Estatistika Azterketak

eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

Universidad del País Vasco / Euskal Herriko Unibertsitatea

Titulazioa: Kudeaketaren eta Informazio Sistemen Informatikaren Ingeniaritzako Gradua
Saila: Matematika Aplikatua
Irakasgaia: Estatistika Metodoak Ingeniaritzan
Kurtsoa: 2021-2022

Egilea: Arana González, Janire

Edukien aurkibidea

Urtarrila 2022

---

1. Ariketa

Taille® enpresak bulego-materiala egiten du, eta motor elektriko batek eragindako zorrozkailu (txorroskilo)-modelo bat du katalogoan. Motor horren ezaugarri nagusietako bat biraketa-abiadura da, minutuko biratan neurtzen dena (rpm). Enpresa fabrikatzailearen arabera, biraketa-abiaduraren batezbestekoa 70 rpm-koa da. Denda batek zorrozkailu-modelo hori erostea erabaki du, baina eskaera egin aurretik 5 zorrozkailuko lagin bat eskatu du, abiadura, katalogoan nabarmentzen dena dela egiaztatzeko. Dagozkion saiakuntzak egin ondoren, hauxe da dendak erregistratu dituen bost zorrozkailuen biraketa-abiadura:

	Zorrozkailua				
	1	**2**	**3**	**4**	**5**
Biraketa abiadura (rpm)	**67.3**	**71.2**	**68.1**	**70.5**	**70.8**

a) **Zorrozkailuaren motorraren biraketa-abiadurak banaketa normal bat jarraitzen duela suposatuz eta bost zorrozkailuen lagineko datuak erabiliz, kalkulatu biraketa-abiadura horren batezbestekorako konfiantza-tartea %90eko konfiantza mailaz (5 puntu)**

$$I_{\mu}^{1-\alpha} = \left[\bar{x} - t_{4;0.05} \cdot \frac{S}{\sqrt{n}},\ \bar{x} + t_{4;0.05} \cdot \frac{S}{\sqrt{n}}\right]$$

$$\bar{x} = \frac{67.3 + 71.2 + 68.1 + 70.5 + 70.8}{5} = 69.58$$

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = 1.7570$$

$$I_{\mu}^{0.90} = \left[69.58 - qt(0.95, 4) \cdot \frac{1.7570}{\sqrt{5}},\ 69.58 + qt(0.95, 4) \cdot \frac{1.7570}{\sqrt{5}}\right]$$

$$I_{\mu}^{0.90} = \left[69.58 - 2.1318 \cdot \frac{1.7570}{\sqrt{5}},\ 69.58 + 2.1318 \cdot \frac{1.7570}{\sqrt{5}}\right]$$

$$I_{\mu}^{0.90} = [67.9049,\ 71.2551]$$

Denda ez da gustura geratu egindako probarekin, eta beste lagin bat, kasu honetan 6 zorrozkailuz eratutakoa, eskatu dio beste enpresa bati, Crayon® enpresari, eta zorrozkailu horien biraketa abiadurak erregistratu ditu, emaitza hauek lortuz:

	Zorrozkailua					
	1	**2**	**3**	**4**	**5**	**6**
Biraketa abiadura (rpm)	**69.9**	**70.2**	**68.8**	**70.1**	**69.4**	**72.1**

b) **%2ko adierazgarritasun mailaz eta $\sigma_1 = 1.8$ rpm eta $\sigma_2 = 2.1$ rpm direla suposatuz, bigarren enpresan (Crayon®) zorrozkailuen motorren biraketa-abiaduraren batezbestekoa lehenengo enpresan (Taillle®) baino handiagoa dela onartu al daiteke? (7 puntu)**

X_1: "Lehenengo enpresako zorrozkailuen motorren biraketa abiadura rpm-tan" $\longrightarrow X_1 \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$

X_2: "Bigarren enpresako zorrozkailuen motorren biraketa abiadura rpm-tan" $\longrightarrow X_2 \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 < \mu_2 \end{cases}$

$$\bar{x}_2 = \frac{69.9 + 70.2 + 68.8 + 70.1 + 69.4 + 72.1}{6} = 70.08333$$

Onarpen eremua: $[z_{0.02},\ \infty) = [-qnorm(0.98, 0, 1),\ -\infty) = [-2.0537,\ \infty)$

Kontrasterako estatistikoa: $z = \dfrac{\bar{x_1} - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \dfrac{69.58 - 70.08333}{\sqrt{\frac{1.8^2}{5} + \frac{2.1^2}{6}}} = -0.4280$

$-0.4280 \in [-2.0537,\ \infty) \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen da, ezin da onartu bigarren enpresako zorrozkailuen motorren biraketa-abiadura lehenengo enpresako zorrozkailuena baino handiagoa dela.

c) **Kalkulatu aurreko kontrastearen p-balioa. (3 puntu)**

$\alpha_c = P(Z < -0.4280) = 1 - P(Z < 0.428) = 1 - pnorm(0.428, 0, 1) = 1 - 0.6657 = 0.3343$

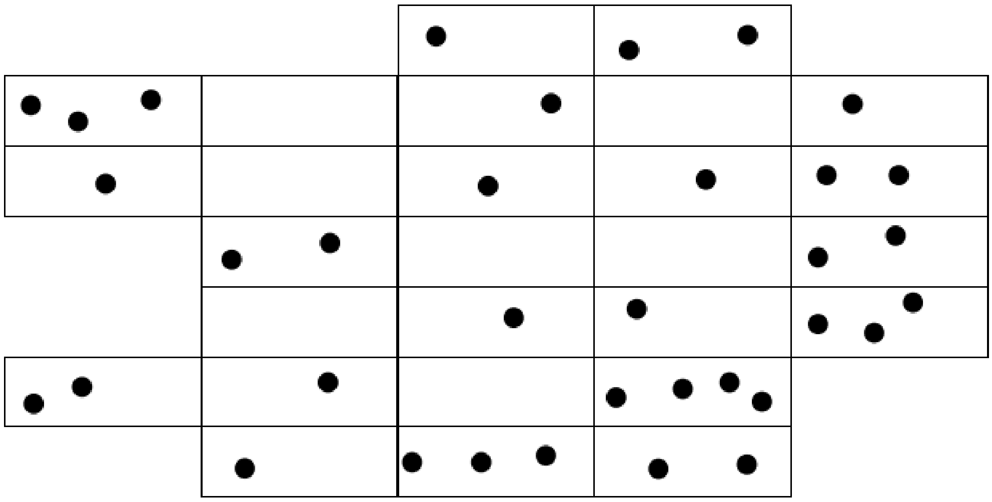
d) **Benetan $\mu_1 - \mu_2 = -1.5$ dela egiaztatzen bada, kalkulatu b) ataleko kontrastean II. motako errorea egiteko probabilitatea. (5 puntu)**

$$z_{0.02} = \frac{\bar{x_1} - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \to \bar{x_1} - \bar{x}_2 = z_{0.02} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = -2.0537 \cdot \sqrt{\frac{1.8^2}{5} + \frac{2.1^2}{6}} = -2.4152$$

$$P(H_0\ onartu|H_0\ gezurra) = P\left(X > -2.4252\right) = P\left(Z > \frac{-2.4252 - (-1,5)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = P(Z > -0.7782) =$$
$$1 - P(Z < 0.7782) = 1 - pnorm(0.7782.0.1) = 0.7818$$

2. Ariketa

Eredu sismiko batek eskualde bateko seismoen epizentroen banaketak planoan Poisson-en banaketa jarraitu beharko lukeela adierazten du. Aditu-talde batek eredu hori betetzen den egiaztatu nahi du. Horretarako, eskualdeko mapa bat irudikatu du, 100 km2 tamainako laukietan banatua, eta epizentroen posizioak puntuekin seinalatu ditu (ikus ondoko irudia):



a) **%5eko adierazgarritasun mailaz, eskualdeko 100 km^2-ko epizentro kopuruak Poisson-en banaketa bat jarraitzen duela onartu al daiteke? (9 puntu)**

X: "Eskualdeko 100 km^2-ko epizentro kopurua" $\longrightarrow X \sim \mathcal{P}(\lambda)$

x_i	0	1	2	3	4
f_i	7	10	6	3	1

Hipotesi kontrastea: $\begin{cases} H_0 : \text{X aldagaiak Poisson-en banaketa jarraitzen du} \\ H_a : \text{X aldagaiak ez du Poisson-en banaketa jarraitzen} \end{cases}$

$$\lambda = \bar{x} = \frac{0 \cdot 7 + 1 \cdot 10 + 2 \cdot 6 + 3 \cdot 3 + 4 \cdot 1}{27} = 1.2963$$

$$P(X=0) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-1.2963} \cdot \frac{1.2963^0}{0!} = 0.2735$$

$$P(X=1) = e^{-1.2963} \cdot \frac{1.2963}{1!} = 0.3546$$

$$P(X=2) = e^{-1.2963} \cdot \frac{1.2963^2}{2!} = 0.2298$$

$$P(X=3) = e^{-1.2963} \cdot \frac{1.2963^3}{3!} = 0.0993$$

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.2735 - 0.3543 - 0.2298 - 0.0993 = 0.0428$$

x_i	o_i	p_i	$e_i = n \cdot p_i$	$(o_i - e_i)^2/e_i$
0	7	0.2735	7.3845	0.0200
1	10	0.3546	9.5742	0.0189
2	6	0.2298	6.2046	
3	3	0.0993	2.6811	0.0001
≥ 4	1	0.0428	1.1556	

$v = 3 - 1 - 1 = 1$

Onarpen eremua: $\left[0,\ \chi^2_{0.05;1}\right] = [0,\ qchisq(0.95, 1)] = [0,\ 3.8415]$

$\sum \frac{(o_i - e_i)^2}{e_i} = 0.039 \in [0,\ 3.8415] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, H_0 onartzen da, aldagaiak Poisson-en banaketa jarraitzen du.

b) Proposatutako zorizko aldagaiak Poisson-en banaketa bat jarraitzen duela suposatuz, zein da 250 km^2-ko azalera duen hiri batean gutxienez bi epizentro egoteko probabilitatea? (5 puntu)

X: "Eskualdeko 250 km^2-ko epizentro kopurua" $\longrightarrow X \sim \mathcal{P}(\lambda = 1.2735 \cdot 2.5 = 3.2408)$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-3.2408} \cdot \left(\frac{3.2408^1}{1!} + \frac{3.2408^0}{0!} \right) = 1 - 0.1660 = 0.8340$$

c) Bestalde, jakina da Richter eskalan 7tik gorako magnitudea duten bi seismoen arteko denborak 30 urteko batezbestekoa duen banaketa esponentziala jarraitzen duela. Kalkula ezazu gehienez hurrengo urtean zehar 7 magnitudetik gorako lurrikara bat gertatzeko probabilitatea, jakinik magnitude horren aurreko lurrikara duela 20 urte gertatu zela. (6 puntu)

X: "Richter eskalan 7tik gorako magnitudea duten bi seismoren arteko denbora urtetan" $\longrightarrow X \sim \varepsilon(\beta = 30)$

Esponentzialaren memoria falta aplikatuz:
$P(Z < 21 | X > 20) = P(X < 1) = 1 - e^{-1/30} = 0.0328$

3. Ariketa

Enpresa batek hiru sail dauzka, A, B eta C, 24, 12 eta 18 langileekin hurrenez hurren. Urtarrileko asteburuetarako lan-taldea antolatzeko, zoriz 10 langile hautatzea erabaki da.

a) Zein da gutxienez 2 langile B sailekoak izateko probabilitatea? (3 puntu)

X: "10 langileko lan-taldea osatzeko B saileko langile kopurua" $\longrightarrow X \sim H(54, 10, 12/54)$

$$P(X \geq 2) = \frac{\binom{12}{2} \cdot \binom{42}{8}}{\binom{54}{10}} = 0.3255$$

b) Lan-talde erdia A saileko langilez osaturik egon behar bada, zein da lan-taldeko beste erdia C saileko langilez osaturik egoteko probabilitatea? (ezin dira A saileko langile gehiago aukeratu) (3 puntu)

X: "Lan taldean geratzen diren bost langileetatik C saileko langile kopurua" $\longrightarrow X \sim H(30, 5, 18/30)$

$$P(X = 5) = \frac{\binom{18}{5} \cdot \binom{12}{0}}{\binom{30}{5}} = 0.0601$$

c) Lan-taldeko langileak edozein modutan konbinatu badaitezke, zenbat langile espero dira A sailekoak izatea? Zein da langile kopuru horren desbiderazio tipikoa? (4 puntu)

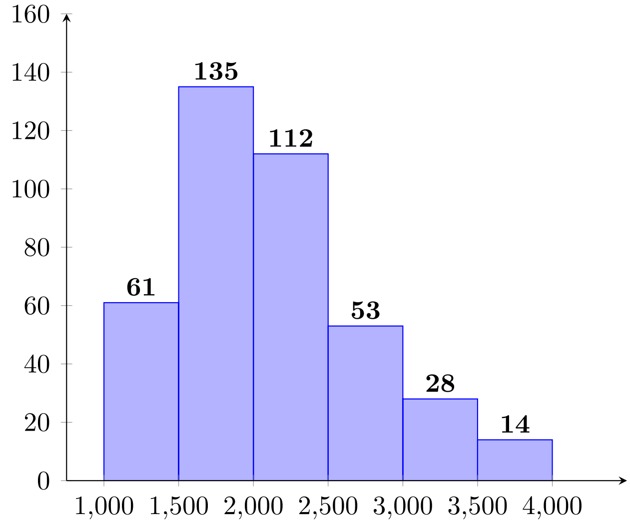
X: "10 langileko lan-taldea osatzeko A saileko langile kopurua" $\longrightarrow X \sim H(54, 10, 24/54)$

$E(X) = n \cdot p = 10 \cdot \frac{24}{54} = 4.4444$ langile espero dira

$$s = \sqrt{n \cdot p \cdot q \cdot \frac{N-n}{N-1}} = \sqrt{10 \cdot \frac{24}{54} \cdot \frac{26}{54} \cdot \frac{54-10}{54-1}} = \sqrt{2.0498} = 1.4317$$

4. Ariketa

Duela gutxi, Europako Batzordeak energia nuklearra energia "berde" gisa sartzea proposatu du, 2045. urtera arte gutxienez. Erabaki hori justifikatzeko, mundu osoko 403 erreaktore nuklearren lagin batek sortutako potentziak bildu dira. Lagin horren datuak ondorengo histograman ageri dira.



a) **Zeintzuk dira erreaktore nuklearrak sortutako potentziaren lehenengo, bigarren eta hirugarren kuartilen balioak? (7 puntu)**

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
[1000, 1500)	1250	61	61	0.1514	0.1514
[1500, 2000)	1750	135	196	0.3350	0.4864
[2000, 2500)	2250	112	308	0.2779	0.7643
[2500, 3000)	2750	53	361	0.1315	0.8958
[3000, 3500)	3250	28	389	0.0695	0.9653
[3500, 4000]	3750	14	403	0.0347	1

$$Q_1 = l_i + \frac{0.25n - F_{i-1}}{f_i} \cdot d_i = 1500 + \frac{0.25 \cdot 403 - 61}{135} \cdot 500 = 1647.2222$$

$$Q_2 = Me = l_i + \frac{0.5n - F_{i-1}}{f_i} \cdot d_i = 2000 + \frac{0.5 \cdot 403 - 196}{112} \cdot 500 = 2024.5536$$

$$Q_3 = l_i + \frac{0.75n - F_{i-1}}{f_i} \cdot d_i = 2000 + \frac{0.75 \cdot 403 - 196}{112} \cdot 500 = 2474.3304$$

b) **Kalkulatu erreaktore nuklearrak sortutako potentziaren moda eta Pearson-en alborapen koefizienteak. (4 puntu)**

$$Mo = l_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}) = \frac{135 - 61}{(135 - 61) + (135 - 112)} \cdot d_i = 1881.4433$$

$$\bar{x} = \frac{1250 \cdot 61 + 1750 \cdot 135 + 2250 \cdot 112 + 2750 \cdot 53 + 3250 \cdot 28 + 3750 \cdot 14}{403} = 2118.4864$$

$$s = \sqrt{\frac{\sum f_i \cdot (x_i - \bar{x})^2}{n}} = \sqrt{398336.9148} = 631.1294$$

$$P = \frac{\bar{x} - Mo}{s} = \frac{2118.4864 - 1881.4433}{631.1294} = 0.3756 > 0 \longrightarrow \text{Eskumarantz alboratua}$$

c) Erreaktore nuklearrak sortutako potentziak banaketa normal bat jarraitzen duela suposatuz, kalkulatu potentzia horren desbiderazio tipikorako konfiantza-tartea %99ko konfiantza mailaz. (6 puntu)

Bariantzarako konfiantza-tartea:

$$I_{\sigma^2}^{0.00} = \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}\right] = \left[\frac{402 \cdot 398336.9148}{qchisq(0.995; 402)}, \frac{402 \cdot 398336.9148}{qchisq(0.005; 402)}\right]$$

$$I_{\sigma^2}^{0.99} = \left[\frac{402 \cdot 398336.9148}{478.7884}, \frac{402 \cdot 398336.9148}{332.7208}\right] = [334451.3772,\ 481278.717]$$

Desbiderazio tipikorako konfiantza-tartea:

$$I_{\sigma^2}^{0.99} = \left[\sqrt{334451.3772},\ \sqrt{481278.717}\right] = [578.3177,\ 693.7425]$$

d) Energia-iturri guztiek sortutako batezbesteko potentzia 10500 MW bada, %5eko adierazgarritasun mailaz, erreaktore nuklearrek sortutako batezbesteko potentzia gutxienez potentzia horren guztiaren %20 dela baieztatu al daiteke? (Erreaktore nuklearrek sortutako potentziak banaketa normala jarraitzen duela suposatuz) (8 puntu)

X: "Erreaktore nuklearrek sortutako batezbesteko potentzia MW-tan"

$$\text{Hipotesi kontrastea: } \begin{cases} H_0 : \mu = 0.2 \cdot 10500 = 2100 \\ H_a : \mu < 0.2 \cdot 10500 = 2100 \end{cases}$$

Onarpen eremua: $[t_{\alpha;n-1},\ \infty) = [qt(0.95, 402),\ \infty) = [1.6486,\ \infty)$

Kontrasterako estatistikoa: $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{2118.4864 - 2100}{631.1294/\sqrt{403}} = 0.5880$

$0.5880 \notin [1.6486,\ \infty) \longrightarrow$ Estatistikoa ez dago onarpen eremuaren barne, beraz, ezin da baieztatu gutxienez potentziaren %20-a erreaktore nuklearrek sortzen dutenik.

R-ko komandoak

- pbinom(2,10,0.25)=0.5256
- pbinom(1,15,0.3387)=0.0176
- pnorm(0.4280,0,1)=0.6657
- pnorm(0.5821,0,1)=0.7198
- pnorm(0.7782,0,1)=0.7818
- pnorm(1.0236,0,1)=0.8470
- pnorm(1.3452,0,1)=0.9107
- pnorm(1.8421,0,1)=0.9673
- pf(0.8571,4,5)=0.4533
- pt(0.3545,10)=0.6348
- pt(0.6310,10)=0.7289
- pchisq(1.2588,8)=0.0040
- pchisq(5.8421,10)=0.1716
- qf(0.98,4,5)=8.2330
- qnorm(0.95,0,1)=1.6449
- qnorm(0.975,0,1)=1.9600
- qnorm(0.98,0,1)=2.0537
- qnorm(0.99,0,1)=2.3263
- qt(0.95,402)=1.6486
- qt(0.95,5)=2.0150
- qt(0.95,4)=2.1318
- qchisq(0.90,2)=4.6052
- qchisq(0.95,1)=3.8415
- qchisq(0.95,3)=7.8147
- qchisq(0.005,402)=332.7208
- qchisq(0.995,402)=478.7884

Ekaina 2021

---

1. Ariketa

Bilboko hornidura-sareko edateko uraren bentzeno-edukia (ppb) aztertzeko espektrofotometro batek gehienez 25 analisi egin ditzake orduko, azterketa bakoitza egiteko behar duen denbora dela eta. Espektrofotometro horrek, batezbeste 21 analisi egiten ditu orduko, eta egunean 10 orduz egoten da martxan. Era berean, jakina da edateko uraren bentzeno-edukiak banaketa normala duela, batezbestekoa 30 ppb eta desbiderazio tipikoa 2 ppb izanik.

a) **Kalkula ezazu makina saturatzeko eta ordu batean analisi gehiago egin ezin izateko probabilitatea. (6 puntu)**

X: "Espektrofotometroak ordu batean egindako analisi kopurua" $\longrightarrow X \sim \mathcal{P}(\lambda = 21)$

$\lambda > 18 \longrightarrow$ Normalera bihurtu $\longrightarrow X \sim \mathcal{N}(\lambda, \sqrt{\lambda}) \longrightarrow X \sim \mathcal{N}(21, \sqrt{21})$

Orduko gehienez 25 analisi egin ahal baditu, 25 analisi baino gehiago baditu saturatu egingo da.

$P(X > 25) \overset{zuz}{=\!=} P(X > 25.5) = P\left(Z > \dfrac{25.5 - 21}{\sqrt{21}}\right) = P(Z > 0.9820) = 1 - pnorm(0.9820, 0, 1) = 1 - 0.8370 = 0.1630$

b) **Zein da egun batean gutxienez 220 ur-laginen analisia egiteko probabilitatea? (7 puntu)**

Ordu batean 21 analisi egiten baditu batezbeste, egun batean 210 egingo ditu ($\lambda = 21 \cdot 10$ ordu).

X: "Espektrofotometroak egun batean egindako analisi kopurua" $\longrightarrow X \sim \mathcal{P}(\lambda = 210)$

$\lambda > 18 \longrightarrow$ Normalera bihurtu $\longrightarrow X \sim \mathcal{N}(\lambda, \sqrt{\lambda}) \longrightarrow X \sim \mathcal{N}(210, \sqrt{210})$

$P(X \geq 220) \overset{zuz}{=\!=} P(X > 219.5) = P\left(Z > \dfrac{219.5 - 210}{\sqrt{210}}\right) = P(Z > 0.6556) = 1 - pnorm(0.6556, 0, 1) = 1 - 0.7440 = 0.2560$

c) **Ura edangarria izan dadin bentzeno-edukiaren muga 32 ppb bada, zein da 5 analisiko laginaren bentzeno-edukiaren batezbestekoak muga hori gainditzeko probabilitatea? (7 puntu)**

X: "Uraren bentzeno edukia ppb-tan neurtuta"

Populazioaren batezbestekoa eta desbiderazio tipikoa ezagunak direnez, $\bar{X}$ aldagaiak $\mathcal{N}\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$ banaketa jarraitzen du. Hortaz, X $\sim \mathcal{N}\left(30, \dfrac{2}{\sqrt{5}}\right)$

$P(\bar{X} > 32) = P\left(Z > \dfrac{32 - 30}{2/\sqrt{5}}\right) = P(Z > 2.2361) = 1 - pnorm(2.2361, 0, 1) = 1 - 0.9873 = 0.0127$

2. Ariketa

Hainbat azterketa egin ondoren, txanpon baten zabalerak honen egokitasunean eragin handia duela jakina da, hau da, aurpegia ateratzeko probabilitatea eta gurutzea ateratzeko probabilitatea berdinak izatearen gainean eragina dauka. Txanponen zabaleraren batezbestekoa 2 mm-tik gorakoa bada, baztertu egin behar dira, ez bailirateke zuzenak izango. Azken hilabetean EBZak sortutako txanponak zuzenak direla egiaztatzeko, 120 txanponeko lagina hartu da, eta horien batezbesteko zabalera 2.03 mm-koa eta desbiderazio tipikoa 0.30 mm-koa direla kalkulatu da.

a) **%5eko adierazgarritasun mailaz, azken hilabetean EBZak ekoiztutako txanponak egokiak kontsideratu al daitezke? (8 puntu)**

X: "EBZak ekoiztutako txanponen zabalera mm-tan neurtuta"

Bariantza ez dugunez ezagutzen, kuasibariantza estimatuko dugu:

$$S^2 = \frac{\sigma^2 \cdot n}{n-1} = \frac{0.3^2 \cdot 120}{119} = 0.0908 \longrightarrow S = 0.3012$$

Hortaz, X aldagaiak $\mathcal{N}\left(\mu, \frac{\mathcal{S}}{\sqrt{n}}\right)$ banaketa jarraitzen du: X $\sim\mathcal{N}\left(2, \frac{0.3012}{\sqrt{120}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : \mu = 2 \\ H_a : \mu > 2 \end{cases}$

Onarpen eremua: $(-\infty, qnorm(0.95, 0, 1)] = (-\infty, 1.6449]$

Kontrasterako estatistikoa: $z = \dfrac{\bar{x} - \mu_0}{\mathcal{S}/\sqrt{n}} = \dfrac{2.03 - 2}{0.3012/\sqrt{120}} = 1.0909$

$1.0909 \in (-\infty, 1.6449] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen da, txanponak egokia kontsideratu daitezke.

b) **Hipotesi-kontrastearen p-balioa kalkulatu. (5 puntu)**

p-balioa, hipotesi nulua onartzeko adierazgarritasun maila maximoa bezala definitzen da.

$\alpha_c = P\,(Z > 1.0909) = 1 - pnorm(1.0909, 0, 1) = 1 - 0.8623 = 0.1377$

Hala ere, txanponen zabaleraren batezbestekoa egiaztatzeaz gain, beste saiakuntza bat egin nahi da, txanponak aurpegi baterantz kargatuta ez daudela egiaztatzeko, zehazki, gurutzerantz kargaturik ez daudela egiaztatzeko. Hipotesi hori probatzeko, zoriz aukeratutako txanpon bat n aldiz jaurtitzen da, eta baldintza hauek ezarri nahi dira:

- **Hipotesi nulua errefusatzeko probabilitatea benetan hau egia izanik, gehienez $\alpha =$ %5ekoa izan behar da.**
- **Hipotesi alternatiboa errefusatzeko probabilitatea benetan p 0.5-etik 0.1 edo gehiagotan aldentzen denean, hau da p $\geq$ 0.6 denean, gehienez 0.05 izan behar da.**

c) **Baldintza hauekin, egin beharreko txanponaren jaurtiketak zehaztu eta erabaki-araua enuntziatu. (12 puntu)**

$\hat{p}$: "Txanpona jaurtitzean gurutzea ateratzen den proportzioa" $\longrightarrow \hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{pq}{n}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : p = 0.5 \\ H_a : p > 0.5 \end{cases}$

Erabaki araua zehaztu ahal izateko, lehenengo jakin behar da non dagoen erabaki muga, hortaz, aurreko bi puntuetan zehaztutako kasuak betetzen dituen muga kalkulatu behar da.

Hipotesi nulua egia dela suposatuz, aldagaiak honako banaketa jarraituko luke: $\hat{p} \sim \mathcal{N}\left(0.5, \dfrac{0.5}{\sqrt{n}}\right)$

$$P(H_0 \text{ errefusatu} \mid H_0 \text{ egia}) = P(\hat{p} > a) = P\left(Z > \frac{a - 0.5}{0.5/\sqrt{n}}\right) = 0.05$$

$$\frac{a - 0.5}{0.5/\sqrt{n}} = qnorm(0.95, 0, 1) = 1.6449 \longrightarrow a - 0.5 = 1.6449 \cdot \frac{0.5}{\sqrt{n}} \longrightarrow \boxed{a = 0.5 + 1.6449 \cdot \frac{0.5}{\sqrt{n}}}$$

Hipotesi alternatiboa egia balitz, aldagaiak honako banaketa jarraituko luke: $\longrightarrow \hat{p} \sim \mathcal{N}\left(0.6, \sqrt{\dfrac{0.24}{n}}\right)$

$$P(H_0 \text{ onartu} \mid H_0 \text{ gezurra}) = P(\hat{p} \leq a) = P\left(Z \leq \frac{a - 0.6}{\sqrt{0.24/n}}\right) = 0.05$$

$$\frac{a - 0.6}{\sqrt{0.24/n}} = qnorm(0.05, 0, 1) = -qnorm(0.95, 0, 1) = -1.6449 \longrightarrow a - 0.6 = -1.6449 \cdot \frac{0.4899}{\sqrt{n}} \longrightarrow$$

$$\boxed{a = 0.6 - 1.6449 \cdot \frac{0.4899}{\sqrt{n}}}$$

Lortutako bi ekuazioekin sistema bat osatuz:

$$\begin{cases} a = 0.5 + 1.6449 \cdot \dfrac{0.5}{\sqrt{n}} \\ \\ a = 0.6 - 1.6449 \cdot \dfrac{0.4899}{\sqrt{n}} \end{cases} \Longrightarrow 0.5 + 1.6449 \cdot \frac{0.5}{\sqrt{n}} = 0.6 - 1.6449 \cdot \frac{0.4899}{\sqrt{n}} \longrightarrow \frac{0.5}{\sqrt{n}} + \frac{0.4899}{\sqrt{n}} = \frac{0.1}{1.6449}$$

$$\sqrt{n} = \frac{0.9899 \cdot 1.6449}{0.1} \longrightarrow n = 16.2829^2 = 265.1317 \longrightarrow \boxed{n \sim 266}$$

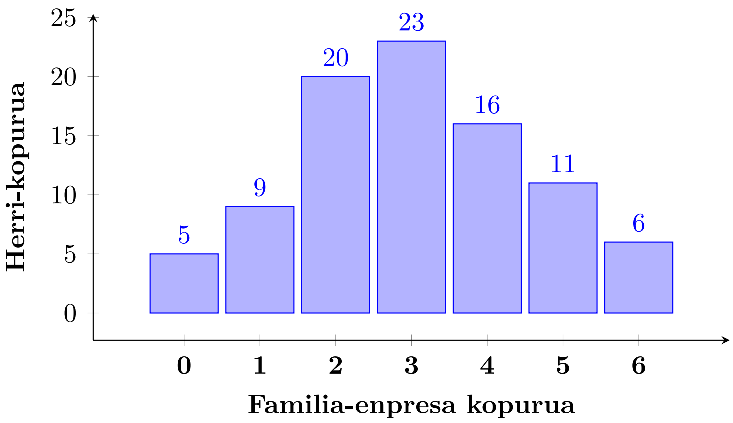
$$a = 0.5 + 1.6449 \cdot \frac{0.5}{\sqrt{266}} \longrightarrow a = 0.5504$$

Erabaki araua:

- 266 jaurtiketa edo gehiago egiten badira ($n \geq 266$), ateratako gurutze proportzioa $\hat{p} \leq 0.5504$ izango da, hortaz, H_0 onartuko litzateke.
- 266 jaurtiketa baino gutxiago egiten badira ($n < 266$), ateratako gurutze proportzioa $\hat{p} > 0.5504$ izango da, hortaz, H_0 errefusatu eta H_a onartuko litzateke.

3. Ariketa

Eusko Jaurlaritzako Ekonomiaren Garapen, Jasangarritasun eta Ingurumen Sailak 1000 biztanletik beherako herrietan kokatutako familia-enpresen kopuruak banaketa binomiala jarraitzen duen ala ez aztertu nahi du. Horrela, banaketa hori jarraitzen duela egiaztatzen bada, enpresa horien ikuskapena askoz ere modu eraginkorragoan egingo da. Hori egiaztatzeko, Euskadiko 90 herritan dauden familiaenpresen kopurua kontabilizatu da. Lortutako emaitzak barra-grafiko honetan ageri dira.



%10eko adierazgarritasun-mailaz, onartu al daiteke 1000 biztanletik beherako herrietan kokatutako familia-enpresen kopuruak banaketa binomial bat jarraitzen duela? (15 puntu)

X: "1000 biztanletik beherako herrietan kokatutako familia-enpresen kopurua" $\longrightarrow X \sim Bin(6, p)$

Hipotesi kontrastea: $\begin{cases} H_0 : \text{X aldagaiak banaketa binomiala jarraitzen du} \\ H_a : \text{X aldagaiak ez du banaketa binomiala jarraitzen} \end{cases}$

"p" parametroa ez dugunez ezagutzen, estimatzen dugu:

$$\hat{p} = \frac{\bar{x}}{m} = \frac{(0 \cdot 5 + 1 \cdot 9 + 2 \cdot 20 + 3 \cdot 23 + 4 \cdot 16 + 5 \cdot 11 + 6 \cdot 6)/90}{6} = \frac{273/90}{6} = 0.5056$$

$$P(X = 0) = \binom{6}{0} \cdot 0.5056^0 \cdot 0.4944^6 = 0.0146$$

$$P(X = 1) = \binom{6}{1} \cdot 0.5056^1 \cdot 0.4944^5 = 0.0896$$

$$P(X = 2) = \binom{6}{2} \cdot 0.5056^2 \cdot 0.4944^4 = 0.2291$$

$$P(X = 3) = \binom{6}{3} \cdot 0.5056^3 \cdot 0.4944^3 = 0.3124$$

$$P(X = 4) = \binom{6}{4} \cdot 0.5056^4 \cdot 0.4944^2 = 0.2396$$

$$P(X = 5) = \binom{6}{5} \cdot 0.5056^5 \cdot 0.4944^1 = 0.098$$

$$P(X = 6) = \binom{6}{6} \cdot 0.5056^6 \cdot 0.4944^0 = 0.0167$$

x_i	o_i	p_i	$e_i = n \cdot p_i$	$(o_i - e_i)^2/e_i$
0	5	0.0146	$\boxed{1.314}$	2.278
1	9	0.0896	$\boxed{8.064}$	
2	20	0.2291	20.619	0.0186
3	23	0.3124	28.116	0.9309
4	16	0.2396	21.564	1.4356
5	11	0.098	$\boxed{8.82}$	4.3187
6	6	0.0167	$\boxed{1.503}$	

$v = k - r - 1 = 5 - 1 - 1 = 3 \longrightarrow$ Onarpen eremua: $[0, \chi^2_{0.1;3}] = [0, qchisq(0.90, 3)] = [0, 6.2514]$

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = 2.278 + 0.0186 + 0.9309 + 1.4356 + 4.3187 = 8.0509 \notin [0, 6.2514]$$

Estatistikoa onarpen eremutik kanpo dago, hortaz, H_0 errefusatu eta H_a onartzen dugu. Aldagaiak ez du banaketa binomiala jarraitzen.

4. Ariketa

Izan bedi X banaketa normal bat jarraitzen duen zorizko aldagai bat, desbiderazio tipikoa 0.5-ekoa duena eta batezbestekoa ezezaguna. Hala ere, banaketaren batezbesteko horrek haue-

tako balioetako bat hartu dezake, $\boldsymbol{\mu_1 = 3,\ \mu_2 = 3.2,\ \mu_3 = 3.4,\ \mu_4 = 3.6,\ \mu_5 = 3.8}$, ondorengo probabilitateekin:

$$\begin{aligned}\boldsymbol{P(\mu = \mu_1) = 0.1}\\ \boldsymbol{P(\mu = \mu_2) = 0.2}\\ \boldsymbol{P(\mu = \mu_3) = 0.4}\\ \boldsymbol{P(\mu = \mu_4) = 0.2}\\ \boldsymbol{P(\mu = \mu_5) = 0.1}\end{aligned}$$

Populazio horretatik zoriz elementu bat hartzen da, x , eta $\boldsymbol{x < 3.5}$ da. Emaitza hau kontuan harturik, kalkulatu zorizko aldagai horren batezbestekoa $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$, edo $\boldsymbol{\mu_4}$ izateko probabilitatea. (15 puntu)

$$P(x < 3.5|\mu = \mu_1) = P\left(Z < \frac{3.5-3}{0.5}\right) = P(Z < 1) = 0.8413$$

$$P(x < 3.5|\mu = \mu_2) = P\left(Z < \frac{3.5-3.2}{0.5}\right) = P(Z < 0.6) = 0.7257$$

$$P(x < 3.5|\mu = \mu_3) = P\left(Z < \frac{3.5-3.4}{0.5}\right) = P(Z < 0.2) = 0.5793$$

$$P(x < 3.5|\mu = \mu_4) = P\left(Z < \frac{3.5-3.6}{0.5}\right) = P(Z < -0.2) = 1 - P(Z < 0.2) = 1 - 0.5793 = 0.4207$$

$$P(x < 3.5|\mu = \mu_5) = P\left(Z < \frac{3.5-3.8}{0.5}\right) = P(Z < -0.6) = 1 - P(Z < 0.6) = 1 - 0.7257 = 0.2743$$

$P(x < 3.5) = P(\mu = \mu_1) \cdot P(x < 3.5|\mu = \mu_1) + P(\mu = \mu_2) \cdot P(x < 3.5|\mu = \mu_2) + P(\mu = \mu_3) \cdot$
$P(x < 3.5|\mu = \mu_3) + P(\mu = \mu_4) \cdot P(x < 3.5|\mu = \mu_4) + P(\mu = \mu_5) \cdot P(x < 3.5|\mu = \mu_5) =$
$= 0.1 \cdot 0.8413 + 0.2 \cdot 0.7257 + 0.4 \cdot 0.5793 + 0.2 \cdot 0.4207 + 0.1 \cdot 0.2743 = 0.5726$

$$P(\mu = \mu_2|x < 3.5) = \frac{P(\mu = \mu_2) \cdot P(x < 3.5|\mu = \mu_2)}{P(x < 3.5)} = \frac{0.2 \cdot 0.7257}{0.5726} = 0.2535$$

$$P(\mu = \mu_3|x < 3.5) = \frac{P(\mu = \mu_3) \cdot P(x < 3.5|\mu = \mu_3)}{P(x < 3.5)} = \frac{0.4 \cdot 0.5793}{0.5726} = 0.4047$$

$$P(\mu = \mu_4|x < 3.5) = \frac{P(\mu = \mu_4) \cdot P(x < 3.5|\mu = \mu_4)}{P(x < 3.5)} = \frac{0.2 \cdot 0.4207}{0.5726} = 0.1469$$

$P(\mu = \mu_2 \cup \mu = \mu_3 \cup \mu = \mu_4) = P(\mu = \mu_2|x < 3.5) + P(\mu = \mu_3|x < 3.5) + P(\mu = \mu_4|x < 3.5) =$
$= 0.2535 + 0.4047 + 0.1469 = 0.8051$

R-ko komandoak

- pnorm(25,220,0.5)=0.7660
- pnorm(0.2,0,1)=0.5793
- pnorm(0.5,0,1)=0.6915
- pnorm(0.6,0,1)=0.7257
- pnorm(0.6556,0,1)=0.7440
- pnorm(0.8729,0,1)= 0.8086
- pnorm(0.9820,0,1)=0.8370
- pnorm(1,0,1)=0.8413
- pnorm(1.0909,0,1)=0.8623
- pnorm(1.4578,0,1)=0.9276
- pnorm(2.2361,0,1)=0.9873
- pnorm(2.4823,0,1)=0.9935
- pt(0.5,119)=0.6910
- pt(1.2503,119)=0.8932
- qnorm(0.1,0,1)=-1.2816
- qnorm(0.80,0,1)=0.8416
- qnorm(0.95,0,1)=1.6449
- qnorm(0.975,0,1)=1.9600
- qnorm(0.995,0,1)=2.5758
- qt(0.15,120)=-1.0409
- qt(0.90,120)=1.2886
- qt(0.9,6)=1.4398
- qt(0.975,120)=1.9799
- qf(0.9,120,120)=1.2646
- qchisq(0.1,5)=1.6103
- qchisq(0.90,3)=6.2514
- qchisq(0.90,5)=9.2364
- qchisq(0.95,5)=11.0705

Abendua 2020

---

1. Ariketa

Kutxa batek 9 bola ditu; batzuk zuriak dira eta beste batzuk gorriak. Zoriz 5 bola ateratzen dira, birjarpenik gabe, bi bola zuri eta hiru gorri lortuz.

a) **Emaitza hori lortu arren, kutxako hasierako konposizioan bola zurien kopurua bola gorriena baino handiagoa izateko probabilitatea kalkulatu. Emandako datuekin bateragarriak diren kutxako konposizio guztiek probabilitate bera dutela onartzen da. (8 puntu)**

A: "Bi bola zuri eta hiru gorri ateratzea"
B_i: "i bola zuri egotea"

9 bola izanda, kutxako hasierako konposizioan ezinezkoa da B_0 eta B_1 egotea, jada baitakigu bi bola zuri atera direla kutxatik. Eta baita B_9, B_8 eta B_7, jada 3 bola gorri atera baitira.

$$P(A|B_2) = \frac{\binom{2}{2}\cdot\binom{7}{3}}{\binom{9}{5}} = \frac{1\cdot 35}{126} = \frac{35}{126}$$

$$P(A|B_3) = \frac{\binom{3}{2}\cdot\binom{6}{3}}{\binom{9}{5}} = \frac{3\cdot 20}{126} = \frac{60}{126}$$

$$P(A|B_4) = \frac{\binom{4}{2}\cdot\binom{5}{3}}{\binom{9}{5}} = \frac{6\cdot 10}{126} = \frac{60}{126}$$

$$P(A|B_5) = \frac{\binom{5}{2}\cdot\binom{4}{3}}{\binom{9}{5}} = \frac{10\cdot 4}{126} = \frac{40}{126}$$

$$P(A|B_6) = \frac{\binom{6}{2}\cdot\binom{3}{3}}{\binom{9}{5}} = \frac{15\cdot 1}{126} = \frac{15}{126}$$

$$P(B_5\cup B_6|A) = \frac{P(B_5)\cdot P(A|B_5) + P(B_6)\cdot P(A|B_6)}{P(A))}$$

Konbinazio guztiek probabilitate bera dutenez:

$$P(B_5\cup B_6|A) = \frac{P(A|B_5)+P(A|B_6)}{P(A|B_2)+P(A|B_3)+P(A|B_4)+P(A|B_5)+P(A|B_6)} =$$

$$= \frac{40/126+15/126}{35/126+60/126+60/126+40/126+15/126} = \frac{55}{210} = 0.2619$$

Bestalde, jakina da bola bat edo bestea hartzeko probabilitatea bola horien dentsitatearen araberakoa dela hein handi batean, dentsitate txikienekoek kutxaren goikaldean kokatzeko joera baitute, eta, beraz, hartzeko errazagoak dira. Bolen dentsitateak banaketa normala jarraitzen

duela onartzen da, desbiderazio tipikoa 0.14 g/mL izanik bola zurientzat, eta 0.17 g/mL bola gorrientzat. Une jakin batean, kutxaren barnen ondorengo bola-konposizio hau dago, eta bakoitzaren dentsitatea taulan ageri da:

	Dentsitatea (g/mL)				
Bola zuria	**1.38**	**1.34**	**1.41**	**1.36**	**1.48**
Bola gorria	**1.42**	**1.38**	**1.45**	**1.39**	

b) **%3ko adierazgarritasun-mailaz, zein bola motak edukiko du edozein kutxatik ateratzeko probabilitaterik handiena? (6 puntu)**

X_1: "Bola zurien dentsitatea g/mL-tan neurtuta" $\longrightarrow X_1 :\sim \mathcal{N}\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right) = \mathcal{N}\left(\mu_1, \frac{0.14}{\sqrt{5}}\right)$

X_2: "Bola gorrien dentsitatea g/mL-tan neurtuta" $\longrightarrow X_2 \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right) = \mathcal{N}\left(\mu_2, \frac{0.17}{\sqrt{4}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 < \mu_2 \end{cases}$

$\bar{x}_1 = \dfrac{1.38 + 1.34 + 1.41 + 1.36 + 1.48}{5} = 1.394$

$\bar{x}_2 = \dfrac{1.42 + 1.38 + 1.45 + 1.39}{4} = 1.41$

Onarpen eremua: $[qnorm(0.03, 0, 1),\ \infty) = [-1.8808,\ \infty)$

Kontrasterako estatistikoa: $z = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \dfrac{1.394 - 1.41}{\sqrt{\frac{0.14^2}{5} + \frac{0.17^2}{4}}} = \dfrac{-0.016}{0.1056} = -0.1515$

$-0.1515 \in [-1.8808,\ \infty) \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen da. Txanpon zuriek gorriek baino dentsitate handiagoa dute, hortaz, gorriek dute ateratzeko probabilitaterik handiena.

c) **Bola zurien dentsitatearen batezbestekoa 1.41 g/mL dela eta gorriena 1.42 g/mL dela kontsideratzen da. Zein izan beharko litzateke kutxa batean egon beharko liratekeen bola zurien kopurua, hauen batezbesteko dentsitatea bola gorrien dentsitatearen batezbestekoa baino handiagoa izateko %90eko probabilitatearekin? (6 puntu)**

X_1: "Bola zurien dentsitatea g/mL-tan neurtuta" $\longrightarrow X \sim \mathcal{N}\left(1.41, \frac{0.14}{\sqrt{n_1}}\right)$

$P(\bar{x}_1 > 1.42) = P\left(Z > \dfrac{1.41 - 1.42}{0.14/\sqrt{n_1}}\right) = 0.9$

$\dfrac{1.41 - 1.42}{0.14/\sqrt{n_1}} = qnorm(0.90, 0, 1) = 1.2816 \longrightarrow -0.01 = 1.2816 \cdot \dfrac{0.14}{\sqrt{n_1}} \longrightarrow n_1 = \left(\dfrac{0.1794}{-0.01}\right)^2$

$n_1 = 321.9297 \longrightarrow \boxed{n_1 \sim 322}$

2. Ariketa

Desinfektatzaileak merkaturatzen dituen enpresa batek ziurtatzen du ekoizten duen hidrogelak 70 º-ko batezbesteko alkohol-graduazioa duela (70 mL alkohol produktuaren 100 mL bakoitzeko). Hala ere, kanpo-ikuskaritzako enpresa batek ez du produktu horren homologazioa eman nahi, graduazioa txikiagoa dela uste baitu. Horretarako, 120 hidrogel-ontzi aztertu ditu, eta lagin horren batezbestekoa 69.3 º-koa dela eta bariantza 7.2 $^{\text{o}(2)}$ dela egiaztatu du. Alkoholaren graduazioak

banaketa normal bat jarraitzen duela kontsideratzen da, honen desbiderazio tipikoa 2.7 º-koa izanik.

a) **%2ko adierazgarritasun-mailaz, kanpo-ikuskaritzako enpresaren susmoak egiazkoak al dira? (6 puntu)**

X: "Hidrogelaren graduazioa °-tan neurtuta" $\longrightarrow X \sim \mathcal{N}\left(70, \dfrac{2.7}{\sqrt{120}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : \mu = 70 \\ H_a : \mu < 70 \end{cases}$

Onarpen eremua: $[qnorm(0.02, 0, 1), \infty) = [-2.0537, \infty)$

Kontrasterako estatistikoa: $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{69.3 - 70}{2.7/\sqrt{120}} = -2.8400$

$-2.8400 \notin [-2.0537, \infty) \longrightarrow$ Estatistikoa onarpen eremutik kanpo dago, beraz, hipotesi nulua errefusatu eta alternatiboa onartuko dugu. Hortaz, kanpo-ikuskaritzako enpresaren susmoak egiazkoak dira.

b) **Aurreko kontrastearen p-balioa kalkulatu. (4 puntu)**

p-balioa, hipotesi nulua onartzeko adierazgarritasun maila maximoa bezala definitzen da.

$\sigma_c = P(Z < -2.8400) = pnorm(-2.8400, 0, 1) = 0.0023$

c) **Zein izan beharko litzateke laginaren tamaina %99ko konfiantza-mailaz, errorea 0.25 º baino txikiagoa izan zedin? (4 puntu)**

$$I_\mu^{1-\alpha} = \left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] \longrightarrow 1 - \alpha = P\left(|\bar{x} - \mu| \leq z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = P(|\bar{x} - \mu| \leq \varepsilon)$$

$$\varepsilon = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = z_{0.995} \cdot \frac{2.7}{\sqrt{n}} = 2.5758 \cdot \frac{2.7}{\sqrt{n}} \leq 0.25 \longrightarrow n \geq \left(\frac{6.9547}{0.25}\right)^2 = 773.8856 \longrightarrow \boxed{n \geq 774}$$

d) **Benetan hidrogelaren ahalko-graduazioa enpresa ekoizleak esandakoa baino %1 txikiagoa dela baieztatzen bada, kalkulatu a) ataleko kontrastean II. motako errorea egiteko probabilitatea eta kontrastearen potentzia zehaztu. (6 puntu)**

Hidrogelaren batezbestekoa %1 jaisten bada, honako batezbestekoa dugu orain: $70 - 70 \cdot 0.01 = 69.3$.

Beraz, X aldagaiak honako banaketa jarraitzen du: $\mathcal{N}\left(69.3, \dfrac{2.7}{120}\right)$

H_0 onartzeko limitea: $\bar{x} = \mu + z_{0.995} \cdot \dfrac{\sigma}{\sqrt{n}} = 70 + (-2.0537) \cdot \dfrac{2.7}{\sqrt{120}} = 69.4938$

$$P(H_0 \text{ onartu} \mid H_0 \text{ gezurra}) = P(\bar{x} < 69.2652) = P\left(Z < \frac{69.4938 - 69.3}{2.7/\sqrt{120}}\right) = pnorm(0.7863) = 0.7842$$

3. Ariketa

Ohiko ikuskapen batean, hiru marka komertzialeko hainbat gailuren seinalearen intentsitatea neurtu da. Neurketa horietan, gailuen seinalearen intentsitatea taula honetan ikusten den bezala kalifikatu da:

Intents. \ Marka	A	B	C	**Guztira**
Txarra	22	24	20	**66**
Onargarria	50	42	60	**152**
Ona	18	22	16	**56**
Bikaina	6	12	8	**26**
Guztira	**96**	**100**	**104**	**300**

%5eko adierazgarritasun-mailaz, gailuen intentsitatea marka komertzialaren araberakoak diren edo ez konprobatu. (15 puntu)

$$\text{Hipotesi kontrastea: } \begin{cases} H_0 : \text{“Intentsitatea eta Marka independenteak dira”} \\ H_a : \text{“Intentsitatea eta Marka ez dira independenteak”} \end{cases}$$

e_ij	A	B	C
Txarra	21.1200	22.0000	23.8800
Onargarria	48.6400	50.6667	52.6933
Ona	18.9200	18.6667	19.4133
Bikaina	8.3200	9.6667	9.0133

$$\sum_{i=1}^{4}\sum_{j=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = 6.3533$$

$v = (4-1)\cdot(3-1) = 6 \longrightarrow$ Onarpen eremua: $\left[0, \chi^2_{0.05;6}\right] = [0, qchisq(0.95, 6)] = [0, 12.5916]$

$\displaystyle\sum_{i=1}^{4}\sum_{j=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} \in [0, 12.5916] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, hortaz, hipotesi nulua onartzen dugu. Bi aldagaiak independenteak dira.

4. Ariketa

Postontzi batean 100 gutun gordetzen dira txanda bakoitzean. Gutun hauek jasotzen dituen postariak bi gutun hartzeko jarraibideak ditu, birjarpenik gabe, eta biek frankeoa zuzena badute, besteak onartuko ditu beste egiaztapenik gabe.

Aztertutako bi gutunek frankeoa zuzena dutela suposatuz eta gaizki frankeatutako 0, 1, 2, 3, 4, 5, 6... gutuna agertzeak 5, 4, 3, 2, 1, 0, 0... zenbakiekiko proportzionalak diren probabilitateak dituztela onartuz, hurrenez hurren.

Kalkula ezazu zein probabilitaterekin baiezta daitekeen 100 gutuneko loteak frankeoa akastuna duen gutun bat baino gehiago ez duela. (8 puntu)

X: “Frankeo zuzena duten karta kopura”
Y: “Gaizki frankeatutako gutun kopurua”

Gaizki frankeatutako gutunak agertzearen probabilitatea:

$$\left.\begin{array}{l} P(Y=0)=5p \\ P(Y=1)=4p \\ P(Y=2)=3p \\ P(Y=3)=2p \\ P(Y=4)=1p \end{array}\right\} \Longrightarrow 5p+4p+3p+2p+1p=1 \longrightarrow p=1/15$$

$$P(X=2) = \frac{5}{15}\cdot\frac{\binom{100}{2}\cdot\binom{0}{0}}{\binom{100}{2}} + \frac{4}{15}\cdot\frac{\binom{99}{2}\cdot\binom{1}{0}}{\binom{100}{2}} + \frac{3}{15}\cdot\frac{\binom{98}{2}\cdot\binom{2}{0}}{\binom{100}{2}} + \frac{2}{15}\cdot\frac{\binom{97}{2}\cdot\binom{3}{0}}{\binom{100}{2}} + \frac{1}{15}\cdot\frac{\binom{96}{2}\cdot\binom{4}{0}}{\binom{100}{2}} =$$

$$= \frac{5}{15}\cdot 1 + \frac{4}{15}\cdot\frac{98}{100} + \frac{3}{15}\cdot\frac{98\cdot 97}{100\cdot 99} + \frac{2}{15}\cdot\frac{97\cdot 96}{100\cdot 99} + \frac{1}{15}\cdot\frac{96\cdot 95}{100\cdot 99} + = \frac{5}{15} + \frac{392}{1500} + \frac{28518}{148500} + \frac{18624}{148500} + \frac{9120}{148500} =$$
$$= 0.9735$$

$$P(Y \le 1 \cap X = 2) = \frac{5}{15}\cdot\frac{\binom{100}{2}\cdot\binom{0}{0}}{\binom{100}{2}} + \frac{4}{15}\cdot\frac{\binom{99}{2}\cdot\binom{1}{0}}{\binom{100}{2}} = \frac{5}{15}\cdot 1 + \frac{4}{15}\cdot\frac{98}{100} = 0.5947$$

$$P(Y \le 1 \mid X = 2) = \frac{P(Y \le 1 \cap X = 2)}{P(X = 2)} = \frac{0.5947}{0.9735} = \boxed{0.6109}$$

5. Ariketa

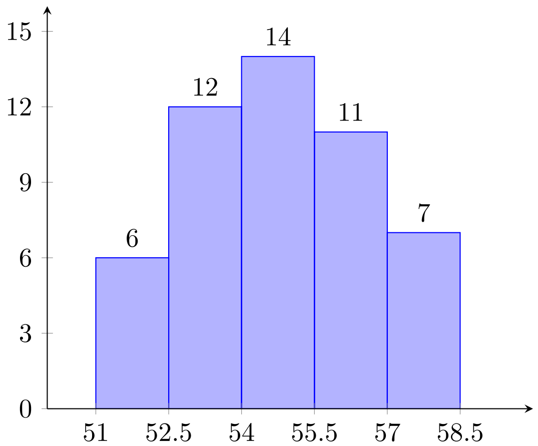
Barne-ikerketa bat egiteko, telefono-konpainia bateko bezeroek egunero deskargatutako gigabyte (GB) kopuruaren datuekin zorizko lagin bakun bat jaso da. Datu horiek ondorengo taula honetan daude:

51.00	**51.56**	**51.73**	**52.02**	**52.22**	**52.35**	**52.56**	**52.89**	**52.94**	**53.08**
53.10	**53.39**	**53.43**	**53.55**	**53.73**	**53.77**	**53.82**	**53.95**	**54.13**	**54.15**
54.20	**54.25**	**54.30**	**54.46**	**54.55**	**54.65**	**54.89**	**54.96**	**55.01**	**55.13**
55.21	**55.33**	**55.57**	**55.77**	**55.88**	**56.00**	**56.15**	**56.25**	**56.35**	**56.52**
56.60	**56.78**	**56.91**	**57.05**	**57.27**	**57.35**	**57.50**	**57.73**	**58.16**	**58.50**

a) Zabalera bereko bost klase dituen maiztasun-taula bat eraiki eta maiztasun-taulako klase berdinak dituen histograma bat irudikatu. (3 puntu)

Min=51.00
Max=58.50
Tartea=58.50-51.00=7.5
$d_i = 7.5/5 = 1.5$

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
[51.0, 52.5)	51.75	6	6	0.12	0.12
[52.5, 54.0)	53.25	12	18	0.24	0.36
[54.0, 55.5)	54.75	14	32	0.28	0.64
[55.5, 57.0)	56.25	11	43	0.22	0.86
[57.0, 58.5]	57.75	7	50	0.14	1



b) Behin datuak tartekaturik daudela, kalkulatu ondorengo estatistiko deskribatzaileak: kuartilak (Q1, Q2 eta Q3), kuartilarteko-heina eta moda. (2 puntu)

$$Q_1 = l_i + \frac{25n/100 - F_{i-1}}{f_i}\cdot d_i = 52.5 + \frac{0.25\cdot 50 - 6}{12}\cdot 1.5 = 53.3125$$

$$Q_2 = l_i + \frac{50n/100 - F_{i-1}}{f_i}\cdot d_i = 54 + \frac{0.25\cdot 50 - 18}{14}\cdot 1.5 = 54.75$$

$$Q_3 = l_i + \frac{75n/100 - F_{i-1}}{f_i}\cdot d_i = 55.5 + \frac{0.75\cdot 50 - 32}{11}\cdot 1.5 = 56.25$$

$$RIQ = Q_3 - Q_1 = 56.25 - 53.3125 = 2.9375$$

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}\cdot d_i = 54 + \frac{0.04}{0.04 + 0.06}\cdot 1.5 = 54.6 \;\; ; \;\; \Delta_1 = \frac{14}{50} - \frac{12}{50} = 0.04 \;\; ; \;\; \Delta_2 = \frac{14}{50} - \frac{11}{50} = 0.06$$

c) **Onartu al daiteke, %10eko adierazgarritasun-mailaz, telefono-konpainiako bezeroek egunero deskargatutako GB kopuruak banaketa normal bat jarraitzen duela? (Laginetik beharrezkoak diren estatistikoak erabili jadanik tartekatuta dauden datuetatik abiatuz) (5 puntu)**

X: “Telefono-konpainiako bezeroek egunero deskargatutako GB kopurua”

Hipotesi kontrastea: $\begin{cases} H_0: \text{“X aldagaiak banaketa normala jarraitzen du”} \\ H_a: \text{“X aldagaiak ez du banaketa normala jarraitzen”} \end{cases}$

$\bar{x} = \dfrac{\sum x_i \cdot f_i}{n} = 54.78 \qquad s^2 = \dfrac{\sum (x_i - \bar{x})^2 \cdot f_i}{n} = 3.3741 \longrightarrow s = 1.8369$

$P(X < 51) = P\left(Z < \dfrac{51 - 54.78}{1.8368}\right) = P(Z < -2.0578) = 1 - pnorm(2.0578, 0, 1) = 1 - 0.9802 = 0.0198$

$P(51 < X < 52.5) = P\left(\dfrac{51 - 54.78}{1.8369} < Z < \dfrac{52.5 - 54.78}{1.8369}\right) = P(Z < -1.2412) - P(Z < -2.0578) =$

$\text{P}(Z < 2.0578) - P(Z < 1.2412) = pnorm(2.0578, 0, 1) - pnorm(1.2412, 0, 1) = 0.9802 - 0.8927 = 0.0875$

$P(52.5 < X < 54) = P\left(\dfrac{52.5 - 54.78}{1.8369} < Z < \dfrac{54 - 54.78}{1.8369}\right) = P(Z < -0.4246) - P(Z < -1.2412) =$

$\text{P}(Z < 1.2412) - P(Z < 0.4246) = pnorm(1.2412, 0, 1) - pnorm(0.4246, 0, 1) = 0.8927 - 0.6644 = 0.2283$

$P(54 < X < 55.5) = P\left(\dfrac{54 - 54.78}{1.8369} < Z < \dfrac{55.5 - 54.78}{1.8369}\right) = P(Z < 0.392) - P(Z < -0.4246) =$

$\text{P}(Z < 0.392) - (1 - P(Z < 0.4246)) = pnorm(0.3920, 0, 1) - (1 - pnorm(0.4246, 0, 1)) = 0.6525 - (1 - 0.6644) = 0.3169$

$P(55.5 < X < 57) = P\left(\dfrac{55.5 - 54.78}{1.8369} < Z < \dfrac{57 - 54.78}{1.8369}\right) = P(Z < 1.2086) - P(Z < 0.392) =$

$\text{P}(Z < 1.2086) - P(Z < 0.392) = pnorm(1.2086, 0, 1) - pnorm(0.3920, 0, 1) = 0.8866 - 0.6525 = 0.2341$

$P(57 < X < 58.5) = P\left(\dfrac{57 - 54.78}{1.8369} < Z < \dfrac{58.5 - 54.78}{1.8369}\right) = P(Z < 2.0252) - P(Z < 1.2086) =$

$\text{P}(Z < 2.0252) - P(Z < 0.2086) = pnorm(2.0252, 0, 1) - pnorm(1.2086, 0, 1) = 0.9786 - 0.8866 = 0.092$

$P(X > 58.5) = P\left(Z > \dfrac{58.5 - 54.78}{1.8369}\right) = 1 - P(Z < 2.0252) = 1 - pnorm(2.0252, 0, 1) = 1 - 0.9786 = 0.0214$

x_i	o_i	p_i	$e_i = n \cdot p_i$	$(o_i - e_i)^2/e_i$
< 51	0	0.0198	$\boxed{0.99}$	0.0753
51.75	6	0.0875	$\boxed{4.375}$	
53.35	12	0.2283	11.515	0.0204
54.75	14	0.3169	15.845	0.2148
56.25	11	0.2341	11.705	0.0425
57.75	7	0.092	$\boxed{4.6}$	0.312
> 58.5	0	0.0214	$\boxed{1.07}$	

$v = k - r - 1 = 5 - 2 - 1 = 2$

Onarpen eremua: $\left[0, \chi^2_{0.90;2}\right] = [0, qchisq(0.90, 2)] = [0, 4.6052]$

$\sum \dfrac{(o_i - e_i)^2}{e_i} = 0.665 \in [0, 4.6052] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen dugu. Telefono-konpainiako bezeroek egunero deskargatutako GB kopuruak banaketa normala jarraitzen du.

d) %95eko konfiantza-mailaz, kalkulatu telefono-konpainiako bezeroek egunero deskargatutako GB kopuruaren batezbestekorako konfiantza-tartea. (2 puntu)

R-ko komandoak

- pbinom(3,9,0.5556)=0.1573
- pnorm(-2.8400,0,1)=0.0023
- pnorm(0.3920,0,1)=0.6525
- pnorm(0.4246,0,1)=0.6644
- pnorm(0.7863,0,1)=0.7842
- pnorm(0.9210,0,1)=0.8215
- pnorm(1.2086,0,1)=0.8866
- pnorm(1.2412,0,1)=0.8927
- pnorm(2.0252,0,1)=0.9786
- pnorm(2.0578,0,1)=0.9802
- pnorm(2.5581,0,1)=0.9947
- pt(-0.2347,7)=0.4106
- pt(2.5123,119)=0.9933
- qnorm(0.02,0,1)=-2.0537
- qnorm(0.03,0,1)=-1.8808
- qnorm(0.90,0,1)=1.2816
- qnorm(0.975,0,1)=1.9600
- qnorm(0.995,0,1)=2.5758
- qt(0.03,7)=-2.2409
- qt(0.1,49)=-1.2991
- qt(0.95,49)=1.6766
- qt(0.975,49)=2.0096
- qchisq(0.90,2)=4.6052
- qchisq(0.95,3)=7.8147
- qchisq(0.95,6)=12.5916
- qchisq(0.975,6)=14.4495

Urtatrila 2020

1. Ariketa

Etxebizitza batean hiru giltzatako daude: A, B eta C. Lehenengoak bost giltza ditu, bigarrenak zazpi eta hirugarrenak bederatzi, eta soilik giltzatako bakoitzeko giltza batek trastelekuko atea irekitzen du. Zoriz giltzatako bat aukeratzen da eta bertatik giltza bat atea irekitzeko.

a) **Zein izango da atea irekitzen duen giltzarekin asmatzeko probabilitatea? (2 puntu)**

A: "Giltza A giltzatakoa izatea" $\longrightarrow$ 5 giltza
B: "Giltza B giltzatakoa izatea" $\longrightarrow$ 7 giltza
C: "Giltza B giltzatakoa izatea" $\longrightarrow$ 9 giltza
D: "Giltza trastelekua irekitzea"

$$P(D) = P(A) \cdot P(D|A) + P(B) \cdot P(D|B) + P(C) \cdot P(D|C) = \frac{1}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{1}{7} + \frac{1}{3} \cdot \frac{1}{9} = 0.1513$$

b) **Zein izango da hirugarren giltzatakoa harturik aukeratutako giltzak atea ez irekitzeko probabilitatea? (Puntu 1)**

$$P(\bar{D}|C) = 1 - P(D|C) = 1 - \frac{1}{9} = \frac{8}{9} = 0.8889$$

c) **Eta aukeratutako giltza egokia bada, zein izango da lehenengo giltzatakoa (A) izateko probabilitatea? (2 puntu)**

$$P(A|D) = \frac{P(A \cap D}{D} = \frac{P(A) \cdot P(D|A)}{P(D)} = \frac{\frac{1}{3} \cdot \frac{1}{5}}{0.1513} = \frac{0.0667}{0.1513} = 0.4408$$

2. Ariketa

Artikulu zehatz baten eguneroko eskaria, ondorengo dentsitate funtzioa duen zorizko aldagai bat bezala kontsideratu daiteke:

$$\text{f(x)} = \begin{cases} \mathbf{\frac{1}{8}} & \mathbf{0 < x \leq 4} \\ \mathbf{\frac{12 - x}{64}} & \mathbf{4 < x \leq 12} \\ \mathbf{0} & \textbf{Beste kasuetan} \end{cases}$$

Eguneroko irabaziak eskariaren menpe daude ondorengo funtzioak adierazten duen moduan:

$$\textbf{Irabazia} = \begin{cases} \textbf{-5} & \textbf{eskaria 2 baino gutxiagokoa denean} \\ \textbf{5} & \textbf{eskaria 2 eta 4 bitartean dagoenean} \\ \textbf{10} & \textbf{eskaria 4 eta 8 bitartean dagoenean} \\ \textbf{15} & \textbf{eskaria 8 eta 12 bitartean dagoenean} \end{cases}$$

Kalkulatu:

a) **Edozein egunetan, eskaria 10 baino handiago izateko probabilitatea. (3 puntu)**

X: "Artikulu zehatz baten eguneroko eskaria" $\longrightarrow X \sim U(X)$

$$P(X > 10) = \int_{10}^{12} \frac{12-x}{64}\, dx = \frac{1}{64} \cdot \left[12x - \frac{x^2}{2}\right]_{10}^{12} = \frac{1}{64} \cdot \left[\left(12 \cdot 12 - \frac{12^2}{2}\right) - \left(12 \cdot 10 - \frac{10^2}{2}\right)\right] =$$
$$= \frac{72-70}{64} = 0.0313$$

b) **Eskaria 3 baino txikiagoa izateko probabilitatea. (3 puntu)**

$$P(X < 3) = \int_0^3 \frac{1}{8}\, dx = \left[\frac{x}{8}\right]_0^3 = 0.375$$

c) **Itxarotako eguneroko eskaria. (4 puntu)**

$$E(X) = \int_0^4 x \cdot \frac{1}{8}\, dx + \int_4^{12} x \cdot \frac{12-x}{64}\, dx = \frac{1}{8} \cdot \left[\frac{x^2}{2}\right]_0^4 + \frac{1}{64} \cdot \left[\frac{12x^2}{2} - \frac{x^3}{3}\right]_4^{12} = 4.3333$$

d) **Itxarotako eguneroko irabazia. (5 puntu)**

$$Irabazia = -5 \cdot \int_0^2 \frac{1}{8}\, dx + 5 \cdot \int_2^4 \frac{1}{8}\, dx + 10 \cdot \int_4^8 \frac{12-x}{64}\, dx + 15 \cdot \int_8^{12} \frac{12-x}{64}\, dx = -5 \cdot \left[\frac{x}{8}\right]_0^2 + 5 \cdot \left[\frac{x}{8}\right]_2^4 +$$
$$+10 \cdot \frac{1}{64} \cdot \left[12x - \frac{x^2}{2}\right]_4^8 + 15 \cdot \frac{1}{64} \cdot \left[12x - \frac{x^2}{2}\right]_8^{12} = -5 \cdot 0.25 + 5 \cdot 0.25 + 10 \cdot \frac{1}{64} \cdot 24 + 15 \cdot \frac{1}{64} \cdot 8 = 5.625\text{€}$$

3. Ariketa

Madrilen ospatutako klimaren azken goi-bileran onartutako araudia betetzeko, Alemaniako ibilgailu marka ezagun batek ibilgailu hibridoko gama berri bat kaleratzea erabaki du. Ibilgailu hauek araudia bete dezaten neurtzen den aldagaietako bat, 10 nm baino gutxiagoko diametroa duten partikulen emisioa da. Lehenengo proba batean, 7 ibilgailuren emisioak neurtu dira, ondorengo emaitzak (mg/km) lorturik:

Partikulen emisioa (mg/km)						
1.88	1.95	2.01	2.10	1.96	1.89	2.08

Aurreko neurketek banaketa normal bat jarraitzen dute. Lehenengo hiru ataletarako, neurketetan erabilitako tresnaren dokumentazio teknikoaren arabera, neurketen zehaztasuna $\sigma = 0.1$ mg/km dela kontsideratu. %5-eko adierazgarritasun mailaz:

a) **Ibilgailu hibridoko gama berriaren emisioaren batezbestekorako konfiantza-tartea kalkulatu. Konfiantza-tartearen zabalera maximoa 0.01 mg/km-koa izateko laginean beharrezkoa liratekeen ibilgailu kopurua kalkulatu. (4 puntu)**

$$\bar{x} = \frac{1.88 + 1.95 + 2.01 + 2.10 + 1.96 + 1.89 + 2.08}{7} = 1.9814$$

$$I_{\mu}^{0.95} = \left[\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[1.9814 - 1.96 \cdot \frac{0.1}{\sqrt{7}}, 1.9814 + 1.96 \cdot \frac{0.1}{\sqrt{7}}\right] = [1.9073,\ 2,0555]$$

$$Zabalera = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.01 \longrightarrow n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{0.01}\right)^2 = \left(2 \cdot 1.96 \cdot \frac{0.1}{0.01}\right)^2 = 1536.64 \longrightarrow n \sim 1537$$

b) **Partikulen emisioak 2.00 mg/km-ko limitea gainditzen badu, ihes-hodian nano-iragazki bat jartze beharrezkoa da, ibilgailuaren prezioa handituz. Nano-iragazki horren erabilpena beharrezkoa den edo ez kontrastatu. (4 puntu)**

X: "Ibilgailuen partikulen emisioa mg/km-tan neurtuta" $\longrightarrow X \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Hipotesi kontrastea: $\begin{cases} H_0 : \mu = 2.00 \\ H_a : \mu > 2.00 \end{cases}$

Onarpen eremua: $(-\infty,\ qnorm(0.95, 0, 1)] = (-\infty,\ 1.6449]$

Kontrasterako estatistikoa: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = z = \dfrac{1.9814 - 2.00}{0.1/\sqrt{7}} = -0.4921$

$-0.4921 \in (-\infty,\ 1.6449] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartuko da, hau da, partikulen emisioa 2.00 mg/km baino txikiagoa da. Hortaz, ez da filtrorik jarri behar.

c) **"b" ataleko kontrasteko p-balioa kalkulatu. (4 puntu)**

p-balioa, hipotesi nulua onartzeko adierazgarritasun maila maximoa bezala definitzen da.

$\sigma_c = P(Z > -0.4921) = P(Z < 0.4921) = pnorm(0.4921, 0, 1) = 0.6887$

d) **"b" ataleko kontrastean, II motako errorea egiteko probabilitatea kalkulatu, partikulen emisioaren batezbestekoa 2.10 mg/km-koa izanik. (4 puntu)**

$X \sim \mathcal{N}\left(2.1, \dfrac{0.1}{\sqrt{7}}\right)$

H_0 onartzeko limitea: $\bar{x} = \mu + z_{0.95} \cdot \dfrac{\sigma}{\sqrt{n}} = 2 + 1.6449 \cdot \dfrac{0.1}{\sqrt{7}} = 2.0622$

$P(H_0 \text{ onartu} \mid H_0 \text{ gezurra}) = P(X > 2.0622|\mu = 2.1) = P\left(Z > \dfrac{2.0622 - 2.1}{0.1/\sqrt{7}}\right) = P(Z > -1.0001) =$
$P(Z < 1.0001) = pnorm(1.0001, 0, 1) = 0.8414$

e) **Aurreko behaketak kontuan harturik, emisioen azterketa egiten ari den ingeniariak hauen neurketen bariantza benetan $\sigma^2 = 0.01\ mg^2/kg^2$-koa dela, edo aitzitik handiagoa dela konprobatu nahi du. Beharrezkoa den hipotesi-kontrastea egin, neurketatan erabilitako tresnaren bariantzak espezifikazio teknikoak betetzen dituela zehazteko. (4 puntu)**

Hipotesi kontrastea: $\begin{cases} H_0 : \sigma^2 = 0.01 \\ H_a : \sigma^2 < 0.01 \end{cases}$

$S^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1} = 0.0074$

Onarpen eremua: $\left[\chi^2_{n-1;\alpha},\ \infty\right) = [qchisq(0.95, 6), \infty) = [12.5916,\ \infty)$

Kontrasterako estatistikoa: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} = \dfrac{6 \cdot 0.0074}{0.01} = 4.4686$

$4.4687 \notin [12.5916,\ \infty) \longrightarrow$ Estatistikoa onarpen eremutik kanpo dago, beraz, hipotesi nulua errefusatu eta hipotesi alternatiboa onartzen da. Neurketen bariantza benetan 0.01 mg^2/km^2 baino txikiagoa da.

4. Ariketa

Kalkuluko kurtso berdin batean, gai bat hiru metodo ezberdinen bitartez azaldu egin da hiru ikasle talde ezberdinei. Azterketa berdina egin ondoren, ondorengo emaitzak lortu dira: G: gutxiegi, N: nahikoa, O: oso ondo eta B: bikain. Emaitzak ondorengo taulan agertzen dira:

Metodoa / Emaitzak	a	b	c	**Guztira**
G	22	24	20	**66**
N	50	42	60	**152**
O	18	22	16	**56**
B	6	12	8	**26**
Guztira	**96**	**100**	**104**	**300**

Cuadro 4.1: OHARRA: “Guztira” zutabeak ez datoz enuntziatuan, ariketa egiteko gehitu dira.

%5-eko adierazgarritasun mailaz, emaitzak erabilitako metodoarekiko independenteak diren edo ez konprobatu. (15 puntu)

Hipotesi kontrastea: $\begin{cases} H_0 : \text{“Emaitzak eta erabilitako metodoa independenteak dira”} \\ H_a : \text{“Emaitzak eta erabilitako metodoa ez dira independenteak”} \end{cases}$

e_ij	a	b	c
G	21.1200	22.0000	23.8800
N	48.6400	50.6667	52.6933
O	18.9200	18.6667	19.4133
B	8.3200	9.6667	9.0133

$$\sum_{i=1}^{4}\sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 6.3533$$

$v = (4-1) \cdot (3-1) = 6$

Onarpen eremua: $\left[0,\ \chi^2_{0.05;6}\right] = [0,\ qnorm(0.95, 6)] = [0,\ 12.5916]$

$\displaystyle\sum_{i=1}^{4}\sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \in [0,\ 12.5916] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen da. Azterketetan lortutako emaitzak erabilitako metodoarekiko independenteak dira.

5. Ariketa

Patata frijituko banaketa-enpresa batek kalitate-kontrol bat egiten du, eta hartarako 350 poltsako zorizko lagin bakuna hartzen du, ondorengo emaitzak lortuz:

Pisua gramotan	Poltsa kopurua
0-50	20
50-100	65
100-150	100
150-200	95
200-250	60
250-300	10

a) %5-eko adierazgarritasun mailaz, esan al daiteke patata poltsen pisuak banaketa normal bat jarraitzen duela? (6 puntu)

X: "Patata poltsen pisua gramotan" $\longrightarrow X \sim \mathcal{N}(\mu, \sigma)$

Hipotesi kontrastea: $\begin{cases} H_0 : \text{X aldagaiak banaketa normala jarraitzen du} \\ H_a : \text{X aldagaiak ez du banaketa normala jarraitzen} \end{cases}$

μ eta σ parametroak ez ditugunez ezagutzen, estimatzen ditugu:

$$\mu = \bar{x} = \frac{25 \cdot 20 + 75 \cdot 65 + 125 \cdot 100 + 175 \cdot 95 + 225 \cdot 60 + 275 \cdot 10}{350} = 145$$

$$\sigma = s = \sqrt{\frac{\sum f_i \cdot (x_i - \bar{x})^2}{n}} = 60.5923$$

$$P(0 < X < 50) = P\left(\frac{0 - 145}{60.5923} < Z < \frac{50 - 145}{60.5923}\right) = P(Z < -1.5679) - P(Z < -2.3930) =$$

$$P(Z < 2.3930) - P(Z < 1.5679) = pnorm(2.3930, 0, 1) - pnorm(1.5679, 0, 1) = 0.9916 - 0.9415 = 0.0501$$

$$P(50 < X < 100) = P\left(\frac{50 - 145}{60.5923} < Z < \frac{100 - 145}{60.5923}\right) = P(Z < -0.7427) - P(Z < -1.5679) =$$

$$P(Z < 1.5679) - P(Z < 0.7427) = pnorm(1.5679, 0, 1) - pnorm(0.7427, 0, 1) = 0.9415 - 0.7712 = 0.1703$$

$$P(100 < X < 150) = P\left(\frac{100 - 145}{60.5923} < Z < \frac{150 - 145}{60.5923}\right) = P(Z < 0.0825) - P(Z < -0.7427) =$$

$$P(Z < 0.0825) - (1 - P(Z < 0.7427)) = pnorm(0.0825, 0, 1) - (1 - pnorm(0.7427, 0, 1)) =$$

$$0.5329 - 0.2288 = 0.3041$$

$$P(150 < X < 200) = P\left(\frac{150 - 145}{60.5923} < Z < \frac{200 - 145}{60.5923}\right) = P(Z < 0.9077) - P(Z < 0.0825) =$$

$$pnorm(0.9077, 0, 1) - pnorm(0.0825, 0, 1) = 0.8180 - 0.5329 = 0.2851$$

$$P(200 < X < 250) = P\left(\frac{200 - 145}{60.5923} < Z < \frac{250 - 145}{60.5923}\right) = P(Z < 1.7329) - P(Z < 0.9077) =$$

$$pnorm(1.7329, 0, 1) - pnorm(0.9077, 0, 1) = 0.9584 - 0.8180 = 0.1404$$

$$P(250 < X < 300) = P\left(\frac{250 - 145}{60.5923} < Z < \frac{300 - 145}{60.5923}\right) = P(Z < 2.5581) - P(Z < 1.7329) =$$

$$0.9947 - 0.9584 = 0.0363$$

x_i	o_i	p_i	$e_i = n \cdot p_i$	$(o_i - e_i)^2/e_i$
25	20	0.0501	17.535	0.3465
75	65	0.1703	59.605	0.4883
125	100	0.3041	106.435	0.3891
175	95	0.2851	99.785	0.2295
225	60	0.1404	49.14	2.4001
275	10	0.0363	12.705	0.5759

$v = k - r - 1 = 6 - 2 - 1 = 3$

Onarpen eremua: $\left[0, \chi^2_{0.05;3}\right] = [0, qchisq(0.95, 3)] = [0, 7.8147]$

$\sum \frac{(o_i - e_i)^2}{e_i} = 4.4294 \in [0, 7.8147] \longrightarrow$ Estatistikoa onarpen eremuaren barne dago, beraz, hipotesi nulua onartzen da: patata poltsen pisuak banaketa normala jarraitzen du.

b) Emaitzen taularen goialdeko %15-a kontsideratzen bada, zein da patata poltsak eduki beharreko pisu minimoa? (3 puntu)

$[l_i, l_{i+1})$	f_i	F_i	h_i	H_i
[0, 50)	20	20	0.0572	0.0572
[50, 100)	65	85	0.1857	0.2429
[100, 150)	100	185	0.2857	0.5286
[150, 200)	95	280	0.2714	0.8000
[200, 250)	60	340	0.1714	0.9714
[250, 300]	10	350	0.0286	1

$$P_{85} = l_i + \frac{85n/100 - F_{i-1}}{f_i} \cdot d_i = 200 + \frac{0.85 \cdot 350 - 280}{60} \cdot 50 = 214.5833$$

c) Emaitzen taularen goialdeko %50-a kontsideratzen bada, zein da patata poltsak eduki beharreko pisu minimoa? (3 puntu)

$$Me = l_i + \frac{50n/100 - F_{i-1}}{f_i} \cdot d_i = 100 + \frac{0.5 \cdot 350 - 85}{100} \cdot 50 = 145$$

d) Taulako emaitzak kontuan hartuz, zein da patata poltsaren pisurik ohikoena? (4 puntu)

$$\Delta_1 = \frac{f_i}{n} - \frac{f_{i-1}}{n} = \frac{100}{350} - \frac{65}{350} = \frac{35}{350}$$
$$\Delta_2 = \frac{f_i}{n} - \frac{f_{i+1}}{n} = \frac{100}{350} - \frac{95}{350} = \frac{5}{350}$$
$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i = 100 + \frac{35/350}{35/350 + 5/350} \cdot 50 = 143.75$$

e) Taulako emaitzak kontuan hartuz, zein da datuen %90-a harturik, probabilitate berdineko aldeak uzten dituen tartea? (4 puntu)

$$\left.\begin{array}{l} P_5 = l_i + \frac{5n/100 - F_{i-1}}{f_i} \cdot d_i = 0 + \frac{0.05 \cdot 350 - 0}{20} \cdot 50 = 43.75 \\ P_{95} = l_i + \frac{95n/100 - F_{i-i}}{f_i} \cdot d_i = 200 + \frac{0.95 \cdot 350 - 280}{60} \cdot 50 = 243.75 \end{array}\right\} \Longrightarrow [43.75,\ 243.75]$$

R-ko komandoak

- ppois(4,6)=0.2851
- pnorm(0.0825,0,1)=0.5329
- pnorm(0.4921,0,1)=0.6887
- pnorm(0.7427,0,1)=0.7712
- pnorm(0.9077,0,1)=0.8180
- pnorm(1.0001,0,1)=0.8414
- pnorm(1.5679,0,1)=0.9415
- pnorm(1.7329,0,1)=0.9584
- pnorm(2.3930,0,1)=0.9916
- pnorm(2.5581,0,1)=0.9947
- qnorm(0.9,0,1)=1.2816
- qnorm(0.95,0,1)=1.6449
- qnorm(0.975,0,1)=1.9600
- qt(0.95,3)=2.3534
- qt(0.95,6)=1.9432
- qt(0.95,7)=1.8946
- qchisq(0.05,3)=0.3518
- qchisq(0.95,3)=7.8147
- qchisq(0.95,6)=12.5916
- qchisq(0.975,6)=14.4495