ESTATITIKA METODOAK INGENIARITZAN

5. Laginketara sarrera

Universidad del País Vasco Euskal Herriko Unibertsitatea

5. Laginketara sarrera

Universidad del País Vasco Euskal Herriko Unibertsitatea

5. Laginketara sarrera

5.5 Estatistikoak

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.

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. Indibiduoen 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).

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

5.2 Zorizko laginketa

5.2.2 Zorizko laginketa geruzatua

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

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

Adibidea

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

Universidad del País Vasco Euskal Herriko Unibertsitatea

5.2 Zorizko laginketa

5.2.3 Zorizko laginketa konglomeratua

Aurreko metodoetan populazioko elementuak aukeratzen genituen, orain berriz populazioko elementuek sortzen dituzten talde batzuk aukeratuko 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.

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

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,..., X_n denotatuko ditugunak, zorizko aldagai askeak eta berdinki banatuak dira. Hauen banaketa eta populazioaren banaketa berdina izanik.
4. Hautaketa itzulerarik gabe 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.

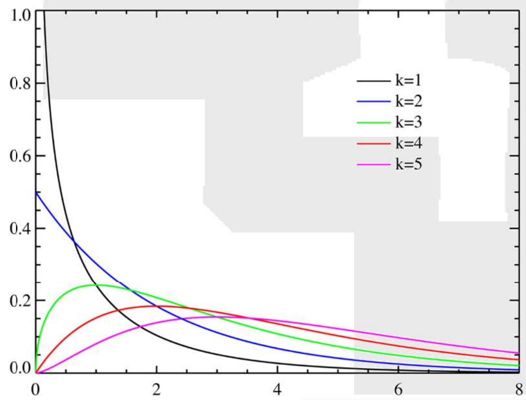
5.4 Laginketa-teoriako banaketak

5.4.1 Pearson-en χ^2 banaketa: χ_n^2

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

$$Y \sim \chi_n^2$$

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



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≥30 denean, banaketa normalaren bidez hurbil daiteke:

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

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

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

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. (Taula)

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

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

5.4.2 Student-en t banaketa: t_n

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

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

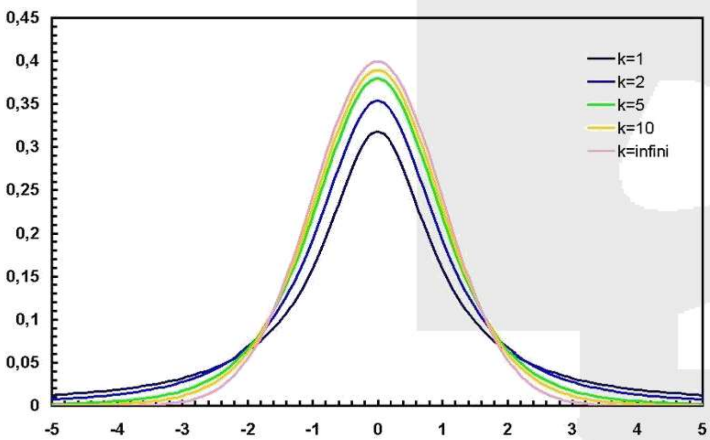
t zorizko aldagaiak, n askatasun-graduko **Student-en t banaketa** du: $t \sim t_n$

Universidad del País Vasco Euskal Herriko Unibertsitatea

5.4 Laginketa-teoriako banaketak

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

5.4.2 Student-en t banaketa: t_n



Zenbakitzailean eta izendatzailean debiderazio tipikoaz zatituz:

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

Universidad del País Vasco Euskal Herriko Unibertsitatea

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≥30 denean, banaketa normal tipifikatuaren bidez hurbil daiteke:

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

v. Batezbestekoa: $\mu = 0$

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

Universidad del País Vasco Euskal Herriko Unibertsitatea

5.4 Laginketa-teoriako banaketak

5.4.3 Snedecor-en F banaketa

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

$X \sim \chi^2_{n_1}$ eta $Y \sim \chi^2_{n_2}$ Orduan, $F = \dfrac{X/n_1}{Y/n_2}$

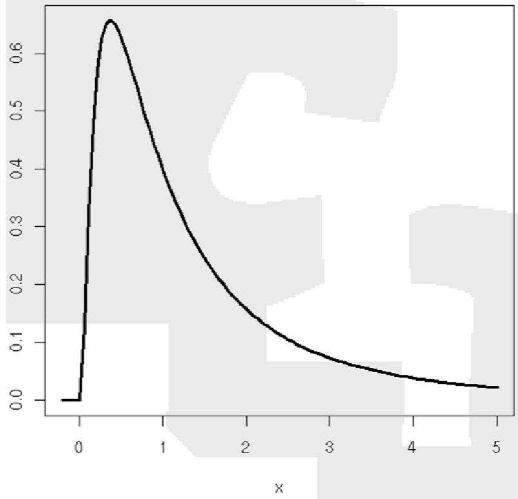
Zorizko aldagaiak, n_1 eta n_2 askatasun graduko Snedecor-en F banaketa du:

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

Universidad del País Vasco Euskal Herriko Unibertsitatea

5.4 Laginketa-teoriako banaketak

5.4.3 Snedecor-en F banaketa



i. $[0,+\infty)$ tartean definiturik dago

ii. Definizio-eremuan jarraitua da

iii. Ez da simetrikoa.

iv. Batezbestekoa: $\mu = \dfrac{n_2}{n_2 - 2} \quad (n_2 > 2)$

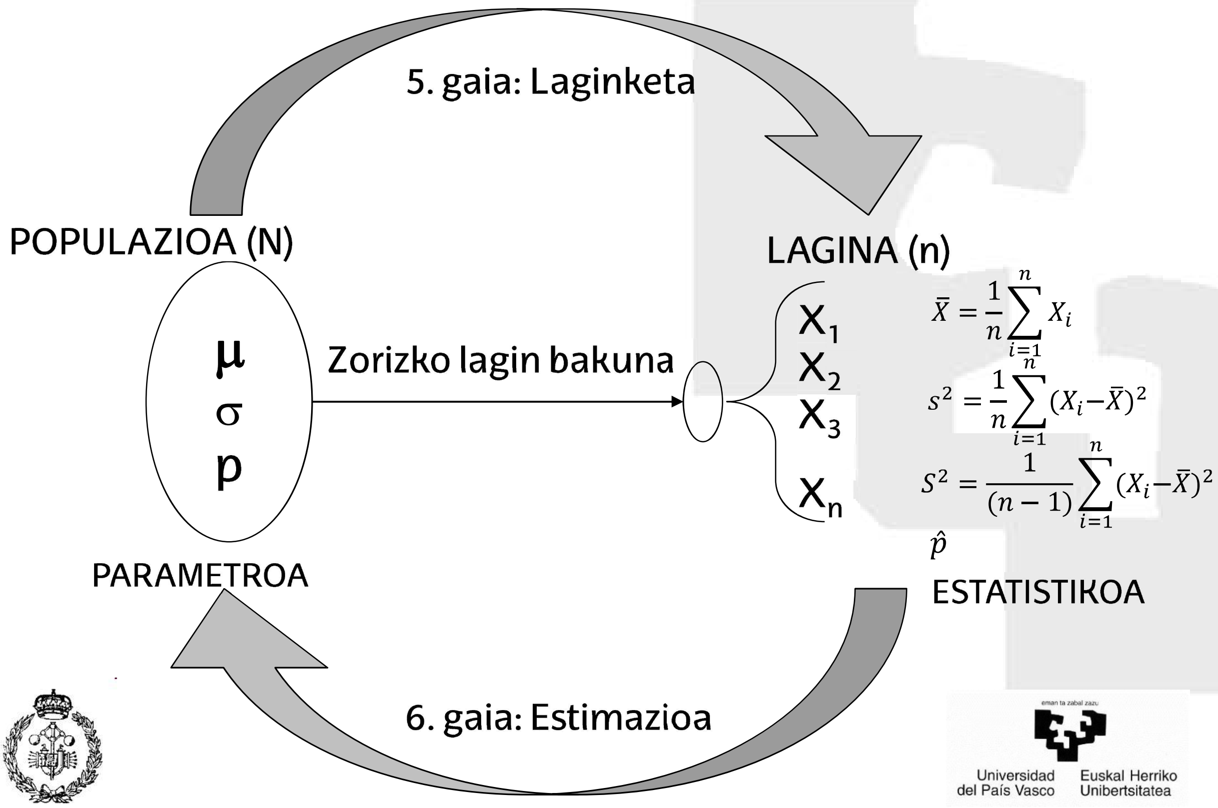
v. Bariantza: $\sigma^2 = \dfrac{2n_2^{\,2}(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$

vi. $F_{1-\alpha;\, n_1, n_2} = \dfrac{1}{F_{\alpha;\, n_2, n_1}}$

Adibidea

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

5.5 Estatistikoak



5.5 Estatistikoak

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 ezaguna	$n > 100$	$\bar{X} \sim N\left(\mu, \frac{S}{\sqrt{n}}\right)$ *edo* $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$

5.5 Estatistikoak

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 $\sigma_1 \neq \sigma_2$		$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \sim t_\nu$ non $\nu = \frac{\left(\frac{S_1^2}{n}+\frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n+1}+\frac{\left(\frac{S_2^2}{m}\right)^2}{m+1}} - 2$ OHARRA: Formula honetan kuasibariantza dira S guztiak
$\bar{X}_1 - \bar{X}_2$	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 Estatistikoak

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	$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \sqrt{\frac{p_1q_1}{n} + \frac{p_2q_2}{m}}\right)$

5.5 Estatistikoak

Populazioa eta lagina
Zorizko laginketa
Orokortasunak
Laginketa-teoriako zenbait banaketa
Estatistikoak

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, lore bedi laginaren batezbestekoa gehienez 2 izateko probabilitatea.