VADEMECUM. ESTATISTIKA ETA DATUEN ANALISIA. II garren ZATIA

Asoziazioa: aldagai kualitatiboen arteko erlazioa

abiapuntua: kontingentzia-taula

maizt. enpiriko	a	b	c	a+b+c (bazter maizt)
	d	e	f	d+e+f (bazter maizt)
bazter maizt	a+d	b+e	c+f	n

independentzia: asoziazio-eza
% berdinak errenkadetan eta zutabeetan

portzentajeak: asoziazio esploratzeko

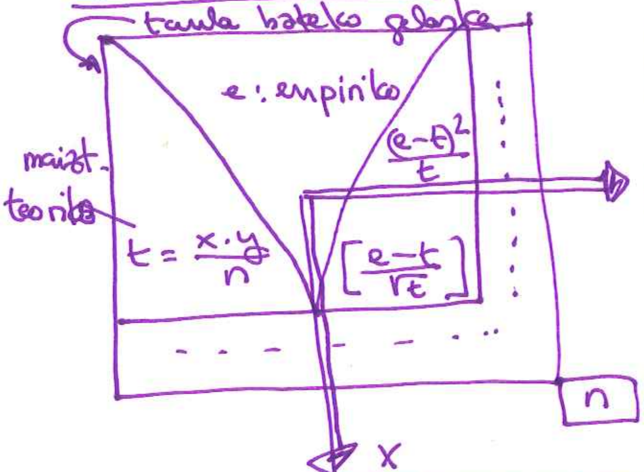
$\left[\frac{a}{a+d}, \frac{d}{a+d}\right] \cdot 100, \left[\frac{b}{b+e}, \frac{e}{b+e}\right], \ldots$

errenkadako aldagaiarenak

$\left[\frac{a}{a+b+c}, \frac{b}{a+b+c}, \frac{c}{a+b+c}\right], \ldots$

zutabeko aldagaiarenak.

Asoziazio-neurriak



$\chi^2 = \sum \frac{(e-t)^2}{t}$ $[0, \infty)$

khi karratu; $\sum$ → gelaxka guztiak; [0, ∞) → ezin da interpretatu

phi $\varphi = \sqrt{\frac{\chi^2}{n}}$	0-1 balioak; 2x2 taulak soilik
$C = \sqrt{\frac{\chi^2}{\chi^2+n}}$ kontingentzia koef	0-1 taula karratuak; $C_{max} = \sqrt{\frac{m-1}{m}}$ taula ez karratuak; C/C_{max}
$V = \sqrt{\frac{\chi^2}{n(m-1)}}$ Cramerren V	taula guztiak

λ (lambda) (PRE neurria)

indep? dep?

indep. kontuan hartu gabe, zenbat errore? → (m) (erroreak guztira)

indep. kontuan harturik, indep. A bada, zenbat errore? (a)
B bada, zenbat errore? (b)
C bada, zenbat errore? (c)

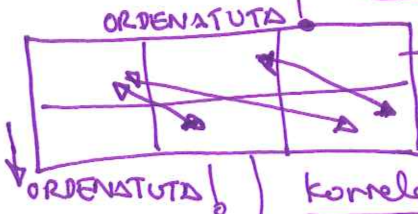
indep: A → Σa; B → Σb; C → Σc; ...

$\lambda = \frac{m - (a+b+c+\ldots)}{m}$ → [zenbat errore gutxiago]

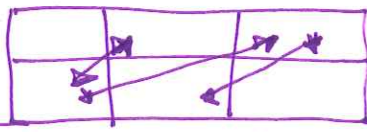
Goodman eta Kruskalen gamma

$\gamma = \frac{k-d}{k+d}$ → int. <0 → x↑ y↓; >0 → x↑ y↑

bi aldagai **ordinal**



→ k konkordantziak: gezi puntak bidertu eta batu



d diskordantziak gezi puntak bidertu eta batu

Korrelazioa: aldagai kuantitatiboen arteko erlazioa

$S_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n} = \frac{\sum xy}{n} - \bar{x}\bar{y}$ (kobariantza)

x	y	xy
Σ	Σ	Σ

$S_{xy} \in [-\infty, \infty]$ └ sendotasuna ez.
<0 → x↑ y↓; >0 → x↑ y↑

r_{xy} korrelazio koef lineala! $= \frac{S_{xy}}{S_x S_y}$

x	y	xy	x²	y²
Σ	Σ	Σ	Σ	Σ

$r_{xy} \in [-1, 1]$ (<0 → x↑ y↓; >0 → x↑ y↑)

ANTZEKO IKERKETEN ERRESERBAEN
asoc. neurriak, |γ| eta |r_{xy}| nola interpretatu?
0-0'3 → asoz edo kor ahula
0'3-0'6 → asoz edo kor ertaina
>0'6 → asoz kor sendoa

r_{xy} dikotomiko - kuantitatibo

0 1 nahi den eran → r_{xy} int { x : 0 → 1 ; y : ↑↓ ?

r_{xy} dikotomiko - dikotomiko

0 1 0 1 → nahi den eran

r_{xy} int { x : 0 → 1 ; y : 0 → 1 ? x ↑ y ↑ ; 1 → 0 ? x ↑ y ↓

korrelazio partziala :

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

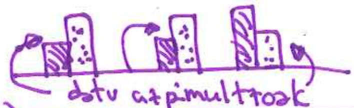
↓ baztertzen den aldagaia.

alegiazko korrelazioa : sendoa baina fundamentu gabea

kuantitatibo - kualitatibo erlazio estatistikoa

indep? dep? independentearen araberako multzoak egin eta horien arabera bestearen datu-azpimultzoak aztertu.

kualitatibo



kuantitatibo → nolakoa den: } histograma, puntu diagrama, kaxa-diagrama, $\bar{x}$

datu azpimultzo bakoitzeko

korrelazio matrizea

- simetrikoa
- diagonala 1

> x = c(...)
> y = c(...)
> data.frame(x, y) = izena
> cor(izena, izena)

ERREGRESIOA

(1) datuak : x, y

zuzena : $\hat{y} = a + bx$

aurresonak

$b = \frac{S_{xy}}{S_y^2}$; $a = \bar{y} - b\bar{x}$

x	y	xy	x^2
Σ	Σ	Σ	Σ

(2) $e = y - \hat{y}$ erroreak

PROP : $\Sigma e = 0$

$\Sigma y = \Sigma \hat{y}$

R : >zuzena = lm(y ~ x)

zuzena $ residuals
zuzena $ fitted.values

Doikuntzaren egokitasuna

(3) PROP : $S_y^2 = S_{\hat{y}}^2 + S_e^2$

Totala ; azaldua ; azaldu gabea

x	y	xy	x^2	y^2
Σ	Σ	Σ	Σ	Σ

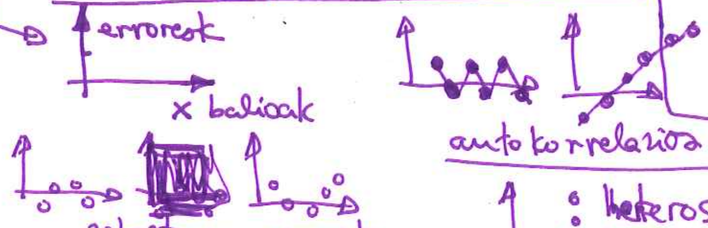
$\hat{y}$	$\hat{y}^2$	e^2
Σ	Σ	Σ

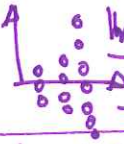
$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2}$$

mugatze koef

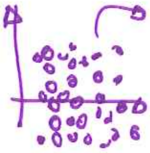
[0,1] → > 0'6 → egokitasun handia

(4) Erroreen diagramak (diagnostikoa)





→ σ^2 desberdina



→ σ^2 berdintsua

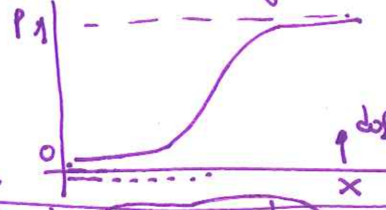
homoskedastikotasuna

logit eredua (erregresio logistikoa)

aldagai dep : probabilitatea edo portzentajea (p)
indep : dosia

$$\hat{y} = \ln \frac{p}{1-p} = a + bx$$

↓ dosia



→ datuen joera alderantzizkoa bada ezezkoaren edo aurkakoaren probabilitatea egin.

doikuntza ez linealak

kurbam — linealizatu → zuzenean

[illegible] zuk — balondu — a b

indep → v ; dep → w

ln v, 1/v → indep (x)

dep (y) ← w, ln w

→ R^2 BETI x eta y-rekin dd