MEDIE

		definizione	distribuzioni di freq.
armonica	m-1	$\dfrac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$	$\dfrac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k}\frac{1}{x_i}n_i}$
geometrica	m0	$\sqrt[n]{\prod_{i=1}^{n} x_i}$	$\sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}}$
aritmetica	m1	$\dfrac{\sum_{i=1}^{n} x_i}{n}$	$\dfrac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$
quadratica	m2	$\sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n}}$	$\sqrt{\dfrac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i}}$

nb :
la media geometrica
risulta più semplice
passando ai logaritmi

$$\log m_0 = \frac{\sum_{i=1}^{n} \log x_i}{n} \qquad \log m_0 = \frac{\sum_{i=1}^{k} (\log x_i) n_i}{\sum_{i=1}^{k} n_i}$$

nb : $m_{-1} \le m_0 \le m_1 \le m_2$

INDICI DI VARIABILITÀ

		definizione	distribuzioni di freq.	proprietà
scostamento semplice medio dalla media aritmetica	S_{m1}	$\frac{\sum_{i=1}^{n}\lvert x_i - m_1\rvert}{n}$	$\frac{\sum_{i=1}^{k}\lvert x_i - m_1\rvert n_i}{\sum_{i=1}^{k} n_i}$	$\sum_{i=1}^{n}(x_i - m_1) = 0$
scostamento semplice medio dalla mediana	S_{Me}	$\frac{\sum_{i=1}^{n}\lvert x_i - Me\rvert}{n}$	$\frac{\sum_{i=1}^{k}\lvert x_i - Me\rvert n_i}{\sum_{i=1}^{k} n_i}$	$\sum_{i=1}^{n}\lvert x_i - Me\rvert = \text{minimo}$
scostamento quadratico medio	σ	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - m_1)^2}{n}}$	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - m_1)^2 n_i}{\sum_{i=1}^{k} n_i}}$	$\sum_{i=1}^{n}(x_i - m_1)^2 = \text{minimo}$

centri	$\sqrt[r]{\sum_{i=1}^{n}\lvert x_i - C_r\rvert^r} = \text{minimo}$ r = 1 ⇨ C_1 = Me r = 2 ⇨ C_2 = m_1	$\sum_{i=1}^{k}\lvert x_i - C_0\rvert^0 n_i = \text{minimo}$ quando C_0 è X corrispondente al n_{max} (= moda, r = 0)

nb : $S_{Me} \le S_{m_1} \le \sigma$

nb : varianza $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - m_1)^2}{n}$ nb : devianza $n\sigma^2 = \sum_{i=1}^{n}(x_i - m_1)^2$

proprietà della varianza

$$\begin{cases} \sigma^2 = M(x - M(x))^2 \quad \text{oppure} \quad M(x^2) - (M(x))^2 \quad \text{con M valor medio} \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 - (A - m_1)^2 \\ \sigma^2 = m_2^2 - m_1^2 \qquad \text{se A = 0} \end{cases}$$

differenza media semplice	$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert x_i - x_j\rvert}{n(n-1)}$	differenza quadratica semplice	${}_2\Delta = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n(n-1)}}$
differenza media con ripetizione	$\Delta_r = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert x_i - x_j\rvert}{n^2}$	differenza quadratica con ripetizione	${}_2\Delta_r = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n^2}}$

nb : $({}_2\Delta_r)^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n^2} = \frac{1}{n^2}\left(2\sum x_i^2 - 2\left(\sum x_i\right)^2\right) = 2\cdot\left(\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2\right) = 2\sigma^2$

indice di concentrazione del Gini: $G = \sum_{i=1}^{n-1}(P_i - Q_i) = \frac{n-1}{2} - \sum_{i=1}^{n-1} Q_i$

rapporto di concentrazione del Gini: $g = \frac{\sum_{i=1}^{n-1}(P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} = 1 - \frac{2}{n-1}\sum_{i=1}^{n-1} Q_i$

dove :
- *Qi è il rapporto fra l'ammontare del carattere posseduto dalle i unità più povere e l'ammontare complessivo del carattere*
- *Pi è il rapporto delle i untà più povere sul totale delle unità*

considerando la curva diconcetrazione, nasce un'altro rapporto di concentrazione del Gini:

$$R = \frac{\frac{1}{2} - \frac{1}{2}\sum_{i=0}^{n-1}(P_{i+1} - P_i)(Q_{i+1} - Q_i)}{\frac{1}{2}} = 1 - \sum_{i=0}^{n-1}(P_{i+1} - P_i)(Q_{i+1} - Q_i)$$

dove si considera il triangolo di base e altezza unitarie come termine di confronto rispetto l'area che delimitata fra la spezzatta di concentrazione e il segmento di equidistribuzione (v. somma di trapezi)

MOMENTI

		definizione	distribuzioni di freq.
momento di origine A e ordine r se A = 0 momenti dall'orignine se A = m_1 momenti centrali m	${}_A m_r$	$\frac{\sum_{i=1}^{k}(x_i - A)^r}{n}$	$\frac{\sum_{i=1}^{k}(x_i - A)^r n_i}{\sum_{i=1}^{k} n_i}$

r	A = 0	A = m_1	
1	$\frac{\sum_{i=1}^{n} x_i}{n} = m_1$	$\sum_{i=1}^{n}(x_i - m_1) = 0$	primo momento
2	$\frac{\sum_{i=1}^{n} x_i^{\,2}}{n} = m_2^2$	$\frac{\sum_{i=1}^{n}(x_i - m_1)^2}{n} = \sigma^2$	secondo momento
3	$\frac{\sum_{i=1}^{n} x_i^{\,3}}{n} = m_3^3$	$\frac{\sum_{i=1}^{n}(x_i - m_1)^3}{n}$	terzo momento
4	$\frac{\sum_{i=1}^{n} x_i^{\,4}}{n} = m_4^4$	$\frac{\sum_{i=1}^{n}(x_i - m_1)^4}{n} = curtosi$	quarto momento nb : nella distrib. di Gauss $\overline{m}_4 = 3\sigma^4$
r	$\sqrt[r]{\frac{\sum_{i=1}^{n} x_i^{\,4}}{n}} = m_r$	detta anche media potenziata di ordine r	

$$\overline{m}_r = \frac{\sum_{i=1}^{n}(x_i - m_1)^r}{n} = \frac{\sum_{i=1}^{n}\sum_{s=0}^{r}\binom{r}{s} x_i^{\,r-s}(-m_1)^s}{n} = \sum_{s=0}^{r}\binom{r}{s} m_{r-s}(-m_1)^s$$

momento dall'origine

r = 2 $\quad \overline{m}_2 = m_2^2 - m_1^2 \ (= \sigma^2)$ ⇦ *medie*

$\quad m_2 - m_1^2$ ⇦ *momenti*

r = 3 $\quad \overline{m}_3 = m_3^3 - 3m_2^2 m_1 + 2m_1^3$ ⇦ *medie*

$\quad m_3 - 3m_2 m_1 + 2m_1^3$ ⇦ *momenti*

TRASFORMAZIONI LINEARI

$y = a + bx \Rightarrow M(y) = M(a + bx) = M(a) + M(bx) = a + bM(x)$

$\sigma^2(y) = M(y - m(y))^2 = b^2 M(x - M(x))^2 = b^2\sigma^2(x) \Rightarrow$

dato che $\sigma^2 = \overline{m}_2$ sarà $\overline{m}_2(y) = b^2\,\overline{m}_2(x) \Rightarrow \overline{m}_r(y) = b^r\,\overline{m}_r(x)$

$z = a + bx + cy \Rightarrow M(z) = a + bM(x) + cM(y)$

$\sigma^2(z) = M(a + bx + cy - a - bM(x) - cM(x))^2 = M\left[b\left(x - M(x)\right) + c\left(y - M(y)\right)\right]^2 =$

$= b^2 M(x - M(x))^2 + c^2 M(y - M(y))^2 + 2bcM(x - M(x))(y - M(y)) = b^2\sigma^2(x) + c^2\sigma^2(y) + 2bc\,\mathrm{cov}(xy)$

$y = a + bx_1 + cx_2 + ... + nx_n \Rightarrow$ *posti* $a = 0,\ b = c = ... = n = 1$ *si ha* $y = x_1 + x_2 + ... + x_n$

$M(y) = M(x_1) + M(x_2) + ... + M(x_n)$

$\sigma^2(y) = \sigma^2(x_1) + \sigma^2(x_1) + ... + \sigma^2(x_1) + \mathrm{cov}(x_1x_2) + \mathrm{cov}(x_1x_3) + ... + \mathrm{cov}(x_ix_j) + ... + \mathrm{cov}(x_{n-1}x_n)$

nb : nella variabile binomiale le covarianze sono tutte nulle perchè le variabili sono stocasticamente indipendenti

INDICI DI VARIABILITÀ (continua)

INDICI DI VARIABILITÀ (sono >0)	$\frac{S_{m_1}}{m_1}$	$\frac{S_{Me}}{Me}$	$\frac{\sigma}{m_1}$	$\frac{\sigma^2}{m_1^{\ 2}}$	
INDICI NORMALIZZATI (sono 0<I>1)	$\frac{S_{m_1}}{S_{m_1\max}}$	$\frac{S_{Me}}{S_{Me\max}}$	$\frac{\sigma}{\sigma_{\max}}$	$\frac{\sigma^2}{\sigma_{\max}^{\ 2}}$	$\frac{\Delta}{\Delta_{\max}}$

nb : *dicesi coefficiente di variazione :* $\frac{\sigma}{m_1}\cdot 100$

nb : *dicesi indice normalizzato, in generale :* $I_{rel} = \frac{I_{effettivo} - I_{\min}}{I_{\max} - I_{\min}}$

nb : *nelle distibuzioni massimanti, cioè dove il fenomeno è concentrato in un solo elemento, si ha :*

$$\begin{cases} S_{m_1\max} = \dfrac{2m_1(n-1)}{n} \\ \sigma_{\max}^{\ 2} = m_1^{\ 2}(n-1) \\ \sigma_{\max} = m_1\sqrt{(n-1)} \end{cases}$$

INDICI ADIMENSIONALI DI ASIMMETRIA

NB :	$m_3 > 0$	simm dx o positiva	moda < Me < m_1
	$m_3 = 0$	simm	coindidono tutte
	$m_3 < 0$	simm sx o negativa	m_1 < Me < moda

(Pearson)	$\frac{moda - m_1}{\sigma}$	simm dx	> 0
		simm	$= 0$
		simm sx	< 0
Pearson	$\beta_1 = \frac{(\overline{m}_3)^2}{(\sigma^2)^3}$	simm	$= 0$
		no simm	> 0
Fisher	$\gamma_1 = \frac{\overline{m}_3}{\sigma^3}$	simm dx	> 0
		simm	$= 0$
		simm sx	< 0
Pearson	$\beta_2 = \frac{\overline{m}_4}{\sigma^4}$	$= 3$	distrib. normale
		> 3	distrib. leptocurtica (curtosi ipernormale)
		< 3	distrib. platicurtica (curtosi iponormale)
Fisher	$\gamma_2 = \beta_2 - 3$	$= 0$	distrib. normale
		> 0	distrib. leptocurtica (curtosi ipernormale)
		< 0	distrib. platicurtica (curtosi iponormale)

VARIABILI DOPPIE

covarianza

$$\frac{\sum_{i=1}^{n}(x_i - m_1(x))(y_i - m_1(y))}{n} = M(xy) - M(x)M(y)$$

se > 0 relazione diretta ↗
se < 0 relazione indiretta ↘
se = 0 indipendenza lineare (random o U)

retta di regressione

$y = f(x)$ valori empirici
$\hat{y} = a + bx$ valori teorici

condizioni :
1) $\sum y_i = \sum \hat{y}_i$
2) $\sum (y_i - y_i)^2 =$ minimo

sviluppo seconda condizione

$g(a,b) = \sum (y_i - a - bx_i)^2$

derivate parziali poste a sistema :

$$\begin{cases} -2\sum (y_i - a - bx_i) = 0 \\ -2\sum (y_i - a - bx_i)x_i = 0 \end{cases} \Rightarrow \begin{cases} \sum y_i = na + b\sum x_i \\ \sum x_i y_i = a\sum x_i + b\sum x_i^2 \end{cases}$$

da cui :

$$b = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{M(xy) - M(x)M(y)}{m_2^2 - m_1^2} = \frac{cov(x,y)}{\sigma^2(x)}$$

$$a = \frac{\sum y_i}{n} - b\frac{\sum x_i}{n} = M(y) - bM(x)$$

NB :
tutto ciò se y dipendente, cioè con x indipendente

NB :
se cov(x,y)=0, allora entrambe le rette di regressione (y=a+bx, x=a+by) sono parallele agli assi del sistema, allora c'è indipendenza fra le due variabili

coefficiente di correlazione lineare

$$\rho = \sqrt{b_{y/x} \cdot b_{x/y}} = \frac{cov(x,y)}{\sigma(x) \cdot \sigma(y)}$$

se = 0 indipendenti
se = 1 dipendenti positiviamente
se = -1 dipendenti negativamente

NB : relazione perfetta se
$\hat{y}_i - y_i \ \forall i$
$\sum (y_i - \hat{y}_i)^2 = 0$

varianza totale

$$\sigma^2(y) = \frac{\sum (y_i - M(y))^2}{n} = \frac{\sum (y_i - \hat{y}_i)^2}{n} + \frac{\sum (\hat{y}_i - M(y))^2}{n}$$

varianza totale = var. residua + var. spiegata
varianza totale = var. *within* + var. *between*

coefficiente di determinazione

$$\rho^2 = 1 - \frac{varianza\ residua}{varianza\ totale} = \frac{varianza\ spiegata}{varianza\ totale}$$

se = 0 $\sigma^2(y/x) = \sigma(y)$, cioè $\hat{y} = M(y)$, x ininfluente
se = 1 $\sigma^2(y/x) = 0$, cioè $\hat{y}_i = y_i \ \forall i$, relazione perfetta

varianza residua

$$\sigma^2(y/x) = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

sviluppando e sostituendo poi a e b, si ottiene :

$$\sigma^2(y/x) = \sigma^2(y) + b^2\sigma^2(x) - 2b\,cov(x,y) = \sigma^2(y) - \frac{cov^2(x,y)}{\sigma^2(x)}$$

da cui si ha :

$$1 - \frac{\sigma^2(y/x)}{\sigma^2(y)} = \frac{cov^2(x,y)}{\sigma^2(x) \cdot \sigma^2(x)} = \rho^2$$

che è il coefficiente di determinazione ;

inoltre : $\sigma^2(y/x) = \sigma^2(y) - \rho^2\sigma^2(y)$ quindi la varianza spiegata è : $\rho^2\sigma^2(y) = b^2\sigma^2(x) = \frac{cov^2(x,y)}{\sigma^2(x)}$

per ridurre la varianza residua :

$\hat{y} = a + bx_i + cx_i^2 \qquad g(a,b,c) = \sum (y_i - a - bx_i - cx_i^2)^2$

derivate parziali poste a sistema :

$$\begin{cases} -2\sum (y_i - a - bx_i - cx_i^2) = 0 \\ -2\sum (y_i - a - bx_i - cx_i^2)x_i = 0 \\ -2\sum (y_i - a - bx_i - cx_i^2)x_i^2 = 0 \end{cases} \Rightarrow \begin{cases} \sum y_i = na + b\sum x_i + c\sum x_i^2 \\ \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \\ \sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \end{cases}$$

retta di regressione, quindi di II grado, allora coefficiente di determinazione ρ^2 (diventa R^2) di II grado, perché varianza residua di II grado ; se la rettà di regressione è di III grado, lo sarà anche il coefficiente di determinazione R^2 e la varianza residua, e così via.

nb : $\rho^2 \le R_{II}^2 \le R_{III}^2$

NELLA TAVOLA A DOPPIA ENTRATA …

covarianza

$$\frac{\sum_{i=1}^{r} \sum_{j=1}^{c} (x_j - m_1(x))(y_i - m_1(y))n_{ij}}{N}$$

medie condizionate

$$\frac{\sum_{i=1}^{r} y_i n_{ij}}{n_{\bullet j}} = m(y/j)$$

$$\frac{\sum_{j=1}^{c} x_j n_{ij}}{n_{i\bullet}} = m(x/i)$$

media delle medie condizionate = media marginale

$$m(y) = \frac{\sum_{j=1}^{c} m(y/j)\, n_{\bullet j}}{\sum_{j=1}^{c} n_{\bullet j}} = \frac{\sum_{j=1}^{c} \sum_{i=1}^{r} y_i n_{ij}}{n} = \frac{\sum_{i=1}^{r} y_i n_{i\bullet}}{n}$$

$$m(x) = \frac{\sum_{i=1}^{r} m(x/i)\, n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}} = \frac{\sum_{j=1}^{c} \sum_{i=1}^{r} x_j n_{ij}}{n} = \frac{\sum_{j=1}^{c} x_j n_{\bullet j}}{n}$$

varianza totale

$$\sigma^2(y) = \frac{\sum_{i=1}^{r} (y_i - M(y))^2 n_{i\bullet}}{n} = \frac{\sum_{j=1}^{c} \sum_{i=1}^{r} (y_i - \hat{y}_i)^2 n_{ij}}{n} + \frac{\sum_{j=1}^{c} (\hat{y}_i - M(y))^2 n_{i\bullet}}{n}$$

varianza totale = varianza residua + varianza spiegata
varianza totale = varianza *within* + varianza *between*

coefficiente di determinazione

$$\eta^2 = 1 - \frac{\text{varianza residua}}{\text{varianza totale}} = \frac{\text{varianza spiegata}}{\text{varianza totale}}$$

NB :
la varianza residua che interessa η^2 è minore di quella che interessa ρ^2, quindi sarà $\eta^2 \ge \rho^2$;
se $\eta^2 = \rho^2$ allora le medie condizionate sono tutte allineate sulla stessa retta