
INGENIARITZAKO METODO ESTATISTIKOAK

5. Laginketara sarrera



emari ta zabal zazu:



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

5. Laginketara sarrera

5.1 Populazioa eta lagina

5.2 Zorizko laginketak

5.2.1 Zorizko laginketa bakuna

5.2.2 Zorizko laginketa geruzatua

5.2.3 Zorizko laginketa konglomeratua

5.2.4 Zorizko laginketa sistematikoa

5.3 Orokortasunak

5.3.1 Estatistikoak

5.4 Laginketa-teoriako banaketak

5.4.1 Pearson-en χ^2 banaketa

5.4.2 Student-en t banaketa

5.4.3 Snedecor-en F banaketa



5. Laginketara sarrera

5.5 Estatistikoen banaketak

- 5.5.1 Batezbestekoaren lagin-banaketa
- 5.5.2 Bi laginen batezbestekoen arteko kenduraren banaketa
- 5.5.3 Bariantzaren lagin-banaketa
- 5.5.4 Kuasibariantzaren arteko zatiduraren lagin-banaketa
- 5.5.5 Laginketaren proportzioaren banaketa
- 5.5.6 Bi laginen proportzioen arteko kenduraren banaketa



5.1 Populazioa eta lagina

Inferentziaren bidez laginketatik ondorioztatutako emaitzak populaziora orokor daitezke, konfiantza-maila zehatz batez.

Laginketa egokia izateko, laginetatik lortutako emaitzak populazio osoaren adierazgarriak izan behar dute. Gai honetan laginketaren oinarriak aztertuko ditugu.

Populazioa

Aztertu nahi den multzoa. Azterketa estatistikoa multzo honetan gauzatuko da. (N)

Lagina

Populazioaren azpimultzo bat da. (n)

Lagina populazioaren adierazgarria izan behar da. Hau da, lagina populazioari buruzko informazioa lortzeko erabiltzen dugunez, laginaren azpimultzo adierazgarria izan behar da.



emari ta zabal zazu:



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

5.1 Populazioa eta lagina

Adibidea

1) Demagun Bilbo hirian bizi diren biztanleen osasunari buruzko ikerketa egin nahi dugula.

Populazioa: Bilboko biztanle guztiak

Lagina:

- Bilboko auzo bakoitzean bizi den familia bat hartuz lortzen den azpimultzoa. ✓
- Hirugarren adineko taldea. ✗

Adibide honetan ikusi den bezala lagina populazioaren adierazgarria izan behar da.

Bi elementu garrantzitsu daude:

1. Tamaina
2. Indibiduen aukeraketa



5.2 Zorizko laginketa

Laginketa bat zorizko laginketa dela esaten da lagineko indibiduo guztiak zorian aukeratzen direnean eta, beraz, populazioko indibiduo guztiek aukeratzeko probabilitate bera dutenean (a priori).

Populazioa eta lagina

Zorizko laginketa

Orokortasunak

Laginketa-teoriako banaketak

Estatistikoen banaketak



5.2 Zorizko laginketa

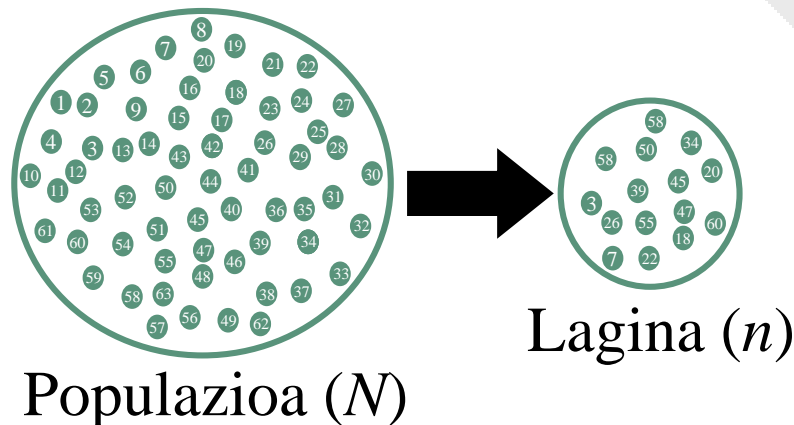
5.2.1 Zorizko laginketa bakuna

Metodorik sinpleena da. Lagin bat lortzeko, populazioko elementuak zenbakitu eta laginak izan behar dituen n elementuak zoriz aukeratzen dira.

Adibidea

2) Demagun ikastetxe batean 1300 ikasle daudela, eta 100 ikasleek sortutako lagina zorizko laginketa sinplea erabiliz eraiki nahi dugula:

1. Ikasle bakoitzari zenbaki bat ematen diogu
2. 1,2,...,1300 zenbakiren artean 100 zoriz aukeratzen ditugu



emari ta zabal zazu:



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

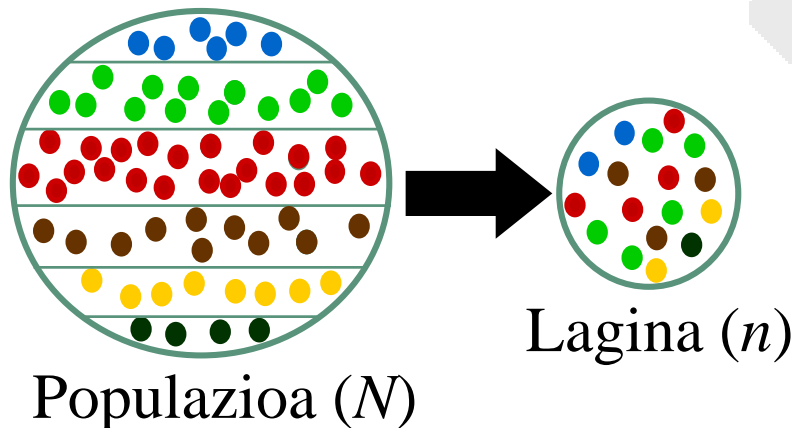
5.2 Zorizko laginketa

5.2.2 Zorizko laginketa geruzatua

Populazioa geruzetan (taldeetan) banatzean oinarritzen da. Ondoren, N tamainako populaziotik n tamainako lagina lortu nahi bada, geruza bakoitzetik n_1, n_2, \dots, n_k (k geruza badaude) elementu aukeratzen dira, $n_1 + n_2 + \dots + n_k = n$ izanik.

Adibidea

- 3) Ikastetxe batean ikasleen altuera aztertu nahi da. Ikasleen altuera printzipioz adinarekin erlazionatuta dagoenez, azterketa zorizko laginketa geruzatua erabiliz egingo da.



5.2 Zorizko laginketa

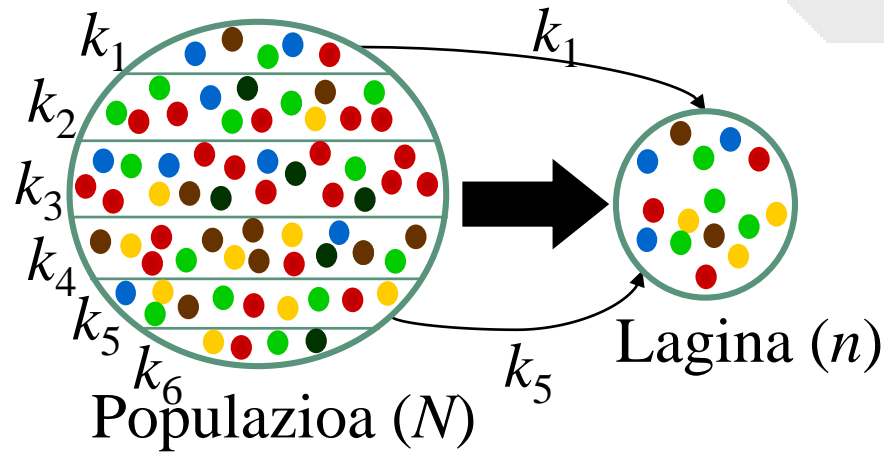
5.2.3 Zorizko laginketa konglomeratua

Aurreko metodoetan populazioko elementuak aukeratzeko genituen, orain berriz populazioko elementuek sortzen dituzten talde batzuk aukeratzeko ditugu, talde hauei konglomeratuak deritze. Zorizko laginketa konglomeratua bi pausutan oinarritzen da:

1. Zoriz konglomeratu batzuk aukeratu (laginaren tamaina kontuan izanik)
2. Aukeratutako konglomeratuen elementu guztiak kontuan hartzen dira.

Adibidea

- 4) Lurralde desberdinetan kokatua dagoen denda-kate handi batean lan egiten duten langileen iritzia ezagutu nahi dugunean.



emari la zabal zazu:



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

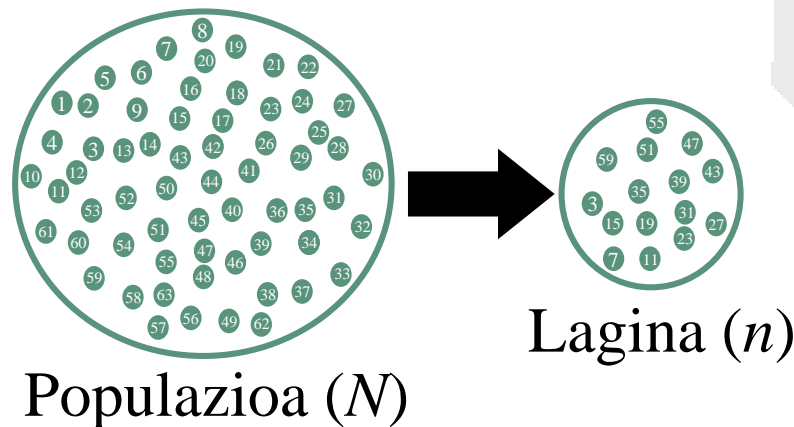
5.2 Zorizko laginketa

5.2.4 Zorizko laginketa sistematikoa

Zorizko laginketa sistematikoan, populazioaren elementu bat zoriz aukeratu ondoren (a_1), elementu honetatik abiatuz lagineko beste elementu guztiak k periodoa erabiliz aukeratzen dira.

Adibidea

- 5) Demagun ikastetxe batean 1300 ikasle daudela eta 100 ikaslek sortutako lagina zorizko laginketa sistematikoa erabiliz eraiki nahi dugula. Hirugarren ikaslea hartu eta 12-naka hurrengoak hartzen ditutgu.



5.3 Orokortasunak

Hipotesiak

1. Populazioak N elementu ditu eta n tamainako zorizko lagin bakuna (z.l.b.) kontsideratuko dugu.
2. Populazioko elementuak zorizko aldagai independenteak dira, euren banaketak berdinak izanik. (berdinki banatuak)
3. Era berean, zorizko lagin bakuneko elementuak, X_1, X_2, \dots, X_n denotatuko ditugunak, zorizko aldagai askeak eta berdinki banatuak dira. Hauen banaketa eta populazioaren banaketa berdina izanik.
4. Hautaketa itzulerarekin egiten da.



5.3 Orokortasunak

5.3.1 Estatistikoak

Laginaren edozein funtziori estatistikoa deritzo. Garrantzitsuenak:

i. Laginaren batezbestekoa:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ii. Laginaren bariantza:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

iii. Laginaren kuasibariantza:

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$



5.3 Orokortasunak

Oharra

Estatistikoak zorizko aldagaiak dira, horregatik, hurrengo atalean estatistiko hauen lagin-banaketak aztertuko ditugu.

Adibidea

- 6) Demagun Bilboko biztanleek osatzen duten populazioa dugula eta biztanle bakoitzak zenbat eskumuturreko erloju dituen aztertu nahi dugula.



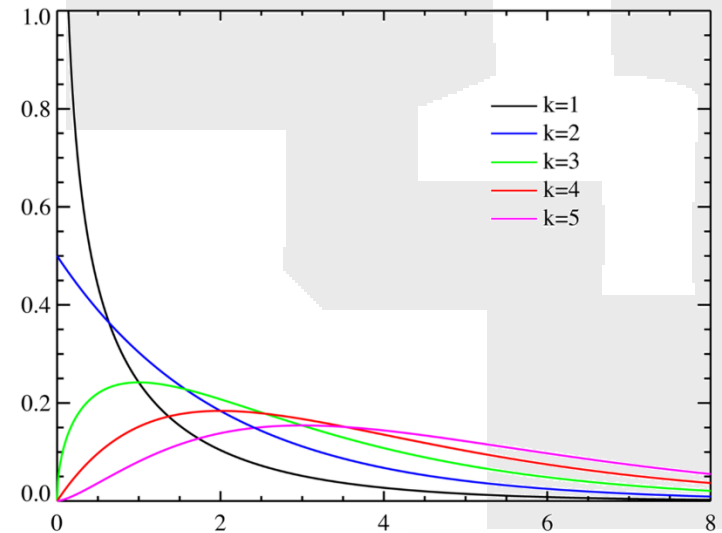
5.4 Laginketa-teoriako banaketak

5.4.1 Pearson-en χ^2 banaketa: χ_n^2

Izan bitez X_1, X_2, \dots, X_n zorizko aldagai askeak eta berdinki banatuak, euren banaketa $N(0,1)$ izanik. Ondorioz, $Y = X_1^2 + X_2^2 + \dots + X_n^2$ zorizko aldagaia n askatasun-graduak χ^2 (khi karratu) banaketa du.

$$Y \sim \chi_n^2$$

$$f(y) = \begin{cases} \frac{y^{(n/2)-1}}{2^{(n/2)} \Gamma\left(\frac{n}{2}\right)} e^{(-y/2)} & y > 0 \\ 0 & y \leq 0 \end{cases}$$



5.4 Laginketa-teoriako banaketak

5.4.1 Pearson-en χ^2 banaketa: χ_n^2

χ_n^2 banaketaren propietateak

- i. $[0, +\infty)$ tartean definiturik dago
- ii. Definizio-eremuan jarraitua da
- iii. Ez da simetrikoa
- iv. Askatasun gradua $n \geq 30$ denean, banaketa normalaren bidez hurbil daiteke:

$$Y \sim \chi_n^2 : \quad \sqrt{2Y} \cong N(\sqrt{2n-1}, 1)$$

- v. Batezbestekoa: $\mu = n$
- vi. Bariantza: $\sigma^2 = 2n$



5.4 Laginketa-teoriako banaketak

5.4.1 Pearson-en χ^2 banaketa: χ_n^2

Adibidea

7) Kalkula ezazu $P(\chi_{12}^2 \leq 20)$ probabilitatea.

Teorema

Izan bitez X, Y bi zorizko aldagai $X \sim \chi_n^2$ eta $Y \sim \chi_m^2$ izanik. Orduan, $X + Y$ zorizko aldagaiak χ_{n+m}^2 banaketa du.



5.4 Laginketa-teoriako banaketak

5.4.2 Student-en t banaketa: t_n

Izan bitez X, X_1, X_2, \dots, X_n zorizko aldagai askeak eta berdinki banatuak, euren banaketa $N(0, \sigma)$ izanik. Orduan:

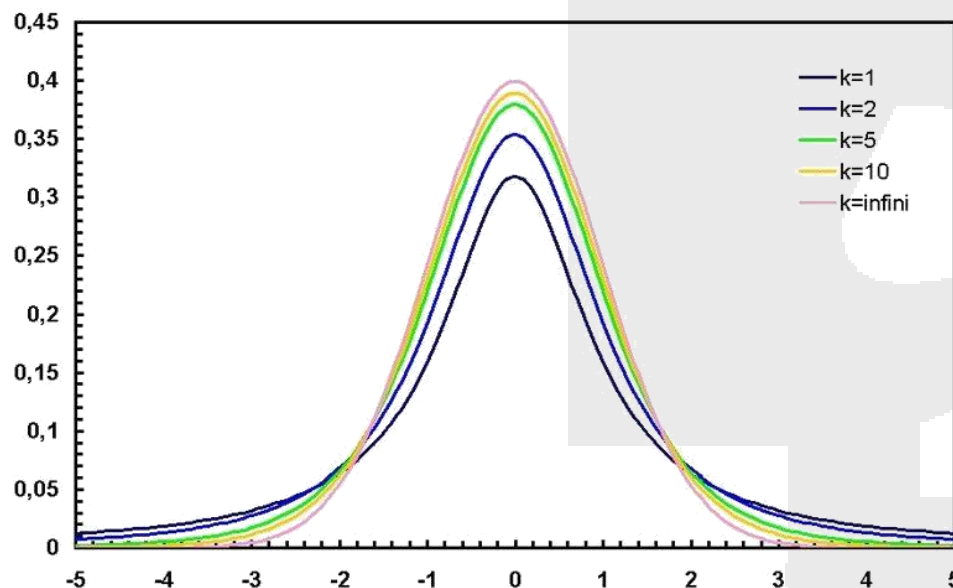
$$T = \frac{X}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}$$

T zorizko aldagaiak, n askatasun-graduak Student-en t banaketa du: $T \sim t_n$



5.4 Laginketa-teoriako banaketak

5.4.2 Student-en t banaketa: t_n



Zenbakitzailean eta izendatzailean debiderazio tipikoaz zatituz:

$$T = \frac{\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}}} = \frac{Z}{\sqrt{\chi_n^2 / n}}, \quad Z \sim N(0,1) \text{ izanik}$$



5.4 Laginketa-teoriako banaketak

5.4.2 Student-en t banaketa: t_n

Student-en t banaketaren propietateak

- i. $(-\infty, +\infty)$ tartean definiturik dago
- ii. Definizio-eremuan jarraitua da
- iii. Batezbestekoarekiko simetrikoa da.
- iv. Kanpai itxura du eta $n \geq 30$ denean, banaketa normal tipifikatuaren bidez hurbil daiteke:

$$n \geq 30 : t_n \approx N(0,1)$$

- v. Batezbestekoa: $\mu = 0$

- vi. Bariantza: $\sigma^2 = \frac{n}{n-2} (n > 2)$



5.4 Laginketa-teoriako banaketak

5.4.3 Snedecor-en F banaketa: $F_{n_1; n_2}$

Izan bitez X eta Y zorizko aldagai askeak $\chi_{n_1}^2$ eta $\chi_{n_2}^2$ banaketa dutenak, hurrenez hurren.

$$X \sim \chi_{n_1}^2 \text{ eta } Y \sim \chi_{n_2}^2 \quad \text{Orduan,} \quad F = \frac{X/n_1}{Y/n_2}$$

Zorizko aldagaiak, n_1 eta n_2 askatasun graduak Snedecor-en F banaketa du:

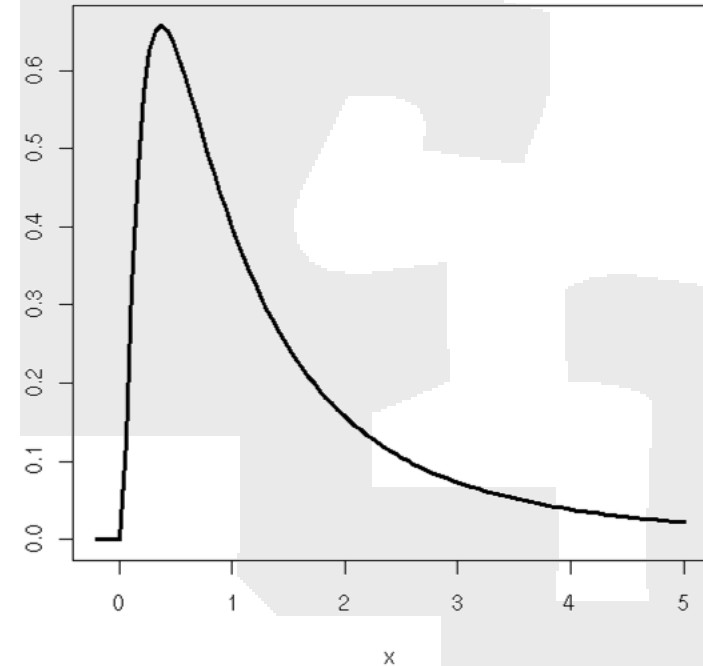
$$F = \frac{X/n_1}{Y/n_2} \sim F_{n_1, n_2}$$



5.4 Laginketa-teoriako banaketak

5.4.3 Snedecor-en F banaketa: $F_{n_1; n_2}$

- i. $[0, +\infty)$ tartean definiturik dago
- ii. Definizio-eremuan jarraitua da
- iii. Ez da simetrikoa.
- iv. Batezbestekoa: $\mu = \frac{n_2}{n_2 - 2} \quad (n_2 > 2)$
- v. Bariantza: $\sigma^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$
- vi. $F_{1-\alpha; n_1, n_2} = \frac{1}{F_{\alpha; n_2, n_1}}$



Adibidea

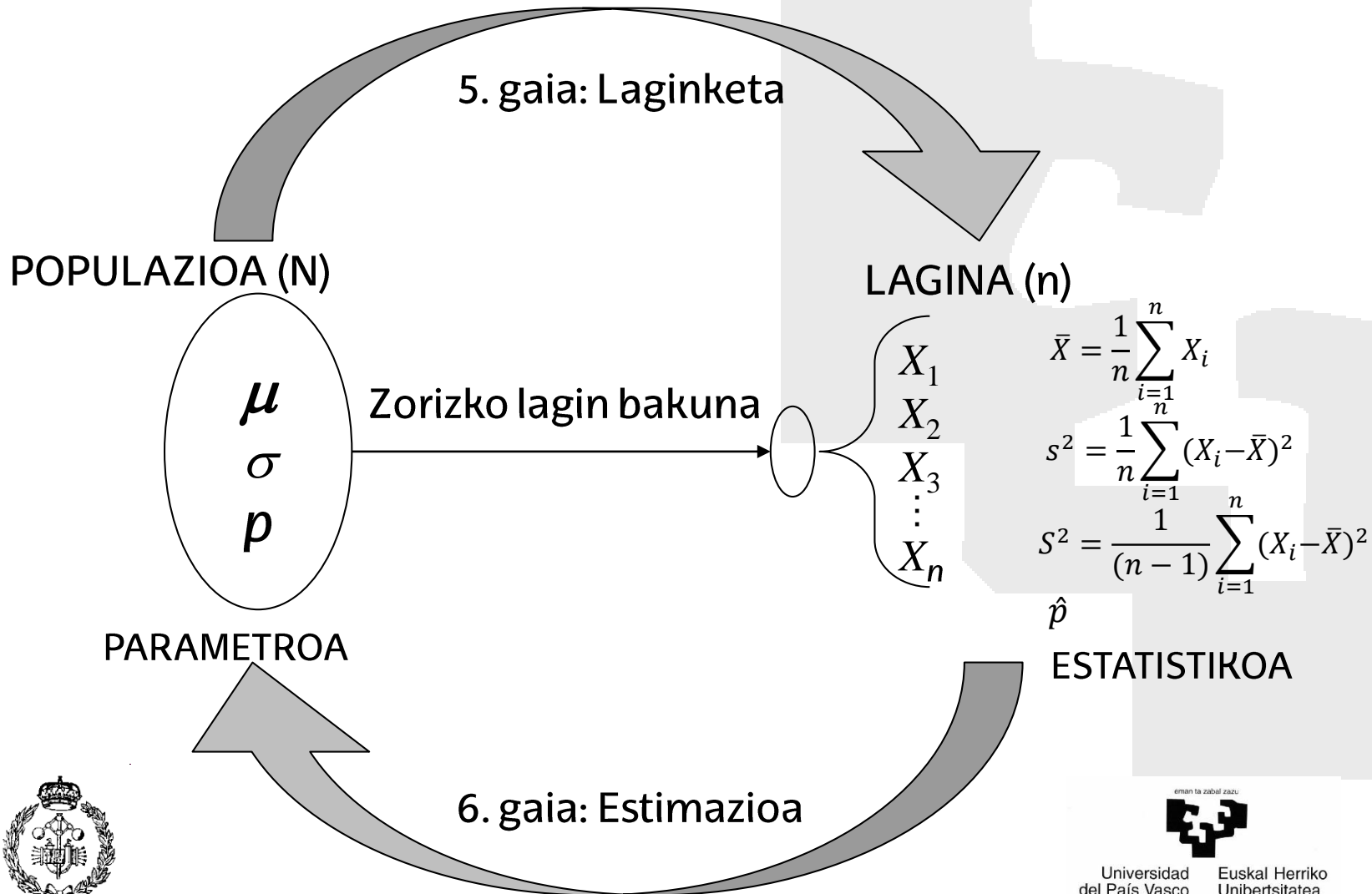


8) Kalkula ezazu $F_{0.975; 8, 12} \left(F_{8, 12; 0.975} \right)$



5.5 Estatistikoaren banaketak

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako banaketak
Estatistikoaren banaketak



5.5 Estatistikoaren banaketak

Populazioa eta lagina

Zorizko laginketa

Orokortasunak

Laginketa-teoriako banaketak

Estatistikoaren banaketak

Estatistikoa	Populazioa	Lagina	Estatistikoaren banaketa
\bar{X}	Normala σ ezaguna		$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ edo $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$
\bar{X}	Normala σ ezezaguna		$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$
\bar{X}	Edozein σ ezaguna	$n > 30$	$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ edo $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$
\bar{X}	Edozein σ ezezaguna	$n > 100$	$\bar{X} \sim N\left(\mu, \frac{S}{\sqrt{n}}\right)$ edo $Z = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim N(0,1)$



5.5 Estatistikoaren banaketak

Estatistikoa	Populazioa	Lagina	Estatistikoaren banaketa
$\bar{X}_1 - \bar{X}_2$	Normalak independenteak σ_1 , σ_2 ezagunak		$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$
$\bar{X}_1 - \bar{X}_2$	Normalak independenteak σ_1 , σ_2 ezezagunak $\sigma_1 = \sigma_2$		$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
$\bar{X}_1 - \bar{X}_2$	Normalak independenteak σ_1 , σ_2 ezezagunak $\sigma_1 \neq \sigma_2$		$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim t_v$ <div> non <div> $v = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n+1} + \frac{\left(\frac{S_2^2}{m}\right)^2}{m+1}} - 2$ </div> <div>OHARRA: Formula honetan kuasibariantza dira S guztiak</div> </div>
$\bar{X}_1 - \bar{X}_2$	Edozein independenteak σ_1 , σ_2 ezagunak	$n > 15$ $m > 15$	$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$
$\bar{X}_1 - \bar{X}_2$	Normalak independenteak σ_1 , σ_2 ezezagunak	$n > 100$ $m > 100$	$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}\right)$

5.5 Estatistikoaren banaketak

Populazioa eta lagina

Zorizko laginketa

Orokortasunak

Laginketa-teoriako banaketak

Estatistikoaren banaketak

Estatistikoa	Populazioa	Estatistikoaren banaketa
s^2	Normala μ ezaguna	$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \sim \chi_n^2$
s^2	Normala μ ezezaguna	$\frac{ns^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

Estatistikoa	Populazioa	Estatistikoaren banaketa
$\frac{S_1^2}{S_2^2}$	Normalak	$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{n-1, m-1}$

Estatistikoa	Lagina	Estatistikoaren banaketa
\hat{p}	$n > 100$	$\hat{p} \approx N\left(p, \sqrt{\frac{pq}{n}}\right)$

Estatistikoa	Lagina	Estatistikoaren banaketa
$\hat{p}_1 - \hat{p}_2$	$n > 100$ $m > 100$	$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}\right)$

5.5 Estatistikoan banaketak

Adibidea

- 9) Altzairuzko xaflen gogortasuna neurtzen duen koefizienteak $N(1.6, 0.3)$ banaketa du. Zoriz, altzairuzko bost xafla hartu dira.
- a) Kalkula bedi zoriz hartutako altzairuzko bost xaflen batezbesteko gogortasuna gehienez 1.5 izateko probabilitatea.
 - b) Demagun orain σ^2 ezezaguna dela. Zoriz aukeratutako altzairuzko bost xaflen gogortasuna neurtzen duen koefizienteen bariantza 0.2 dela jakinik, lor bedi laginaren batezbestekoa gehienez 2 izateko probabilitatea.



5.5 Estatistikoan banaketak

Adibidea

- 10) Hiri bateko airearen kalitatea neurtzeko bi estaziotan sufre dioxidoaren hileko batezbestekoak aztergai dira. Zoriz A estaziotik 10 hilabetetako batezbesteko sufre dioxido mailak eta B estaziotik 12 hilabetetako batezbesteko sufre dioxido mailak hartu dira. A estazioko 10 datuen desbiderazio tipikoa 2 eta B estazioko 12 datuen desbiderazio tipikoa 1 dira. Sufre dioxidoaren mailak A estazioan $N(15, \sigma_A)$ banaketa jarraitzen du eta B estazioan $N(13, \sigma_B)$ banaketa jarraitzen du, bariantza ezezagunak eta ezberdinak direla onartuz.

Demagun bi aldagaiak elkarrekiko independenteak direla. Kalkula bedi A estazioko 10 datuz eta B estazioko 12 datuz osaturiko laginak hartzerakoan, A estazioko hileko batezbesteko sufre dioxido mailak B estazioko hileko batezbesteko sufre dioxido maila unitate batean baino gehiagoan gainditzeko probabilitatea.



5.5 Estatistikoen banaketak

Adibidea

- 11) LV agentziak egindako azken azterketen arabera, gaur egun populazioaren %90ek sakeleko telefonoa erabiltzen du. Zoriz 200 pertsonako zorizko lagin bakuna hartu da.
- a) Kalkula bedi 200 pertsona horietatik sakelako telefonoa erabiltzen dutenen proportzioa gutxienez 0.85 izateko probabilitatea.
 - b) Zein da sakelako telefonoen erabiltzaileen benetako proportzioaren eta lagineko erabiltzaileen proportzioaren arteko diferentzia gehienez %2 izateko probabilitatea?



5.5 Estatistikoen banaketak

Adibidea

12) Luxuzko hotel batek hilero jasotzen duen bezero-kopuruak banaketa normalari jarraitzen dio, hilero bezero-kopuruaren desbiderazio tipikoa 30 bezerokoa izanik. Zoriz, bost hilabetetan jasotako bezero kopuruen datuak jaso dira.

- a) Zein da bost hilabetetako lagin batean luxuzko hotelak hilero jasotako bezero-kopuruaren kuasibariantza gutxienez $64 (\text{bezero})^2$ -koa izateko probabilitatea?
- b) Zein da 5 hilabeteko laginaren kuasibariantza populazioaren bariantzaren gutxienez bikoitza izateko probabilitatea?

