Studio della dipendenza di una variabile quantitativa da una variabile qualitativa (il rapporto di correlazione)

	Task Description	Due Date	Status
☐	Studio asimmetrico del legame tra due variabili: l'indipendenza in media come condizione più debole di indipendenza.		
☐	Uso della proprietà di decomposizione della varianza (devianza) per misurare la dipendenza in media di una variabile quantitativa da una variabile qualitativa: l'indice η^2.		
☐			
☐			
☐			
☐			
☐			
☐			
☐			
☐			

Analisi simmetrico] associazione, connessione, correlazione, interdipendenza

X		Y		
qualitativa		qualitativa		χ^2, Φ^2, V
X	$\longleftrightarrow$	Y		
quantitativa		quantitativa		σ_{XY}, ρ_{XY}

Analisi asimmetrico] dipendenza

X		Y		
qualitativa		quantitativa		$\eta^2_{Y\mid X}$
X	$\longrightarrow$	Y		
esplicativa (indipendente)		dipendente		
quantitativa		quantitativa		regressione

IPOTESI DI INDIPENDENZA ASSOLUTA

le distribuzioni condizionate

$X \mid Y = y_j \qquad j = 1, \dots, c$

e

$Y \mid X = x_i \qquad i = 1, \dots, r$

sono "simili"

→ profili riga uguali

→ profili colonna uguali

Se una delle due variabili è numerica posso rilassare questa condizione richiedendo l'uguaglianza non delle distribuzioni ma soltanto di un suo parametro di sintesi

INDIPENDENZA PARAMETRICA

caso più comune: la media

INDIPENDENZA IN MEDIA

le distribuzioni condizionate possono essere diverse ma devono avere la stessa media (che per la proprietà di baricentricità è anche la media marginale)

MEMO: DECOMPOSIZIONE DELLA VARIANZA

$$\boxed{\sigma^2 = \sigma^2_{INT} + \sigma^2_{EST}}$$

$$\boxed{\frac{\sum (x_i-\mu)^2}{n}}$$

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\mu)^2$$

$$\sum_{i=1}^{g}\frac{\sigma_i^2\, n_i}{n}$$

media delle varianze di gruppo

$$\sum_{i=1}^{g}\frac{(\bar{x}_i-\mu_x)^2\, n_i}{n}$$

varianza delle medie di gruppo rispetto alla media generale

entrambe pesate per le numerosità dei vari gruppi

MEMO: DECOMPOSIZIONE DELLA DEVIANZA

$$\boxed{Dev(x) = Dev_{INT} + Dev_{EST}}$$

$$\underbrace{\sum (x_i - \mu)^2}$$

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \mu)^2$$

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

misura di variabilità interna ai gruppi

$$\sum_{i=1}^{g} (\bar{x}_i - \mu_x)^2 n_i$$

devianza delle medie di gruppo rispetto alla media generale (pesata per le numerosità dei gruppi)

Associazione tra caratteri:
La dipendenza di una variabile da una mutabile

variabile qualitativa

variabile quantitativa

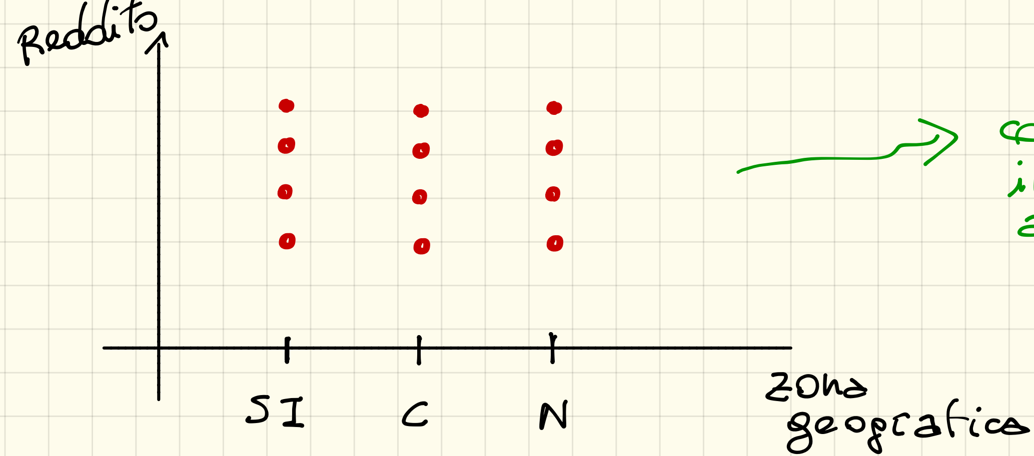
Reddito | Nord

Reddito (Y) / Zona (X)	10-15 mila	15-20 mila	20-25 mila	25-30 mila	
Nord		7	34	5	*46*
Centro	1	18	5	1	*25*
Sud e Isole	31	1			*32*
	32	*26*	*39*	*6*	*103*

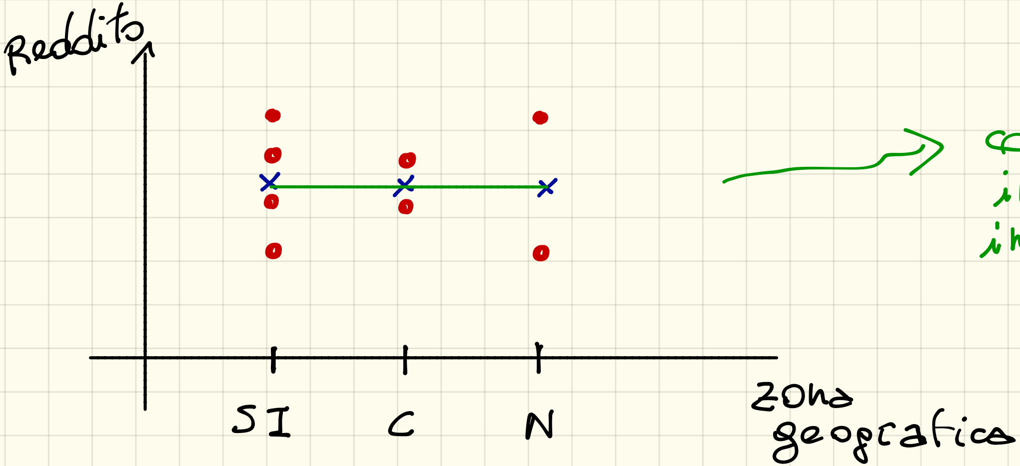
Redd. | Centro

Reddito / Sud

Reddito



condizione di indipendenza assoluta



condizione di indipendenza in media

Le informazioni per l'analisi della dipendenza in media sono nella varianza (devianza) esterna

$$\sigma^2_{EST} = 0 \quad \text{se e solo } \bar{x}_i = \mu_x \quad \forall i = 1, \dots, G$$

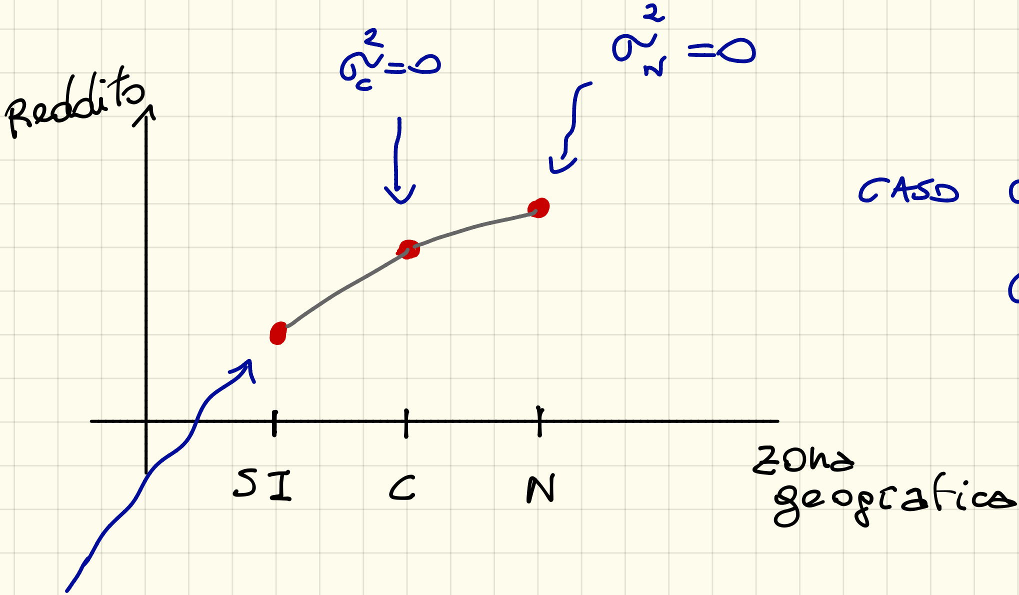
$$\| \\ Dev_{EST}$$

(condizione di indipendenza in media)

quanto più questi due indicatori sono vicini allo zero tanto più forte è il segnale di assenza di legame in media

↳ per avere un indice normalizzato:

$$\eta^2_{Y|X} = \frac{\sigma^2_{EST}}{\sigma^2} = \frac{Dev_{EST}}{Dev(Y)}$$



CASO $\sigma^2_{INT} = 0$

$\sigma^2_{EST} = \sigma^2$

$\eta^2 = 1$

max dipendenza in media

tutte le province del SI hanno lo stesso livello di reddito

$$\sigma^2_{SI} = 0 = Dev_{SI}$$

Associazione tra caratteri:
La dipendenza di una variabile da una mutabile

Reddito (**Y**) / Zona (**X**)	10-15 mila	15-20 mila	20-25 mila	25-30 mila	
Nord		7	34	5	*46*
Centro	1	18	5	1	*25*
Sud e Isole	31	1			*32*
	32	*26*	*39*	*6*	*103*

M(Y|X=N)	22,3	$\bar{y}$ nord
M(Y|X=C)	18,7	$\bar{y}$ centro
M(Y|X=S)	12,7	$\bar{y}_{SI}$
M(Y)	18,4	μ_Y
Dev(Y|X=N)	297,83	le tre componenti della devianza interna; $(\bar{y}_N - \mu_Y)^2 n_{Nord}$
Dev(Y|X=C)	214,00	
Dev(Y|X=S)	24,22	
Dev(B)	1751,33	devianza esterna (tra i gruppi o between)
Dev(TOT)	2287,38	Dev(Y)
Eta quadro	0,77	Dev(B)/Dev(Y)

NOTA:

$$0 \le \eta^2 \le 1$$

$0 \rightarrow$ Dev. esterna $= 0$

$1 \rightarrow$ Dev. est. $=$ Dev. Tot

ovvero: Dev. int. $= 0$

Posso fare la stessa analisi anche quando X è una variab. quantitativa

"teoricamente" potrei calcolare sia:

$\eta^2_{Y|X}$ che $\eta^2_{X|Y}$ (valori diversi)

logicamente le variabili giocano un ruolo diverso per cui, tranne casi particolari, si studia la relazione in un solo verso

inoltre, al fine di sfruttare la natura quantitativa di entrambe le variabili conviene sfruttare una tecnica diversa

REGRESSIONE