eman ta zabal zazu

EUSKAL HERRIKO UNIBERTSITATEA
UNIVERSIDAD DEL PAIS VASCO

BILBOKO INGENIARITZA ESKOLA

ESCUELA DE INGENIERÍA DE BILBAO

MATEMATIKA APLIKATUA

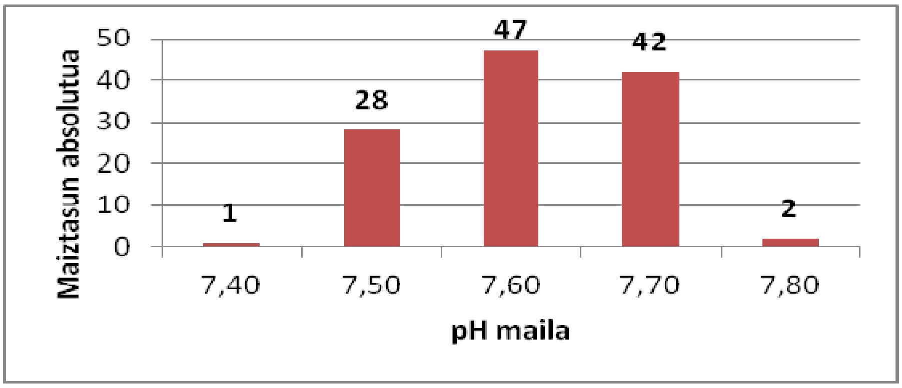
INDUSTRIA INGENIARITZA TEKNIKOKO ATALA
SECCIÓN INGENIERÍA TÉCNICA INDUSTRIAL

INGENIARITZAKO METODO ESTATISTIKOAK

LEHENENGO DEIALDIKO **EBAZPENA:** 2016-2017 KURTSOA

Kalifikazio data: 2017ko ekainak 6 (arratsaldeko 17:00etan, G.A.U.R.-en)
Berrikuspen data: 2017ko ekainak 9 (goizeko 10:00etan, 7I1 gelan)

Uda sasoian mendian dagoen herri bateko udal-igerilekua betetzeko ur alkalinoa ($pH > 7$) duen lurpeko iturburu bat erabiltzen da. Urari osasunarentzako kaltegarria ez den klorazio-prozesu bat aplikatu behar zaio, gaixotasunen eragile diren mikroorganismoak deuseztatzeko. Hala ere, klorazio-prozesu hau eraginkorra izateko igerilekuko uraren pH -a $pH \in (6.5, 8)$ izan behar da. Are gehiago, $pH > 7.6$ denean, kloroaren efektua ez da eraginkorra izango, eta pH -a erregulatzeko produkturen bat gehitu beharko zaio. Horrela, eta erabilitako uraren jatorria kontutan izanda, betetze-prozesuan alkalinotasun maila neurtzeko orduko batezbeste 4 datu hartzen dira Poisson-en banaketa bat erabiliz. Hurrengo irudian egindako zorizko laginketan neurtutako pH -en balioen maiztasun absolutuak agertzen dira:



1. ARIKETA

Neurtutako pH mailekin, zehaztu arrazoituz:

(1.) Laginean datua atipikorik existitzen den, eta irudikatu kutxa-diagrama. (4 puntu)

Balio atipikoak lortzeko, kuartilak eta kuartilarteko heina kalkulatuko ditugu lehendabizi:

$$\boxed{lagina = \{x_i\} \quad non \quad i = 1, 2, \cdots, 120}$$

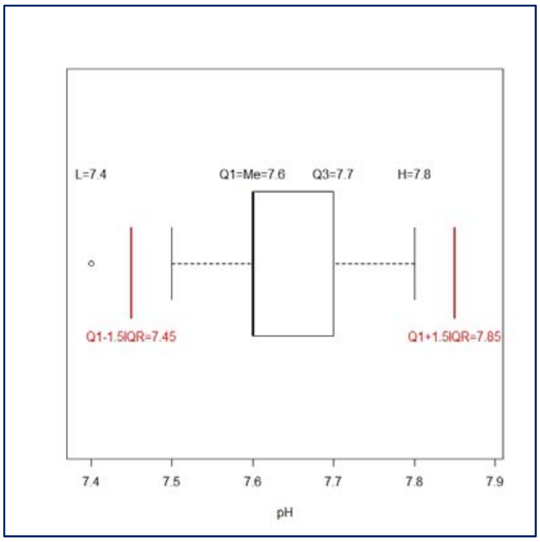
- balio minimoa: $L = \min\{x_i\} = 7.4$
- balio maximoa: $H = \max\{x_i\} = 7.8$
- mediana: $Me = \dfrac{x_{60} + x_{61}}{2} = 7.6$
- lehenengo kuartila: $Q_1 = \dfrac{x_{30} + x_{31}}{2} = 7.6$
- hirugarren kuartila: $Q_3 = \dfrac{x_{90} + x_{91}}{2} = 7.7$
- $IQR = Q_3 - Q_1 = 7.7 - 7.6 = 0.1$

Ondore, barne hesiak kalkulatuko ditugu:

- $m_1 = Q_1 - 1.5 \cdot IQR = 7.6 - 0.15 = 7.45$
- $m_3 = Q_3 + 1.5 \cdot IQR = 7.7 + 0.15 = 7.85$

Beraz, balio atipiko bezala kontsideratuko dugu: $\boxed{x_1 = 7.4 \notin [7.45, 7.85]}$

Laginari dagokion Tukey-ren kutxa-diagrama edo *boxplot*-a, *R*-rekin, honakoa da:



1. Irudia. Lortutako laginaren kutxa diagrama

(2.) Igerilekua beteta ikusi nahi duen ume baten arrazonamendua egokia den: “*orduko 4 neurketetako batezbestekoa errespetatu bada, igerilekua betetzeko behar izan den denbora 30 ordukoa da*”. (2 puntu)

Igerilekuaren kapazitatea eta bera betetzeko erabilitako ur-fluxua jakin gabe, nekez jakin dezake umeak igerilekua betetzeko behar den denbora.

Umearen arrazoibide intuitiboa honakoa dela ematen du: igerilekuaren betetze-prozesuan 120 datu hartu badira, eta orduko 4 datuko batezbestekoarekin, orduan igerilekua betetzeko 30 ordu behar izan dira:

$$\frac{120\, neurketa}{4 \; {}^{neurketa}\!/_{ordu}} = 30\, ordu$$

Hala ere, igerilekua betetzeko behar den denbora ezin da horrela zehaztu, datuak zoriz hartzen baitira, Poisson-en banaketa jarraituz. Beraz, umearen arrazonamendua ez da zuzena. Gainera, ordu zehatz batean ez dago zergatik 4 datu hartu beharrik, izan ere, hau gertatzeko probabilitatea, $X =$ *”ordu batean jasotako neurketa kopurua”* izanik, hauxe da:

$$P(X = 4) = 0.1954 \quad non \quad X \sim \mathscr{P}(4)$$

Honetaz gain, $T=$ *”igerilekua betetzeko behar den denbora, ordutan”* zorizko aldagai jarraitua definitzen bada, balio konkretu bat hartzeko probabilitatea zero da:

$$P(T = 30) = 0$$

(3.) Igerilekua betetzeko 30 ordu baino gehiago behar izateko probabilitatea, igerilekua betetzeko erabiltzen den ur-fluxuak (*l*/*s*-tan) $N(6.5, 0.5)$ banaketa jarraitzen badu, eta igerilekuaren bolumena $750\, m^3$ -koa bada. (4 puntu)

1. ERA

Izan bedi S := "*Ur-fluxua litro/segundu-tan*" zorizko aldagaia

$$S \sim N\left(6.5\, {}^{l}/_{s}, 0.5\, {}^{l}/_{s}\right)$$

Igerilekua 30 ordutan betetzeko fluxu maximoaren limitea honako izango litzateke:

$$s_m = \frac{750 m^3}{30\, h} \frac{1\, h}{3600\, s} \frac{1000 l}{1\, m^3} = 6.9\hat{4}$$

Honela, igerilekua betetzeko 30 ordu baino gehiago behar izateko probabilitatea:

$$P(T > 30) = P(S < 6.94)$$

S tipifikatuz:

$$Z = \frac{S - 6.5}{0.5} \sim N(0,1) \Rightarrow P(T > 30) = P\left(Z < \frac{6.94 - 6.5}{0.5}\right) = P(Z < 0.8889)$$

Taula erabiliz, interpolatu gabe, balio hau 0.8133 litzateke (R erabiliz 0.81296), beraz, gutxi-gorabehera:

$$P(T > 30h) \approx 81.3\%$$

2. ERA

Izan bedi S = "*ur-fluxua litro/segundu-tan*" zorizko aldagaia:

$$S \sim N(6.5, 0.5)$$

Unitateak aldatuz, H = "*ur-kantitatea litro/ordu-tan*" zorizko aldagaia daukagu:

$$H \sim N(23400, 1800)$$

Igerilekua betetzeko behar diren 30 orduak kontutan hartuz, Y = "*ur-kantitatea 30 ordutan, litrotan*" zorizko aldagaia definitzen da, non, banaketa normalaren batukortasun-propietatea erabiliz (zorizko aldagai baten barientza eta itxaropen matematikoaren propietateen ondorio):

$$X_i \sim N(\mu, \sigma),\ 1 \le i \le n \text{ ,n zorizko aldagai independentе: } \rightarrow \sum_{i=1}^{n} X_i \sim N\left(\mu = \sum_{i=1}^{n} \mu_i,\ \sigma = \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

Honakoa ondoriozta daiteke:

$$Y = \sum_{i=1}^{30} H_i \quad \Rightarrow \quad Y \sim N(702000, 54000)$$

Igerilekua betetzeko 30 ordu baino gehiago behar izateko probabilitatea kalkulatu behar da, emandako ur-fluxuarekin. Igerilekuaren kapazitatea $750 m^3 = 750000 l$ denez, zehaztutako denbora tartean ur-kantitate hori ez lortzeko probabilitatea kalkulatzen da:

$$P(Y \le 750000)$$

Aldagaia tipifikatuz:

$$\boxed{P(Y \le 750000) = P\left(Z \le z = \frac{750000 - 702000}{54000}\right) = P(Z \le z = 0.8889) = 0.81297}$$

30 ordutan $750 m^3 = 750000 l$ ez lortzeko probabilitatea % 81.30 da.

2. ARIKETA

Izan bedi X = *"ordu batean neurtutako uraren pH mailaren datu kopurua"* zorizko aldagaia.

(1.) Irudikatu eta taula gisa jarri X zorizko aldagaiaren probabilitate funtzioa. (2.5 puntu)

X zorizko aldagai diskretuak Poisson-en probabilitate banaketa jarraitzen du, bere parametroa batazbestekoa izanik, hau da:

$$E[X] = \lambda = 4 \quad \Rightarrow \quad \boxed{X \sim \mathscr{P}(4)}$$

X zorizko aldagai diskretu baten probabilitate funtzioa honela definitzen da:

$$\boxed{p(x_i) = P(X = x_i) \quad \forall x_i}$$

Poisson banaketantzako, formula honakoa da:

$$\boxed{P(X = i) = \left.\frac{e^{-\lambda}\cdot\lambda^i}{i!}\right|_{\lambda=4} = \frac{e^{-4}\cdot 4^i}{i!}}$$

Enuntziatuan emandako script-a erabiliz (ikusi 2. irudia) eta bertan ez dauden balioak kalkulatuz, eskatutako taula lortuko da:

- $P(X=0) = \frac{e^{-4}\cdot 4^0}{0!} = 0.0183$
- $P(X=1) = \frac{e^{-4}\cdot 4^1}{1!} = 0.0733$
- $P(X=2) = \frac{e^{-4}\cdot 4^2}{2!} = 0.1465$
- $P(X=3) = \frac{e^{-4}\cdot 4^3}{3!} = 0.1954$
- $P(X=4) = \frac{e^{-4}\cdot 4^4}{4!} = 0.1954$

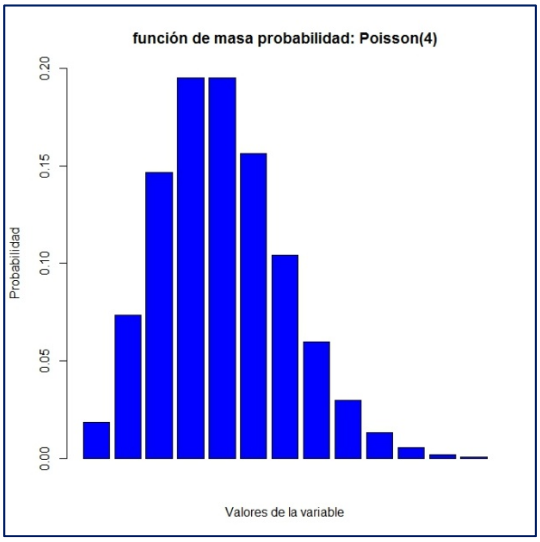
```
>    X<-seq(5,12)
>    prob<-dpois(X,4)
>    data.frame(X,prob)
   X         prob
1  5 0.1562934519
2  6 0.1041956346
3  7 0.0595403626
4  8 0.0297701813
5  9 0.0132311917
6 10 0.0052924767
7 11 0.0019245370
8 12 0.0006415123
```

2. Irudia. Azterketaren enuntziatuan emandako Script-a

X zorizko aldagaiaren probabilitate funtzioa					$X \sim \mathscr{P}(4)$
x_i		x_i		x_i	
0	0.0183	5	0.1563	10	0.0053
1	0.0733	6	0.1042	11	0.0019
2	0.1465	7	0.0595	12	0.0006
3	0.1954	8	0.0298		
4	0.1954	9	0.0132		

1. Taula. X zorizko aldagaiaren probabilitate funtzioa

Eskatutako adierazpen grafikoa, R-rekin eginez:



3. Irudia. X zorizko aldagaiaren probabilitate funtzioa

(2.) Kalkulatu:

a. Bi ur neurketen artean 20 minutu baino gehiago igarotzeko probabilitatea. (3 puntu)

Izan bitez honako zorizko aldagaiak:

- $X =$" *ordu batean neurtutako uraren pH mailaren datu kopurua*"
- $T =$"*bi neurketen arteko denbora, ordutan*"

$X \sim \mathscr{P}(4)$ denez, orduan: $\boxed{T \sim \varepsilon(4)}$

Eskatutako probabilitatea hau da:

$$P\left(T > \frac{1}{3}\right) = 1 - P\left(T \leq \frac{1}{3}\right) = e^{-\frac{4}{3}} \quad \Rightarrow \quad \boxed{P\left(T > \frac{1}{3}\right) = 0.2636}$$

b. Ordu batean batezbestekoa baino neurketa gehiago hartzeko probabilitatea. (1.5 puntu)

$X \sim \mathscr{P}(4)$ denez, eta gainera, $E[X] = \lambda = 4$:

$$P(X > 4) = 1 - P(X \leq 4) = 1 - \sum_{i=0}^{4} P(X = i) \quad \Rightarrow \quad \boxed{P(X > 4) = 0.3712}$$

---

c. Orduro zehazki 4 neurketa egiteko probabilitatea igerilekuaren betetze-prozesu guztian zehar. (3 puntu)

Igerilekua betetzeko r ordu behar direla suposatuz $(r > 0)$ eta ordu ezberdinetan jasotako neurketak independenteak direla kontutan hartuz, eskatutako probabilitatea honela calculatzen da:

$$\boxed{\prod_{j=1}^{r} P(X_j = 4) = P(X_1 = 4) \cdot P(X_2 = 4) \cdot \quad \ldots \quad \cdot P(X_r = 4) = (0.1954)^r}$$

Igerilekua betetzeko 30 ordu baino gehiago behar direla dakigunez, $r=30$ denean honako balioa lortzen da:

$$\prod_{j=1}^{30} P\left(X_{30}=4\right)=\left(0.1954\right)^{30}=5.32\times10^{-22}$$

r-ren balioa handitzen den heinean, eskatutako probabilitatea txikiagotu egiten da, ia nulua dela kontsidera daiteke.

---

3. ARIKETA

Laginketan lortutako datuak kontutan hartuz:

(1.) Lortu igerilekuko uraren pH mailaren batezbestekoaren eta desbiderazio tipikoaren estimazio puntuala. (2 puntu)

Laginaren tamaina: $n=120$

Populazioaren batezbestekoaren estimazio puntuala bere estimatzaile alboragabea den laginaren batezbestekoaren bidez lortzen da; beraz:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{5} n_i\cdot x_i=\frac{913.6}{120}\quad\Rightarrow\quad\boxed{\overline{x}=7.6133}$$

Populazioaren desbiderazio tipikoaren estimazio puntuala bere estimatzaile alboragabea den laginaren desbiderazio tipikoaren bidez lortzen da; beraz:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2\cdot n_i=\frac{0.7987}{119}=0.0067\quad\Rightarrow\quad\boxed{S=\sqrt{S^2}=0.0819}$$

(2.) Erabaki % 98ko konfidantza mailaz, kloroa eraginkorra izateko eta igerilekuko urari pH -a murrizteko produkturen bat gehitu behar zaion. (3 puntu)

urari pH -a murrizteko produkturen bat gehitu behar zaio erabaki behar da. Hau $pH>7.6$ denean egin behar da. Beraz, $\boxed{\alpha=0.02}$ adierazgarritasunarekin, alde bakarreko hipotesi kontrastea egitea proposatzen da, honako hipotesiekin,

- Hipotesi nulua: $\boxed{H_0:\mu\le 7.6=\mu_0}$
- Hipotesi alternatiboa: $\boxed{H_1:\mu>7.6=\mu_0}$

Laginaren tamaina $n=120\ge 30$ denez, eta populazioaren desbiderazio tipikoa ezezaguna, kontrastearen estatistikoa honakoa da:

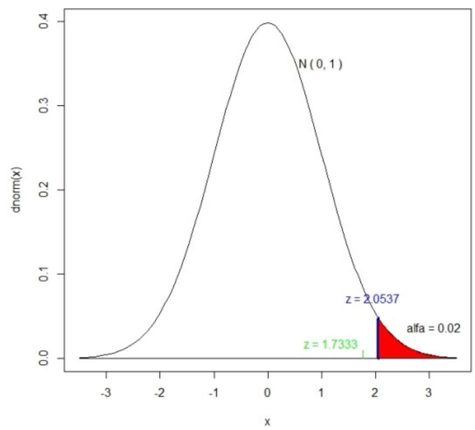
$$\overline{X}\sim N\left(\mu_0,\frac{S}{\sqrt{n}}\right)\quad\Rightarrow\quad\boxed{\overline{X}\sim N\left(7.6,0.0075\right)}$$

Tipifikatuz: $$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}=Z\sim N\left(0,1\right)$$

Eskualde kritikoa *EK*, emandako adierazgarritasunarekin, honakoa da: $\boxed{EK=\left\{Z>z_\alpha=2.0537\right\}}$

Emandako R-ko funtzioa erabili da: `qnorm(0.02)=-2.0537`

Kontrastearen estatistikoaren balioa: $\boxed{z = \dfrac{\overline{x} - \mu_0}{S/\sqrt{n}} = 1.7733}$



4. irudia. Kontrastearen eskualde kritikoa

$\boxed{z = 1.7733 \notin RC = \left\{ Z > z_\alpha = 2.0537 \right\}}$ dagoenez, lagineko datuen arabera ez dago ebidentzia estatistiko nahikorik hipotesi nulua errefusatzeko, emandako adierazgarritasunarekin $(\alpha = 0.02)$.

Beraz, uraren pH maila murrizteko produkturik ez gehitzeko erabakia hartzen da.

(3.) Kalkulatu aurreko kontrastearen p-balioa eta interpretatu lortutako balioa. (2.5 puntu)

p-balioa honela definitzen da:

$$p - balioa = P\left(\overline{X} > 7.6133 \,|\, H_0 \; egia\right) = P\left(\overline{X} > 7.6133 \,|\, \overline{X} \sim \left(7.6, 0.0075\right)\right)$$

Tipifikatu ondoren:

$$\boxed{p - balioa = P\left(Z > 1.7333 \,|\, Z \sim \left(0,1\right)\right) = 0.0381}$$

Emandako R-ko funtzioa erabili da: `pnorm(1.7333)=0.9585`

$0.01 \leq p - balioa \leq 0.05$ denez, lortutako laginarekin ebidentzia handiak daude H_0 errefusatu ahal izateko, izan ere adierazgarritasun maila $\boxed{\alpha = 0.0381}$ da. Hau da, hipotesi nulua errefusa daiteke, gutxi gorabehera, % 96ko konfidantza mailarekin.

Kasu honetan, uraren pH maila murrizteko produktua gehitzea beharrezkoa izango zen, honela klorazio-prozesua eraginkorra izateko.

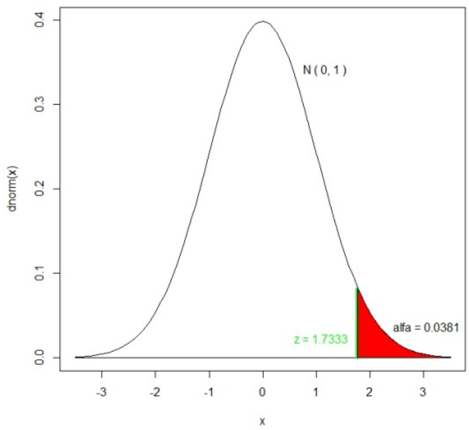


Figura 5. irudia. p-balioaren adierazpena

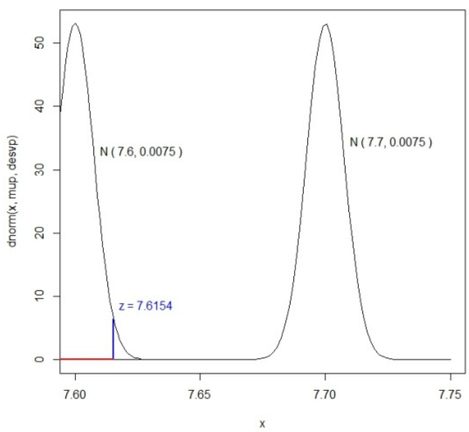
(4.) Kalkulatu kontrastearen potentzia, uraren pH -aren batezbestekoa $\mu = 7.7$ denean. (2.5 puntu)

Kontrastearen potentzia $\boxed{1-\beta}$ bezala kalkulatzen da, non

$$\beta = P\left(II.\, motako\ errorea\right) = P\left(H_0\ onartu \,|\, H_0\ gezurra\right) = P\left(\bar{X} < 7.6154 \,|\, \bar{X} \sim N\left(7.7\,,\, 0.0075\right)\right)$$

Tipifikatuz: $\beta = P\left(Z < -11.28 \,|\, Z \sim N\left(0,1\right)\right) = 8.24 \times 10^{-30}$

Orduan, kasu honetan: $\boxed{1-\beta \approx 1}$



6. irudia. Kontrastearen potentziaren adierazpidea
