldagaia, laginaren tamaina $n = 11$ egun delarik.

$$\begin{array}{l|l} \mathbf{1} & 0 \\ \mathbf{1} & 22 \\ \mathbf{1} & 4 \\ \mathbf{1} & 666 \\ \mathbf{1} & 8 \\ \mathbf{2} & 00 \\ \mathbf{2} & 2 \end{array}$$

Joera-zentraleko estatistikoak: batezbestekoa, mediana eta moda kalkulatu ditugu. Batezbestekoa: $\overline{x} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{\sum_{i=1}^{7} x_i f_i}{11} = \frac{176}{11} = 16$ $\mathrm{l/m^2}$. Mediana: $H_3 < 0.50 < H_4 \Rightarrow Me = x_4 = 16$ $\mathrm{l/m^2}$. Eta moda: $max f_i = f_4 = 3 \Rightarrow Mo = x_4 = 16$ $\mathrm{l/m^2}$. Beraz, banaketa guztiz simetrikoa da, grafikoa ikusten denez eta $\overline{x} = Me = Mo$ izateagatik.

b) Egun euritsuenetariko %20a 80. pertzentila eta balioen maximoaren artean egongo da. $H_5 < 0.80 < H_6 \Rightarrow p_{80} = x_6 = 20$ $\mathrm{l/m^2}$. Beraz, 20 $\mathrm{l/m^2}$ eta 22 $\mathrm{l/m^2}$ balioen artean.

c) Ur kantitatearen populazio-batezbestekoaren puntu-estimazioa $\hat{\mu} = \overline{x} = 16$ $\mathrm{l/m^2}$-koa da eta populazio-bariantzaren puntu-estimazioa $\hat{\sigma}^2 = s_{n-1}^2 = s_n^2 \cdot \frac{11}{10} = \left(\frac{\sum_{i=1}^{11} x_i^2}{11} - (\overline{x})^2 \right) \cdot \frac{11}{10} = \left(\frac{2960}{11} - 16^2 \right) \frac{11}{10} = 13.0909 \frac{11}{10} = 14.4$ $\mathrm{l/m^2}$-koa da.

d) Kalkula dezagun %95-eko konfiantza-tartea populazio-batezbestekorako. Populazio normala, σ ezezaguna eta lagin tamaina txikia denez, $\varepsilon = t_{0.025;10} \frac{s_{n-1}}{\sqrt{11}} = 2.228 \frac{3.7947}{\sqrt{11}} = 2.5491$ eta

$$I_{\mu}^{0.95} = (16 \mp 2.5491) = (13.4508,\ 18.5492).$$

Beraz, 15 $\mathrm{l/m^2}$ izatea posiblea da, $15 \in I_{\mu}^{0.95}$ baita.

e) Kalkula dezagun %95-eko konfiantza-tartea populazio-desbideratze estandarrerako. Populazio normala denez,

$$I^{0.95}_{\sigma^2} = \left(\frac{10 s^2_{n-1}}{\chi^2_{0.025;10}}, \ \frac{10 s^2_{n-1}}{\chi^2_{0.975;10}} \right) = (\frac{10 \cdot 14.4}{20.4830}, \ \frac{10 \cdot 14.4}{3.247}) =$$

$$= (7.03, \ 44.35)$$

eta

$$I^{0.95}_{\sigma} = (2.6515, \ 6.6595).$$

Hortaz, ezin da ondorioztatu desbideratze estandarra 5 l/m^2 baino txikiagoa denik.

f) μ populazio-batezbestekoa $\overline{X}$ lagin-batezbestekotik goitik edo behetik gehienez 0.05 l/m^2-tan desberdina izatea $|\mu - \overline{X}| \leq 0.05$ adierazten da, beraz, gertaera honen probabilitatea eskatzen digutenez,

$$P(|\mu - \overline{X}| \leq 0.05) = P(-0.05 \leq \mu - \overline{X} \leq 0.05) =$$

$$= P(-0.05 \leq \overline{X} - \mu \leq 0.05) =$$

$$= P\Big(\frac{-0.05}{s_{n-1}/\sqrt{11}} \leq \frac{\overline{X} - \mu}{s_{n-1}/\sqrt{11}} \leq \frac{0.05}{s_{n-1}/\sqrt{11}} \Big) = \Big[\frac{\overline{X} - \mu}{s_{n-1}/\sqrt{n}} \approx t_{n-1} \Big] =$$

$$= P(-0.0437 \leq t_{10} \leq 0.0437) = [simetria] = 1 - 2P(t_{10} > 0.0437) =$$

$$= [interpolatuz] = 1 - 2 \cdot 0.4832 = 0.0336$$

g) Kalkulatu behar dugu e balioa non $P(|\mu - \overline{X}| \leq e) = 0.99$ betetzen den. Hau da,

$$0.99 = P(-e \leq \overline{X} - \mu \leq e) =$$

$$= P\Big(\frac{-e}{s_{n-1}/\sqrt{11}} \leq \frac{\overline{X} - \mu}{s_{n-1}/\sqrt{11}} \leq \frac{e}{s_{n-1}/\sqrt{11}} \Big) =$$

$$= P\Big(\frac{-e}{s_{n-1}/\sqrt{11}} \leq t_{10} \leq \frac{e}{s_{n-1}/\sqrt{11}} \Big) = 1 - 2P(t_{10} > \frac{e}{1.1441})$$

$$\Leftrightarrow P(t_{10} > \frac{e}{s_{n-1}/\sqrt{11}}) = 0.005 \Leftrightarrow t_{0.005;10} = \frac{e}{s_{n-1}/\sqrt{11}} \Leftrightarrow$$

$$e = t_{0.005;10} \cdot s_{n-1}/\sqrt{11} = 3.169 \cdot 1.1441 = 3.6257 l/m^2$$

5. Emisioak aztertzeko zentro batean aztertutako ibilgailu guztien %60-k azterketa gainditzen du. Demagun azterketa gainditzeko gertaerak askeak direla, kalkula itzazu ondoko probabilitateak:

a) Aztertutako hurrengo 5 ibilgailuek azterketa gainditzeko probabilitatea.

b) Aztertutako hurrengo 5 ibilgailuen artean, gutxienez batek azterketa ez gainditzeko probabilitatea.

c) Aztertutako hurrengo 5 ibilgailuen artean, zehazki bik azterketa gainditzeko probabilitatea.

d) Aztertutako hurrengo 5 ibilgailuen artean, gutxienez batek azterketa gainditzeko probabilitatea.

e) Egunero 50 ibilgailu aztertzen dira. Zein da erdiek baino gehiagok azterketa gainditzeko probabilitatea?

Ebazpena:

Izan bedi X ='*5 ibilgailuen artean, azterketa gainditzen duten kopurua*' zorizko aldagaia, $Bin(5,\ 0.6)$ banaketari darraiola argi dago.

a) $P(X=5) = \binom{5}{5} p^5 q^0 = 0.6^5 = 0.0778$

b) $P(X \leq 4) = 1 - P(X > 4) = 1 - P(X = 5) = 1 - 0.0778 = 0.9222$

c) $P(X=2) = \binom{5}{2} p^2 q^3 = 10 \cdot 0.6^2 \cdot 0.4^3 = 0.2304$

d) $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \binom{5}{0} p^0 q^5 = 1 - 0.4^5 = 0.9898$

e) Izan bedi Y ='*50 ibilgailuen artean, azterketa gainditzen duten kopurua*' zorizko aldagaia, $Bin(50,\ 0.6)$ banaketari darraiola argi dago. Eskatutako probabilitatea $P(Y > 25)$ kalkulatzeko ondoko hurbilketa erabiliko dugu:
$Bin(50,\ 0.6) \approx \mathcal{N}(50 \cdot 0.6,\ \sqrt{50 \cdot 0.6 \cdot 0.4}) = \mathcal{N}(30,\ 3.4641)$. Beraz, tipifikatuz, $P(Y > 25) = P(\frac{Y-30}{3.4641} > \frac{25-30}{3.4641}) \approx P(Z > -1.44) = 1 - P(Z > 1.44) = 1 - 0.0749 = 0.9251$

6. Kamioien motorraren pieza baten bizitza (ordutan) X zorizko aldagai jarraitua da, bere dentsitate funtzioa ondokoa delarik:

$$f(x) = \begin{cases} \frac{k}{x^2}, & \text{baldin } x \geq 1000 \\ 0, & \text{baldin } x < 1000 \end{cases}$$

a) Kalkula ezazu k-ren balioa eta $F(x)$ banaketa funtzioa.

b) Zein da motorrak gutxienez 2.000 ordu irauteko probabilitatea?

c) Zenbat izan behar da x_0, erosleak x_0 ordu baino lehen pieza aldatzeko beharrik ez izateko %70-eko probabilitatearekin?

Ebazpena:

a) Izan bedi X ='*piezaren bizitza (ordutan)*' z.a.j.
f funtzioak dentsitate-funtzioa izateko ondokoa bete behar du: i) $f(x) \geq 0$ eta ii) $1 = \int_{-\infty}^{\infty} f(x)\ dx$. Beraz, $k \geq 0$ eta $1 = \int_{-\infty}^{\infty} f(x)\ dx = \int_{-\infty}^{1000} 0\ dx + \int_{1000}^{\infty} \frac{k}{x^2}\ dx =$

$= -\frac{k}{x}]_{1000}^{\infty} = \frac{k}{1000} \Rightarrow k = 1000$
F banaketa-funtzioa: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt.$
• Baldin $x < 1000$, $F(x) = \int_{-\infty}^{x} 0\ dt = 0$
• Baldin $x \geq 1000$, $F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_{-\infty}^{1000} 0\ dt + \int_{1000}^{x} \frac{1000}{t^2}\ dt = -\frac{1000}{x}]_{1000}^{x} = 1 - \frac{1000}{x}$
Beraz,

$$F(x) = \begin{cases} 1 - \frac{1000}{x}, & \text{baldin } x \geq 1000 \\ 0, & \text{baldin } x < 1000 \end{cases}$$

b) Izan bedi $B =$ '*motorrak gutxienez 2.000 ordu irautea*' gertaera.

$P(B) = P(X \geq 2000) = \int_{2000}^{\infty} \frac{1000}{x^2}\ dx = -\frac{1000}{x}]_{2000}^{\infty} = 0.5$

c) Izan bedi D ondoko gertaera $D =$ '*erosleak x_0 ordu baino lehen pieza aldatzeko beharrik ez izatea*', hots, $D =$ '*piezaren bizitza x_0 ordu baino gehiago edo berdin izatea*'.

$0.70 = P(D) = P(X \geq x_0) = \int_{x_0}^{\infty} \frac{1000}{x^2}\ dx = -\frac{1000}{x}]_{x_0}^{\infty} = \frac{1000}{x_0} \Rightarrow$

$x_0 = 1428.5714$ ordu.

7. Tabakoari buruzko ikerketa baten arabera, zigarro baten zupada bakoitzean arnasten den nikotina banaketa normalari darraio. Bi markako zigarroen egindako aurreko ikerketak direla eta ondokoa ezaguna da: A markan behatutako X_A zupada bakoitzeko arnastutako nikotinaren itxaropen matematikoa 0.2 mg eta desbideratze estandarra 0.03 mg dira eta B markako kasuan, ordez, behatutako X_B nikotinaren itxaropen matematikoa 0.1 mg eta desbideratze estandarra 0.02 mg dira. Demagun bi marken arteko askatasuna. Erantzun itzazu ondoko galderak:

a) Zein da A markako hartutako lagin baten zupadaka nikotina 0.15 mg baino handiagoa izateko probabilitatea? Zein izango zen aurreko probabilitatea B markarako?

b) Defini dezagun $Y = \frac{1}{2}(X_A + X_B)$. Zein da aldagai berriaren banaketa? Kalkula itzazu banaketaren parametroak.

c) Zigarro marka bakoitzetik lagin bat kontsideratzen bada, kalkula ezazu lagin biren batezbestekoa 0.22 mg baino handiagoa izateko probabilitatea.

d) Kalkula ezazu A markako hartutako lagin baten eta B markako hartutakoaren arteko diferentzia 0.12 mg baino handiagoa izateko probabilitatea.

Ebazpena:

Izan bitez $X_A =$'A markako zigarro zupada bakoitzeko arnastutako nikotina' eta $X_B =$'B markako zigarro zupada bakoitzeko arnastutako nikotina' zorizko aldagaiak.
Dakigunez, $X_A : \mathcal{N}(0.2,\ 0.03)$ eta $X_B : \mathcal{N}(0.1,\ 0.02)$.

a) $P(X_A > 0.15) = [tipifikatuz] = P(\frac{X_A-0.2}{0.03} > \frac{0.15-0.2}{0.03}) = \quad P(Z > -1.67) = 1 - P(Z > 1.67) = 1 - 0.0475 = 0.9525$ eta

$$P(X_B > 0.15) = [tipifikatuz] = P(\frac{X_B-0.1}{0.02} > \frac{0.15-0.1}{0.02}) = \quad P(Z > 2.5) = 0.0062$$

b) Alde batetik, X_A eta X_B banaketa normal askeak direnez, orduan edozein konbinazio lineal, partikularki, Y aldagaia banaketa normalari darraio. Bestalde, $E(Y) = \frac{E(X_A)+E(X_B)}{2} = 0.15$ mg eta $Var(Y) = \frac{Var(X_A)+Var(X_B)}{4} = 0.000325$ mg ^{2}. Beraz,

$$Y : \mathcal{N}(0.15,\ 0.0180)$$

c) $P(Y > 0.22) = [tipifikatuz] = P(\frac{Y-0.15}{0.0018} > \frac{0.22-0.15}{0.0018}) = \quad P(Z > 3.89) = 0.0000481$

d) Alde batetik, X_A eta X_B banaketa normal askeak direnez, orduan edozein konbinazio lineal, partikularki, $X_A - X_B$ banaketa normalari darraio. Bestalde, $E(X_A - X_B) = E(X_A) - E(X_B) = 0.1$ mg eta $Var(X_A - X_B) = Var(X_A) + Var(X_B) = 0.0013$ mg^2. Beraz,

$$X_A - X_B : \mathcal{N}(0.1,\ 0.0361)$$

eta $P(X_A - X_B > 0.12) = [tipifikatuz] = P(\frac{X_A-X_B-0.1}{0.0361} > \frac{0.12-0.1}{0.0361}) = P(Z > 0.56) = 0.2877$

8. Adituen arabera, pertsona altuek baxuek baino energia gutxiago kontsumitzen dute. Esperimentu batean 4 pertsona altu eta 4 pertsona baxu kontsideratu ziren, bere beste ezaugarri guztiak oso antzekoak izanik (adina, sexua, pisua eta abar), eta jarduera fisiko sendoa egitean kontsumitutako energia (orduko kaloriatan) neurtu zen, ondoko taulan agertzen delarik:

Baxuak	1189	840	1020	980
Altuak	853	900	733	785

a) Kalkula ezazu, %90eko konfiantza-mailaz, batezbesteko energiaren diferentziarako estimatzaile-tartea.

b) Interpreta ezazu emaitza.

c) Estimazio-prozesua baliogarria izateko, zeintzuk dira baldintza beharrezkoak?

Ebazpena:

a) Izan bitez, $X_1 =$' baxuek kontsumitutako energia' eta $X_2 =$' altuek kontsumitutako' zorizko aldagaiak. Datuetan oinarrituta, $n_1 = 4,\ \overline{x}_1 = 1007.25, s_1 = 143.656; n_2 = 4,\ \overline{x}_2 = 817.75, s_2 = 73.627$. Lagin tamaina txikiak eta bariantzak ezegunak direnez, beraien arteko berdintasuna aztertuko dugu:

$$I^{0.90}_{\sigma_1^2/\sigma_2^2} = \left(\frac{s_1^2/s_2^2}{F_{0.05;3,3}},\ \frac{s_1^2/s_2^2}{F_{0.95;3,3}}\right) = \left(\frac{3.8069}{9.28},\ 3.8069 \cdot 9.28\right) =$$

$$= (0.4102,\ 35.3280)$$

Hortaz, $1 \in I^{0.90}_{\sigma_1^2/\sigma_2^2}$ dagoenez, $\sigma_1^2 = \sigma_2^2$, berdintasuna ezin da errefusatu %90eko konfiantza mailaz. Beraz, $I^{0.90}_{\mu_1-\mu_2} = (\overline{x}_1 - \overline{x}_2 \mp \varepsilon)$ non $\varepsilon = t_{0.05;6} s_p \sqrt{1/4 + 1/4} = 1.943 \cdot 114.1443/\sqrt{2} = 156.8238$ den. Izan ere, $s_p^2 = \frac{3s_1^2+3s_2^2}{6} = \frac{s_1^2+s_2^2}{2} = 13028.9115$ baita. Horrela,

$$I^{0.90}_{\mu_1-\mu_2} = (189.5 \mp 156.8238) = (32.6762,\ 346.3238)$$

b) %90eko konfiantza mailaz, bi multzoen benetako batezbesteko energien diferentzia aurreko tartearen barne dago. 0ren eskuinaldean dagoenez, $\mu_1 - \mu_2 > 0$, hau da, $\mu_1 > \mu_2$, jarduera fisiko sendoa egitean, pertsona baxuen kontsumitutako batezbesteko energia pertsona altuena baino handiagoa dela ondoriozta daiteke %90eko konfiantza mailaz, diferentzia 32 cal/h baino handiago eta 347 cal/h baino txikiago izanik.

c) Baldintza beharrezkoak: banaketa normalak eta askeak. Normaltasuna eta askatasuna hipotesi-kontrasteen bidez aztertu beharko genituzke.

9. Laborategi batean plastiko berri baten propietateak aztertzen ari direnez, plastiko berri eta zaharren artean distira konparatu nahi da. Demagun plastikoen distira banaketa normalari darraiola eta ondoko informazioa lortzen dela:

Plastiko berria	Plastiko zaharra
$n_1 = 16$	$n_2 = 25$
$\overline{x}_1 = 9.48$	$\overline{x}_2 = 9.46$
$s_{n-1,1} = 0.53$	$s_{n-1,2} = 0.25$

a) Kalkulatu %99-ko konfiantza mailako estimatzaile tartea plastiko berria eta zaharren distira aldagaiaren bariantzen zatidurarako.

b) Zeintzuk dira bete behar diren baldintzak aurreko tartea erabili ahal izateko?

c) Zer ondoriozta daiteke?

d) Kalkulatu %99-ko konfiantza mailako estimatzaile tartea plastiko berria eta zaharren batezbesteko distira aldagaiaren diferentziarako.

e) Zeintzuk dira bete behar diren baldintzak aurreko tartea erabili ahal izateko?

f) %99-ko konfiantza-mailaz, plastiko berria eta zaharren batezbesteko distira aldagaiaren arteko desberdintasunak adierazgarriak direla ondoriozta dezakegu?

Ebazpena:

a) Izan bitez X_1 = '*plastiko berriaren distira*' eta X_2 = '*plastiko zaharraren distira*' zorizko aldagaiak.

$$I^{0.99}_{\sigma_1^2/\sigma_2^2} = \left(\frac{s^2_{n-1,1}/s^2_{n-1,2}}{F_{0.005;15,24}},\ \frac{s^2_{n-1,1}/s^2_{n-1,2}}{F_{0.995;15,24}} \right) =$$

$$= \left(\frac{4.4944}{3.25},\ 4.4944 \cdot 3.79 \right) = (1.3829,\ 17.0338)$$

b) Baldintzak: Populazio normalak eta askeak.

c) Ondorioa: $I^{0.99}_{\sigma_1^2/\sigma_2^2} \subset (1,\ \infty) \Rightarrow \frac{\sigma_1^2}{\sigma_2^2} > 1 \Rightarrow \sigma_1^2 > \sigma_2^2$. Beraz, %99-eko konfiantza mailaz populazioen bariantzak desberdinak direla ondoriozta dezakegu. Are gehiago, plastiko berrien distiraren bariantza plastiko zaharrenena baino handiago da.

d) $I^{1-\alpha}_{\mu_1-\mu_2} = (\overline{x}_1 - \overline{x}_2 - \varepsilon,\ \overline{x}_1 - \overline{x}_2 + \varepsilon) = (0.02 \mp \varepsilon)$

$$\varepsilon = t_{\alpha/2;g} \cdot \sqrt{\frac{s^2_{n-1,1}}{n_1} + \frac{s^2_{n-1,2}}{n_2}}$$

$$g \approx 19 \Rightarrow t_{0.005;19} = 2.861 \Rightarrow \varepsilon = 0.4052$$

$$I^{0.99}_{\mu_1-\mu_2} = (-0.3852,\ 0.4252)$$

e) Baldintzak: Bariantza desberdineko eta ezezaguneko populazio normalak eta askeak eta gutxienez lagin tamaina bat *txikia* da (kasu honetan biak dira: $n_1 = 16 < 30$ eta $n_2 = 25 < 30$).

f) Ondorioa: %99-eko konfiantza-mailaz plastiko berri eta zaharren batezbesteko distiraren arteko desberdintasunak adierazgarriak ez direla ondoriozta dezakegu. Izan ere, $\mu_1 - \mu_2 = 0$ ezin da errefusatu, $0 \in I^{0.99}_{\mu_1-\mu_2}$ dagoelako, hau da, $\mu_1 = \mu_2$. Eta ezin da ondorioztatu bataren distira bestea baino altuagoa edo baxuagoa denik.

10. Enpresa batek ibai batean kutsatzen duen ala ez aztertzeko egindako ikerketan, bi motako neurketak konparatu ziren. Enpresatik urrun hartutako 733 ur-lagin artean, 13 kutsatuta zeuden. Enpresatik hurbil hartutako 742 ur-lagin artean, 29 kutsatuta zeuden.

a) Erabil itzazu datu hauek estimatzeko enpresatik urrun egondako uretan kutsadura egoteko probabilitatea %95-eko konfiantza mailaz.

b) Aurreko proportzioa estimatu nahi badugu %95-eko konfiantza-maila erabiliz, zenbat ur-lagin kontuan hartu behar izango ditugu proportzioaren estimatzailearen errorea %1 baino txikiagoa izateko? Konpara ezazu emaitza jatorrizko datuarekin eta azalpena eman.

c) Datuetan oinarrituta, enpresak ibaia kutsatzen duela ondoriozta daiteke? Hau da, enpresatik hurbil kutsadura altuagoa da enpresatik urrun baino? Erabil ezazu %95-eko konfiantza-tartea.

d) Ondorio berdina ateratzen al da %99-eko konfiantza-tartea erabiliz? Zein da konfiantza-tarte luzeagoa? Zergatik? Egiazta ezazu erantzuna.

Ebazpena:

Izan bitez X_1 = '*enpresatik urrun kutsatutako ur-ale kopurua*' eta X_2 = '*enpresatik hurbil kutsatutako ur-ale kopurua*' izeneko zorizko aldagaiak. Argi dago banaketa binomialak direla, p_1 enpresatik urrun kutsatutako ur-ale proportzioa eta p_2 enpresatik hurbil kutsatutako ur-ale proportzioa direlarik, bere estimazio puntualak $\hat{p}_1 = \frac{13}{733} = \%1.77$ eta $\hat{p}_2 = \frac{29}{742} = \%3.91$ izanik.

a) Kalkula dezagun %95eko konfiantza-tartea p_1 proportziorako. p_1 populazio-proportzioa $\hat{p}_1$ lagin-proportzioaren bidez estimatzen da, $\hat{p}_1 = \frac{13}{733} = 0.0177$. Laginaren tamaina $n_1 = 733$ ur-ale.

$$\varepsilon = z_{0.025}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 1.96\sqrt{\frac{0.0177 \cdot 0.9823}{733}} = 0.0096$$

$$I_{p_1}^{0.95} = (0.0177 \mp 0.0096) = (0.0081,\ 0.0273)$$

Beraz, enpresatik urrun egondako uretan kutsadura egoteko probabilitatea %0.81 eta %2.73 tartean dagoela ondoriozta daiteke %95eko konfiantza-mailaz.

b) Baldin $\varepsilon = 0.01$ bada, $n_1 = (\frac{1.96}{0.05})^2 \cdot 0.0177 \cdot 0.9823 = 667.93$. Beraz, 668 ur-ale aztertu beharko ziren, proportzioa estimatzean errorea %1 baino txikiagoa izateko. Aurreko datuetan, laginaren tamaina handiagoa zen, noski, lagin-errorea %0.96-koa, txikiagoa.

c) Kalkula dezagun %95-eko konfiantza-tartea proportzioen diferentziarako.

$$\varepsilon = z_{0.025} \cdot \sqrt{\frac{0.0177 \cdot 0.9823}{733} + \frac{0.0391 \cdot 0.9609}{742}} = 1.96 \cdot 0.0087 =$$
$$= 0.0171$$

Hortaz,

$$I_{p_1-p_2}^{0.95} = (-0.0214 \mp 0.0171) = (-0.0385,\ -0.00043)$$

Konfiantza-tartea 0-ren ezkerraldean dagoenez, diferentzia negatiboa dela ondorioozta daiteke %95eko konfiantza-mailaz, $p_1 - p_2 < 0$, hots, $p_1 < p_2$. Enpresatik hurbil egondako uretan kutsadura egoteko probabilitatea enpresatik urrun baino handiagoa da. Beraz, enpresak ibaian eragina duela ondorioozta daiteke.

d) Kasu honetan, $\varepsilon = z_{0.005} \cdot 0.0087 = 2.58 \cdot 0.0087 = 0.0224$ eta

$$I^{0.99}_{p_1-p_2} = (-0.0214 \mp 0.0224) = (-0.0438,\ 0.0010)$$

Konfiantza-tarte barruan 0 dagoenez, ezin da ondorioztatu diferentzia positiboa edo negatiboa denik, hots, berdintasuna ezin da deuseztatu %99ko konfiantza-mailaz, $p_1 - p_2 = 0$, hots, $p_1 = p_2$. Ondorioa ez da berdina. Enpresatik urrun eta hurbil egondako kutsatutako ur-lagin proportzioa berdina da. α txikiagoa erabiltzean $z_{\alpha/2}$ puntu-kritikoa handiagoa da eta horrela, lagin-errorea ere bai, horregatik konfiantza-tartearen luzera aldez aurretik handiagoa geratzea espero genuen.