8.Gaia

Inferentzia Estatistikoa:

Estimazioa

Sarrera

Adibidea

Komunikazio sare batean 0 bidalita 1 seinalea jaso daiteke zarata dela eta.
Zein proportziotan gertatzen da?

Lehenik eta behin egoera formalizatuko dugu.

Populazioa: Komunikazio sarean bidalitako edo bidaliko diren 0 seinale guztiak.

$X=0 \Rightarrow$ 0 bidali eta 0 jaso
$X=1 \Rightarrow$ 0 bidali eta 1 jaso

Ondorioz, Bernoulliren banaketa erabili dezakegu.

$X \sim Bernoulli(p)$ Baina p ez dakigu zenbat den

Horrela, datuetan begiratuta aterako dugu p

Datuak

$x_1=0 \quad x_2=1$
$x_3=0 \quad x_4=0$ $\quad \hat{p}=1/4$

Estimazioa p-ren hurbilketa bat egitean datza, datuetatik abiatuta.

Estatistika parametrikoa

Egoera orokorra:

$X \sim f(x, \theta)$ non θ parametro ezezaguna den

Populazioa eta lagina

Aurreko adibidera itzuliko gara

Populazio-eredua	Lagina
$X \sim Bernoulli(p)$	$X_1 \quad X_2 \quad \dots \quad X_n$ $\downarrow \quad \downarrow \quad \quad \quad \downarrow$ $0 \quad 1 \quad \dots \quad 0$

Kaso honetan p ezezaguna dugu, baina badakigu laginak ez digula informazio osoa emango.

Definizioa $X_1, X_2, \dots, X_n$ zorizko aldagaiek zorizko lagin bakuna osatzen dute baldin eta:

i) X_i guztiak elkarrekiko askeak dira.

ii) X_i guztiak berdinki banatuak dira.

Estimatzaileak eta beren propietateak

Definizioa. Izan bedi $X_1, X_2, \dots, X_n$ zorizko lagin bakuna. Aldagai horien gainean osatutako funtzio erreal bati estatistiko esaten zaio.

$$T = T(X_1, \dots, X_n)$$

Adibidea

1) $X \sim f(x, \theta)$ $\quad X_1, \dots, X_n$ (zorizko aldagai bakuna)

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \qquad T(X_1, \dots, X_n)$$

2)

Estatistikoa deskribatzaileratzeko erabilitako edozein estatistiko.

$$T(X_1, \dots, X_n) \sim \text{Banaketa} \qquad \text{Lagin Banaketa}$$

$$\boxed{X_i \sim f}$$

Oharrak:

Estatistikoa zorizko aldagaia da.

Estatistikoaren banaketari lagin-banaketa esaten zaio.

Estatistikoaren lagin-banaketa, jatorrizko populazioaren banaketaren (X_i banaketa) eta lagin-tamainaren menpekoa da.

Adibidea

Demagun populazioaren $X \sim N(\mu, \sigma^2)$ eredua egokia dela. Izan bedi $X_1, X_2, \dots, X_n$ bertatik ateratako zoriazko lagin bakuna. Nolakoa izango da $\bar{X}$ren lagin-banaketa?

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$X_i \sim N(\mu, \sigma^2)$ non $i = 1, \dots, n$

i)

$$E\bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{1}{n}\sum_{i=1}^{n}\mu =$$

$$= \frac{n \cdot \mu}{n} = \mu$$

$$\boxed{\begin{array}{l} EX_i = \mu \\ VAR(X_i) = \sigma^2 \end{array}}$$

ii)

$$VAR(\bar{X}) = VAR\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} VAR\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} VAR(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 =$$

$$= \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Ondorioz, $\bar{X}$ren lagin banaketa honela geratuko da.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

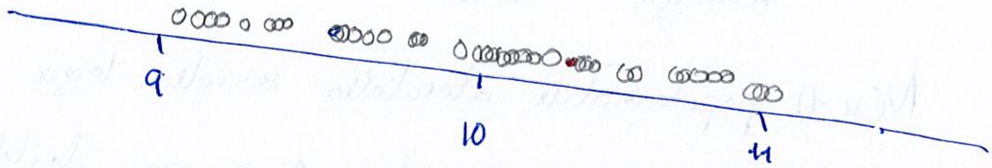
Demagun hau dela eskuartean dugun kasu berezia

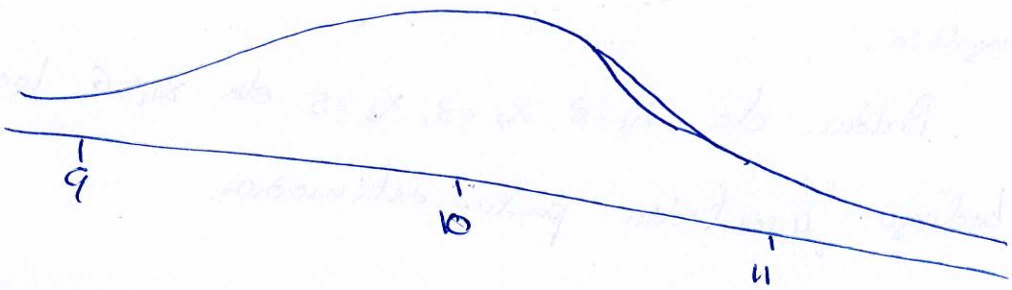
Populazioa: $N(10, 2^2)$

1. lagina 10.0 13.9 11,3 8.3 12.5 6.3 10.0 7.4 14.0 9.4 $\bar{x} = 10.3$

⋮

n. lagina 9.8 7.9 12.6 7.3 12.2 8.5 11.5 8.2 9.8 7.7 $\bar{x} = 9.5$





Orain lehenago kasu orokorretik kasu partikularra aztertuko dugu.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$ kasu orokorra

Populazioa: $N(10, 2^2)$

$$\bar{X} \sim N\left(10, \frac{2^2}{10}\right)$$ lagin-banaketa

Desbiderazio-estandarra kalkulatzeko.

$$SD(\bar{X}) = \sqrt{VAR(\bar{x})} = \sqrt{\frac{2^2}{10}} = \frac{2}{\sqrt{10}} = 0'63$$

Definizioa. Demagun θ parametro ezezaguna duen populaziotik ateratako $X_1, X_2, \ldots, X_n$ zorizko lagin bakuna dugula. $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ estimatzaileak har ditzakeen balioak θ parametroa hurbiltzeko erabili ditzakegu; $\hat{\theta}$ ri θ parametroaren puntu-estimatzailea dela esango dugu.

Adibidea

Demagun $X_1, \ldots, X_n$

$N(\mu, 1)$ populaziotik ateratako zoriko lagin bakuna.

μ estimatu nahi dugu eta $\hat{\mu} = \bar{X}$ izan daiteke estimatzaile egokia.

Baldin eta, $x_1 = 3$, $x_2 = 3$, $x_3 = 5$ eta $x_4 = 6$ lagina baldin badugu μ-rentzako puntu-estimazioa.

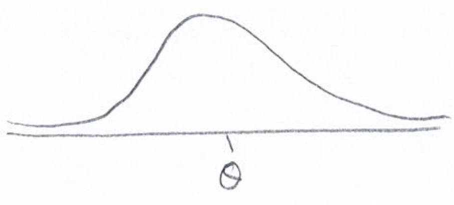
$$\hat{\mu} = \frac{3+3+5+6}{4} = 4'25$$

Adibidea

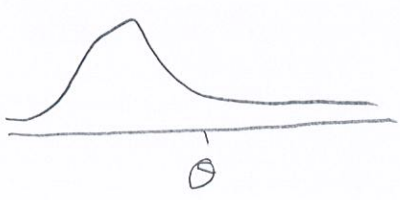
Demagun $\hat{\theta}$ eta θ^* θ-rentzako estimatzaileak ditugu. Zein da hobea?

$\hat{\theta}$ren itxaropena θ izango da baina θ^*rena ez, grafikoetan ikusi dezakegun bezala ondorioz, $\hat{\theta}$ hobea da

$E(\hat{\theta}) = \theta$



$E(\theta^*) \neq \theta$



Definizioa. Demagun θ parametro ezezaguna dela eta $\hat{\theta}$ bere estimatzailea. $\hat{\theta}$ estimatzaile alboragabea dela esango dugu baldin eta

$$E(\hat{\theta}) = \theta \text{ bada.}$$

Propietateak. Izan bedi $X_1, \dots, X_n$ zorizko lagin bakuna, $EX_i = \mu$ eta $VAR(X_i) = \sigma^2$ direlarik $(i = 1, \dots, n)$.

i) $\hat{\mu} = \bar{X}$ estimatzaile alboragabea da μ-rentzat.

$$E(\bar{X}) = \mu$$

ii) $\hat{\sigma}^2 = S_n^2$ lagin-bariantza estimatzaile alboratua da σ^2-rentzat.

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

iii) $\hat{\sigma}^2 = S^2$ kuasibariantza estimatzaile alboragabea da σ^2-rentzat.

$$E(S^2) = \sigma^2$$

Bi estimatzaile desberdin daude σ^2-rentzat, zein da egokiena?

i) Frogatu $\hat{\mu} = \bar{X}$ estimatzaile alboragabea dela μ-rentzat.

$E(\bar{x}) = \mu$

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n \cdot \mu}{\mu}$$

$E(aX) = a\,EX$

$E(X+Y) = EX + EY$

$VAR(aX) = a^2 VAR(X)$

$VAR(X+Y) = VAR(X) + VAR(Y)$ Askeak

$VAR(X) = EX^2 - (EX)^2$

(ii) frogatu $\hat{\sigma}^2 = S^2$ lagin-bariantza estimatzaile alboratua da σ^2-rentzat

$$E(S_n^2) = \frac{n-1}{n}\sigma^2 \Rightarrow E(S_n^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right) = \frac{1}{n}\sum_{i=1}^{n} E(x-\bar{x})^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x^2 - 2x\cdot\bar{x} + \bar{x}^2) = \frac{1}{n}\sum_{i=1}^{n}\left(EX^2 - 2E(x\cdot\bar{x}) + E\bar{X}^2\right) =$$

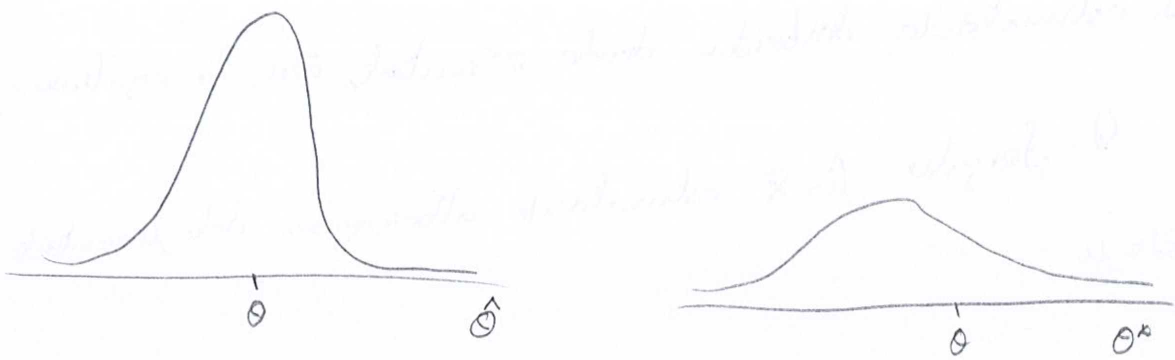
$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2n\bar{x}\bar{x} + n\bar{x}^2\right)\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - n\bar{x}^2\right)\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(Ex_i^2 - E\bar{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2}\right) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} =$$

$$= \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{\sigma^2(n-1)}{n} < \sigma^2$$

$$\boxed{\begin{array}{l} E(x_i^2) = VAR(x_i) + E(x_i)^2 = \sigma^2 + \mu^2 \\ E(\bar{x})^2 = VAR(\bar{x}) + E(\bar{x})^2 = \frac{\sigma^2}{n} + \mu \end{array}}$$

θ Parametro ezezaguna



$E(\hat{\theta}) = \theta$

$E(\theta^*) \neq \theta$ $\qquad$ $VAR(\hat{\theta}) < VAR(\theta^*)$

Definizioa. Izan bitez $\hat{\theta}$ eta θ^*, θ parametro ezezagunaren bi estimatzaile alboragabe. $\hat{\theta}$ estimatzailearen θ^* estimatzailearekiko eraginkortasuna honela neurtzen da.

$$\frac{VAR(\theta^*)}{VAR(\hat{\theta})}$$

Baldin eta zatidura >1 bada, $\hat{\theta}$ estimatzailea θ^* baino eraginkorragoa da.

Estimatzailearen desbideratze estandarrari errore estandarra esaten zaio sarri.

Egoera arruntak

X aldagai kuantitatiboa

Parametro interesgarriak	Estimatzaileak
$E(X)=\mu$	$\hat{\mu}=\bar{X}$
$VAR(X)=\sigma^2$	$\hat{\sigma}^2=S^2$

Lagin-banaketak

Lagin-batezbestekoa $\bar{X}$

$$E\bar{X}=\mu \quad VAR(\bar{X})=\frac{\sigma^2}{n} \quad SD(\bar{X})=\frac{\sigma}{\sqrt{n}}$$

Gainera, lagina handia ($n\geq 30$) bada:

$$\boxed{\bar{X}\approx N(\mu,\sigma^2/n)}$$

Kuasibariantza, S^2

$$E(S^2)=\sigma^2, \quad (*) \ \frac{(n-1)S^2}{n} \sim \chi^2_{n-1}$$

(*) Baldin eta jatorrizko populazioa normala $N(\mu,\sigma^2)$ bada.

Adibidea

$n=40$ $X \equiv$ "Programa baten exekuzio-denbora"

$\hat{\mu}=\bar{X}$ estimazioa kalkulatu.

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{18368}{40} = 459'2 \ \text{(puntu-estimazioa)}$$

Errore estandarra $SD(\bar{X})$ kalkulatu

$$SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{105.48}{\sqrt{40}} = \boxed{16'7}$$

$S=\sqrt{S^2}=105.47$

$S^2=\frac{n}{n-1}S_n^2=11125.34$

$S_n^2=\frac{1}{n}\sum_{i=1}^{n} x_i^2-\bar{x}^2=$

$=\frac{8868474}{40}-\left(\frac{18368}{40}\right)^2=$

$=10847.21$

X aldagai bitarra

$p=P(X=1)$ izanik.

Parametro interesgarria	Estimatzailea
$E(x)=p$	$\hat{p}=\bar{X}$

Lagin-banaketa

Lagin-proportzioa $\hat{p}$

$$E(\hat{p}) = p, \quad VAR(\hat{p}) = \frac{pq}{n}, \quad SD(\hat{p}) = \sqrt{\frac{pq}{n}} \text{ errore estandarra}$$

Gainera, lagin handia bada ($n\hat{p} \geq 5$ eta $n\hat{q} \geq 5$)

$$\boxed{\hat{p} \approx N(p, pq/n)}$$

Adibidea

$n = 150$

$$X_i = \begin{cases} 0 & \text{PZU ez alditibaldita} \\ 1 & \text{alditibaldita} \end{cases}$$

123 tan alditibaldita.

$\hat{p} = \bar{X}$ kalkulatuko da

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{123}{150} = 0'82$$

$\hat{p} = 0'82$

Errore estandarra

$$SD(\bar{X}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0'82 \cdot 0'18}{150}} = 0'031$$

$SD(\hat{p}) = 0'031$

Konfiantza-tarteak

Adibidea

$X \sim N(\mu, V)$ populazioa adierazteko, eta Vren balioa badakigu zein den.

$X \sim N(\mu, V)$

μren estimazioa: $\hat{\mu} = \bar{X} \Rightarrow \bar{X} \sim (\mu, \frac{V}{n})$

Ondorioz, $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{V}{n}}} \sim N(0,1) \Rightarrow P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1-\alpha \Rightarrow$

$$\Rightarrow P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\sqrt{\frac{V}{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha \Rightarrow$$

$$\Rightarrow P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{V}{n}} \leq \bar{X}-\mu \leq z_{\frac{\alpha}{2}}\sqrt{\frac{V}{n}}\right) = 1-\alpha \Rightarrow$$

$$\Rightarrow P\left(\underbrace{\bar{X} - z_{\alpha/2}\sqrt{\frac{V}{n}}}_{t_1} \leq \mu \leq \underbrace{z_{\alpha/2}\sqrt{\frac{V}{n}}}_{t_2}\right) = 1-\alpha$$

Definizioa. Izan bedi $X_1, \dots, X_n$ θ parametro ezezaguna duen populazio jakin batetik ateratako zoriazko lagin bakuna. Demagun bi estatistiko ditugula, $t_1(X_1, \dots, X_n)$ eta $t_2(X_1, \dots, X_n)$ non

$$P(t_1(X_1, \dots, X_n) < \theta < t_2(X_1, \dots, X_n)) = 1-\alpha$$

den finkatutako balioak $0 < \alpha < 1$ izanik. Laginean behatutako balioez egindako $(t_1(x_1, \dots, x_n), t_2(x_1, \dots, x_n))$ tarteari %$(1-\alpha) \cdot 100$eko konfiantza mailako θrako konfiantza tartea esaten zaio. $I_\theta^{1-\alpha} = (t_1(x_1, \dots, x_n), t_2(x_1, \dots, x_n))$

Adibidea

$X \sim N(\mu, V)$ baina V balioa ez dakigu zein den

μ-ren estimazioa kalkulatu.

$$\hat{\mu} = \bar{X} \Rightarrow \bar{X} \sim N\left(\mu, \frac{V}{n}\right)$$

$$z = \frac{\bar{X} - \mu}{\sqrt{\frac{V}{n}}} \sim N(0,1)$$

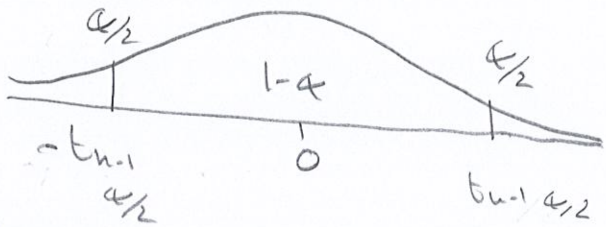
$$\hat{V} = S^2 \Rightarrow \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Student-en banaketa

t banaketa

Soilik n txikia bada.

$$\frac{\frac{\bar{X} - \mu}{\sqrt{V/n}}}{\sqrt{\frac{(n-1) S_c^2}{V} \Big/ n-1}} \quad \Bigg| \quad \frac{N(0,1)}{\sqrt{\chi^2_{n-1} / n-1}} \sim t_{n-1}$$



$$P(t_{n-1, \alpha/2} \le t_{n-1} \le t_{n-1, \alpha/2}) = 1-\alpha$$

$$P\left(-t_{n-1, \alpha/2} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1, \alpha/2}\right) = 1-\alpha$$

$$P\left(\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$I_\mu^{1-\alpha} = \left(\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \ \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right)$$

8. Gaia

Inferentzia Estatistikoa

Estimazioa Ariketak

6. Ariketa

$X_1, X_2, \ldots, X_n$ zorizko lagina

$$f(x) = \frac{1}{2}(1+\theta x) \quad -1 \leq x \leq 1$$

$\hat{\theta} = 3\bar{x}$ estimatzaile alboragabea θ-rentzat

$$E\bar{x} = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n}\frac{\theta}{3} = \frac{n\theta}{n \cdot 3} = \frac{\theta}{3}$$

$$\boxed{Ex_i = \int_{-1}^{1} x f(x)\,dx = \int_{-1}^{1} \frac{x}{2} + \frac{\theta x^2}{2}\,dx = \left. x^2 + \frac{\theta x^3}{6}\right]_{-1}^{1} = 1 + \frac{\theta}{6} - \left(1 - \frac{\theta}{6}\right) = \frac{2\theta}{6} = \frac{\theta}{6}}$$

$$E3\bar{x} = \frac{3\theta}{3} = \theta$$

1. Ariketa

$n = 150$

$X \equiv$ "piezaren urratu-ko puntua"

$X \sim P(\lambda)$

ⓐ Irentzako estimatzaile-zentratu.

$\hat{\lambda} = \bar{x}$

$$E(\hat{\lambda}) = E(\bar{x}) = E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{n\lambda}{n} = \lambda$$

Behidetako datuen araberako estimazioa

$$E(\hat{\lambda}) = \frac{1}{n}\sum x_i = \frac{1}{150}\cdot 312 = 2'11$$

b)

Errore estandarra

$$SD(\hat{\lambda}) = \sqrt{\frac{\lambda}{n}} =$$

$$VAR(\hat{\lambda}) = VAR(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} VAR(\bar{X}) = \frac{n\cdot\lambda}{n^2} = \frac{\lambda}{n}$$

$$SD(\hat{\lambda}) = \sqrt{VAR(\hat{\lambda})} = \sqrt{\frac{\lambda}{n}}$$

Errore estandararen estimazioa

$$SD(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{2'11}{150}} = 0'1186$$

2. Ariketa

$n=80$ $\quad$ akastunak $=12$

a) osagaien akatsik gabeko proportzioaren estimazioa

$$\hat{p} = \bar{X} \Rightarrow \bar{X} = \frac{1}{80}\sum_{i=1}^{n} x_i = \frac{68}{80} = 0'85$$

b)

40 sistema

3. Ariketa

X ≡ "Ogillo karbono-hidrato edukia"

a)

$$\hat{\mu}_x = \bar{X} \Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{921'82}{12} = 76'82$$

b)

$$SD(\hat{\mu}_x) = \sqrt{\frac{\mu}{n}} = \sqrt{\frac{76'82}{12}} = 2'53$$

$$VAR(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Ex_i = \frac{n \cdot \mu}{n^2} = \frac{\mu}{n}$$

c)

$$\hat{\mu}_x - \hat{\mu}_y = \bar{X} - \bar{Y}$$

$$E(\hat{\mu}_x - \hat{\mu}_y) = E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = E\left(\frac{1}{n}\sum x_i\right) - E\left(\frac{1}{n}\sum y_i\right) =$$

$$= \frac{1}{n}\sum Ex_i - \frac{1}{n}\sum Ey_i = \frac{n \cdot \mu_x}{n} - \frac{n \cdot \mu_y}{n} = \mu_x - \mu_y$$

d)

$n = 14$

$\bar{y} = 74'28$

Puntu-estimazioa.

$$\hat{\mu}_x - \hat{\mu}_y = 76'82 - 74'28 = 2'54$$

e)

$$SD(\hat{\mu}_x - \hat{\mu}_y) = 0'05$$

$$VAR(\bar{X} - \bar{Y}) = VAR(\bar{X}) - VAR(\bar{Y}) =$$

8.GAIA.
INFERENTZIA ESTATISTIKOA: ESTIMAZIOA

1. Sarrera

Adibidea:

Komunikazio sare batean 0 seinalea bidaltzean 1 seinalea jaso daiteke zareta dela eta.

Zein proportziotan gertatzen da hori?

Galdera horri erantzun egokia eman ahal izateko egoera hori formalizatzen saiatuko gara.

Populazioa: Komunikazio sarean bidalitako/biliko diren 0 seinale guztiak

$x = 0 \Rightarrow$ 0 bidali eta 0 jaso.

$x = 1 \Rightarrow$ 0 bidali eta 1 jaso.

$X \sim$ Bernoulli(p)

Baina p ez dakigu zeatz mehazt zenbat den

<u>Datuak:</u>

$x1 = 0,\ x2 = 1,$
$x3 = 0,\ x4 = 0$ $\qquad \hat{p} = 1/4$

Inferentzia
- Estimazioa
  "p" hurbiltzea (8 Gaia)
- Hipotesi-probak
  p-ri buruzko susmoak datuek berresten duten ala ez erabaki (9 Gaia)

<u>Estatistika parametrikoa</u>

Egoera orokorra:

$X \sim f(x, \theta)$ non parametro ezezaguna θ den.

2. Populazioa eta lagina

Lehengo adibidera itzuliz,

Populazioa-Eredua	Lagina
$X \sim \text{Bernoulli}(p)$	$\begin{array}{cccc} X_1 & X_2 & \dots & X_n \\ \downarrow & \downarrow & & \downarrow \\ 0 & 1 & \dots & 0 \end{array}$

Egoera honetan, p, hurbildu/estimatu nahi dugun parametroa da. Hasieratik badakigu laginak ez digula informazio osoa emango. Beraz:

1. Nola atera lagina?

2. Nola atera populazioari buruzko informazioa laginetik?

3. Zein konfiantza-maila dute?

Definizioa: $X_1, X_2, \dots, X_n$ zorizko aldagaiek *zorizko lagin bakuna* osatzen dute baldin eta:

(i) X_i guztiak elkarrekiko askeak dira.

(ii) X_i guztiak berdinki banatuak dira.

3. Estimatzaileak eta beren propietateak

Definizioa: Izan bedi $X_1, X_2, \ldots, X_n$ zorizko lagin bakuna. Aldagai horien gainean osaturiko funtzio errealari *estatistikoa* esaten zaio.

$$T = T(X_1, \ldots, X_n)$$

Adibideak:

$X \sim f(x, \theta)$ $X_1, \ldots, X_n$ (zorizko lagin bakuna)

1)

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \qquad T(X_1, \ldots, X_n)$$

2)

Estadistika deskribatzaileetako erabilitako edozein estadistiko.

$T(X_1, \ldots, X_n) \sim$ Banaketa (lagin banaketa)

$\boxed{X_i \sim f}$

Oharrak:

- Estatistikoa zorizko aldagaia da.
- Estatistikoaren banaketari *lagin-banaketa* esaten zaio.

  Orohar estatistikoaren lagin-banaketa 2 faktoreren menpekoa da: jatorrizko populazioren banaketa ($X_i$en banaketa) eta lagin-tamaina (n).

Adibidea:

Demagun populazioan $X \sim N(\mu, \sigma^2)$ eredua egokia dela. Izan bedi $X_1, X_2, \ldots, X_n$ bertatik ateratako zorizko lagin bakuna. Nolakoa izango da $\bar{X}$ren lagin-banaketa?

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$E x_i = \mu$

$VAR(x_i) = \sigma^2$

$X_i \sim N(\mu, \sigma^2)$ non $i = 1, \ldots, n$

i)

$$E\bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E x_i = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu$$

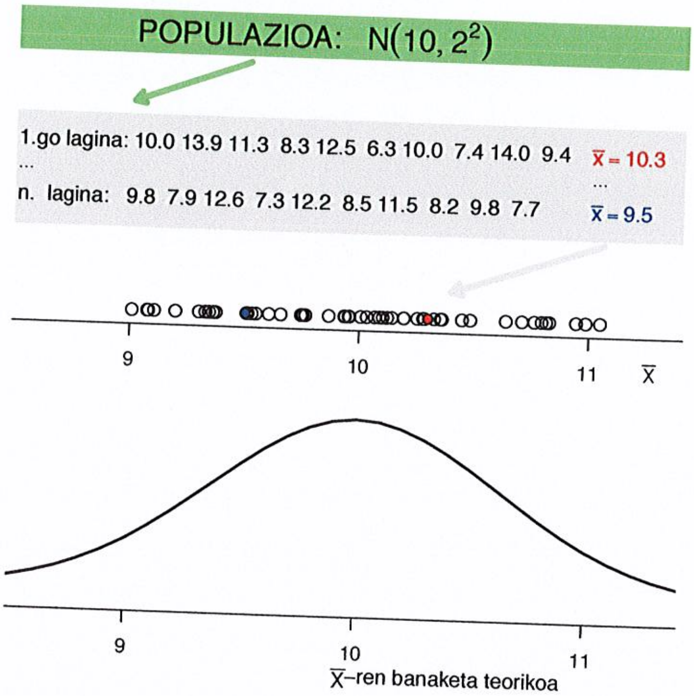
ii)

$$VAR(\bar{X}) = VAR\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} VAR\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} VAR(x_i) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Ondorioz, $\bar{X}$ren lagin banaketa honela geratuko da.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Demagun hau dela esku artean dugun kasu berezia:



Orain lengo kasu orokoretik kasu partikularra aztertuko dugu.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$ (kasu orokorra)

Populazioa: $N(10, 2^2)$

$$\bar{X} \sim N\left(10, \frac{2^2}{10}\right)$$ (lagin-banaketa)

Desbideratze-estandarra kalkulatuko dugu.

$$SD(\bar{X}) = \sqrt{VAR(\bar{X})} = \sqrt{\frac{2^2}{10}} = \frac{2}{\sqrt{10}} = 0'63$$

Definizioa: Demagun θ parametro ezezaguna duen populaziotik ateratako $X_1, X_2, \dots, X_n$ zorizko lagin bakuna dugula. $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ estimatzaileak har ditzakeen balioek θ parametroa hurbildu nahi dutenean, $\hat{\theta}$-ri θ parametroaren *puntu-estimatzailea* dela esango diogu.

Adibidea-1:

Demagun $X_1, \dots, X_n$ $N(\mu, 1)$ popolaziotik ateratako zorizko lagin bakuna. μ estimatu nahi dugu eta $\hat{\mu} = \bar{X}$ izan daiteke estimatzaile egokia.

Baldin eta $x_1 = 3$, $x_2 = 3$, $x_3 = 5$ eta $x_4 = 6$ lagina behatu badugu μ-rentzako puntu-estimazioa

$$\hat{\mu} = \frac{3+3+5+6}{4} = 4'25$$

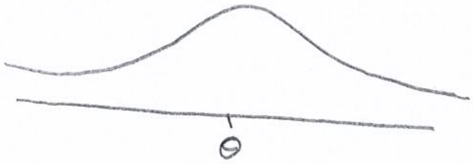
Adibidea-2:

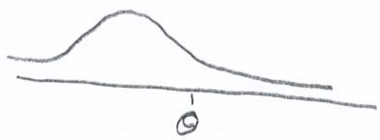
Demagun $\hat{\Theta}$ eta Θ^* Θ-rentzako estimatzaileak ditugu. Zein da hobea?

$\hat{\Theta}$-ren itxaropena Θ izango da baina Θ^*-rena ez, grafikoetan ikusi dezakegun bezala. Ondorioz, $\hat{\Theta}$ hobea da.

$E(\hat{\theta}) = \theta$

$E(\theta^*) \neq \theta$





Definizioa: Demagun θ parametro ezezaguna dela eta $\hat{\theta}$ bere estimatzailea. $\hat{\theta}$ estimatzaile *alboragabea* dela esango dugu baldin eta

$$E(\hat{\theta}) = \theta \quad \text{bada.}$$

PROPIETATEAK:

Izan bedi $X_1, \ldots, X_n$ zorizko lagin bakuna, $E\,X_i = \mu$ eta $VAR(X_i) = \sigma^2$ direlarik ($i = 1, \ldots, n$).

(I) $\hat{\mu} = \bar{X}$ estimatzailea alboragabea da μrentzat. Hau da,

$$E(\bar{X}) = \mu.$$

(II) $\hat{\sigma}^2 = S_n^2$, lagin-bariantza estimatzaile alboratua da σ^2rentzat. Zehazki,

$$E(S_n^2) = \frac{n-1}{n}\sigma^2.$$

(III) $\hat{\sigma}^2 = S^2$, kuasibariantza estimatzaile alboragabea da σ^2rentzat. Hau da,

$$E(S^2) = \sigma^2.$$

Bi estimatzaile desberdin ditugu σ^2 parametroa estimatzeko. Aurreko propietateetan oinarrituz, zein uste duzu dela estimatzailea egokiena? Zergatik?

$E(aX) = aEX$

$E(X+Y) = EX + EY$

$VAR(aX) = a^2 VAR(X)$

$VAR(X+Y) = VAR(X) + VAR(Y)$

X,Y askeak

$VAR(X) = EX^2 - (EX)^2$

i) frogatu $\hat{\mu} = \bar{X}$ estimatzaile alboragabea dela μ-rentzat, $E(\bar{X}) = \mu$

$$E(\bar{X}) = \mu \Rightarrow E(\bar{X}) = E\left(\frac{1}{n}\sum_i^n x_i\right) = \frac{1}{n}E\left(\sum_i^n x_i\right) =$$

$$= \frac{1}{n}\sum E x_i = \frac{1}{n}\sum_i^n \mu = \frac{n\mu}{n} = \mu.$$

ii) frogatu $\hat{\sigma}^2 = S^2$ lagin-bariantza estimatzaile alboratua da σ^2-rentzat

$$E(S_n^2) = \frac{n-1}{n}\cdot\sigma^2 \Rightarrow E(S_n^2) = E\left(\frac{1}{n}\sum_{i=0}^n (X-\bar{X})^2\right) = \frac{1}{n}E\left(\sum_{i=0}^n (X-\bar{X})^2\right) =$$

$$= \frac{1}{n}\sum_{i=0}^n \left(E(X-\bar{X})^2\right) = \frac{1}{n}\sum_{i=0}^n \left(EX^2 - 2E(X\cdot\bar{X}) + E(\bar{X})^2\right) =$$

$$= E\left[\frac{1}{n}\left[\sum x_i^2 - 2n\bar{x}\cdot\bar{x} + n\bar{x}^2\right]\right] = E\left[\frac{1}{n}\sum x_i^2 - \bar{x}^2\right] = \frac{1}{n}\sum E(x_i)^2 - E(\bar{x})^2 =$$

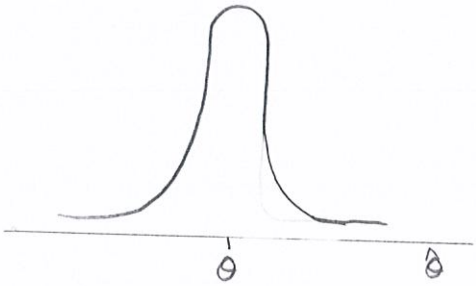
$n\bar{X} = \sum x_i$

$$= \frac{1}{n}\sum(\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 < \sigma^2$$

$E(x_i^2) = VAR(x_i) + (E(x_i))^2 = \sigma^2 + \mu^2$

$E(\bar{x}^2) = VAR(\bar{X}) + (E\bar{x})^2 = \frac{\sigma^2}{n} + \mu^2$

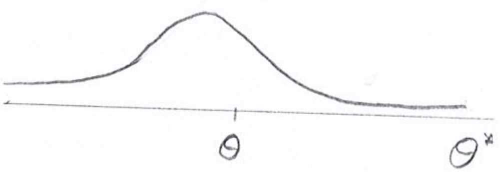
θ Parametro ezezaguna



$E(\hat{\theta}) = \theta$

$E(\theta^*) = \theta$

$VAR(\hat{\theta}) < VAR(\theta^*)$



Adibidea:

Definizioa: Izan bitez $\hat{\theta}$ eta θ^*, θ parametro ezezagunaren bi estimatzaile alboragabe. $\hat{\theta}$ estimatzailearen θ^* estimatzailearekiko *eraginkortasuna* honela neurtzen da:

$$\frac{VAR(\theta^*)}{VAR(\hat{\theta})}.$$

Baldin eta zatidura hau > 1 bada, $\hat{\theta}$ estimatzailea θ^* estimatzailea baino *eraginkorragoa* dela esango dugu.

Sarri, estimatzailearen desbideratze estandarrari *errore estandarra* esaten zaio.

4. Egoera arruntak

1. X **aldagai kuantitatiboa**

Parametro interesgarriak	Estimatzaileak
$E(X) = \mu$	$\hat{\mu} = \bar{X}$
$VAR(X) = \sigma^2$	$\hat{\sigma}^2 = S^2$

Lagin-banaketak

- Lagin-batez bestekoa, $\bar{X}$.

$$E(\bar{X}) = \mu\,, \quad VAR(\bar{X}) = \frac{\sigma^2}{n}\,, \quad SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} \text{ (errore estandarra)}$$

Gainera, lagina handia ($n \geq 30$) bada: $\bar{X} \approx N(\mu, \sigma^2/n)$.

- Kuasibariantza, S^2.

$$E(S^2) = \sigma^2\,, \quad (*)\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

(*)Baldin eta jatorrizko populazioa normala $N(\mu, \sigma^2)$ bada.

Adibidea: Programa baten exekuzio-denborak zenbatekoak diren aztertu nahi da. Horretarako hainbat datu multzori programa aplikatu eta behatu diren exekuzio denborak hauek izan dira.

335	568	392	452	479	503	516	398	443	343
316	551	636	363	344	486	518	497	501	310
540	434	415	493	595	405	365	361	433	472
645	332	559	408	455	476	289	561	788	391

Zer esan dezakezu batez besteko exekuzio denboren inguruan?

$X \equiv$ 'Denbora' (kuantitatiboa)

$EX = \mu$

$\hat{\mu} = \bar{X}$, estimazioa:

$n = 40$ $\quad \sum x_i = 18368 \Rightarrow \bar{X} = 459'2$ (puntu-estimazioa)

Errore estandarra

$SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

(Errore estandarraren estimazioa) $SD(\bar{X}) = \frac{S}{\sqrt{40}} \underset{*}{=} 16'7$

$* S^2 = \frac{40}{39} S_n^2 = 11125,34$

$S_n^2 = \frac{1}{n}\sum x_i^2 - (\bar{X})^2 = 8868474 - \left(\frac{18368}{40}\right)^2$

2. X **aldagai bitarra**

$p = P(X = 1)$ izanik,

Parametro interesgarria	Estimatzailea
$E(X) = p$	$\hat{p} = \bar{X}$

Lagin-banaketa

- Lagin-proportzioa, $\hat{p}$.

$$E(\hat{p}) = p\,, \quad VAR(\hat{p}) = \frac{pq}{n}\,, \quad SD(\hat{p}) = \sqrt{\frac{pq}{n}} \text{ (errore estandarra)}$$

Gainera, lagina handia bada ($n\hat{p} \geq 5$ eta $n\hat{q} \geq 5$):

$$\hat{p} \approx N(p, pq/n).$$

Adibidea:

Demagun PZU baten aktibitatea neurtu nahi dela eta horretarako $n = 150$ behaketa egin direla. Behaketa bakoitzean (i. behaketa) $X_i = \begin{cases} 0 & \text{PZU ez aktiboa,} \\ 1 & \text{PZU aktiboa,} \end{cases}$ kodetu da. Behaketa denak denboran zehar nahiko tarterekin egin dira hauek elkarrekiko askeak izan zitezen eta 123 neurketean PZUa aktibo zegoela ikusi da. Azter ezazu zein proportziotan egoten den aktibo PZUa.

$EX = p$

$\hat{p} = \bar{X} \Rightarrow \hat{p} = \frac{1}{n}\sum x_i = 0'82$

$\hat{SD}(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0'82 \cdot 0'18}{150}} = \underline{0'031}$

5. Konfiantza-tarteak

Adibidea:

Demagun $X \sim N(\mu, v)$ eredua dugula populazioa adierazteko eta v **balioa badakigula zein den**. Helburua, μ batezbestekoaren inguruko informazioa lortzea da.

$X \sim N(\mu, v)$ Baina ez dakigu zenbat den v

μ-ren estimazioa: $\hat{\mu} = \bar{X}$ non $\bar{X} \sim N(\mu, \frac{v}{n})$

$$z = \frac{\bar{X} - \mu}{\sqrt{\frac{v}{n}}} \sim N(0,1) \Rightarrow P\left(-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha \Rightarrow$$

$$\Rightarrow P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sqrt{\frac{v}{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha \Rightarrow$$

$$\Rightarrow P\left(\underbrace{\bar{X} - z_{\alpha/2}\sqrt{\frac{v}{n}}}_{T_1} \leq \mu \leq \underbrace{\bar{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{v}{n}}}_{T_2}\right) = 1-\alpha$$

Definizioa: Izan bedi $X_1, \ldots, X_n$ θ parametro ezezaguna duen populazio jakin batetik ateratako zorizko lagin bakuna. Demagun baditugula bi estatistiko $T_1(X_1, \ldots, X_n)$ eta $T_2(X_1, \ldots, X_n)$ non:

$$P(T_1(X_1, \ldots, X_n) < \theta < T_2(X_1, \ldots, X_n)) = 1 - \alpha$$

den ($0 < \alpha < 1$, finkaturiko balioa izanik).

Laginean behatutako balioez egindako $(T_1(x_1, \ldots, x_n)\,,\, T_2(x_1, \ldots, x_n))$ tarteari *%$(1-\alpha)$100eko konfiantza-mailako θrako konfiantza-tartea* esaten zaio.

$I_\theta^{1-\alpha} = (T_1(x_1, \ldots, x_n)\,,\, T_2(x_1, \ldots, x_n))$ eran adieraziko dugu.

Adibidea:

Demagun $X \sim N(\mu, v)$ eredua dugula populazioa adierazteko baina v **balioa ez dakigula zein den**. Helburua, μ batezbestekoaren inguruko informazioa lortzea da. Zer egin dezakegu?

$X \sim N(\mu, v)$ baina ez dakigu μ eta v zenbat diren.

μ-ren estimazioa kalkulatuko dugu,

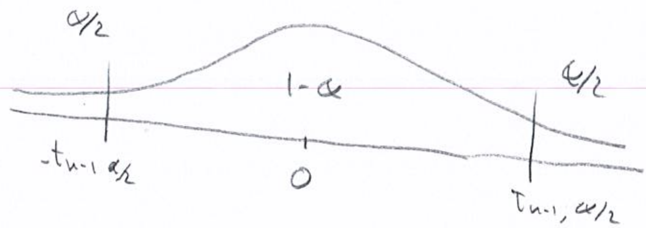
$$\hat{\mu} = \bar{X} \Rightarrow \bar{X} \sim N\left(\mu, \frac{v}{n}\right)$$

$$z = \frac{\bar{X} - \mu}{\sqrt{v/n}} \sim N(0,1)$$

$$\hat{v} = S^2 \Rightarrow \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Student-en t banaketa   soilik n txikia bada

$$\frac{\frac{\bar{X} - \mu}{\sqrt{v/n}}}{\sqrt{\frac{(n-1)S^2}{v}/(n-1)}} \quad \Bigg| \quad \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}} \sim t_{n-1}$$



$$P(-t_{n-1, \alpha/2} \le t_{n-1} \le t_{n-1, \alpha/2}) = 1 - \alpha$$

$$P\left(-t_{n-1, \alpha/2} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1, \alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$I_\mu^{1-\alpha} = \left(\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right)$$

Laburpena

Parametroa	Baldintzak	Puntu estimazioa	Lagin banaketa	%$(1-\alpha)$100ko K.T.
μ	• σ ezaguna • Pop. normala edo $n > 30$	$\bar{x}$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left(\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$
μ	• σ ezezaguna • $n > 30$	$\bar{x}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$	$\left(\bar{x} \mp z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$
μ	• σ ezezaguna • Eredu normala	$\bar{x}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\left(\bar{x} \mp t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right)$
σ^2	• Eredu normala	S^2	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right)$
p	• Bernoulli eredua/bin. • $n\hat{p} \geq 5$ eta $n\hat{q} \geq 5$	$\hat{p}$	$\frac{\hat{p}-p}{\sqrt{pq/n}} \approx N(0,1)$	$\left(\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

Notazioa:

$$z_\alpha \ : \ P(Z > z_\alpha) = \alpha$$
$$t_{n,\alpha} : P(t_n > t_{n,\alpha}) = \alpha$$
$$\chi^2_{n,\alpha} : P(\chi^2_n > \chi^2_{n,\alpha}) = \alpha$$

Adibidea:

Udaleko Gazteria sailak agertutako datuen arabera, gazteek gurasoen etxea uzteko adina (X) aldagai normaltzat har daiteke. Oraintsu gurasoen etxea utzi duten 16 gazterekin egindako ikerketa batean, 28.1 urteko batez bestekoa lortu da eta 3ko desbideratze estandarra.

- Zein da ikerketaren arabera udal hartako gurasoen etxea uzteko adina, batez bestean, %99ko konfiantzaz?
- Zein da desbideratze estandarraren balioa gutxi gorabehera, %99ko konfiantzaz?

a) $X \equiv$ 'adina'

$X \sim N(\mu, \sigma^2)$ baina ezezagunak dira bi balioak

$\boxed{n=16 \quad \bar{X}=28'1 \quad S=3}$

$$I_\mu^{0'99} = \left(\bar{X} \mp \underbrace{t_{n-1,0'005}}_{} \frac{S}{\sqrt{n}}\right) = \left(28'1 - 2'95 \cdot \frac{3}{\sqrt{16}}, \ 28'1 + 2'95 \cdot \frac{3}{\sqrt{16}}\right) =$$

$t_{16-1} = 15$

$t_{0'995}$

$= (25'9, 30'3)$

b) $$I_{\sigma^2}^{0'99} = \left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \ \frac{(n-1)S^2}{\chi^2_{n+1,\alpha/2}}\right) = \left(\frac{15 \cdot 3^2}{32'8}, \ \frac{15 \cdot 3^2}{4'66}\right) = (4'11, 29'35)$$

$I_\sigma^{0'99} = (2'02, 5'4)$

5. Gaia

Inferentzia Estatikoa

Estimazioa

Ariketak

6. Ariketa

$x_1, x_2, \dots, x_n$

$f(x) = \frac{1}{2}(1+\theta x), \quad -1 \leq x \leq 1 \quad (-1 \leq \theta \leq 1$ izanik$)$

$$\mathcal{E}(\hat{\theta}) = \theta \Rightarrow \mathcal{E}(3\bar{X}) = 3\mathcal{E}\bar{X} = 3\mathcal{E}\left(\frac{1}{n}\sum_1^n X_i\right) = \frac{3}{n}\mathcal{E}\left(\sum_1^n X_i\right) = \frac{3}{n}\sum_1^n \mathcal{E}X_i =$$

$$= \frac{3}{n}\sum_1^n \frac{\theta}{3} = \frac{3}{n}\cdot\frac{n\theta}{3} = \theta$$

$$\mathcal{E}X_i = \int_{-1}^{1} x_i \frac{1}{2}(1+\theta x_i)\,dx_i = \int \frac{x_i}{2} + \frac{\theta x_i^2}{2}\,dx_i = x_i^2 + \frac{\theta x_i^3}{6}\Big]_{-1}^{1} = 1 + \frac{\theta}{6} - \left(1 - \frac{\theta}{6}\right) =$$

$$= \frac{2\theta}{6} = \frac{\theta}{3}$$

2. Ariketa

$n = 80$

12 akastunak

a) Akatsik gabeko osagaien estimazioa

$\hat{\lambda} = \frac{80-12}{80} = 0'85$ izango da estimazioa

b)

$\frac{80}{2} = 40$ Sistema

$\hat{\lambda}^2 = 0'85 \cdot 0'85 = 0'7225$

1. Ariketa

Populazioa: $X \equiv$ 'urratu kop'

$X \sim P(\lambda)$ Baina ez dakigu zenbat den λ.

$X_i \sim P(\lambda)$ non $X_1, \dots, X_n$ $n = 150$

$\hat{\lambda} = \bar{X}$

a)

$E(\hat{\lambda}) = \lambda$

$$E(\hat{\lambda}) = E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$

$\bar{X}$ estimatzaile zentratua da λ-rako.

Estimazioa:

$$\hat{\lambda} = \frac{1}{150}\sum_{i=1}^{150} x_i = 2'11$$

$\hat{\lambda} \approx 2'11$ puntu-estimazioa

b)

$$SD(\hat{\lambda}) = ? = \sqrt{\frac{\lambda}{n}}$$

$$VAR(\hat{\lambda}) = VAR(\bar{X}) = VAR\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} VAR\left(\sum_{i=1}^{n} x_i\right) \underset{\text{askeak}}{=} \frac{1}{n^2}\sum_{i=1}^{n} VAR(x_i) =$$

$$= \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$$

$X \sim P(\lambda)$

$VAR(X) = \lambda$

$$\widehat{SD}(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{150}} = \sqrt{\frac{2'11}{150}} = 0'118$$

~~9. Ariketa~~

~~$n = 1000$~~

~~Konfiantza-tartea: 0'95~~

10. Ariketa

$n = 250$ informatikari

113 ez menperatu

$\hat{p} = \frac{113}{250} = 0'452 \Rightarrow n\hat{p} \geq 5$

$n\hat{q} \geq 5$

$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \left(\hat{p} \pm z_{\frac{0'05}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

$z_\alpha : P(z > z_\alpha) = \alpha$

$I_p^{0'95} = \left(0'452 \pm 1'96\sqrt{\frac{0'452 \cdot 0'548}{250}}\right) = (0'39, 0'514)$

11. Ariketa

110 txirrindulari

$\bar{x} = 0'81$

$SD(x) = 0'34 \Rightarrow VAR(x) = 0'1156 = \sigma^2$

$\bar{X} \sim N(0'81, 0'001)$

$I_\mu^{0'99} = \left(\bar{X} \pm t_{n-1, 0'005} \cdot \frac{S}{\sqrt{n}}\right) = \left(0'81 \pm 2'63 \cdot \frac{0'34}{\sqrt{110}}\right) =$

$= (0'725, 0'895)$

3 Ariketa

A ogi mota $\longrightarrow n=12$ (zoriztko lagina)

a)

$X \sim N(\mu, \sigma^2)$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \mu_x$$

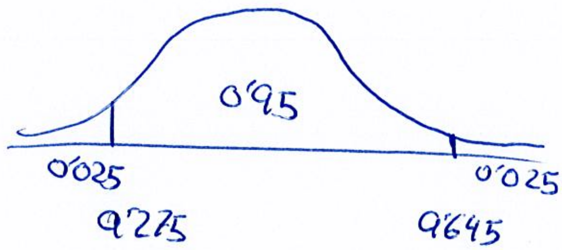
12. Ariketa

$n = 40$

$\bar{x} = 9'46$

$S = 0'58$

Konfiantza tartea: 0'95



$$I_{\mu}^{0'95} = \left(\bar{x} \pm t_{n-1,\,0'025} \cdot \frac{S}{\sqrt{n}}\right) = \left(9'46 \pm 2'02 \cdot \frac{0'58}{\sqrt{40}}\right) =$$

$$= (9'275,\ 9'645)$$

18 Ariketa

a)

$$\hat{p} = \frac{31}{40} = 0'775$$

$$\hat{SD}(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0'775 \cdot 0'225}{40}} = 0'066$$

$$I_{\mu}^{0'95} = \left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \left(0'775 \pm 1'96\sqrt{\frac{0'775 \cdot 0'225}{40}}\right) =$$

$$= (0'646,\ 0'904)$$

b)

$$\text{Luzera} = 0'02 = 2 \cdot z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0'02^2 = 4\, z^2_{\alpha/2} \frac{\hat{p}\hat{q}}{n} \Rightarrow n = \frac{4 z^2_{\alpha/2} \cdot \hat{p}\hat{q}}{0'02^2}$$

$$n = \frac{4 \cdot 1'96^2 \cdot 0'174}{0'02^2} = 6698'29\ \cancel{6}669$$

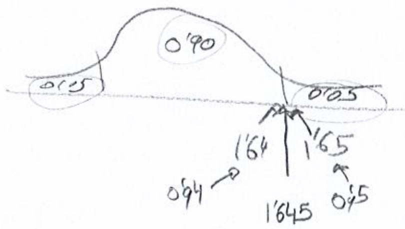
$$n \geq 6699$$

13. Ariketa

a)

200 bezeroK kreditu txartelarekin erosketak
136k interesak ordaindu.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{136}{200} = 0'68 \Rightarrow \text{puntu-estimazioa}$$



$n\hat{p} \geq 5 \Rightarrow 136 \geq 5$

$n\hat{q} \geq 5 \Rightarrow 200 \cdot (1-0'68) \geq 5 \Rightarrow 64 \geq 5$

$$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \left(\hat{p} \pm z_{0'05}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

$z_a \Rightarrow P(z > z_a) = \alpha$

$$I_p^{0'90} = \left(0'68 \pm 1'645\sqrt{\frac{0'68\, 0'32}{200}}\right) = (0'626, 0'734) \quad \hat{p} = 0'68$$

$$I_p^{0'90} = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Luzera $0'05 = 2 z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$0'05^2 = 4 z^2_{\alpha/2}\frac{\hat{p}\hat{q}}{n} \Rightarrow \frac{0'05^2}{4} = z^2_{\alpha/2}\frac{\hat{p}\hat{q}}{n} \Rightarrow n = \frac{4 z^2_{\alpha/2}\hat{p}\hat{q}}{0'05^2}$$

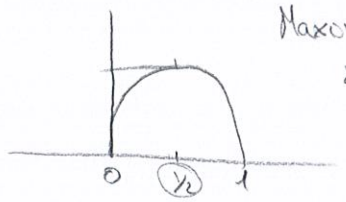
1) $\hat{p} = 0'68 \Rightarrow \hat{p}\hat{q} = 0'68 \cdot 0'32 \Rightarrow n = 942.13 \Rightarrow n \geq 943$

Aurretiko informazioa behar

2) Aurreko informaziorik gabe

$\hat{q}\hat{p} = \hat{p}(1-\hat{p}) \leq \frac{1}{4}$

$f(x) = x(1-x) \quad x \in (0,1)$



Maximoa
$x = \frac{1}{2} \Rightarrow f(x) = \frac{1}{4}$

Kasurik txarrenean $\hat{p}(1-\hat{p}) = \frac{1}{4}$

$$n = \frac{4 \cdot z^2_{\alpha/2} \cdot \frac{1}{4}}{L^2} = \frac{z^2_{\alpha/2}}{L^2} = 1082'41$$

$n \geq 1083$

9. GAIA: INFERENTZIA ESTATISTIKOA: HIPOTESI-TESTAK

Adibidea 1:

Etxe jakin bateko liburu elektronikoaren bateria % 100n kargatzeko behar den denborak banaketa normala jarraitzen duela ikusi da. Zehazki, $X \sim N(\mu, 0{,}3^2)$. Etxeak dionez, batez beste 4.5 ordu behar dira bateria guztiz kargatzeko baina zalantzak daude eta hori horrela den ala ez argitzeko zorizko lagin bat hartzea erabaki da. Behaturiko datuak azpialdeko taulan bildu dira, horien arabera, liburu elektronikoa ekoizten duen etxeak arrazoirik ez duela onar al daiteke?

4.8	4.5	4.7	4.7	4.8	4.7	4.8	4.4	4.5	5.1
5.0	4.3	4.3	4.6	5.1	5.0	4.6	5.1	4.5	4.9
4.9	4.5	4.8	4.6	4.7	4.5	4.5	4.3	4.1	4.6

$n = 30 \quad \sum x_i = 139{,}9$

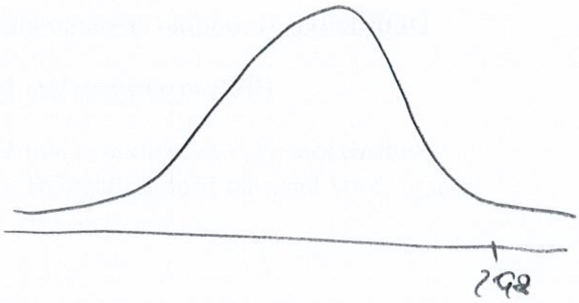
X = "denbora" Hipotesia: $\begin{cases} H_0 & \mu = 4.5 \\ H_1 & \mu \neq 4.5 \end{cases}$

$X \sim N(\mu, 0'3^2)$ Erabakia hartu

$\bar{X} \sim N(\mu, \frac{0'3^2}{n}) \Rightarrow Z = \frac{\bar{X} - \mu}{0'3/\sqrt{n}} \sim N(0,1)$

Ho egiazkoa balitz

$Z = \frac{\bar{X} - 4'5}{0'3/\sqrt{30}} \sim N(0,1)$



Laginean

$Z_{behe} = \frac{\bar{X} - 4'5}{0'3/\sqrt{n}} = \frac{\frac{139'9}{30} - 4'5}{0'3/\sqrt{30}} = 2'98$

$p = P(Z > 2'98) \cdot 2 = 0'002882$

Erabakia hartzeko

1) Laginean ikusi duguna arraroa da, baina lagin arraro bat suertatu zaigulako da Ho egiazkoa.

2) Laginean ikusi duguna arraroa da Ho egiazkoa, dela pentsatzen jarraitzeko. Ho errefusatu, H1 onartu.

1. Oinarrizko definizio eta kontzeptuak

Definizioa: Populazioaren gainean ezarritako susmoa *hipotesi estatistikoa* da. Hipotesi testetan bi hipotesi kontrajartzen dira:

- *hipotesi nulua* (H_0) eta
- *ordezko hipotesia* (H_1).

Hipotesi-testa burutzeak bi hipotesi horien artean bat aukeratzea esan nahi du, beti ere, lagineko datuek erakusten dutena oinarri harturik. Gainera, prozedura honela ezartzen da: H_0 hipotesia egiazkotzat hartzen da, baldin eta datuek ez badute kontrakoa erakusten.

Hipotesi-testari dagozkion egoera posibleak

		H_0 egiazkoa da	H_0 ez da egiazkoa
Erabaki	H_0 onartu	ZUZENA	II motako errorea
posiblea	H_0 errefusatu	I motako errorea	ZUZENA

Definizioa: I motako erroreari testaren *adierazgarritasun-maila* (α) esaten zaio.

$$\alpha = P(H_0 \text{ errefusatu } | H_0 \text{ egiazkoa})$$

Definizioa: II motako errorearekin lotuta testaren *ahalmena* honela definitzen da:

$$P(H_0 \text{ errefusatu } | H_0 \text{ faltsua}) = 1 - P(H_0 \text{ onartu } | H_0 \text{ faltsua})$$

Definizioa: H_0 egiazkotzat harturik, laginean behatutako emaitza bezain arraroa edo arraroagoa zoriz lortzeko probabilitateari *p-balioa* esaten zaio.

2. Hipotesi-testa egiteko urratsak

Demagun populazioan ezezaguna den θ parametroaren gainean hipotesi-testa egin nahi dugula.

1. Hipotesiak zehaztu. Oro har bi mota daude:

   a) Alde batekoa

   1) Ezkerretikoa $\begin{cases} H_0: & \theta \geq \theta_0 \\ H_1: & \theta < \theta_0 \end{cases}$

   2) Eskuinetikoa $\begin{cases} H_0: & \theta \leq \theta_0 \\ H_1: & \theta > \theta_0 \end{cases}$

   b) Alde bikoa $\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$

   Oharra: θ_0 balioa erabiltzaileak kasu bakoitzean finkatuko duen balioa da eta erreparatu $\theta = \theta_0$ kasua beti hipotesi nuluan barneratzen dela.

2. Adierazgarritasun-maila, α, finkatu.

3. Estatistiko egokia aukeratu, T, eta H_0 **egiazkoa dela suposatuz** laginean estatistikoak hartzen duen balioa kalkulatu: T_{beha}

4. p-balioa kalkulatu.

5. p-balioa eta α adierazgarritasun-maila konparatu eta horren arabera erabakia hartu:

   - $p \leq \alpha$ bada, H_0 errefusatu. Hau da, H_0 faltsua dela erabaki.
   - $p > \alpha$ bada, H_0 ezin dugu errefusatu. Hau da, H_0 faltsua dela erabakitzeko ez dugu nabaritasunik.

   p-balioaren interpretaziorako orientagarri:

   - p-balioa $\leq 0{,}001$ bada, emaitzak H_0ren aurka guztiz adierazgarriak direla esaten da.
   - $0{,}001 <$ p-balioa $\leq 0{,}01$ bada, emaitzak H_0ren aurka oso adierazgarriak direla esaten da.
   - $0{,}01 <$ p-balioa $\leq 0{,}05$ bada, emaitzak H_0ren aurka adierazgarriak direla esaten da.
   - p-balioa $> 0{,}05$ bada, emaitzak ez dira adierazgarriak. Hau da, behatzen diren desberdintasunak zorzikoak izan daitezke eta ez dago H_0 errefusatzeko nabaritasunik.

Adibidea:
Bi txanpon ditugu baina orekatuta ez ote dauden zalantza dugu.

- Lehen txanpona 17 aldiz bota ondoren 5 aldiz bakarrik atera da aurpegia.
- Bigarren txanpona 100 aldiz bota dugu baina 29 aldiz bakarrik atera da aurpegia.

Datu horien arabera, txanponak ez orekatuak daudela esateko nahikoa nabaritasun bai al dugu?

Lehen txanpona

$H_0 : p = 0'5$

$H_1 : p \neq 0'5$

Laginaren informazioa:

$n = 17$ aurpegi kopurua $= 5$

$\hat{p} = \frac{5}{17} = 0'29$

Adierazgarritasun-maila: $\alpha = 0'05$

$\hat{p} \approx N(0'5, \frac{0'5^2}{17})$ H_0 pean

$(n\hat{p}, n\hat{q} \geq 5)$

$z = \frac{\hat{p} - 0'5}{0'5/\sqrt{17}} \sim N(0,1)$

$z_{behat} = \frac{0'29 - 0'5}{0'5/\sqrt{17}} = -1'7$

P. balioa

$p = 2P(z > 1'7) = 0'089$

erabakia hartzeko:

$0'089 > 0'05 \Rightarrow H_0$ ez baztertu

Bigarren txanpona

$H_0 : p = 0'5$

$H_1 : p \neq 0'5$

Laginaren informazioa:

$n = 100$ aurpegiak 29

$\hat{p} = \frac{29}{100} = 0'29$

Adierazgarritasun maila: $\alpha = 0'05$

$z = \frac{\hat{p} - 0'5}{0'5/\sqrt{100}} = -4'2$

P balioa $p = 2P(z > 4'2) = 3 \cdot 10^{-5}$

erabakia hartzeko:

$3 \cdot 10^{-5} < 0'05 \Rightarrow H_0$ baztertu

3. Lagin bakarrerako hipotesi-test arruntenak

- **X aldagai kuantitatiboa. $EX = \mu$ batez bestekorako hipotesi-testa.**

  Baldintzak:

  1. σ ezaguna.
  2. Populazioa normala edo $n > 30$.

<table>
<tr><th>Hipotesiak</th><th>Estatistikoa eta bere banaketa H_0 pean</th><th>p-balioa</th></tr>
<tr><td>$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$</td><td rowspan="3">$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

$z_{beha} = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$</td><td>$p = 2P(Z > |z_{beha}|)$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu \geq \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$</td><td>$p = P(Z < z_{beha})$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu \leq \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$</td><td>$p = P(Z > z_{beha})$</td></tr>
</table>

- **X aldagai kuantitatiboa. $EX = \mu$ batez bestekorako hipotesi-testa.**

  Baldintzak:

  1. σ ezezaguna.
  2. $n > 30$.

<table>
<tr><th>Hipotesiak</th><th>Estatistikoa eta bere banaketa H_0 pean</th><th>p-balioa</th></tr>
<tr><td>$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$</td><td rowspan="3">$Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \approx N(0,1)$

$z_{beha} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$</td><td>$p = 2P(Z > |z_{beha}|)$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu \geq \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$</td><td>$p = P(Z < z_{beha})$</td></tr>
<tr><td>$\begin{cases} H_0: & \mu \leq \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$</td><td>$p = P(Z > z_{beha})$</td></tr>
</table>

- **X aldagai kuantitatiboa. $EX = \mu$ batez bestekorako hipotesi-testa.**

  Baldintzak:

  1. σ ezezaguna.
  2. Populazioa normala.

Hipotesiak	Estatistikoa eta bere banaketa H_0 pean	p-balioa
$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$	$t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$p = 2P(t_{n-1} > \lvert t_{beha} \rvert)$
$\begin{cases} H_0: & \mu \geq \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$		$p = P(t_{n-1} < t_{beha})$
$\begin{cases} H_0: & \mu \leq \mu_0 \\ H_1: & \mu > \mu_0 \end{cases}$	$t_{beha} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$p = P(t_{n-1} > t_{beha})$

- **X aldagai kuantitatiboa. $EX = \sigma^2$ bariantzarako hipotesi-testa.**

  Baldintzak:

  1. Populazio normala.

Hipotesiak	Estatistikoa eta bere banaketa H_0 pean	p-balioa
$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 \neq \sigma_0^2 \end{cases}$	$\dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$	$s^2 > \sigma_0$ bada, $p = 2P(\chi^2 n - 1 > K_{beha})$ $s^2 < \sigma_0$ bada, $p = 2P(\chi^2 n - 1 < K_{beha})$
$\begin{cases} H_0: & \sigma^2 \geq \sigma_0^2 \\ H_1: & \sigma^2 < \sigma_0^2 \end{cases}$		$p = P(\chi^2_{n-1} < K_{beha})$
$\begin{cases} H_0: & \sigma^2 \leq \sigma_0^2 \\ H_1: & \sigma^2 > \sigma_0^2 \end{cases}$	$K_{beha} = \dfrac{(n-1)s^2}{\sigma_0^2}$	$p = P(\chi^2_{n-1} > K_{beha})$

- **X aldagai bitarra. $P(X=1)=p$ proportziorako hipotesi-testa.**

  Baldintzak:

  1. $n\hat{p}\,, n\hat{q} \geq 5$ $(\hat{q} = 1-\hat{p})$

Hipotesiak	Estatistikoa eta bere banaketa H_0 pean	p-balioa
$\begin{cases} H_0: & p = p_0 \\ H_1: & p \neq p_0 \end{cases}$	$Z = \dfrac{\hat{p}-p_0}{\sqrt{p_0 q_0/n}} \approx N(0,1)$	$p = 2P(Z > \lvert z_{beha}\rvert)$
$\begin{cases} H_0: & p \geq p_0 \\ H_1: & p < p_0 \end{cases}$		$p = P(Z < z_{beha})$
$\begin{cases} H_0: & p \leq p_0 \\ H_1: & p > p_0 \end{cases}$	$z_{beha} = \dfrac{\hat{p}-p_0}{\sqrt{p_0 q_0/n}}$	$p = P(Z > z_{beha})$

4. Hipotesi-testen erabilerarako adibideak

- **1. adibidea**

  Full adder deituriko zirkuituak, bi zenbaki bitar emanik, horien batura eraikitzen du. Batura eraikitzeko behar duen denbora sarreren araberakoa eta zirkuituaren kapazitatearen araberakoa da. Ondorioz, batura eraikitzeko behar den denbora zorizko aldagaia da. Izan bedi T denbora hori eta demagun bere banaketa normala dela $\sigma^2 = 9$ ns^2 delarik. Honelako 10 zirkuitu zoriz aukeratu eta batura eraikitzeko ondorengo denborak behatu dira:

  52.4, 52.4, 53.3, 54.7, 54.9, 55.7, 57.0, 57.4, 57.5, 57.6.

  Batez beste 50 ns baino denbora gehiago behar dutela esan al daiteke % 5 eko adierazgarritasun-mailarekin? Adierazi ongi zein diren ezarriko dituzun hipotesiak eta hartuko duzun erabakia.

- **2. Adibidea**

  Pintura marka bat ekoizten duen enpresa batetako marketing departamentua pinturak lehortzen pasatzen duen denbora ezagutzean interesatua dago. Jakin badakite, lehortze-denborak 81 $min.^2$ bariantzadun banaketa normala jarraitzen duela. Marketing departamentuak dionez, batezbesteko lehortze-denbora 75 baino txikiagoa bada, publizitate kanpaina berezi bat jarriko du martxan. 25 tamainuko lagin batean $\bar{x} = 72{,}3$ minututakoa izan dela behatu da. $\alpha = 0{,}01$ finkatuz, martxan jarriko al da publizitate kanpaina berezi hori?

- **3. Adibidea**

  Autoen pneumatikoen marrazkien higadura-denbora aztertu nahi da. Ikertzaileek badakite higadura-denboraren desbideratze estandarra 1500 ordukoa dela baina batezbesteko higadura-denbora 20000 ordu baino handiagoa den ikusi nahi dute. Horretarako 130 pneumatiko zoriz aukeratu dira eta beraien batezbesteko higadura-denbora 20376 ordukoa izan da. Plantea ezazu hipotesi-test egokia eta esan ezazu % 1 eko adierazgarritasun-mailaz ia pneumatikoen batezbesteko higadura-denbora 20000 ordu baino handiagoa den.

- **4. Adibidea**

  Datuak bistaratzeko aplikazio bat garatu dugu eta aplikazioak memoriara kargatzeko behar duen denbora (X) aztertu nahi dugu. $n = 50$ aldiz kargatu ondoren, 0.91 segundoko batezbestekoa eta 0.19 segundoko desbideratze estandarra behatu dira. Aplikazioa merkaturatu ahal izateko, aplikazioa kargatzeko batez besteko denbora 1 segundo baino gehiago ez dela ikustea da gure helburua. Gure helbururako hipotesi kontraste egokia plantea ezazu. Zein estatistikotan oinarrituko zara testa burutzeko? Adierazgarritasun-maila $\alpha = 0{,}05$ finkaturik, kalkula ezazu p-balioa eta har ezazu dagokion erabakia.

- **5. Adibidea**

  Zerbitzu informatikoak eskaintzen dituen enpresa bat beste eraikuntza baten bila dabil instalazio berriei leku emateko. Eraikuntza berria oraingo tokitik gertu egotea behar dute langileen eta produktuen higikortasuna baimentzeko. Eguerdian, trafiko gehien dagoenean, toki batetik bestera joateko behar den denborak banaketa normala jarraitzen duela badakite. Printzipioz egokia izan daitekeen eraikuntza bat topatu dute eta berau aukeratuko dute baldin eta eraikuntza zaharretik berrirako batezbesteko denbora 20 minutu baino gehiago ez bada. Erabakia hartzen laguntzeko 10 bidaia egin dituzte eta hauek dira behaturiko datuak: $\sum_i x_i = 216$, $\sum_i x_i^2 = 4921$. $\alpha = 0{,}05$ adierazgarritasun-mailarekin, eraikin berri hau aukeratuko al dute?

- **6. Adibidea**

Aurreko adibidearekin jarraituz, eraikuntza zaharretik berrirako denboraren bariantza $25min.^2$-takoa dela esan al dezakegu, $\alpha = 0{,}05$ adierazgarritasun-mailaz?

- **7. Adibidea**

Gozokietan erabiltzen den A substantzia kantitate handiegitan jateak ondorio kaltegarriak eduki ditzake eta horregatik merkatuan dagoen gozokien azterketa bat egin da. Datuen normaltasuna onartuz eta behaturiko lagina ondorengoa dela jakinik, gozokien A substantzia edukiaren desbideratze estandarra % 4 koa dela esan al dezakegu % 5 eko adierazgarritasun-mailaz?

8.4	11.9	10.9	10.6	4.6	4.9	10.7	18.5	11.3	10.7	10.6	7.4

- **8. Adibidea**

Pakete eta fardelak premiaz entregatzen dituen enpresa batek, goizeko 9tarako jasotzen dituen fardelen 90 %a gutxienez, eguerdiko 12tarako entregatuta duela esaten du (hiri barruan egin beharrekoetara mugatuz). 225 fardeletatik 80 % bakarrik entregatu badu garaiz, $\alpha = 0{,}01$ adierazgarritasun-mailaz esan al daiteke enpresak bere hitza betetzen duela?

5. Bi laginetarako hipotesi-test arruntenak

Orain arte lagin batean bakarrik erreparatu gara eta konfiantza-tarte eta hipotesi-testak parametro bakar baten gainean egin ditugu (μ, σ^2 edo p). Orain gure lana ikusitako metodo horiek zabaltzea izango da, bi populazio desberdinetako parametroak konparatu ahal izateko.

1 populazioan	2 populazioan
$X_1 \sim f_1$	$X_2 \sim f_2$
$EX_1 = \mu_1$	$EX_2 = \mu_2$
$VAR(X_1) = \sigma_1^2$	$VAR(X_2) = \sigma_2^2$

Populazio bakoitzak bere ezaugarriak izanik, bakoitzetik n_i tamainako zorizko lagin bakuna ($i = 1, 2$) atera duguneko kasuan kokatuko gara.

- **X_1 eta X_2 aldagai kuantitatiboak. $EX_1 = \mu_1$ eta $EX_2 = \mu_2$ batezbestekoen konparaketarako hipotesi-testa.**

  Baldintzak:

  1. σ_1 eta σ_2 ezezagunak baina berdinak ($\sigma_1 = \sigma_2$)
  2. Populazioak normalak.
  3. Bi laginak elkarrekiko askeak

Hipotesiak	Estatistikoa eta bere banaketa H_0 pean	p-balioa
$\begin{cases} H_0: & \mu_1 - \mu_2 = 0 \\ H_1: & \mu_1 - \mu_2 \neq 0 \end{cases}$	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$ (*)	$p = 2P(t_{n_1+n_2-2} > \lvert t_{beha}\rvert)$
$\begin{cases} H_0: & \mu_1 - \mu_2 \leq 0 \\ H_1: & \mu_1 - \mu_2 > 0 \end{cases}$		$p = P(t_{n_1+n_2-2} < t_{beha})$
$\begin{cases} H_0: & \mu_1 - \mu_2 \geq 0 \\ H_1: & \mu_1 - \mu_2 < 0 \end{cases}$	$t_{beha} = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}}$	$p = P(t_{n_1+n_2-2} > t_{beha})$

(*) $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

Oharra: Test honek $t - test$ izena hartzen du

- **X_1 eta X_2 aldagai bitarrak. $P(X_1 = 1) = p_1$ eta $P(X_2 = 1) = p_2$ proportzioen konparaketarako hipotesi-testa.**

  Baldintzak:

  1. $n_i\hat{p}_i\,, n_i\hat{q}_i \geq 5\ ,\ i = 1, 2\ (\hat{q}_i = 1 - \hat{p}_i)$
  2. Bi laginak elkarrekiko askeak

Izan bitez $p_0 = p_1 - p_2$ eta $\hat{p}_0 = \frac{n_1}{n_1+n_2}\hat{p}_1 + \frac{n_2}{n_1+n_2}\hat{p}_2$.

Hipotesiak	Estatistikoa eta bere banaketa H_0 pean	p-balioa
$\begin{cases} H_0: & p_1 - p_2 = 0 \\ H_1: & p_1 - p_2 \neq 0 \end{cases}$	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0(1/n_1 + 1/n_2)}} \approx N(0,1)$	$p = 2P(Z > \lvert z_{beha}\rvert)$
$\begin{cases} H_0: & p_1 - p_2 \geq 0 \\ H_1: & p_1 - p_2 < 0 \end{cases}$		$p = P(Z < z_{beha})$
$\begin{cases} H_0: & p_1 - p_2 \leq 0 \\ H_1: & p_1 - p_2 > 0 \end{cases}$	$z_{beha} = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0(1/n_1 + 1/n_2)}}$	$p = P(Z > z_{beha})$

6. Hipotesi-testen erabilerarako adibideak - 2 lagin

- **1. adibidea**

  Informatikarien soldatak aztertu nahi dira, bereziki emakumeen eta gizonezkoen bazte besteko soldatak berdinak diren ala ez aztertu nahi da. Horretarako $n_1 = 25$ emakumezko eta $n_2 = 30$ gizonezko informatikarien laginetatik urteko soldatak (mila euro) jaso dira. Hauek dira behatutako emaitzak:

Informatikaria	Lagin-tamainuak	Batezbestekoak	Desbideratze estandarrak
Emakumea	25	25.5	3.6
Gizonezkoa	30	28.7	3.7

  Datu horien arabera, emakumezko eta gizonezkoen urtekoa batez besteko soldatak desberdinak direla esan al daiteke? Adieraz ezazu zein adierazgarritasun maila erabili duzun.

  Egin duzun test hori zuzena izateko, zorizko aldagaien banaketaren gainean balditzaren bat bete beharko litzateke??

- **2. adibidea**

  Lenguaia naturalaren azterketa itzultzaile automatikoak eraikitzeko guztiz beharrezkoa da. Zure ikerketa taldeak dagoeneko eraikita dagoen itzultzaile baten gainean (Z itzultzailea) aldaketa batzuk egin ditu eta horrela itzultzaile berria (B itzultzailea) eraiki. Bi itzultzaileen eraginkortasuna konparatzeko laginak hartu dira: $n_1 = 350$ esalditatik Z itzultzaileak 210 egoki itzuli ditu eta $n = 380$ esalditatik B itzultzaileak 240 esaldi egoki itzuli ditu. Datu horiek eskutan, esan dezakegu itzultzaile berria zaharra baino eraginkorragoa dela?

5. Adibidea

$X \equiv$ "Eraikin zaharberritze berriztatze denbora"

$X \sim N($

Hipotesia: $\begin{cases} H_0: \mu \leq 20 \\ H_1: \mu > 20 \end{cases}$

$n = 10$

$\sum x_i = 216$

$\sum x_i^2 = 4921$

$$t = \frac{\bar{x} - 20}{S/\sqrt{10}} \sim t_{n-1}$$

adierazgarritasun-maila: $\alpha = 0'05$

$$S = \sqrt{S^2} = 5'33$$

$$S_n^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = \frac{4921}{10} - 21'6^2 = 25'54$$

$$S^2 = \frac{n}{n-1} S_n^2 = 28'37$$

$$t = \frac{21'6 - 20}{5'33/\sqrt{10}} = \frac{1'6}{5'33/\sqrt{10}} = 0'95$$

P balioa

$$p = P(t_{n-1} > t_{balio}) = p = P(t_9 > 0'95) \Rightarrow P(t_9 > 0'883) = 0'2$$

$0'1 < p < 0'2$ $\qquad$ $P(t_9 > 1'38) = 0'1$

$\alpha = 0'05$ $\qquad$ $\alpha < p \Rightarrow H_0$ onartu

8. Adibidea

$p = \%90$

$n = 225 \Rightarrow 0'8 \Rightarrow 180$ fardel

$\alpha = 0'01$

Hipotesia: $\begin{cases} H_0: & p \geq p_0 = 0'9 \\ H_1: & p \not\geq p_0 = 0'9 \end{cases}$

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 \, q_0 / n}} \sim N(0,1)$$

$$z_{behatua} = \frac{0'8 - 0'9}{\sqrt{0'9 \cdot 0'1 / 225}} = -5$$

p balioa: $p = P(z < -5) = P(z > 5) < P(z > 4'09) = 1 - 0'99997843 =$
$= 2.157 \cdot 10^{-5}$

$\alpha = 0.01 > p$ balioa $\Rightarrow$ H_0 errefusatu

2. Adibidea 6. Hipotesi - testak

Z itzultzailea $\Rightarrow n_1 = 350$ 210 egoki

B itzultzailea $\Rightarrow n_2 = 380$ 240 egoki

$\hat{p}_1 = \frac{210}{350} = 0'6$

$\hat{p}_2 = \frac{240}{380} = 0'632$

$p_0 = p_1 - p_2 =$

Hipotesia = $\begin{cases} H_0: p_1 - p_2 \geq 0 \\ H_1: p_1 - p_2 < 0 \end{cases}$

$\hat{p}_0 = \frac{n_1}{n_1+n_2}\hat{p}_1 + \frac{n_2}{n_1+n_2}\hat{p}_2 =$

$= \frac{350}{730} \cdot 0'6 + \frac{380}{730} \cdot 0'63 =$

$= 0'616$

$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0\left(1/n_1 + 1/n_2\right)}} = \frac{-0'032}{\sqrt{0'616 \cdot 0'384\left(\frac{1}{350}+\frac{1}{380}\right)}} =$

$= -0'833$

$p = P(Z < -0'833) = 1 - P(Z < 0'833) =$

$= 1 - 0'7967 = 0'2033$

p balioa $\Rightarrow 0'2033 > \alpha = 0'05$ H_0 onartu

9. Gaia Ariketak

6. Ariketa

11 lurretan

Benetako kantitatea: 548 mg/kg

$X \equiv$ "fosforo kantitatea"

$\bar{X} = 587$

$SD(x) = 10$

Hipotesia: $\begin{cases} H_0: \mu = 548 \\ H_1: \mu \neq 548 \end{cases}$

$X \sim N(\mu, 10^2)$

$\bar{X} \sim N(\mu, \frac{10^2}{n}) \Rightarrow z = \frac{\bar{X} - \mu}{10/\sqrt{n}} \sim N(0,1)$

Ho egiazkoa balitz

$z = \frac{\bar{X} - 548}{10/\sqrt{11}} \sim N(0,1)$

Laginean

$z_{behatua} = \frac{\bar{X} - 548}{10/\sqrt{11}} = \frac{587 - 548}{10/\sqrt{11}} = 12'93$

$p = P(z > 12'93) \cdot 2 = 0$

7. Ariketa

$n = 16$ alambre

$\bar{X} = 2160$

$SD(x) = 30$

$\bar{X}_1 = 2150$

a)

Hipotesia: $\begin{cases} H_0: \mu \leq 2150 \\ H_1: \mu > 2150 \end{cases}$

Hipotesia: $\begin{cases} H_0: \sigma^2 \leq \sigma_0^2 \end{cases}$

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$$t_{beha} = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{2160 - 2150}{30/\sqrt{16}} = 1'25$$

P balioa

$$p = P(t_{15} > 1'25) =$$

8. Ariketa

$p \equiv$ "mugara kelduz curretile apurtzen diren piezak"

$p > 0'2$

a) $n=60 \Rightarrow 15$ apurt.

$\hat{p} = \frac{15}{60} = 0'25$

Hipotesia: $\begin{cases} H_0: & p \leq 0'2 \\ H_1: & p > 0'2 \end{cases}$

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$$

$$z_{balio} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0'25 - 0'2}{\sqrt{\frac{0'2 \cdot 0'8}{60}}} = 0'968$$

$p = P(z > 0'968) = P(z > 0'968) \geq P(0'768) = 0'7823$

P balioa

b)

9. Ariketa

22 41 22 27 23 35 30

33 24 22 28 22 24

$$\bar{X} = \frac{1}{n}\sum x_i = \frac{363}{13} = 27'92$$

Hipotesia $\begin{cases} H_0: \mu \geq 25 \\ H_1: \mu < 25 \end{cases}$

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

$$= \frac{1}{n}\sum x_i^2 - \bar{x}^2 =$$

$$= 808'85 - 779'53 =$$

$$= 29'32$$

$$t_{3eha} = \frac{27'92 - 25}{S/\sqrt{n}} = \frac{27'92-25}{5'64/\sqrt{13}} =$$

$$= 1'87$$

$$S^2 = \frac{13}{12} 29'32 = 31'77$$

$$S = 5'64$$

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$p = P(t_{13} < 1'87) = \cancel{0'96}$

$P(t_{12} > 21'78)$

$P(t_{12} > 2'18$

15. Ariketa

$n = 16$

162.9 180.8 184.8 189.8 194.8 200.2 201.9 206.9

202.2 208.4 226.3 222.7 228.5 232.4 239.8 258.6

$X \sim N(0,1)$

Konfiantza-tartea: ~~0~~'95

$\bar{X} = 209'25$

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

$$= \frac{1}{n}\sum x_i^2 - \bar{x}^2 =$$

$$= 44542'11 - 43995'0625 =$$

$$= 547'04$$

~~$\bar{X} - \mu$~~ $\left(\bar{x} \mp t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}\right)$

$$I_\mu^{0'95} = \left(209'25 \mp t_{15,0'025} \frac{23'39}{\sqrt{16}}\right) = \quad S = \sqrt{S^2} = 23'39$$

$$= \left(209'25 \mp 2'13 \cdot \frac{23'39}{\sqrt{16}}\right) = (197'23, 222'21)$$

17. Ariketa

~~bizitt~~ 2.000 M $\rightarrow$ 3 M

$n = 1000$

300 hilda $\Big\}\ \hat{p} = \frac{300}{1000} = 0'3$

$$I_{p}^{0'90} = \left(\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \left(0'3 \mp z_{0'05}\sqrt{\frac{0'3 \cdot 0'7}{1000}}\right) =$$

$$= \left(0'3 \mp 1'645 \cdot \sqrt{\frac{0'3 \cdot 0'7}{1000}}\right) = (0'276, 0'324)$$

28. Ariketa

$$f(x) = \begin{cases} \frac{1}{2} e^{-\lambda} \frac{\lambda^x}{x!} + \frac{1}{2} e^{-\mu} \frac{\mu^x}{x!} & x = 0,1,2,\dots \\ 0 & \text{bestela} \end{cases}$$

a) Probabilitate-funtzioa izateko $\cancel{f(x)=}$ $\sum^n f(x) = 1$

$$\sum f(x) = \sum^n \frac{1}{2} e^{-\lambda} \frac{\lambda^x}{x!} + \frac{1}{2} e^{-\mu} \frac{\mu^x}{x!} = \frac{1}{2} \sum e^{-\lambda} \frac{\lambda^x}{x!} + e^{-\mu} \frac{\mu^x}{x!}$$

b)

c) $EX = \sum x f(x)$

7. Ariketa

$$f(x) = \begin{cases} 2(1 - 1/x^2) & 1 \le x \le 2 \\ 0 & \text{bestela} \end{cases}$$

a) Banaketa-funtzioa

$\cancel{F(x) =}$ $F(x) = \begin{cases} 2\left(x + \frac{1}{x} - 2\right) & 1 \le x \le 2 \\ 0 \end{cases}$

$$\int_1^x \frac{dx}{x^2} = \int_1^x x^{-2} dx = -\frac{x^{-1}}{\cancel{2}} = -\frac{1}{x}$$

~~F(x) = P(X ≤ x~~

$$F(x) = \int_1^x 2\left(1 - \frac{1}{x^2}\right) dx = 2\int_1^x 1 - \frac{1}{x^2} dx = 2\left(\int_1^x dx - \int_1^x \frac{dx}{x^2}\right) = 2x + \frac{2}{x}\Big|_1^x = 2x + \frac{2}{x} - 2 - 2 =$$

$$= 2x + \frac{2}{x} - 4$$

b)

$$EX = \int_1^2 x f(x) = 2\int_1^2 x \cdot \frac{x}{x^2} dx = 2\int_1^2 x - \frac{1}{x} dx =$$

$$= 2\left[\frac{x^2}{2} - \ln x\right]_1^2 = 2\left[\frac{4}{2} - \ln 2 - \frac{1}{2} + \ln 1\right] =$$

$$= 1'61$$

$$VAR(x) = EX^2 - \mu^2 =$$

$$= 2'66 - 2'59 = 0'05$$

$$EX^2 = \int_1^2 x^2 f(x) = 2\int_1^2 x^2 - 1 dx = \frac{2x^3}{3} - 2x \Big|_1^2 =$$

$$= \frac{16}{3} - 4 - \frac{2}{3} + 2 = \frac{16}{3} - \frac{12}{3} - \frac{2}{3} + \frac{6}{3} = \frac{8}{3} = 2'66$$

21. Ariketa

$$f(x) = \begin{cases} 20x^3(1-x) & 0 \leq x \leq 1 \\ 0 & \text{bestela} \end{cases}$$

a) $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^1 f(x)\,dx = \int_0^1 20x^3 - 20x^4 dx = 5x^4 - 4x^5 \Big|_0^1 =$

$$= 5 - 4 = 1$$

b) $F(x) = \int_0^x f(x)\,dx = \int_0^1 20x^3 - 20x^4 = 5x^4 - 4x^5 \Big|_0^x = 5x^4 - 4x^5$

c) $P(X \leq 2/3) = 0'46$

14. Ariketa

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{non } EX=\mu \text{ eta } VAR(x)=\sigma^2$$

a)

$$|X-\mu| \geq 2\sigma$$

b)

$$P(|X-\tfrac{1}{6}| \geq 2 \cdot 0'18) \leq \frac{1}{4}$$

$$\cancel{\bar{X} = \frac{1}{n}\sum f(x_i) = \frac{1}{6}}$$

$$E(x) = \frac{1}{n}\sum x f(x_i) = \frac{2'64}{6} = 0'44$$

$$VAR(x) = \frac{1}{n}\sum (x-\bar{x})^2 = \frac{1}{n}\sum x^2 - \bar{x}^2 = 0'06 \; 0'032$$

$$\sigma = 0'18$$

c) $P|-1$

13. Ariketa

$P(O_0) = 0'7$

$P(L_0) = 0'2$

$P(E_0) = 0'1$

$X \equiv$ "datorren urtean izango duten irabazia"

a) X ren probabilitate legea

$f(x) = P(X = 80000) = 0'7$

$$f(x_i) \begin{cases} 0'7 & x = 80000 \\ 0'2 & x = 50000 \\ 0'1 & x = 20000 \end{cases}$$

b)

$P(X \leq 50000) = 0'3$

c)

$E_x = \sum x f(x_i) = \cancel{50}\,80000 \cdot 0'7 + 50000 \cdot 0'2 + 20000 \cdot 0'1 =$

$= 68000$ euro

d)

$$f(x_i) \begin{cases} 0'7 & x = 77000 \\ 0'2 & x = 47000 \\ 0'1 & x = \cancel{40000}\ 37000 \end{cases}$$

$E_x = \sum x f(x_i) = 67000$ euro

~~Ez litzaioke ""~~ itxaropenari aurbera ez zaion ongi irteten baina eskaintzen bat egongo balitz irabazitako lukeen dirua askoz handiagoa litzateke.

35. Ariketa

$A_i \equiv$ "i = Hardie edo txilla aitoa"

$P(A_T) = \frac{3}{4}$ $\quad P(G|A_T) = 0'9$

$P(A_H) = \frac{1}{4}$ $\quad P(G|A_H) = 0'6$

$$P(G) = P(A_T) \cdot P(G|A_T) + P(A_H) \cdot P(G|A_H) =$$
$$= 0'675 + 0'15 = 0'825$$

$$P(A_T|G) = \frac{P(G \cap A_T)}{P(G)} = \frac{0'675}{0'825} = 0'82$$

38. Ariketa

Askeak badia $P(A \cap B) = P(A) \cdot P(B)$

$$P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B)$$

$$P(\bar{A} \cap B) = P(B|\bar{A}) \cdot P(\bar{A}) = P(B|\bar{A}) \cdot (1 - P(A)) =$$
$$= P(B|\bar{A}) - P(B|\bar{A}) \cdot P(A)$$

b)

$$P(A_0|F_1) = \frac{P(F_1 \cap A_0)}{P(F_1)} = 0$$

$$P(A_1|F_1) = \frac{P(F_1 \cap A_1)}{P(F_1)} = \frac{0'06}{0'13} = 0'458$$

$$P(F_1) = P(F_1|A_0)\cdot P(A_0) + P(F_1|A_1)\cdot P(A_1) + P(F_1|A_2)\cdot P(A_2) =$$

$$= 0 + 0'2\,0'3 + 0'2\cdot 0'355 = 0'13$$

$$P(F_1 \cap A_1) = P(A_1)\cdot P(F_1|A_1) = 0'3\cdot 0'2 = 0'06$$

$$P(A_2|F_1) = \frac{P(F_2 \cap A_1)}{P(F_1)} = \frac{0'071}{0'13} = 0'54$$

34. Arilketa

$$P(BA \cap BU) = 0'05$$

$$P(BA \cap AU) = 0'079$$

$$P(BAA \cap ABU) = 0'089$$

$$P(AA \cap AU) = 0'782$$

$$P(U_b|A_b) = \frac{P(A_b \cap U_b)}{P(A_b)} = \frac{0'05}{0'129} = 0'39$$

$$P(A_b) = P(A_b \cap U_b) + P(A_b \cap U_a) = 0'05 + 0'079 = 0'129$$

$$P(U_a|A_b) = \frac{P(A_b \cap U_a)}{P(A_b)} = \frac{0'079}{0'129} = 0'61$$

22. Ariketa

$P(M_1) = 0'5 \quad P(M_2) = \cancel{0'5}\,0'3 \quad P(M_3) = 0'2$

$P(\cancel{M_1}\, G|M_1) = 0'25 \quad P(G|M_2) = 0'2 \quad P(G|M_3) = 0'1$

a)

$P(M_1 \cap G) = P(G|M_1) \cdot P(M_1) = 0'25 \cdot 0'5 = 0'125$

$P(G|M_1) = \dfrac{P(M_1 \cap G)}{P(M_1)}$

b) $P(G) = ?$

$P(G) = P(M_1 \cap G) + P(M_2 \cap G) + P(M_3 \cap G) =$

$= P(G|M_1) \cdot P(M_1) + P(G|M_2) P(M_2) + P(G|M_3) P(M_3) = 0'205$

c)

$P(M_1|G) = \dfrac{P(G|M_1) \cdot P(M_1)}{P(G|M_1) \cdot P(M_1) + P(G|M_2) \cdot P(M_2) + P(G|M_3) \cdot P(M_3)} = 0'61$

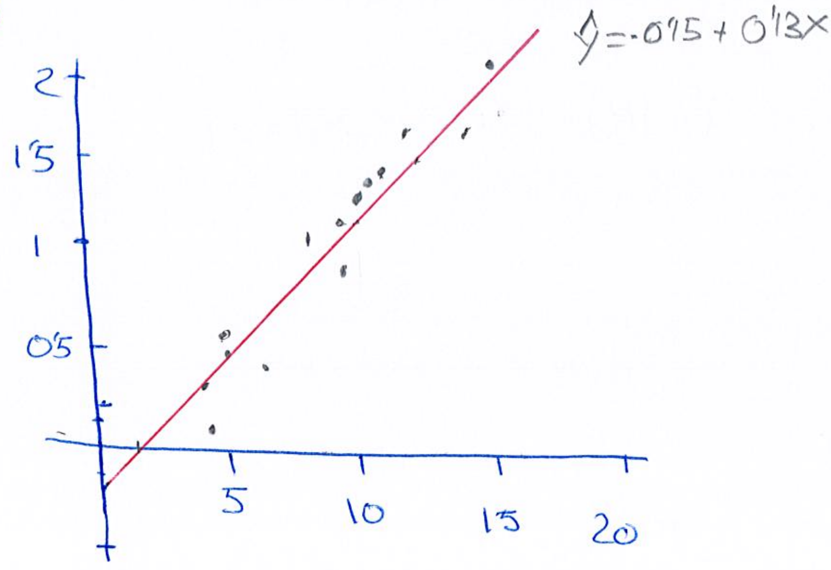
Bayes-en teorema

$P(M_2|G) = \dfrac{P(G|M_2) \cdot P(M_2)}{P(G)} = 0'29$

$P(M_3|G) = \dfrac{P(G \cap M_2)\ \cancel{P(M_3)}}{P(G)} = 0'1$

11. Ariketa

a)



Bai ematen du erlazio zuzena dagoela

b)

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{Sxy}{Sxx} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{\sum x \cdot y - \frac{1}{n}\sum x \sum y}{\sum x^2 - \frac{1}{n}\sum x^2} =$$

$$= \frac{174'5 - \frac{139 \cdot 158}{15}}{1503 - 1288'06} = \frac{174'5 - 146'41}{1503 - 1288'06} = 0'13$$

$$b_0 = \bar{y} - b_1 \bar{x} = 1'05 - 0'13 \cdot 9'26 = -0'15$$

$$\hat{y} = -0'15 + 0'13X \Rightarrow x = \frac{0'15}{0'13} = 1'15$$

22'82

$$r_{xy} = \frac{28'09}{\sqrt{214'94 \cdot 5'43}} = 0'82$$ $r \approx 1$ Korrelazio-koefizientea

dadioa - zuzena

18. Ariketa

a)

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{20}\cdot 1166 = 58'3$$

$$M_e = \frac{55+49}{2} = 52$$

Datuak ordenatzerako egokiena kasu honetan Mediana da, Batezbesteko aritmetikoa ~~altuk agiten~~ bai Sentikorra baita muturreko balioei

b)

$[l_i, l_{i+1})$	n_i	N_i	f_i	F_i	x_i
[20, 40)	7	7	0'35	0'35	30
[40, 60)	5	12	0'25	0'6	50
[60, 80)	6	18	0'3	0'9	70
[80, 215)	2	20	0'1	1	147'5

Sakabanapen absolutua:

$$S_n^2 = \frac{1}{n}\sum n(x_i - \bar{x})^2 \Rightarrow \bar{x} = \frac{1}{n}\sum n x_i = \frac{1125}{20} = 58'25$$

$$S_n^2 = \frac{1}{20}\sum n x_i^2 - \bar{x}^2 =$$

$$= \frac{5830}{20} - \quad = 291'5 - 3451'56 = -3160$$

16. Aribketa

~~Xe~~

1'86 1'88 1'9 1'96 2'36 2'43 2'87 2'91 2'92 2'92

2'99 3'00 3'01 3'03 3'04 3'05 3'06 3'15 3'18 3'20

3'24 3'24 3'29 3'32 3'33 3'34 3'35 3'37 3'45 3'45

3'46 3'49 3'53 3'54 3'57 3'61 3'67 3'75 3'93 3'99

3'99 4'08 4'11 4'12 4'19 4'19 4'28 4'41 4'45 4'47

4'48 4'68 4'89 4'98 5'85

$M_e = 3'36$

$Heina = \cancel{\frac{X_{max}+X_{min}}{2}} = X_{max} - X_{min} = 5'85 - 1'86 =$

$= 3'99$

$RI = Q_3 - Q_1 = 3'5 - 3'02 = 0'48$

$\bar{x} = \frac{1}{n}\sum x_i = \frac{187'27}{55} = 3'4$

$S_n^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = 12'788 - 11'56 =$

$= 1'228$

$S_n = \sqrt{S_n^2} = 1'11$

$CV = \frac{S_n}{|\bar{x}|} = \frac{1'11}{3'4} = 0'33$

15. Anileše

a)

$$\bar{x}_{n+1} = \frac{1}{n+1}\sum^{n+1} x_i = \frac{1}{n+1}\left(\sum^{n} x_i + x_{n+1}\right) = \frac{1}{n+1}\left(n\bar{x} + x_{n+1}\right)$$

$\sum^{n} x_i \to n\bar{x}$

b)

$$(n+1)S^2_{n+1} = nS^2_n + \frac{n}{n+1}\left(\bar{x}_n - x_{n+1}\right)^2$$

$$(n+1)S^2_{n+1} = \cancel{\frac{n+1}{n+1}}\sum^{n+1}(x_i - \bar{x}_{n+1})^2 = \sum^{n+1} x_i^2 - (n+1)(\bar{x}_{n+1})^2 =$$

$$= \sum^{n} x_i^2 + x_{n+1}^2 - \frac{1}{n+1}(n\bar{x} + x_{n+1})^2 =$$

$$= nS_n^2 + n\bar{x} + x_{n+1}^2 - \frac{1}{n+1}(n\bar{x} + x_{n+1})^2 =$$

$$S_n^2 = \frac{1}{n}\sum^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left[\sum^{n} x_i - n\bar{x}\right]$$

$$= nS_n^2 + n\bar{x}^2 + x_{n+1}^2 - \frac{n^2\bar{x}^2 + 2n\bar{x}x_{n+1} + x_{n+1}^2}{n+1} =$$

$$= nS_n^2 + \frac{(n+1)n\bar{x}^2 + (n+1)x_{n+1}^2 - n^2\bar{x}^2 + n\bar{x}x_{n+1} - x_{n+1}^2}{n+1} =$$

$$= nS_n^2 + \frac{\cancel{n^2\bar{x}^2} + n\bar{x}^2 + nx_{n+1}^2 + \cancel{x_{n+1}^2} - \cancel{n^2\bar{x}^2} - 2n\bar{x}x_{n+1} - \cancel{x_{n+1}^2}}{n+1} =$$

$$= nS_n^2 + \frac{1}{n+1}\left(n\bar{x}^2 + nx_{n+1}^2 - 2n\bar{x}x_{n+1}\right) =$$

$$= nS_n^2 + \frac{n}{n+1}\left(\bar{x} - x_{n+1}\right)^2$$

7. Ariketa

$n=10$

S, S, F, S, S, S, F, F, S, S

a) $\hat{p} = \frac{7}{10} = 0'7$

b) $\bar{x} = \frac{1}{n}\sum x_i = \frac{1+1+0+1+1+1+0+0+1+1}{10} = \frac{7}{10} = 0'7$

c) $\hat{p} = \frac{S}{25} = 0'8 \Rightarrow S = 0'8 \cdot 25 = 20$

14. Ariketa

a)

$y_i = x_i - \bar{x} \quad i = 1, 2, \dots, n \qquad S^2_{n,x}$ eta $S^2_{n,y}$

$y_i = x_i - \bar{x} \Rightarrow \frac{1}{n}\sum y_i = \frac{1}{n}\sum x_i - \bar{x} \Rightarrow \bar{y} = \bar{x} - \frac{\cancel{n}\bar{x}}{\cancel{n}} \Rightarrow$

$\Rightarrow \bar{y} = 0$

$y_i = x_i - \bar{x}$

$\Rightarrow \frac{1}{n}\sum(y_i - \cancel{\bar{y}})^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 \Rightarrow S^2_{n,y} = S^2_{n,x}$

b)

$z_i = \frac{x_i - \bar{x}}{S_{nx}} \qquad S^2_{nz}$? eta S_{nz}?

$\Rightarrow \frac{1}{n}\sum z_i = \frac{1}{n}\sum \frac{x_i - \bar{x}}{S_{nx}} \Rightarrow \bar{z} = \frac{\frac{\cancel{S_{nx}}}{n}}{S_{nx}} \sum x_i - \bar{x} \Rightarrow$

$\Rightarrow \bar{z} = \cancel{S_{nx}} \frac{(\bar{x} - \bar{x})}{S_{nx}} = 0$

$\frac{1}{n}\sum(z_i - \cancel{\bar{z}})^2 = \frac{1}{n}\sum \frac{(x_i - \bar{x})^2}{S^2_{nx}} \Rightarrow S^2_{nz} = \frac{S^2_{nx}}{S^2_{nx}} = 1 \qquad \underline{S_{nz} = 1}$

10. Ariketa

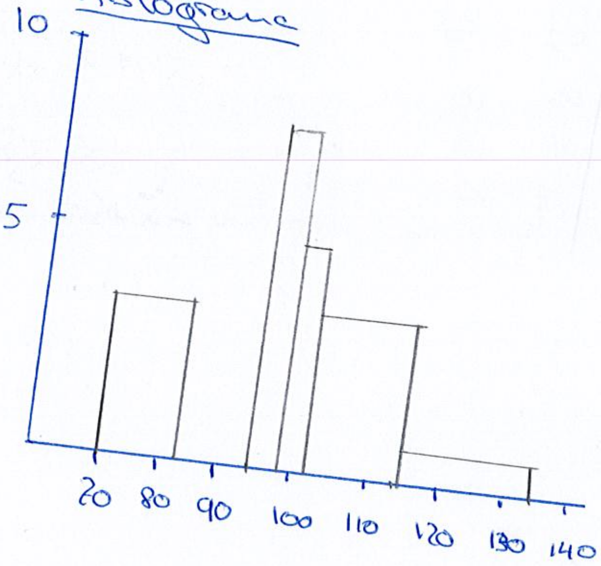
a)

Maiztasun-taula

$[l_i, l_{i+1})$	n_i	N_i	f_i	F_i	x_i	a_i
[70,82)	3	3	0'15	0'15	76	12
[82,94)	0	3	0	0'15	88	12
[94,98)	8	11	0'4	0'55	96	4
[98,102)	5	16		0'8	100	4
[102,114)	3	19		0'95	108	12
[114,134)	1	20		1	124	20

b)

Histograma



c)

$[l_i, l_{i+1})$	n_i	N_i	f_i	F_i	x_i	a_i
[0,4)	1	1	0'05	0'05	2	4
[4,6)	4	5	0'2	0'25	5	2
[6,8)	5	10	0'25	0'5	7	2
[8,10)	6	16	0'3	0'8	9	2
[10,15)	4	20	0'2	1	12'5	5

4 Ariketa

$n = 57$

Maiztasun-taula

$$k = 1 + 3'223 \log_{10} 57 = 6'66 \simeq 7$$

$$Heina: X_{max} - X_{min} = 67$$

$$a_i = \frac{Heina}{k} = \frac{67}{7} \simeq 10$$

$[l_i, l_{i+1})$	n_i	f_i	N_i	F_i	x_i	a_i
[10, 20)	5	0'087	5	0'087	15	10
[20, 30)	19	0'33	24	0'421	25	10
[30, 40)	10	0'175	34	0'596	35	10
[40, 50)	13	0'228	47	0'825	45	10
[50, 60)	4	0'07	51	0'895	55	10
[60, 70)	4	0'07	55	0'965	65	10
[70, 80)	2	0'035	57	1	75	10

Histograma