BILBOKO INGENIARITZA ESKOLA
ESCUELA DE INGENIERÍA DE BILBAO

Departamento de Matemática Aplicada
Edificio II-I,
Paseo Rafael Moreno “Pitxitxi”, 3
48013 Bilbao

INDUSTRIA INGENIARITZA TEKNIKOKO ATALA
SECCIÓN INGENIERÍA TÉCNICA INDUSTRIAL

INGENIARITZAKO METODO ESTATISTIKOAK

UZTAILEKO DEIALDIA (2018-07-04): **AZTERKETA FINALA**

1 ARIKETA

MEVASA fabrikak produktu berri bat kaleratu nahi du, bere instalazioetan duen makinaria erabiliz. Ekoizten dutenaren bariantza 1 m^2-koa dela eta banaketa normal bat jarraitzen duela dakite. 0.4 m-ko batezbestekora doikuntzak egin ondoren, 0.3 m-ko batezbestekoa ekoizten ari direnaren susmoa dute, eta hau ez da onargarria. % 5-eko adierazgarritasun-mailaz eta kontrastearen potentzia % 63.87-koa izanik, kalkulatu:

(1.) Kontrasterako erabili den laginaren tamaina **(5 puntu)**.

Populazio bateko batezbestekoaren kontraste bat da. Izan bedi X=*”produktuaren neurria, metrotan neurtuta”* zorizko aldagaia, non $X \sim N\left(\mu_X = 0.4\, m, \sigma_X = 1\, m\right)$ den. Enuntziatuan emandako informazioa erabiliz kontrastea ondorengoa da:

$$\begin{cases} H_0: & \mu_X \geq \mu_0 = 0.4\, m \\ H_a: & \mu_X < \mu_0 = 0.4\, m \end{cases}$$

Hau da, alde bakarreko kontraste bat da. Kontraste/estimazio honetan:

$$\bar{X} \sim N\left(\mu_0 = 0.4\, m, \frac{\sigma_X}{\sqrt{n}} = \frac{1}{\sqrt{n}}\right)$$

Adierazgarritasun-maila $\alpha = 5\ \%$ denez, eskualde kritikoa hurrengoa da:

$$EC \triangleq (-\infty, \mathcal{L}) / \mathcal{L} = \mu_0 - z_\alpha \cdot \frac{\sigma_X}{\sqrt{n}} = 0.4 - 1.645 \frac{1}{\sqrt{n}}$$

Bestalde, enuntziatuan ematen den potentzia erabiliz II. motako errorea ondorengoa dela lor daiteke:

$$\beta = P\left(II.motako\ errorea\right) = P\left(H_0\ onartu \mid H_0\ gezurra\right) =$$

$$= P\left(\bar{X} \geq L \mid \bar{X} \sim N\left(0.3, \frac{1}{\sqrt{n}}\right)\right) = 1 - 0.6387 = 0.3613$$

Ondorioz:

$$\beta = P\left(II.motako\ errorea\right) = 0.3613 = 1 - P\left(\bar{X} < L \mid \bar{X} \sim N\left(0.3, \frac{1}{\sqrt{n}}\right)\right)$$

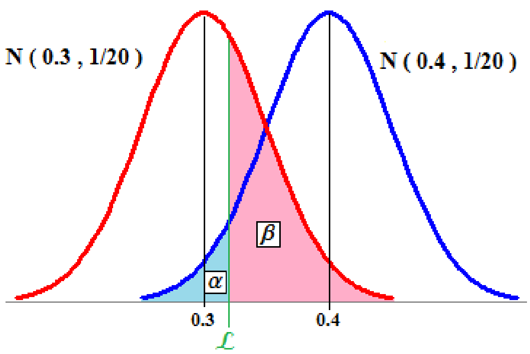
Tipifikatuz:

$$0.3613 = 1 - P\left(Z < \frac{0.4 - 1.645\frac{1}{\sqrt{n}} - 0.3}{\frac{1}{\sqrt{n}}} \mid Z \sim N(0,1)\right) \Rightarrow P\left(Z < \frac{0.1 - 1.645\frac{1}{\sqrt{n}}}{\frac{1}{\sqrt{n}}}\right) = 0.6387$$

Taula erabiliz $F\left(Z \leq z_1 = 0.355\right) = 0.6387$ dela lortzen da. Kalkuluak eginez:

$$\frac{0.1-1.645\frac{1}{\sqrt{n}}}{\frac{1}{\sqrt{n}}}=0.355\Rightarrow 0.1\sqrt{n}=2\Rightarrow n=400$$

Hortaz, lagineko tamaina gutxienez $n=400$ da.



(2.) Zorizko lagin bakun bat aukeratu ondoren lortutako lagineko balioa 0.33 bada, zer hipotesi onar dezakegu? **(3 puntu)**.

Lagina erabiliz eskualde-kritikoa ondorengoa dela lortzen da:

$$EK \triangleq (-\infty,\mathcal{L}) \,/\, \mathcal{L}=\mu_0-z_\alpha\cdot\frac{\sigma_X}{\sqrt{n}}=0.4-1.645\frac{1}{\sqrt{400}}=0.3176\,m$$

Ondorioz, hipotesi nulua onartzeko ebidentzia estatistikoa daude, izan ere:

- $\overline{x}=0.33\,m\notin EK\triangleq(-\infty,\mathcal{L})=(-\infty,0.31775\,m)$
- $\mathcal{L}=0.31775\,m<\overline{x}=0.33\in\ Onarpen\,eremua$

(3.) Aurreko ataleko populazioaren parametrorako konfiantza-tartea kalkulatu, laginaren estatistikoen balioak mantenduz, konfiantza-maila $\alpha=\%99$ eta laginaren tamaina n = 400 izanik **(3 puntu)**.

(2.) ataleko lagina erabiliz, estimatu beharreko populazioko parametroa batezbestekoa dela ondorioztа daiteke, estimatzailea $\hat{\mu}_X=\overline{x}=0.33\,m$ izanik, non $\overline{X}\sim$N$\left(\mu_X\,,\frac{\sigma_X}{\sqrt{n}}=\frac{1}{20}\right)$ den.

Estimazio-tartea ondorengoa da:

$$[l,\mathcal{L}]\triangleq\hat{\mu}_X\pm z_{0.995}\cdot\frac{\sigma_X}{\sqrt{n}}=0.33\pm 2.575\cdot\frac{1}{20}=[0.20125\,m,\,0.45875\,m]$$

Konfiantza-tarte hau $\alpha=\%1$ adierazgarritasun-maila duen bi aldeko kontrasteko onarpen eremuarekin bat dator.

2 ARIKETA

"Casco Viejo/Bilbo Zahar" alda-geltokian 3. lineako metroa hartzeko itxarote-denbora 2 minutuko desbiderazio tipikoko banaketa normala jarraitzen duen X zorizko aldagai bat dela onar dezakegu. Ikasle batek metroa hartuko du baldin eta itxarote-denbora erreala 7 minutu edo txikiagoa bada. Bestela, kotxea hartuko du. Erabaki bat hartu ahal izateko ondorengo proba egiten du: zoriz 9 bidaien itxarote-denbora neurtzen du eta batezbestekoa 8 minutu baino txikiagoa bada tranbia hartuko du, kontrako kasuan kotxea hartuko du.

(1.) Zein da planteaturiko hipotesi-kontrastea? **(1 puntu)**.

$$\begin{cases} H_0: & \mu_X = \mu_0 = 7 \text{ min} \\ H_a: & \mu_X > \mu_0 = 7 \text{ min} \end{cases}$$

Hau da, alde bakarreko kontraste bat da, adierazgarritasun-maila α orokor bat izanik

(2.) Zehaztu proban I. motako errorea egiteko probabilitatea. **(3 puntu)**.

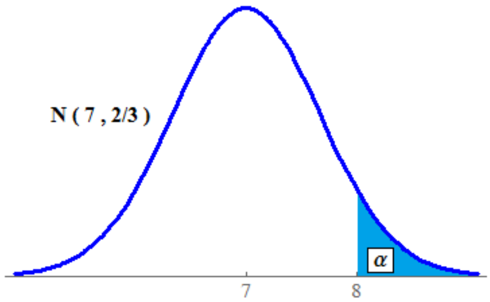
I. motako errorea kalkulatzeko ondorengoa erabiliko da:

$$\overline{X} \sim \mathrm{N}\left(\mu = 7 \text{ min}, \sigma_{\hat{\mu}_X} = \frac{\sigma_X}{\sqrt{n}} = \frac{2}{\sqrt{9}} = 0.6667 \text{ min} \right)$$

$$\alpha = \mathrm{P}(I.motako\, errorea) = \mathrm{P}(H_0 \; errefusatu \,|\, H_0 \; egia) = \mathrm{P}(\overline{X} > 8 \text{ min} \,|\, H_0 \; egia) =$$

$$= \mathrm{P}\left(\overline{X} > 8 \,|\, \overline{X} \sim \mathrm{N}\left(\mu_0 = 7, \sigma_{\hat{\mu}_X} = 0.6667\right)\right) = \mathrm{P}\left(Z > \frac{\overline{x} - \mu_0}{\sigma_{\hat{\mu}_X}} = \frac{8-7}{0.6667} = 1.5 \right) =$$

$$1 - \mathrm{P}(\mathrm{Z} \le \mathrm{z}_1 = 1.5) = 1 - F(\mathrm{Z} = 1.5) = 1 - 0.9332 = 0.0668 \Rightarrow \alpha = 6.68\ \%$$



(3.) Zehaztu II. motako errorea egiteko probabilitatea itxarote-denboraren batezbesteko erreala 10 minutukoa balitz. **(3 puntu)**.

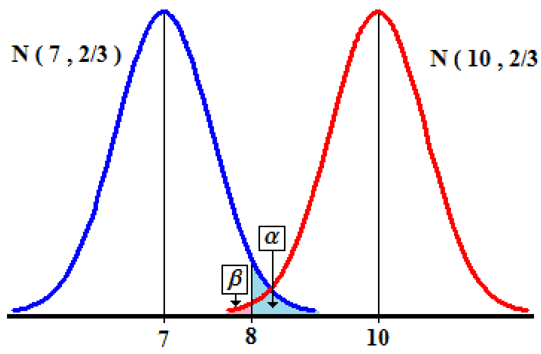
II. motako errorea kalkulatzeko ondorengoa erabiliko da:

$$\overline{X} \sim \mathrm{N}\left(\mu = 10 \text{ min}, \sigma_{\hat{\mu}_X} = \frac{\sigma_X}{\sqrt{n}} = \frac{2}{\sqrt{9}} = 0.6667 \text{ min} \right)$$

$$\beta = \mathrm{P}(II) = \mathrm{P}(H_0 \; onartu \,|\, H_a \; egia) = \mathrm{P}(\overline{X} < 8 \text{ min} \,|\, H_a \; egia) =$$

$$= \mathrm{P}\left(\overline{X} < |\, \overline{X} \sim \mathrm{N}\left(\mu = 10, \sigma_{\hat{\mu}_X} = 0.6667\right)\right) = \mathrm{P}\left(Z < \frac{\overline{x} - \mu}{\sigma_{\hat{\mu}_X}} = \frac{8-10}{0.6667} = -3 \right) = F(\mathrm{Z} = -3) \Rightarrow \beta = \%1.4$$

Kontrastearen potentzia $Potentzia = 1 - \beta = \%99.86$ izanik.



(4.) Zenbat bidaietan (n) ebaluatu beharko du itxarote-denbora, batezbesteko denbora erreala eta n bidai horien batezbesteko denboraren arteko diferentzia ± 1 minutu baino txikiagoa izan dadin, %95 edo gehiagoko probabilitatearekin? **(3 puntu)**.

Estimazioaren errore estandarra (edo desbiderazio tipikoa) $\sigma_{\hat{\mu}_X} = \dfrac{\sigma_X}{\sqrt{n}}$ da, estimazioaren errorea ondorengo eran lor daiteke:

$$P\left(\left|\mu_X - \hat{\mu}\right| \le z_{\alpha/2}\sigma_{\hat{\mu}_X}\right) = 0.95 \Rightarrow P\left(\left|\mu_X - \hat{\mu}\right| \le z_{\alpha/2}\, \sigma/\sqrt{n}\right) = 0.95$$

Ondorioz errorea: $\varepsilon = z_{\alpha/2}\dfrac{\sigma_X}{\sqrt{n}} \le 1 \Rightarrow 1.96\dfrac{2}{\sqrt{n}} = 1 \Leftrightarrow \mathrm{n} = \left(\dfrac{2\times 1.96}{1}\right)^2 = 15.3664$

Hortaz, laginak n=16 elementu izan beharko lituzke.

Beste era batera:

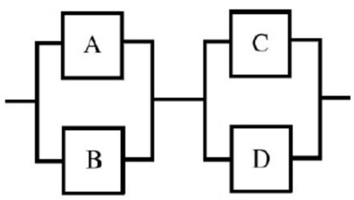
$\overline{X}$ zorizko aldagaiaren banaketa $\overline{X} \sim \mathrm{N}\left(\mu_X, \sigma_{\hat{\mu}_X} = \dfrac{\sigma_X}{\sqrt{n}}\right)$ dela kontuan izanik, eta enuntziatuko informazio erabiliz:

$$\mathrm{P}\left(\overline{X} < \mu_X - 1\right) = 0.025 \Rightarrow \mathrm{P}\left(Z < \frac{(\mu_X - 1) - \mu_X}{\sigma_{\hat{\mu}_X}} = \frac{-1}{2/\sqrt{n}}\right) = 0.025 \Rightarrow \frac{-1}{2/\sqrt{n}} = -1.96 \Rightarrow$$

$$n = (2\times 1.96)^2 = 15.3664$$

3 ARIKETA

4 osagai independentez osatutako sistema batek ondoko eskema jarraitzen du:



X zorizko aldagaiak A osagaiaren bizi iraupena definitzen du, ordutan, bere dentsitate funtzioa ondorengoa izanik

$$\varphi_A(\mathrm{t})=\begin{cases} K\,t & 0\le t<5 \\ \frac{2}{5}-K\,t & 5\le t<10 \\ 0 & \text{beste kasuetan}\end{cases}$$

Oharra: Erantzun guztiak zenbakizko balioa izan behar dute.

$\varphi_A(.)$ funtzioa dentsitate funtzioa izateko **(a)** $\varphi_A(t)\ge 0,\ \forall t\in \mathrm{R}$, eta **(b)** $\int_{-\infty}^{\infty}\varphi_A(\tau)\,\mathrm{d}\,\tau=1$ baldintzak bete behar ditu. Lehenik eta behin hori betetzeko zein baldintza bete behar diren aztertuko da: **(2 puntu)**.

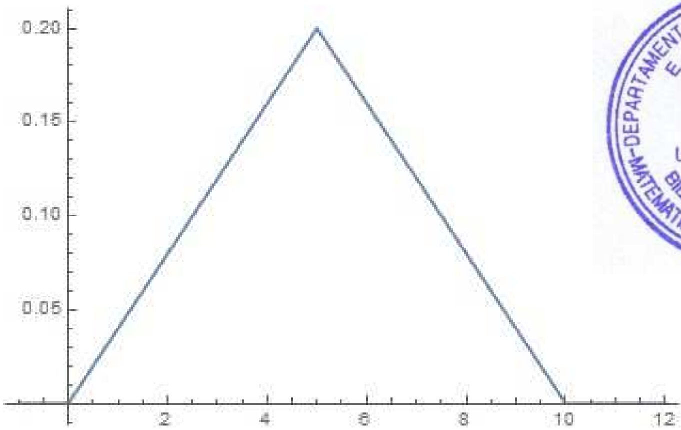
$$\int_{-\infty}^{\infty}\varphi_A(\tau)\,\mathrm{d}\,\tau=\left(\underbrace{\cancel{\int_{-\infty}^{0}}}_{[3]}+\int_0^5+\int_5^{10}+\underbrace{\cancel{\int_{10}^{\infty}}}_{[3]}\right)\varphi_A(\tau)d\tau=\underbrace{\int_0^5 K\,\tau\,d\tau}_{[1]}+\underbrace{\int_5^{10}\left(\frac{2}{5}-K\,\tau\right)d\tau}_{[2]}=1$$

Parte hartzen duten integralak kalkulatuz:

$$\int_0^5 K\,\tau\,d\tau+\int_5^{10}\left(\frac{2}{5}-K\,\tau\right)d\tau=\frac{K}{2}\left(x^2\Big|_0^5\right.+\left(\frac{2}{5}x-\frac{K}{2}x^2\Big|_5^{10}\right.=\frac{K}{2}\left(5^2-0\right)+\left(\frac{2}{5}10-\frac{K}{2}10^2-\frac{2}{5}5+\frac{K}{2}5^2\right)=$$

$$+\left(-\frac{K}{2}10^2-\frac{2}{5}5+\frac{K}{2}5^2\right)=\frac{5^2K}{2}+2-\frac{75K}{2}=1\Rightarrow 1=\frac{50K}{2}\Rightarrow K=\frac{2}{50}=\frac{1}{25}$$

Adierazpen grafikoa ondokoa izanik:



(1.) Zein da A osagaia lehenengo 5 orduetan ez matxuratzeko probabilitatea? **(4 puntu)**.

$$\mathrm{P}(X>5)=1-F_A(5)=1-\mathrm{P}(X<5)=1-\int_0^5\frac{1}{25}\tau\,d\tau=1-\frac{1}{50}\left(x^2\Big|_0^5\right.=1-\frac{1}{50}\left(5^2-0\right)=1-0.50=0.5$$

Beste era batera:

$$\mathrm{P}(X>5)=\int_5^{10}\left(\frac{2}{5}-\frac{1}{25}\tau\right)d\tau=\left(\frac{2}{5}x-\frac{1}{25}x^2\Big|_5^{10}\right.=\left(\frac{2}{5}10-\frac{1}{25}10^2-\frac{2}{5}5+\frac{1}{25}5^2\right)=0.50$$

(2.) C eta D osagai bakoitza, banaka, lehenengo 5 orduetan matxuratzeko probabilitatea 0.1-ekoa bada, kalkulatu gutxienez bi osagaietatik batek denbora tarte horretan era egokian lan egiteko probabilitatea. **(3 puntu)**.

Enuntziatuko informazioa erabiliz, ondorengoa gertaerak definitzen dira:

$$mC \triangleq "C\ osagaiak\ 5\, orduz\, era\, egokian\, lan\, egiten\, du"$$
$$mD \triangleq "D\ osagaiak\ 5\, orduz\, era\, egokian\, lan\, egiten\, du"$$

Gainera: $\mathsf{P}(mC) = \mathsf{P}(mD) = 0.9$. Ondorengoa kalkulatu behar da:

$$\mathsf{P}(mC \cup mD) = \mathsf{P}(mC) + \mathsf{P}(mD) - \mathsf{P}(mC \cap mD) = 0.9 + 0.9 - 0.9 \times 0.9 = 0.99$$

Bilduraren probabilitatearen definizioa erabiliz, eta mC eta mD gertaerak independenteak direla kontuan izanik. (hau da, $\mathsf{P}(mC \cap mD) = \mathsf{P}(mC) \times \mathsf{P}(mD)$).

(3.) B osagaiak 5 ordu baino gehiago era egokian lan egiteko probabilitatea 0.25 da. Kalkulatu sistema osoak (ABCD) lehenengo bost orduetan era egokian lan egiteko probabilitatea. **(3 puntu)**. Kontuan hartu, beharrezkoa denean, aurreko ataletan era egokian lan egiteko kalkulatutako probabilitateak

Izan bitez:

$$mA \triangleq "A\ osagaiak\ 5\, orduz\, era\, egokian\, lan\, egiten\, du"$$
$$mB \triangleq "B\ osagaiak\ 5\, orduz\, era\, egokian\, lan\, egiten\, du"$$
$$Sb1 \triangleq "S1\ osagaiak, A\ eta\ B\ osagaiek\ paraleloan\, sortzen\, dutenak,\ 5\, orduz\, era\, egokian\ lan\, egiten\, du"$$
$$Sb2 \triangleq "S2\, osagaiak, C\ eta\ D\ osagaiek\ paraleloan\, sortzen\, dutenak,\ 5\, orduz\, era\, egokian\ lan\, egiten\, du"$$

Ondorioz: $\mathsf{P}(ABCD) = \mathsf{P}(Sb1 \cap Sb2) = \mathsf{P}(Sb1) \times \mathsf{P}(Sb2)$, Sb1 eta Sb2 osagaiak independenteak baitira. Aurreko ataletatik $\mathsf{P}(mA) = 0.50$ (1. atala) eta $\mathsf{P}(Sb2) = \mathsf{P}(mC \cup mD) = 0.99$ (2. atala) ditugu. Era berean:

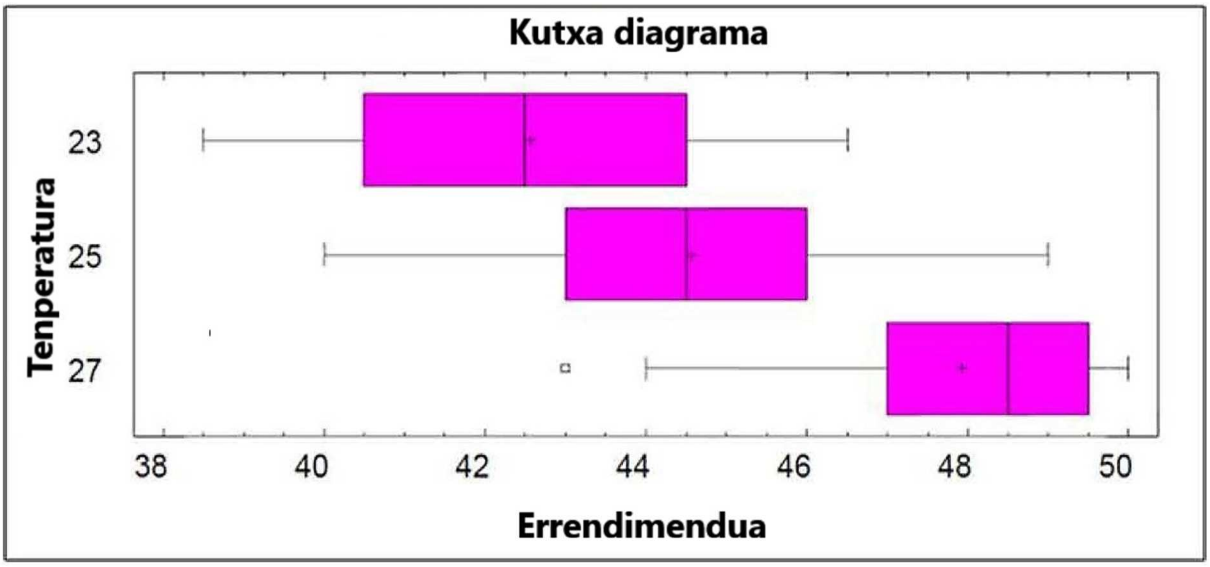
$$\mathsf{P}(Sb1) = \mathsf{P}(mA \cup mB) = \mathsf{P}(mA) + \mathsf{P}(mB) - \mathsf{P}(mA \cap mB) = 0.50 + 0.25 - 0.50 \times 0.25 = 0.625$$

Hortaz:

$$\mathsf{P}(ABCD) = \mathsf{P}(Sb1 \cap Sb2) = 0.99 \times 0.625 = 0.61875$$

4 ARIKETA

LECHES S.L.I. enpresak, hartzidura prozesu baten bidez, gatzagi industriala ekoizten du, 23 ºC-tan. Baldintza hauetan 30 lote lortu ondoren, tenperatura 25 ºC-tara igotzea erabakitzen da. Azkenik, beste 30 lote lortu ondoren, tenperatura 27 ºC-tara igotzen da eta beste 30 lote ekoizten dira. Errendimenduko 90 datu hauekin, ondoren agertzen den Tukey-ren diagrama anizkoitza lortzen da.



Ondorengo galderei modu arrazoituan erantzun:

(1.) 23 ºC-tan errendimenduan lortutako P_{75}-ren balioa 25 ºC-tan lortutako errendimenduaren bigarren kuartilaren balioa baino handiagoa al da? **(2 puntu)**.

75. pertzentila (edo hirugarren kuartila $P_{75} = Q_3$) kutxaren goiko muga da. Ondorioz, T=23ºC denean $P_{75} = Q_3 = 44.5$ da (emandako irudian ikus daiteke). Era berean, T=25ºC denean bigarren kuartila (mediana edo 50. pertzentila) $P_{50} = Q_2 = 44.5$ da (emandako irudia begiratuz). Ondorioz, biak berdinak dira.

(2.) T = 23 ºC eta T = 25 ºC diagramak konparatuz, zeinetan aurkitzen dugu datuen dispertsio handiagoa? **(2 puntu)**.

Heina (R) eta kuartilarteko heina (RIC) sakabanaketa absolutua neurtzen duten estatistikoak dira. Irudia begiratuz ondorengo kalkuluak egin daitezke:

$$R_{T=23°C} = 46.5 - 38.5 = 8 \qquad R_{T=25°C} = 49 - 40 = 9$$
$$RIC_{T=23°C} = 44.5 - 40.5 = 4 \qquad RIC_{T=25°C} = 46 - 43 = 3$$

Ondorioz, T = 23 ºC denean heina txikiagoa da baina RIC handiagoa da. RIC sakabanaketa neurtzen duen estimatzaile sendoagoa dela kontuan izanik, T = 23 ºC kasuan sakabanaketa handiago dela ondoriozta daiteke.

(3.) Hiru kasuetatik, zer kasutan da txikiagoa asimetria koefizientea? **(2 puntu)**.

Erantzuna justifikatzeko Pearson-en asimetrikoa koefizientea erabiliko da:

$$As = \frac{\overline{x} - Mo}{s} \approx \frac{3(\overline{x} - Me)}{s}$$

T = 23 ºC eta T = 25 ºC kasuetan kutxa-diagrama nahiko simetrikoa da, ondorioz, kasu hauetan asimetria koefizientea zerotik hurbil egongo da. Bestalde, T = 27 ºC denean kutxa-diagramari ezker asimetria negatiboa dela ikus daiteke (ezkerreko bibotea eskuinekoa baino luzeagoa da, eta mediana batezbestekoa baino handiagoa). Ondorioz, asimetria koefiziente txikiena T = 27 ºC kasukoa da (negatiboa, alegia).

(4.) Ordenatu, modu arrazoituan, proposatutako hiru laginen kurtosia. **(2 puntu)**

Kutxa-diagramak erabiliz, leptokurtikoagoa denetik (zorrotzagoa denetik), platikurtikoagoa deneraino (leunagoa deneraino) ordena ondorengoa dela ondoriozta daiteke: (1.) T = 27 ºC, (2.) T = 25 ºC, (3.) T = 23 ºC. Hau, RIC txikiagoa delako betetzen da, honek kontzentrazioa altuagoa ematen baitu. Bestalde, heinak lagineko balioak nondik nora mugitzen diren ulertzeko balio du.

(5.) Zer esan daiteke planteatutako hiru egoeren desbiderazio estandarrei buruz? **(2 puntu)**.

Desbiderazio estandarraren (edo tipikoaren) adierazpena $s_X = \left[\frac{1}{n} \sum_{i=1}^{N} f_i \left(x_i - \overline{x} \right)^2 \right]^{1/2}$ denez, balio hau lagineko datuak (serie estatistikoa) erabiliz soilik kalkula daiteke. Baina intuitiboki, aurreko ataletan erabilitako justifikazio bera erabiliz (RIC erabiliz), txikienetik handienera jarriz ordena: (1.) T = 27 ºC, (2.) T = 25 ºC, (3.) T = 23 ºC da.

$RIC_{T=27°C} = 49.5 - 47 = 2.5$ $\qquad$ $RIC_{T=25°C} = 46 - 43 = 3$ $\qquad$ $RIC_{T=23°C} = 44.5 - 40.5 = 4$