BIOESTADISTIKA

1. Estatistika deskribatzailea

- Esperimentuan datuen:
  - bilketa
  - azterketa
  - antolakuntza
  - laburpena

→ Kontzeptuak

Populazioa (azterketaren helburua)

Alea / Unitate est. (indibiduoa)

Lagina + laginaren tamaina (n)

Aldagai estatistikoa

- Diskretua: aldagai kopuru finitua edo infinitu zenbakigarria
- Jarraitua: segmentu baten balio oro

→ Taula estatistikoa

- Aldagaia eta bere balioak txikitik handira.

  Errenkada kopurua = balio ezberdin beste (k)
- Zutabeak
  - Maiztasun absolutua f_i
  - Maiztasun metatua $F_i = f_1 + f_2 \dots + f_i$
  - Maiztasun erlatiboa $h_i = f_i / n$
  - Maiztasun metatu erl. $H_i = h_1 + h_2 \dots h_i$

→ Adierazpen grafikoak

- Izaera kualitatiboa → Barra diagrama ← maiztasuna (altura), ← modalitatea (sexua)

  Sektore diagrama

- Izaera kuantitatiboa → Barra diagrama

  Histograma

  Zurtoin + hosto grafikoa

ESTADISTIKOAK

BATAZBESTEKOA $\bar{x}$	MEDIANA Me	MODA Mo	Pertzent. kuartill.	$R + IR$	Bariantza S_n^2	Desbi. est S_n	Kuasibariantza $S_{n-1}^2 = S^2$	Aldak. koef Ak
$\bar{x} = \frac{\sum_i x_i}{n}$	%50 aldagai $n/2$	Maiztasun handiena	$n/100$ $n/4$ (Q_i)	$X_{max} - X_{min}$ $Q_3 - Q_1$	$\frac{1}{n}\sum_{i=1}^{n} f_i x_i^2 - \bar{x}^2$	$\sqrt{S_n^2}$	$S_n^2 \frac{n}{n-1}$	$\frac{S_n}{\bar{x}} \cdot 100$

Erdigunearen neurriak — Posizio neurri — Sakabanatze neurriak

\+ Alborapena

$$\nabla = \frac{\bar{x} - Mo}{S_n}$$

→ Datu elkartuak

- Datu kopuru handiegia edo datu zehatzak ezagutu ez.

- Metodoa
  - Tarte kopurua definitu (Sturges $k = 1 + 3'322 \log n$)
  - Datu bakoitza talde baten
  - Tarteak berdinak izatea komenigarri
  - Klasearen marka, erdiko puntua x_i
    Tartearen zabalera a_i

→ Estadistikoak
  x_i-ren ordez klase marka (x_i dela beste)
  Mediana, pertzentilak eta moda ezberdin

→ Bi aldagairen distribuzio batera tua

- Diskretua + jarraitua Jarraituaren estatistika deskribatzaileak faktorearen modalitateentzat
- Diskretu + diskretu (2 faktore) kontingentzia taulak, aldagai baten balioak zut. bestearenak er.
- Jarraitua + Jarraitua Erregresioa eta korrelazioa.
  - Puntu hodeia
  - korrelazio koefizientea
  - Erregresio-zuzena

2. Probabilitatea

- Lagin batetik ateratako ondorioak populazio osora hedatzeko tresna matematikoa.

→ Kontzeptuak:

- Zorizko esperimentua: egin aurretik emaitza ezezagun duen esperimentua.
- Lagin espazioa (Ω): Zorizko esperimentuaren emaitza posible guztien multzoa
- Ezinezko egoera ($\emptyset$) + Gertaera segurua (Ω)
- Oinarrizko gertaerak: Elementu bakar batek osatutako gertaerak
- Gertaera konposatuak: + 1'5 elementuz osatuta
- Ω-ren parteak ($\wp(\Omega)$): Zorizko esperimentu bati lotutako gertaera guztiek osatutako multzoa

→ Gertaeren arteko erlazioak

- Bildura $A \cup B$ bietako 1 gutxienez betetzea
- Ebakidura $A \cap B$ A eta B gertaerak batera gertatzen dira
- A parte B $A \subseteq B$ A jazotzen bada B ere bai
- Kontrako gertaera $\overline{A}$ A-rea jazotzea
- Gertaeren diferentzia $A - B$ A jazo, B ez $= A \cap \overline{B}$
- Bateraezintasuna $A \cap B = \emptyset$

→ De Morganen legeak

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

→ Kalkulua eta definizioa

- Zorizko esperimentu bati lotutako gertaeren ziurtasun/gertagarritasuna adierazten duen funtzioa.

$$-P: \begin{array}{l} \wp(\Omega) \longrightarrow [0,1] \\ A \longrightarrow P(A) \end{array} \left\{ \begin{array}{l} P(\Omega) = 1 \\ A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B) \end{array} \right.$$

→ Probabilitatearen propietateak

1º Gertaera osagarriaren propietatea $P(\overline{A}) = 1 - P(A)$

2º Partekotasun eta diferentziaren probabilitatea $A \subseteq B \rightarrow P(A) \leq P(B)$

$$P(B-A) = P(B) - P(A)$$

3º Bilduraren probabilitatea $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4º Bilduraren probabilitatearako goi muga $P(A \cup B) \leq P(A) + P(B)$

Eredu probabilistikoak

1- Zorizko eredu klasikoa

- Lagin espazio finitua $\Omega = \{e_1, e_2, e_3 \dots e_n\}$
- Gertaera orok probabilitate bera $P(\{e_1\}) = P(\{e_2\}) = \dots = 1/n$
- Gertaera sinpleak bateraezinak
- Gertaera konposatuen probabilitatea $P(A) = \dfrac{\text{Aren aldeko emaitza-kopurua}}{\text{emaitza posible-kopurua}}$

2- Zorizko eredu finitua

- Lagin espazio finitua
- Gertaera orok probabilitate bera ez. $P(\{e_i\}) = p_i \quad 1 \leq i \leq n$
- Baldintzak: $0 \leq p_i \leq 1 \;,\; 1 \leq i \leq n$

$$p_1 + p_2 \dots + p_n = 1$$

- Gertaera sinpleak bateraezin
- Gertaera konposatuen probabilitatea $P(A) = P(\{e_{i_1}\}) + P(\{e_{i_n}\}) = p_{i_1} + \dots + p_{i_n}$

Konbinatoria

- Gertaeren emaitza kopurua zenbatzeko teknika
- Multzo bateko azpi-multzo kopurua zenbatzeko
- n faktoriala $(n \in \mathbb{N})$: $n! = n(n-1)\cdot(n-2)\dots 1$
- n gain k konbinazio zenbakia $(k \leq n)$: $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$
- n elementuen permutazioak:
  n elementuz osatutako multzoko elementuen ordenazio posibleak $P_n = n!$

Konbinatoria: definizioak

- $k, n \in \mathbb{N}$ $\;k \leq n$ (errepikapen gabe)
- k-naka hartutako n elementuen konbinazioak $C_n^k = \binom{n}{k}$ (taldeak desberdintzat)
  → aldakuntzak $V_n^k = n(n-1)(n-2)\dots(n-(k-1))$
- Erlazioa $C_n^k \cdot P_k = V_n^k$

- $k, n \in \mathbb{N}$ (errepikapena baimenduz)
- k-naka hartutako n-elementuen errepikatutako konbinazioak $CR_n^k \binom{n+k-1}{k}$

Permutazioak ← → aldakuntzak $VR_n^k = n^k$

i elementua ki aldiz $(k_1 + \dots k_r = n)$

$$P_n^{k_1, k_2 \dots k_r} = \frac{n!}{k_1! \dots k_r!}$$

⊙→ Baldintzazko probabilitatea

- Intuizioa

Gertaera bati buruzko informazioren batek P alda dezake

ADB.

Sexua	Horia	Beltza	Ilea
Arra	10	40	
Emea	40	10	

orokorrean $P(H) = \frac{50}{100} = 0'5$

arren artean $P(H) = \frac{10}{50} = 0'2$

emeen artean $P(H) = \frac{40}{50} = 0'8$

- Definizioa

$P(A) > 0 \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$ (B gertaeraren Arekiko probabilitatea)

• A gertaera finkoa izanik

• $P(\overline{B}|A) = 1 - P(B|A)$ gertaera osagarriaren prob.

• Barnekotasun eta diferentziaren probabilitatea

$B \subseteq C \rightarrow P(B|A) \leq P(C|A)$

$P(C-B|A) = P(C|A) - P(B|A)$

• Bilduraren probabilitatea

$P(B \cup C|A) = P(B|A) + P(C|A) - P(B \cap C|A)$

• Bilduraren probabilitatearen goi muga

$P(B \cup C|A) \leq P(B|A) + P(C|A)$

- Independentzia

• A eta B gertaerak $P(A) > 0$, $P(B) > 0$; A eta B askeak

$P(A \cap B) = P(A) \cdot P(B)$ / $P(A|B) = P(A)$ / $P(B|A) = P(B)$

↳ Probabilitate konposatuaren teorema

$P(A_1 \cap A_2 \ldots \cap A_n) > 0 \rightarrow P(A_1 \cap A_2 \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \ldots \cdot P(A_n|A_1 \cap A_2 \ldots A_{n-1})$

(biderkadura erregela)

- Probabilitate osoaren teorema

• Gertaera sistema osoa: $A_1, A_2 \ldots A_n$ gertaera familia

$P(A_i \cap A_j) = \emptyset \quad \bigcup_{i=1}^{n} A_i = \Omega$

$- P(A_i) > 0$

+

Edozein gertaera → $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$

- Bayes-en teorema

• $A_1, A_2 \ldots A_n$ gertaera familia osoa $P(A_i) > 0$

B edozein gertaera

$$P(A_1|B) = \frac{P(B|A)\,P(A_1)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Zorizko aldagaiak

- Zorizko esperimentua Bihotzeko gaixotasun baten garapena
- Esperimentuari loturiko ezaugarriak Adina, pisua, egunero zigarro kop, gaixotasun aurrek. fam. ...
- Zorizko aldagaia (z.a.) Zorizko esperimentu bati lotutako ezaugarri balioitza

$$X : \Omega \rightarrow R$$

→ Adierazpena

Aldagaia letra larriz (X)

Aldagaiaren balioak letra xehez (x)

- X zorizko aldagai diskretua
  - Balio kopurua finitua edo infinitu zenbakigarria (zki. arruntak)
- X zorizko aldagai jarraitua
  - Segmentu bateko balio guztiak (zki erreal positibo guztiak)

ADB

- Bizitzako lehen 2 urteetan haur batek otitisa duen aldi kopurua (dis.)
- Zortzi seme-alabadun familietako seme kop. (dis)
- Ebakuntza batean bezeroaren galdutako odol kop. (litrotan) (jarr.)
- Istripu bat kausatutakoan lehen anbulantzia heltzeko denbora (jarr.)

→ Zorizko aldagai diskretua

- Probabilitate legea
  - Zorizko aldagai diskretuaren balio bakoitzean gertatze-mailaren neurria ematen duen funtzioa

    $P(X=x)=f(x)$

    - X aldagaiaren heina $\{x_1, x_2 \dots x_n\}$ denean

      $f(x_1)=P(X=x_1), \dots, f(x_n)=P(X=x_n)$

      → X-en beste edozein baliotarako $f(x)=P(X\neq x)=0$
  - $f(x)$ probabilitate-legearen propietateak

    $f(x)\geq 0$ edozein x entzako

    $\sum_x f(x) = \sum_x P(X=x) = 1$

· Banaketa funtzioa

- Xen banaketa funtzioa ($F(x)$): R-n definitutako funtzioa.

Zorizko aldagaiak x en berdin edo gutxiago hartzeko prob.

$$F(x)=P(X\leq x)$$

Edozein x_0 balio finko batentzat

$$F(x_0)=P(X\leq x_0)=\sum_{X\leq x_0} f(x)$$

· $F(x)$ banaketa funtzioaren propietateak

1. $F(-\infty)=0$, $F(\infty)=1$
2. Ez-beherakorra: $a\leq b \Rightarrow F(a)\leq F(b)$
3. Eskuinetik jarraitua: $\lim_{b\downarrow a} F(b)=F(a)$
4. $F(x)$ek

   $\{x_1, x_2 \dots, x_n\}$ puntuetan etenak $f(x_1), f(x_2) \dots f(x_n)$-koak

   Grafikoa eskailera

5. $P(a<X\leq b)=P(X\leq b)-P(X\leq a)=F(b)-F(a)$

ADB

→ Esperimentua: Txanpon 2 aldiz bota

$\Omega=\{AA, AG, GA, GG\}$ (ordena kontuan)

X zorizko aldagaia X ≡ aldagai kopurua

$X: \Omega \to R$

$AA \to 2$

$AG \to 1$

$GA \to 1$

$GG \to 0$

· Probabilitate legea $f(x)=\begin{cases} 1/2 & x=1 \\ 1/4 & x\in\{0,2\} \\ 0 & \text{beste kasuetan} \end{cases}$

· Banaketa funtzioa $F(x)=P(X\leq x)=\begin{cases} 0 & x<0 \\ 1/4 & 0\leq x<1 \\ 3/4 & 1\leq x<2 \\ 1 & x\geq 2 \end{cases}$

→ Esperimentua Beren bizitzako lehen 2 urteetan otitisa

X zorizko aldagaia X ≡ 2 urteetan haur batek otitisa pasatzen duen aldi kopurua.

$\Omega=\{0,1,2,3,4\dots\}$

· Probabilitate legea

X	0	1	2	3	4	5	6
P(X=x)	0'129	0'264	0'271	0'185	0'095	0'039	0'017

· Banaketa funtz?

Zorizko aldagai jarraituak

- X aldagai jarraituaren edozein balioetarako

  $P(X=x)=0$

→ X aldagaiaren $F(x)$ banaketa funtzioa

$\mathbb{R}$-n definitutako funtzioa, aldagaia $\leq x$ izateko prob.

$$F(x)=P(X\leq x)$$

- $F(x)$en propietateak
  1. $F(-\infty)=0,\ F(\infty)=1$
  2. Ez beherakorra: $a\leq b,\ F(a)\leq F(b)$
  3. Jarraitua $\lim_{x\downarrow a} F(x)=F(a)$
  4. $P(X=x)=0$ denez

     $P(a<X<b)=P(a\leq X\leq b)=P(a<X\leq b)=P(a\leq X<b)=F(b)-F(a)$

     $P(X\leq a)=P(X<a)=F(a)$

→ Dentsitate funtzioa ($f(x)$)

- Xen dents. fun. x erreal guztientzat x-eko banaketa funtzioaren deribatua

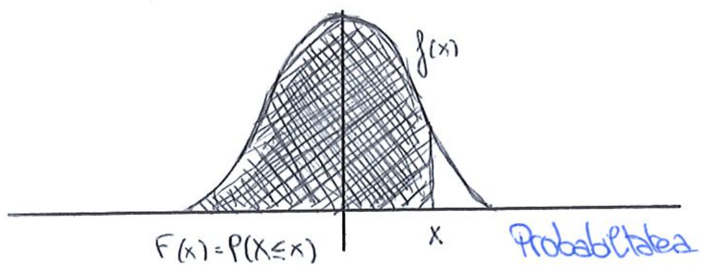
$$f(x)=\frac{d}{dx}F(x)=\lim_{x\to 0}\frac{F(x+h)-F(x)}{h}$$

- $f(x)$en propietateak
  1. $f(x)\geq 0\ \forall x$
  2. $\int_{-\infty}^{\infty} f(x)\,dx=1$

→ $F(x)$ ezagutuz $F'(x)=f(x)$

$f(x)$ ezagutuz $F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)\,dt$

- Banaketa eta dents. funtzioen erlazioa



Probabilitatea = azalera

ADB(banaketa jarraitua)

X ≡ Metro 1eko makila 2tan zatitzean, zati luzeenaren luzera

• Dentsitate-funtzioa $f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{gainerakoetan} \end{cases}$ (0,1) tartean banaketa uniformea

• Banaketa-funtzioa $F(x) = \int_{-\infty}^{x} f(t)dt$

$$F(x) = \begin{cases} 0 & x \le 0 \\ x & x \in (0,1) \\ 1 & x \ge 1 \end{cases}$$

- Zein da zati luzeenaren luzera 80cm baino gehiago izateko prob.?

$P(X > 0'8) = 1 - P(X \le 0'8) = 1 - F(0'8) = 0'2$

- Zein da zati laburrenaren luzera 10-20 cm tartean egoteko prob.?

$P(0'8 \le X \le 0'9) = P(0'8 < X < 0'9) = F(0'9) - F(0'8) = 0'9 - 0'8 = 0'1$

! (a,b) tartean banaketa UNIFORMEA

$X : U(a,b)$

• Dentsitate funtzioa $f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{gainerakoan} \end{cases}$

• Banaketa-funtzioa $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$

ESPONENTZIALA

$X : E(\lambda)$

$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \ge 0 \end{cases}$

$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$

X ≡ Bakterioen biziraldia (orduetan)

• Dentsitate funtzioa $f(x) = \begin{cases} 0 & x < 0 \\ 3e^{-3x} & x \ge 0 \end{cases}$

3h parametroko banaketa esponentziala

• Banaketa funtzioa $F(x) = \int_{-\infty}^{x} f(t)dt$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-3x} & x \ge 0 \end{cases}$$

- Zein da bakterioa 2h baino lehenago hiltzeko prob.? $P(X < 2) = P(X \le 2) = F(2) = 1 - e^{-6} = 0'998$

- Zein da bakterioaren biziraldia 2'7-3h izateko prob.? $P(2'7 < X < 3) = F(3) - F(2'7) = 1'8 \cdot 10^{-4}$

- Zorizko aldagaiaren itxaropena, bariantza eta desbi. est.

- Itxaropen matematikoa (μ)

  X zorizko aldagaia

  → X diskretua $\mu \equiv E(x) = \sum_x x \cdot f(x)$

  x = aldagaiak hartzen dituen balioak

  → X jarraitua $\mu \equiv E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

- $E(x) \equiv$ oreka puntua $\equiv$ probabilitatea - masa kontzentrazioa
- Itxaropen matematikoaren propietateak

  1 $E(X+k) = E(X) + k$

  2 $E(aX) = aE(X)$

  3 $E(X+Y) = E(X) + E(Y)$

  4 X eta Y aldagai askeak badira $E(XY) = E(X)E(Y)$

  - X eta Y diskretuak askeak izateko

    $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$

    X eta Y jarraituak askeak izateko

    $P(\{X \leq a\} \cap \{Y \leq b\}) = P(X \leq a) \cdot P(Y \leq b)$ $\forall a, b \in \mathbb{R}$ bada

  5. $E(k) = k$

- Bariantza (σ^2) eta desbideratze estandarra (σ)

  X zorizko aldagaia $\mu = E(X)$

  $\sigma^2 = E[(X-\mu)^2] = E(X^2) - \mu^2$ $\quad Var(X) = \sigma^2$

  $\sigma = \sqrt{\sigma^2}$

  - Bariantzaren prop.

    1. Beti positiboa edo nulua $\sigma^2 \geq 0$
    2. $Var(X) = 0$ X zorizko aldagaia konstantea bada
    3. $Var(aX) = a^2 Var(X)$
    4. $Var(X+a) = Var(X)$
    5. X eta Y askeak badira

       $Var(X+Y) = Var(X) + Var(Y) = Var(X-Y)$

ADB.

- X ≡ aurpegi kopurua (txanp) diskretua

$\mu = E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) = 1 \cdot 1/2 + 2 \cdot 1/4 = 1$

$E(X^2) = 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) + 2^2 \cdot P(X=2) = 1/2 \cdot 1 + 1/4 \cdot 2^2 = 1'5$

$\sigma^2 = Var(X) = 1'5 - 1^2 = 0'5$

$\sigma = \sqrt{0'5} = 0'707$

- X ≡ bakterioen biziraldia (ordutan) jarraitua

$\mu \equiv E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{0}^{\infty} x \cdot 3e^{-3x} dx = 1/3$

$E(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_{0}^{\infty} x^2 \cdot 3e^{-3x} dx = 2/9$

$\sigma^2 = 2/9 - 1/9 = 1/9$

$\sigma = \sqrt{1/9} = 1/3$

Zenbait banaketa diskretu

- Banaketa binomiala
  - Esperimentua
    - Zorizko esperimentuaren n froga

      Frogen baldintzak:
      - Proba bakoitzean 2 emaitza ezberdin izan daitezke: Arrakasta (A) eta porrota ($\bar{A}$)
      - $p(A) = p$ eta $P(\bar{A}) = 1 - p = q$
      - proba independenteak
    - X ≡ Arrakasta kopurua (Zorizko aldagaia)
    - Lagin espazioa $\Omega = \{0, 1, 2 \ldots n\}$

      $X: Bin(n, p)$
    - Parametroak:

      n proba-kopurua

      p arrakastaren probabilitatea
    - Taulak: 1 taula

- Probabilitate legea

$P(X=r) = \frac{n!}{r!(n-r)!} p^r q^{n-r} = \binom{n}{r} p^r q^{n-3}$

ADB

$X: Bin(10, 0'3)$

$P(X=3) = \binom{10}{3} 0'3^3 \, 0'7^{10-3} = \frac{10!}{3!(10-3)!} 0'3^3 (1-0'3)^{(10-3)} =$

$= 0'267$

- Itxaropena, bariantza, desb. est

$E(X) = np$

$\sigma^2 = npq$

$\sigma = \sqrt{npq}$

• Poisson-en banaketa

- Denbora, luzera edo espazio jarraian gertatutako gertaeren jazotze maila
- Esperimentua

  Unitate tartearen jazotze maila neurtzen da

  → Baldintzak

  • Tarte txiki bakoitzean gertaera jazotzeko probabilitatea konstantea
  • Tarte txiki bakoitzean gertaera jazotzeko probabilitatea = 0
  • Tarte txiki bakoitzean emandako gertaerek ez dute beste tarteengandakoetan eragiten.

  • Lagin espazioa

  $\Omega = \{0, 1, 2 \ldots n\}$

  $X : \mathcal{P}(\lambda)$

- Parametroa λ (itxaropena)

• Probabilitate legea

$P(X=x) = \frac{\lambda e^{-\lambda}}{x!}$ ADB $X : \mathcal{P}(0'7)$ $P(X=3) = \frac{0'7^3 e^{-0'7}}{3!} = 0'0284$

• Itxaropena, σ^2 eta σ

$E(X) = \lambda$

$\sigma^2 = \lambda$

$\sigma = \sqrt{\lambda}$

• Taula: 2.taula

• Banaketa binomialaren hurbilketa $n \geq 20$ $p \leq 0'5$ bada

$Bin(n,p) \approx \mathcal{P}(np)$

Banaketa normala eta aplikazioak

• Naturan maizen agertzen den banaketa jarraitua

$\Omega = (-\infty, +\infty)$

$X : N(\mu, \sigma)$

• Parametroak μ (itxaropena) eta σ (desb. est.)

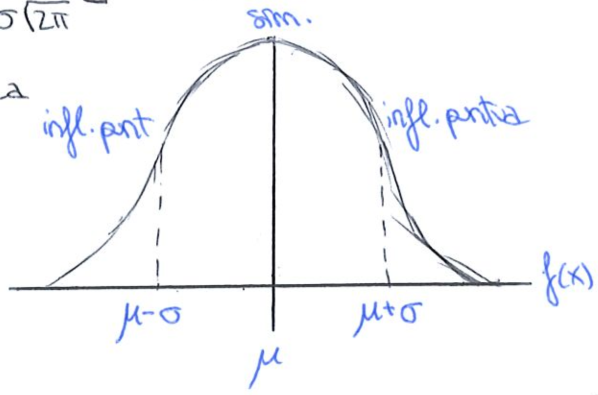
• Dentsitate funtzioa $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{x-\mu}{\sigma}\right)^2}$

• Itx, σ^2, σ → Grafika

$E(X) = \mu$

$Var(X) = \sigma^2$

$DE(X) = \sigma$



- Banaketa normal estandarizatua

$Z : N(0,1)$

- $X : N(\mu, \sigma) \longrightarrow Z = \dfrac{X-\mu}{\sigma} : N(0,1)$

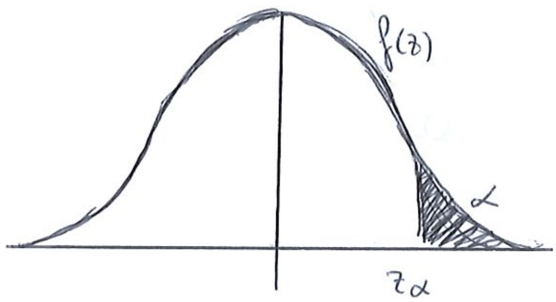
- Erabilpena

$$P(x_1 < X < x_2) = P(x_1 - \mu < X - \mu < x_2 - \mu) = P\left(\frac{x_1-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\right) = P(z_1 < Z < z_2)$$

non $z_1 = \dfrac{x_1-\mu}{\sigma}$ $z_2 = \dfrac{x_2-\mu}{\sigma}$ diren

- Balio kritikoak

$P(Z > z_\alpha) = \alpha$ • Taulak: 3. taula



→ Banaketa normalaren aplikazioak

- Propietatea: Banaketa normalaren konbinazio linealak normalak

$$\left.\begin{array}{l} X_1 : N(\mu_1, \sigma_1) \\ X_2 : N(\mu_2, \sigma_2) \end{array}\right\} \Rightarrow S = X_1 \pm X_2 : N\left(\mu_1 \pm \mu_2, \sqrt{\sigma^2 + \sigma^2}\right)$$

- Limitearen teorema zentrala

$$\left.\begin{array}{l} X_1, X_2 \ldots X_n \\ E(X) = \mu \\ Var(X) = \sigma^2 \end{array}\right\} \Rightarrow S = X_1 + X_2 + \ldots + X_n : N(n\mu, \sigma\sqrt{n}) \Rightarrow \frac{S - n\mu}{\sigma\sqrt{n}} : N(0,1)$$

- Moivre-ren teorema

$$\left.\begin{array}{l} X : Bin(n,p) \\ 0'1 < p < 0'9 \\ n \geq 30 \end{array}\right\} \Rightarrow X \simeq N(np, \sqrt{npq})$$

- Poissonen hurbilketan

$$\left.\begin{array}{l} X : \mathcal{P}(\lambda) \\ \lambda \geq 10 \end{array}\right\} \Rightarrow X \simeq N(\lambda, \sqrt{\lambda})$$

∘∘∘ Estimazioa

- Lagina aztertu eta populazioan ondorioak ateratu

→ Inferentzia estatistikoa

- Emaitzen ondorioak buruz + konfiantza maila neurtu
- Ondorioak interpretatu eta erabakiak hartu
- Zatiak

1. Estimazioa Laginaren estatistikoen bidez parametro ezezagunak hurbildu.
2. Hipotesi-kontrastea Populazioko parametro bateri buruz: hipotesia eman eta hau onartu edo errefusatzeko metodoak ezartzea

∘→ Populazioa eta lagina

- Azterketa estatistikoaren helburu den elementu multzoa (populazioa)

- Zorizko lagina Populazioko elementuen multzoa, aurrez ezarritako erabakirik gabe (zoriz) aukeratua.
- Zorizko lagin bakuna Elementuek prob. berdinarekin hautatuta osatutako multzoa.
- Zorizko zenbakia 0, 1 ... 9 probabilitate berdinarekin hartzen dituen X zorizko aldagaia

$$P(X=0)=P(X=1)=\dots=P(X=9)=1/10$$

- Zorizko zenbakien taula
  - Digitu sorta, non digitu bakoitzak agertzeko P berdina duen
  - Digitu oro askea beste digituekiko.

∘→ Estimazioa

- Laginean lortutako estatistikoen bidez populazioko parametroak hurbiltzea

- Puntu estimazioa Estimazioak parametro bati balioak ematean datza

→ Batazbestekoaren estimazioa

- Populazioaren batazbestekoa: μ
- Laginaren batazbestekoa: $\bar{X}$ (estimatzailea) lagina aukeratu aurretik zorizko aldagaia da
- Laginaren batazbestekoa $\bar{x}$ (estimazioa)
  - zenbakizko balioa
- Bitez X, μ batazbestekoa duenean z.a. laginaren batazbestekoa $\bar{X}=\dfrac{X_1+X_2\dots X_n}{n}$ μ-ren puntu estimatzailea
- $\bar{X}$ μ-ren estimatzaile onena

→ Batazbestekoaren errore estandarra

$\bar{X}$-en desbideratze estandarra $\longrightarrow DE(\bar{X})=\sqrt{Var(\bar{X})}=\sqrt{\dfrac{\sigma^2}{n}}=\dfrac{\sigma}{\sqrt{n}}$

- $\dfrac{\sigma}{\sqrt{n}}$ ezezaguna denean $\dfrac{s}{\sqrt{n}}$-ren bidez estimatzen da s $\sqrt{\text{kuasi}}$.

$$s^2_{n-1}=s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

→ Batazbestekoaren lagin banaketa

• Batazbestekoa μ eta σ^2 bariantzaduna, zoriko lagin tamaina $(X_1, X_2 \dots X_n)$
$n \geq 30$ denean

$$\bar{X} \approx N(\mu, \sigma/\sqrt{n})$$ betetzen da Xren banaketa normala ez izan arren.

ADB (ap.)

→ Batazbestekoaren konfiantza tartea (σ ezaguna)

• %95 eko konfiantza tartea (KT)

$$I_{\mu}^{0'95} = \left(\bar{X} - 1'96 \cdot \frac{\sigma}{\sqrt{n}} , \bar{X} + 1'96 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

↳ Interpretazioa

• Zuzena, n tamainako lagin %95 konf. tarteen %95ak μ barne

• Okerra, μ %95 konf. tartean egoteko $P=0'95$

• % $1-\alpha \times 100$ μren konf. tar.

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

→ Konfiantza tartearen luzera

$$L = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

→ Bariantzaren estimazioa

• Bariantzaren estimatzailea: lagin kuasibariantza

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

• Estimatzaile alboragabea

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

• σ^2 bariantzako populazio normala: lagin kuasibariantzaren banaketa

$$\frac{(n-1) \cdot S^2}{\sigma^2} = \chi^2_{n-1}$$

• Populazio normalaren bariantzarako %$(1-\alpha) \times 100$ KT

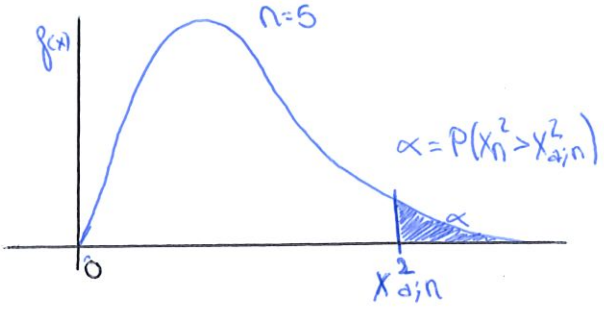
$$I_{\sigma^2}^{1-\alpha} = \left(\frac{(n-1)\, S^2}{\chi^2_{\alpha/2, n-1}} , \frac{(n-1)\, S^2}{\chi^2_{1-\alpha, n-1}}\right)$$

Pearsonen khi-karratu banaketa

- $Z_1, Z_2 \dots Z_n$ $N(0,1)$ banaketaren n zorizko aldagai askeak

$X_n^2 = Z_1^2 + Z_2^2 \dots Z_n^2$ n askatasun graduko khi-karratu banaketa dauka

- Propietateak



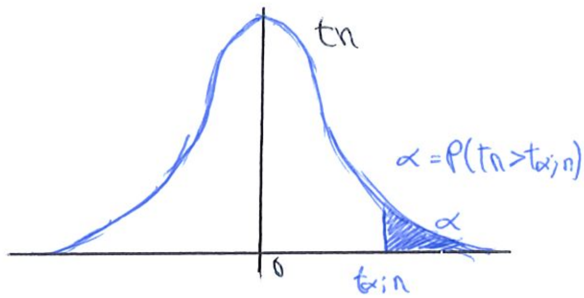
  - Jarraitua
  - Positiboa
  - Ez-simetrikoa
  - Askatasun graduen menpekoa
  - $E(\chi_n^2) = n$ $Var(\chi_n^2) = 2n$
  - $n > 30$ $X_{\alpha,n}^2 = \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$
  - $P(X_n^2 > X_{\alpha,n}^2) = \alpha$ (taulen erabilpena)

Student-en t banaketa

- $Z: N(0,1)$ X_n^2 n askatasun graduko ji-karratu zorizko aldagai askeak izanik

$$\frac{Z}{\sqrt{X_n^2/n}} = t_n$$ t_n aldagaiak n askatasun graduko Studenten banaketa dauka.

- Propietateak



  - $E(t_n) = 0$
  - 0-rekiko simetrikoa
  - $-\infty, \infty$ balioak

$Var(t_n) = \frac{n}{n-2} > 1$

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

$P(t_n > t_{\alpha;n}) = \alpha$ (taulen erabilpena)

Populazio normalaren batezbestekorako konfiantza tartea (σ ezezaguna)

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

Banaketa binomialaren p parametroaren estimazioa

- Puntu estimazioa

$$\hat{p} = \frac{X}{n}$$

- p-ren KT

$$I_p^{1-\alpha} = \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right)$$

- Lagin prop. banaketa

$n \cdot \hat{p} > 5$ $n \cdot \hat{q} > 5$

$\hat{p} \approx N\left(p, \sqrt{\frac{pq}{n}}\right)$

Konf. tar eta lagin tamaina.

$$L = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = 2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$n = 4 \cdot z_{\alpha/2}^2 \cdot \frac{\hat{p} \hat{q}}{L^2} =$$

Hipotesi Kontrasteak

- $X: N(\mu, \sigma)$ eta hipotesia egiaztatu nahi:

  $H_0: \mu = \mu_0$ (hipotesi nulua)

  $H_1: \mu \neq \mu_0$ (hipotesi alternatiboa)

→ Intuizioz H_0 egia bada

- Askotan $\mu_0 \simeq \mu$
- Gutxitan urrun

→ Jokabide araua

- n tamainako lagina aukeratu eta $\bar{x}$-en balioa kalkulatu

$\bar{x}$ μ_0-rekiko hurbil $\Rightarrow \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \in (a,b) = S_0$ (onarpen eskualdea)

$\bar{x}$ μ_0-rekiko urrun $\Rightarrow \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \notin (a,b) \Rightarrow \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \in \overline{S_0} = S_1$ (eskualde kritikoa)

- Errorea

  I motako errorea: H_0 egia denean errefusatu

  II motako errorea: H_0 gezurra denean ez errefusatu

- Planteamentu orokorra
  - H_0 egiaztatu nahi den hipotesia
  - H_1 H_0 ren kontrako hipotesia
  - Adierazgarritasun maila (α) $\alpha = P(\text{I motako errorea}) = P(H_0 \text{ errefusatu} \mid H_0 \text{ egia})$

    (β) $\beta = P(\text{II motako errorea}) = P(H_0 \text{ ez erref.} \mid H_1 \text{ egia})$
  - Ahalmena $(1-\beta) = 1 - P(\text{II motako errorea})$
  - S_0 H_0 errefusatzen ez den estatistikoaren balioek osatutako lekua.
  - S_1 H_0 errefusatzen den estatistikoaren balioek osatutako lekua.
  - p-balioa H_0 egia dela suposatuz, behatutako emaitza ≤ arraroagoa lortzeko prob.

    (p-balioa α rekiko)

    ⇓

    aztergai

→ Jarraitu beharreko urratsak

1. Hipotesia formulatu

• Aldebikoa: $H_0 \Rightarrow \mu = \mu_0$ $H_1 \Rightarrow \mu \neq \mu_0$

• Alde-batekoa (Esk / Ezk): $H_0: \mu \leq \mu_0 \,/\, \mu \geq \mu_0$ $H_1: \mu > \mu_0 \,/\, \mu < \mu_0$

2. Estatistikoa aukeratu

3. p-baloa kalkulatu

• Eskuinekoan Estatistikoaren balotik eskuinerantz dentsitate kurbaren peko azalera

• Ezkerrekoan " " ezkerrerantz " " " "

• Aldebikoan Alde bateko p-baloaren bikoitza.

4. p-baloa eta α konparatu

• p-baloa $< \alpha$ H_0 errefusatu

• p-baloa $> \alpha$ H_0 ez errefusatu

α aurre finkatua, normalean; 0'1, 0'01, 0'05

→ p-baloa zoriz agertzeko txikiegia denean H_0 errefusatu

$0'01 < p.bal < 0'05$ nahiko adierazgarriak

$0'001 < p.bal < 0'01$ oso adierazgarriak

$p.bal < 0'001$ guztiz adierazgarriak

• $0'05 < p.bal < 0'1$ adierazgarritasun estatistikorantz

- Batazbestekorako hipotesi kontrastea (σ ezaguna)

• $X: N(\mu, \sigma)$ $n \geq 30$

- Estatistikoa $z_p = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ H_0 egiazkoa dela suposatuz $z_p \approx N(0,1)$

- Aldebateko ezk. kontrastea $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$ p-baloa $= P(Z < z_p)$

- Aldebateko esk. kontrastea $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$ p-baloa $= P(Z > z_p)$

- Aldebiko kontrastea $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ p-baloa $= 2 \cdot P(Z > |z_p|)$

- Batazbesteko hipotesi kontrastea (σ ezezaguna)

$X: N(\mu, \sigma)$

- Estatistikoa $t_p = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$ H_0 egiazkotzat joz $t_p \approx t_{n-1}$

- Ald1 ezk. kontrastea $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$ p-balioa $= P(t_{n-1} < t_p)$

- Ald1 esk kontrastea $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$ p-balioa $= P(t_{n-1} > t_p)$

- Aldebiko kontrastea $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ p-balioa $= 2P(t_{n-1} > |t_p|)$

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

→ Bariantzarako H.K.

$X: N(\mu, \sigma)$

- Estatistikoa $\chi^2_p = \dfrac{(n-1)s^2}{\sigma_0^2}$ H_0 egiazkotzat joz $\chi^2_p \approx \chi^2_{n-1}$
- Ald1 ezk. kontrastea $H_0: \sigma^2 \geq \sigma_0^2$, $H_1: \sigma^2 < \sigma_0^2$ p-balioa $= P(\chi^2_{n-1} < \chi^2_p)$
- Ald1 esk. kontrastea $H_0: \sigma^2 \leq \sigma_0^2$, $H_1: \sigma^2 > \sigma_0^2$ p-balioa $= P(\chi^2_{n-1} > \chi^2_p)$
- Aldebiko kontrastea $H_0: \sigma^2 = \sigma_0^2$, $H_1: \sigma^2 \neq \sigma_0^2$ p-balioa $= \begin{cases} s^2 \leq \sigma_0^2 \to = 2P(\chi^2_{n-1} < \chi^2_p) \\ s^2 > \sigma_0^2 \to = 2P(\chi^2_{n-1} > \chi^2_p) \end{cases}$

→ HKren ahalmena ($1-\beta$)

- $1-\beta = 1 - P(\text{II motako errorea}) = 1 - P(H_0 \text{ ez errefusatu} \mid H_0 \text{ faltsua}) = P(H_0 \text{ errefusatu} \mid H_0 \text{ faltsua})$

Esk. Ald1. kontrastea	Ezk. Ald1 kontrastea	Aldebiko kontrastea
$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$ $1-\beta = P\left(Z \geq z_\alpha + \dfrac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$	$H_0: \mu \geq \mu_0$ $H_1: \mu = \mu_1 < \mu_0$ $1-\beta = P\left(Z \leq -z + \dfrac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$	$H_0: \mu = \mu_0$ $H_1: \mu = \mu_1 \neq \mu_0$ $1-\beta = P\left(Z \leq -z_{\alpha/2} + \dfrac{(\mu_0 - \mu_1)}{\sigma/\sqrt{n}}\right) + P\left(Z \leq z_{\alpha/2} + \dfrac{(\mu_1 - \mu_0)}{\sigma/\sqrt{n}}\right)$

- Kontrastearen ahalmenean eragiten duten faktoreak
  - Adierazgarritasun maila, α, txikitzean ahalmena txikitu
  - Bataz besteko alternatiboa, $|\mu_0 - \mu_1|$, handitzean ahalmena txikitu
  - Laginaren tamaina, n, handitzean ahalmena handitu.

Laginaren tamaina (n)

- Aldebateko auzerak

$$n = \frac{\sigma^2 (z_\alpha + z_\beta)^2}{(\mu_0 - \mu_1)^2}$$

- Aldebiko auzerak

$$n = \frac{\sigma^2 (z_{\alpha/2} + z_{\beta/2})^2}{(\mu_0 - \mu_1)^2}$$

- Laginaren tamainan eragiten duten faktoreak
  - Adierazgarritasun maila, α, txikitzean n handitu
  - Ahalmena, $1-\beta$, handitzean n handitu
  - Bataz besteko alternatiboa, $|\mu_0 - \mu_1|$, handitzean n txikitu

Konf. tar-en eta HKren arteko erlazioa

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$ (aldebiko kontrastea) α adierazgarritasun maila

- H_0 errefusatu; μ rako $\%(1-\alpha)\times 100$ konf. tar. ak μ_0 ez du barnean
- H_0 ez errefusatu; μ rako $\%(1-\alpha)\times 100$ konf. tar. ak μ_0 barnean

Banaketa binomialaren p parametrorako HK

$X: Bin(n, p) \quad n \cdot \hat{p} > 5 \quad n \cdot \hat{q} > 5$

- Proportziorako HK.
  - Estatistikoa $z_p = \dfrac{\hat{p} - p_0}{\sqrt{p_0 \cdot q_0 / n}}$ H_0 egiazkotzat joz $z_p \approx N(0,1)$
  - Aldebatekoa p-balioa $= P(Z < z_p)$ ezk. $\quad 1-\beta = P\left(Z \leq -z_\alpha \cdot \sqrt{\dfrac{p_0 \cdot q_0}{p_1 \cdot q_1}} + \dfrac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_1 \cdot q_1}}\right)$
    $P(Z > z_p)$ esk.
  - Aldebikoa p-balioa $= 2P(Z > |z_p|)$ $\quad n = \dfrac{(z_\beta \cdot \sqrt{p_1 \cdot q_1} + z_\alpha \cdot \sqrt{p_0 \cdot q_0})^2}{(p_0 - p_1)^2}$

LAGIN BIRAKO INFERENTZIA

- Aldagai jarraitua
  (batazbesteko eta bariantzen konp.)

  - Bi lagin askeren batazbestekoen diferentziaren banaketa

    X_1 (bi populazio aske) $(\mu_1$ (batazb.), σ_1^2 (V)$)$ n_1 tamaina
    X_2 (μ_2, σ_2^2) n_2

  - Kasuak

    1. σ_1^2 eta σ_2^2 ezagunak
       - X_1, X_2 populazio normalak $\rightarrow \bar{X}_1 - \bar{X}_2 : N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
       - $n_1 \geq 30, n_2 \geq 30 \rightarrow \bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
    2. σ_1^2 eta σ_2^2 ezezagunak
       - $n_1 \geq 30, n_2 \geq 30 \rightarrow \bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$

       edo

       - $n_1 < 30$ edo $n_2 < 30$ bariantza ezberdinak

         $\rightarrow \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t_g$

         bariantza berdinak

         $\rightarrow \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx t_{n_1 + n_2 - 2}$

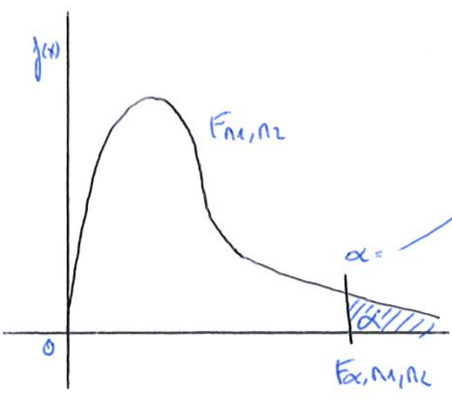
$$\boxed{S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \qquad g = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}}$$

- Fischerren F banaketa

  $\chi^2_{n_1}$ eta $\chi^2_{n_2}$ n_1 eta n_2 askatasun graduko khi^2 banaketa.

  $$F_{n_1, n_2} = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$$

  - Propietateak
    - $[0, +\infty)$ tarteko balioak
    - Simetrikoa ez.
    - n_1 eta n_2 ren pean
    - $P(F_{n_1, n_2} > F_{\alpha, n_1, n_2}) = \alpha$
    - $P(F_{n_1, n_2} > A) = P(F_{n_2, n_1} < 1/A)$
    - $F_{\alpha; n_1; n_2} = \dfrac{1}{F_{1-\alpha; n_2; n_1}}$



∘ Kuasibariantzaen arteko zatiduraren banaketa

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} : F_{(n_1-1),(n_2-1)}$$

Banaketa normalari dagozkion 2 populazio askeren bariantzen konp.

- Estatistikoa $F_p = \frac{S_1^2}{S_2^2}$ H_0 egia dela suposatuz $F_p : F_{(n_1-1);(n_2-1)}$
- Ald1 ezkerreko k. $H_0: \sigma_1^2 \geq \sigma_2^2$, $H_1: \sigma_1^2 < \sigma_2^2$ p. balioa $= P(F_{(n_1-1);(n_2-1)} < F_p)$
- Ald1 eskuineko k. $H_0: \sigma_1^2 \leq \sigma_2^2$, $H_1: \sigma_1^2 > \sigma_2^2$ p. balioa $= P(F_{(n_1-1);(n_2-1)} > F_p)$
- Aldebiko kontrastea $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$ p. balioa
  - $S_1^2 \geq S_2^2$ $p = 2P(F_{(n_1-1)(n_2-1)} > F_p)$
  - $S_1^2 < S_2^2$ $p = 2P(F_{(n_1-1)(n_2-1)} < F_p)$

Bi banaketa askeren batazbestekoen arteko diferentziaren konf. tar.

σ_1^2 eta σ_2^2 ezagunak
→ X_1, X_2 populazio normalak
$n_1 \geq 30$ $n_2 \geq 30$

$$I^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{X}_1-\bar{X}_2 - z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, \bar{X}_1-\bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

→ σ_1^2 eta σ_2^2 ezezagunak
X_1, X_2 populazio normalak

∘ $n_1 \geq 30$, $n_2 \geq 30$ $\quad I^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{X}_1-\bar{X}_2 - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}, \bar{X}_1-\bar{X}_2 + z_{\alpha/2}\ldots\right)$

∘ $n_1 < 30$ edo $n_2 < 30$ → Bariantza berdinak → $I^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{X}_1-\bar{X}_2 - t_{\alpha/2;(n_1+n_2-2)}\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \bar{X}_1-\bar{X}_2+t\ldots\right)$

↳ Bariantza ezberdinak → $I^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{X}_1-\bar{X}_2 - t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}, \bar{X}_1-\bar{X}_2+t\ldots\right)$

$$\boxed{S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \qquad g = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}}$$

⊖▷ Bi banaketa askeren batazbestekoen konparazioraoko HK

A σ_1^2, σ_2^2 ezagunak (X_1, X_2 populazio normalak) $n_1 \geq 30$ $n_2 \geq 30$

- Estatistikoa $Z_p = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ Ho egiazkotzat jo $Z_p \approx N(0,1)$

- Ald1. ezk. k $H_0: \mu_1 \geq \mu_2$, $H_1: \mu_1 < \mu_2$ $\quad$ p bal $= P(Z < Z_p)$

- Ald1 esk. k $H_0: \mu_1 \leq \mu_2$, $H_1: \mu_1 > \mu_2$ $\quad$ pbal $= P(Z > Z_p)$

- Aldebiko. k $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$ $\quad$ pbal $= 2P(Z > |Z_p|)$

B σ_1^2, σ_2^2 ezezagunak (X_1, X_2 pop. normalak) $n_1 \geq 30, n_2 \geq 30$

- Estatistikoa $Z_p = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ Ho egiazkotzat jo $Z_p \approx N(0,1)$

→ 3 kontrasteak A ren berdinak

C σ_1^2, σ_2^2 ezezagunak (X_1 eta X_2 pop. normalak) $n_1 < 30$ edo $n_2 < 30$ ($\sigma_1^2 = \sigma_2^2$)

- Estatistikoa $t_p = \dfrac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ Ho egiazkotzat jo $t_p \approx t_{n_1+n_2-2}$

- Ald1 ezk. k $H_0: \mu_1 \geq \mu_2$, $H_1: \mu_1 < \mu_2$ $\quad$ p. bal $= P(t_{n_1+n_2-2} < t_p)$

- Ald1 esk. k $H_0: \mu_1 \leq \mu_2$, $H_1: \mu_1 > \mu_2$ $\quad$ p. bal $= P(t_{n_1+n_2-2} > t_p)$

- Ald2 k $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$ $\quad$ pbal $= P(t_{n_1+n_2-2} > |t_p|)$

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

D σ_1^2, σ_2^2 ezezagunak (X_1, X_2 pop normalak) $n_1 \leq 30$ edo $n_2 < 30$ ($\sigma_1^2 \neq \sigma_2^2$)

- Estatistikoa $t_p = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ Ho egiazkotzat joz. $t_p \approx t_g$

- Ald1eakk $H_0: \mu_1 \geq \mu_2$, $H_1: \mu_1 < \mu_2$ p.bal $= P(t_g < t_p)$

- Ald1eakk $H_0: \mu_1 \leq \mu_2$, $H_1: \mu_1 > \mu_2$ p.bal $= P(t_p < t_g)$

$$\boxed{g = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}}$$

- Ald2k $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$ p.bal $= 2P(t_g > |t_p|)$

Binakako datuak

- X_1 X_2 normalak baina ez askeak

$$\left.\begin{array}{l} X : N(\mu_X, \sigma_X) \\ Y : N(\mu_Y, \sigma_Y) \end{array}\right\} \Longrightarrow D = X - Y : N(\mu_D, \sigma_D)$$

$\mu_D = E(D) = E(X - X) = \mu_X - \mu_Y$

σ_D ezin dugu kalkulatu, beraz S_D erabiliko dugu.

- μ_D-ren puntu estimazioa $\bar{d}$
- konfiantza tartea $I_{\mu_D}^{1-\alpha} = \left(\bar{d} - t_{\alpha/2;(n-1)} \cdot \dfrac{Sd}{\sqrt{n}}, \bar{d} + t_{\alpha/2} \ldots\right)$

→ HK

- Estatistikoa $t_p = \dfrac{\bar{d}}{Sd/\sqrt{n}}$ Ho egiazkotzat joz $t_p \approx t_{n-1}$

- Ald1eakk $H_0: \mu_D \geq 0$, $H_1: \mu_D < 0$ p.bal $= P(t_{n-1} < t_p)$

- Ald1eakk $H_0: \mu_D \leq 0$, $H_1: \mu_D > 0$ p.bal $= P(t_{n-1} > t_p)$

- Ald2k $H_0: \mu_D = 0$, $H_1: \mu_D \neq 0$ p.bal $= 2P(t_{n-1} > |t_p|)$

Aldagai diskretua (2 kategoria)

- Banaketa binomial askeen p parametroen alderaketa

$X_1 : Bin(n_1, p_1)$
$X_2 : Bin(n_2, p_2)$ askeak

- $(p_1 - p_2)$ ren puntu estimazioa

$$\widehat{(p_1 - p_2)} = \hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

- estimatzailearen banaketa

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

$n \cdot$ (edozein $\hat{p}$ edo $\hat{q}$) > 5 bada

- $(p_1 - p_2)$ konf. tar

$$I^{1-\alpha}_{p_1 - p_2} = \left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha} \ldots\right)$$

- HK

  - Estatistikoa $Z_p = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}}$ H_0 egiaztatzat jo $Z_p \approx N(0,1)$

  - Ald1ezkk $H_0 : p_1 \geq p_2$, $H_1 : p_1 < p_2$ p.bal $= P(Z < z_p)$

  - Ald1eskk $H_0 : p_1 \leq p_2$, $H_1 : p_1 > p_2$ p.bal $= P(Z > z_p)$

  - Ald2k $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$ p.bal $= 2P(Z > |z_p|)$

BARIANTZA ANALISIA

- Faktore bakarreko bariantza-analisia
  - k populazio aske eta bariantza berdinak
    $X: N(\mu, \sigma)$ k populazioetan
    HK betetzen da

    $H_0: \mu_1 = \mu_2 = \mu_3 \cdots \mu_k$

    H_1: Batazbestekoren bat ezberdina

  - Definizioak
    - Karratuen batura totala (KBT) $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2$
    - Tratamenduen karratuen batura (KB(Tr)) $\sum_{i=1}^{k} n_i(\bar{X}_i-\bar{X})^2$
    - Hondar edo erroreen karratuen batura (EKB) $KBT = KB(Tr) + EKB$ $\quad \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2 = EKB$

- Anova taula

Iturria	Askatasun graduak	KB	F	p
Eredua	$k-1$	KB(ered)	F_p	p.bal.
Errorea	$n-k$	EKB		
Totala	$n-1$	KBT		

  - hondar bariantza
    $HB = EKB/(n-k)$
  - estatistikoa
    $$F_p = \frac{KB(ered.)/(k-1)}{EKB/(n-k)}$$
  - p.balioa
    $p.bal = P(F_{k-1;n-k} > F_p)$
    - $p \leq \alpha$ Ho hipotesia errefusatu. X-en batazbestekoa populazioaren menpekoa
    - $p > \alpha$ Ho onart. X en batazbestekoa ez da populazioaren menpekoa.

- Konparaketa anitzak

  Ho errefusatuz gero batazbestekoen arteko diferentzia adierazgarria k populazioen artean
  → zein populaziotan?
  ↳ jakiteko konparaketa anitzak Scheffé / Turkey metodoak

Khi^2-REN APLIKAZIOAK

∘ r errenkada, s zutabe

$$\begin{array}{cccc|c} O_{11} & O_{12} & \cdots & O_{1s} & f_1 \\ O_{21} & \ddots & & \vdots & f_2 \\ \vdots & & & & \vdots \\ O_{r1} & \cdots & & O_{rs} & f_r \\ \hline z_1 & z_2 & \cdots & z_s & n \end{array}$$

$O_{ij} \equiv ij$ klaserako behatutako maiztasuna

$z_{\cdot j} \equiv j$ zutaberako behatutako maiztasuna (ztm - zutabeen total marginala)

$f_{i\cdot} \equiv i$ errenkadarako behatutako maiztasuna (errenkaden total marginala - etm)

$n \equiv$ laginaren tamaina

∘→ Homogeneotasun eta independentzia frogak

∘ Bi faktoreen arteko menpekotasuna aztertzeko

→ Independentzia froga: Total marginal oro zoriztko. Lagin bakar 1.

H_0: A eta B askeak

H_1: A eta B menpekoak

→ Homogeneotasun froga: Tm. batzuk aurrez finkatuak

H_0: Laginak populazio beretik

H_1: Laginak populazio ezberdinetatik

⇝ Planteamendu orokorra

1. r×s kontingentzia taula eraiki (ikus. gora)
2. HK definitu
3. H_0 egia dela suposatuz itxarotako maiztasunak

$$e_{ij} = \frac{f_{i\cdot} \cdot z_{\cdot j}}{n}$$

4. $e_{ij} > 5 \ \forall \ i,j$ betetzen dela egiaztatu

↳ Lortzeko klaseak batu

5. HK egiteko darabilgun estatistikoa kalkulatu

$$\chi^2_p = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{O_{ij}^2}{e_{ij}} - n$$

6. Askatasun gradu kopurua $\equiv (r-1)\cdot(s-1)$
7. p-balioa kalkulatu $p = P(\chi^2_{(r-1)(s-1)} > \chi^2_p)$
8. α ren arabera erabakia hartu
   - $p > \alpha$ H_0 ez errefusatu
   - $p < \alpha$ H_0 errefusatu.

★ Salbuespenak

∘ $e_{ij} > 5$ bete ezean klaseak bildu

∘ Fisherren proba zehatza: klaseak bilduta ere klaseren bat < 5 bada jikaratu ez.

∘ Yatesen zuzenketa

- 2×2 kt.

A.G. = 1

$|O_{11}O_{22} - O_{12}O_{21}| > n/2$

$$\chi^2_Y = \sum_{j=1}^{2} \sum_{i=1}^{2} \frac{\left(|O_{ij} - e_{ij}| - \frac{1}{2}\right)^2}{e_{ij}}$$

• Oharra!

2x2 kontingentzia-taulako kalkuluak errazteko formulak

$$\chi^2_p = \frac{n(o_{11}o_{22}-o_{12}o_{21})^2}{z_{.1}\cdot z_{.2}\cdot f_{1.}\cdot f_{2.}} \qquad \chi^2_Y = \frac{n(|o_{11}o_{22}-o_{12}o_{21}|-n/2)^2}{z_{.1}\cdot z_{.2}\cdot f_{1.}\cdot f_{2.}}$$

$|o_{11}o_{22}-o_{12}o_{21}| \leq n/2$ $\qquad$ $|o_{11}o_{22}-o_{12}o_{21}| > n/2$

→ Doikuntza egokitasunerako proba

• K klase baztertzaile eta exhaustiboetako populaziotik n lagina

H_0: K klaseen probabilitateak $p_1, p_2 \ldots p_k$ dira $(p_1+p_2 \ldots + p_k = 1)$

H_1: ez dira

→ Planteamendu orokorra

1. HK definitu
2. Ho egia dela suposatuz itxarondako maiztasunak kalkulatu
   $e_i = n \cdot p_i$
3. $e_i > 5 \ \forall i$ betetzen
4. HK rako estatistikoa
   $$\chi^2_p = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n$$
5. Askatasun graduak
   • $p_1, p_2 \ldots p_k$ zehaztuta $\equiv k-1$
   • $p_1, p_2 \ldots p_k$ zehazteke $\equiv k-l-1$
6. p balioa
   $p = P(\chi^2_{k-l-1} > \chi_p)$
7. α ren p-rekiko izatearen arabera erabakia

ERREGRESIO LINEALA

- Ale esperimentalaren neurketaren emaitza
  erantzun aldagaia edo aldagai menpekoa (y)

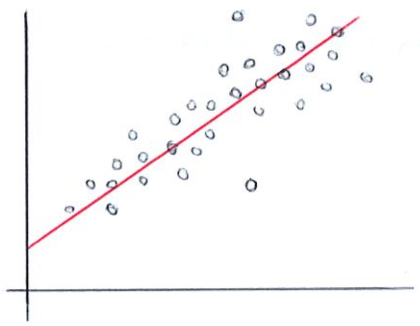
Erantzunean eragin dezakeen argibide aldagaia
  Koaldagaia edo aldagai askea (x)

Helburua x eta y-ren arteko erlazio lineala aztertzea

⊙→ X-en gaineko y-ren erregresio zuzena: $y_i = b_0 + b_1 x_i$

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (eredu teorikoa)

b_0 eta b_1 datuetatik
β_1 eta β_0-ren estimatzaileak



y_i = puntuak
$\hat{y}_i$ = zuzena (puntuak hoberen hurbiltzen dituena)

⊙→ Parametroen estimazioa

Malda $\hat{\beta}_1 = b_1 = \dfrac{\sum_{i=e}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=e}^{n}(x_i-\bar{x})^2}$

Gai askea $\hat{\beta}_0 = b_0 = \bar{y} - b_1\bar{x}$

Hondar bariantza: $\hat{\sigma}^2 = \dfrac{\sum_{i=e}^{n}(y_i-\hat{y}_i)^2}{n-2} = \dfrac{\sum_{i=e} e_i^2}{n-2}$

⊙→ Hipotesi kontrastea (Wald-en testa)

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

- Estatistikoa $T_p = \dfrac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{\sum(x_i-\bar{x})^2}}$

- p-balioa $p = 2P(t_{n-2} > |t_p|)$

Parametroen interpretazioa

- $\hat{\beta}_1 = b_1 \equiv X$ aldagaia $+1$ y aldagaian espero den gehikuntza
- Walden HK X aldagaiak y erantzunean duen eragin linealaren adierazgarritasuna aztertzeko
- $\hat{\beta}_0 = b_0 \equiv X$ aldagaia $=0$ y aldagaiak hartzea espero den balioa

→ $x=0$ koaldagaiaren heinaren barne ez badago, b_0 interpretatzea zentsugabea

ANOVA taula

- Karratuen batura totala $KBT \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \equiv$ aldakortasun totala
- Ereduaren karratuen batura totala $KB(X) \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \equiv X$en bidez azaldutako aldakortasuna
- Erroreen karratuen batura $EKB \equiv \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \equiv$ hondar aldakortasuna, azaldu gabekoa

$$\boxed{KBT = KB(X) + EKB}$$

→ F Kontrastea

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

- Estatistikoa $F_p = \dfrac{KB(X)}{EKB/(n-2)}$
- p.balioa $p = P(F_{1,n-2} > F_p)$

F_p eta p.bal ANOVA taularen eskuinean

Determinazio koefizientea

$$R^2 = \frac{KB(X)}{KBT} = 1 - \frac{EKB}{KBT}$$

R^2 handiagoa, eredu hobea

- Propietateak

Pearsonen korrelazio koefizientearen karratua: $R^2 = r_{xy}^2$

$R^2 \in [0,1]$

Xen gaineko yren err. zuz $\hat{y}_i = b_0 + b_1 x_i$

yren gaineko xen err. zuz $\hat{x}_i = a_0 + a_1 y_i$

$\Big\}\ r_{xy} = r_{yx}$

R^2 x eta y ren arteko erlazio lineala zehazteko

- Hondar bariantza $\hat{\sigma}^2 = \dfrac{\sum (y_i - \hat{y}_i)^2}{n-2} = \dfrac{\sum e_i^2}{n-2}$

Bioest.

- Populazio baten bariantzaren estimazioa Khi karratu
- Banaketa binomialaren p-ren estimazioa banaketa normala $\begin{cases} n\cdot\hat{p} > 5 \\ n\cdot\hat{q} > 5 \end{cases}$ → arrakasta prob. $\frac{\text{arrakasta kop}}{n} = \hat{p}$

→ Estandarizazioa

$X: N(52,6)$ daukagu $\quad z = \frac{x-\mu}{\sigma} = \frac{x-52}{6}$

$N(\mu,\sigma)$

$P(X>50)$ nahi dugu $\quad z_\alpha = \frac{50-52}{6} = -0'33$

$P\left(z > \frac{50-52}{6}\right)$ = minus bada $1 - P\left(z > \left|\frac{5-52}{6}\right|\right)$

s kalkulatzeko

- Batazbestekoaren konf. tar (μ)

- σ^2 ezaguna Ban. normala
- σ^2 ezezaguna studenten ban. $\alpha = P(t_n > t_{\alpha;n})$ $n > 30$ banaketa normala

→ $s^2 = s^2_{n-1} = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ → s lortzeko

★ Binomiala estandarizatzen

- $n \geq 20 \quad p \leq 0'05 \quad Bin(n,p) \approx \wp(np) = \wp(n\cdot p)$
- $n \geq 30 \quad 0'1 < p < 0'9 \quad X: Bin(n,p) \rightarrow X \approx N(np, \sqrt{npq})$

⊙ Ahalmenak

→ Ezkerretik $1-\beta = P\left(Z \leq -z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$

→ Eskuinetik $1-\beta = P\left(Z \geq z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$

→ Aldebikoa $1-\beta = P\left(Z \leq -z_{\alpha/2} \frac{(\mu_0-\mu_1)}{\sigma/\sqrt{n}}\right) + P\left(Z \leq -z_{\alpha/2} + \frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}\right)$

! errorea $= z_{\alpha/2} \frac{S_{n-1}}{\sqrt{n}}$