

FORMULARIO
 Corso di Statistica – CLAMM
 Prof. Patrizia Agati

1. Notazione

- j generica unità statistica osservata
- n numero di unità statistiche osservate
- i generica classe
- k numero di classi
- g generico gruppo
- G numero di gruppi
- n_i frequenza assoluta della classe i -esima
- N_i frequenza assoluta cumulata della classe i -esima
- f_i frequenza relativa della classe i -esima
- F_i frequenza relativa cumulata della classe i -esima
- w_i ampiezza della classe i -esima
- h_i densità di frequenza della classe i -esima
- $x_{i-1} - x_i$ generica modalità intervallare del carattere X
- \hat{x}_i valore centrale della classe $x_{i-1} - x_i$
- $x_h - x_{h-1}$ classe mediana

2. Valori medi e misure di variabilità

	Protocollo elementare	Distribuzione di frequenza	Distrib. freq. per modalità interv.
Media aritmetica $M(X) = \bar{x}$	$\frac{\sum_{j=1}^n x_j}{n}$	$\frac{\sum_{i=1}^k x_i n_i}{n} = \sum_{i=1}^k x_i f_i$	$\frac{\sum_{i=1}^k \hat{x}_i n_i}{n} = \sum_{i=1}^k \hat{x}_i f_i$
	$\bar{x} = \frac{\sum_{g=1}^G \bar{x}_g n_g}{n}$; Se $y = a + bx$, allora $\bar{y} = a + b\bar{x}$		
Media geometrica	$\sqrt[n]{\prod_{j=1}^n x_j}$		
Mediana			$x_{h-1} + \frac{\frac{n}{2} - N_{h-1}}{n_h} (x_h - x_{h-1})$

	Protocollo elementare	Distribuzione di frequenza	Distrib. freq. per modalità interv.
Range	$x_{max} - x_{min}$		
Devianza Dev(X)	$\sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - n\bar{x}^2$	$\sum_{i=1}^k (x_i - \bar{x})^2 n_i = \sum_{i=1}^k x_i^2 n_i - n\bar{x}^2$	$\sum_{i=1}^k (\hat{x}_i - \bar{x})^2 n_i = \sum_{i=1}^k \hat{x}_i^2 n_i - n\bar{x}^2$
Varianza $V(X) = \frac{Dev(X)}{n} = s^2$	$\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 - \bar{x}^2$	$\begin{aligned} \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i &= \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - \bar{x}^2 \\ &= \sum_{i=1}^k (x_i - \bar{x})^2 f_i \\ &= \sum_{i=1}^k x_i^2 f_i - \bar{x}^2 \end{aligned}$	$\begin{aligned} \frac{1}{n} \sum_{i=1}^k (\hat{x}_i - \bar{x})^2 n_i &= \frac{1}{n} \sum_{i=1}^k \hat{x}_i^2 n_i - \bar{x}^2 \\ &= \sum_{i=1}^k (\hat{x}_i - \bar{x})^2 f_i \\ &= \sum_{i=1}^k \hat{x}_i^2 f_i - \bar{x}^2 \end{aligned}$
Coeff. di variazione CV(X)	$\frac{s}{\bar{x}}$		
Indice di concentr. R di Gini	$\frac{\sum_{j=1}^{n-1} (p_j - q_j)}{\sum_{j=1}^{n-1} p_j} = 1 - \frac{2}{n-1} \sum_{j=1}^{n-1} q_j$		

3. Numeri indici

Numeri indici semplici a base fissa: ${}_b I_t = q_t / q_b$

Numeri indici semplici a base mobile: ${}_{t-1} I_t = q_t / q_{t-1}$

Passaggio da base fissa a base mobile: ${}_{t-1} I_t = {}_b I_t / {}_b I_{t-1}$

Passaggio da base mobile a fissa: ${}_b I_t = \prod_{h=b+1}^t {}_{h-1} I_h$ per $t > b$

${}_b I_t = (\prod_{h=t+1}^b {}_{h-1} I_h)^{-1}$ per $t < b$

Cambiamento di base da b a b' : ${}_{b'} I_t = {}_b I_t / {}_b I_{b'}$

Numeri indici complessi	Laspeyres	Paasche	Fisher
Prezzi	${}_0 I_s^{L,p} = \frac{\sum_{j=1}^n p_{sj} q_{0j}}{\sum_{j=1}^n p_{0j} q_{0j}}$	${}_0 I_s^{P,p} = \frac{\sum_{j=1}^n p_{sj} q_{sj}}{\sum_{j=1}^n p_{0j} q_{sj}}$	${}_0 I_s^{F,p} = \sqrt{{}_0 I_s^{L,p} {}_0 I_s^{P,p}}$
Quantità	${}_0 I_s^{L,q} = \frac{\sum_{j=1}^n p_{0j} q_{sj}}{\sum_{j=1}^n p_{0j} q_{0j}}$	${}_0 I_s^{P,q} = \frac{\sum_{j=1}^n p_{sj} q_{sj}}{\sum_{j=1}^n p_{sj} q_{0j}}$	${}_0 I_s^{F,q} = \sqrt{{}_0 I_s^{L,q} {}_0 I_s^{P,q}}$

4. Associazione tra caratteri

Chi-quadrato:

$$\chi^2 = \sum_{i=1}^u \sum_{h=1}^v \frac{(n_{ih} - n_{ih}^*)^2}{n_{ih}^*}$$

Tchuprov:

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(u-1)(v-1)}}}$$

Cramer:

$$C = \sqrt{\frac{\chi^2/n}{\min[(u-1)(v-1)]}}$$

Codevianza:

$$\text{Cod}(X, Y) = \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) = \sum_{j=1}^n x_j y_j - n \bar{x} \bar{y}$$

Covarianza:

$$\text{Cov}(X, Y) = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{n} = \frac{1}{n} \sum_{j=1}^n x_j y_j - \bar{x} \bar{y} = s_{xy}$$

Coefficiente di regressione lineare:

$$b = \frac{\text{Cod}(X, Y)}{\text{Dev}(X)} = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x}$$

Intercetta:

$$a = \bar{y} - b \bar{x}$$

Coefficiente di correlazione lineare:

$$r = \frac{\text{Cod}(X, Y)}{\sqrt{\text{Dev}(X) \text{Dev}(Y)}} = \frac{s_{xy}}{s_x s_y}$$

Devianza di regressione:

$$\text{Dev}(Y)_{\text{regr}} = \sum_{j=1}^n (y_j^* - \bar{y})^2 = b^2 \text{Dev}(X)$$

Devianza di dispersione:

$$\text{Dev}(Y)_{\text{disp}} = \text{Dev}(Y) - \text{Dev}(Y)_{\text{regr}}$$

Indice di determinazione lineare:

$$R^2 = \frac{\text{Dev}(Y)_{\text{regr}}}{\text{Dev}(Y)} = 1 - \frac{\text{Dev}(Y)_{\text{disp}}}{\text{Dev}(Y)} = r^2$$

5. Inferenza statistica induttiva: statistiche-test e intervalli di confidenza

H ₀	n	Statistica-test	Valore teorico	Interv. di confid.
$\mu = \mu_0$	< 30	$\frac{ \bar{x} - \mu_0 }{s/\sqrt{n}}$, dove $s = \sqrt{\frac{Dev(X)}{n-1}}$	t_{n-1}	$\bar{x} \pm t_{\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}$
$\mu = \mu_0$	> 30	$\frac{ \bar{x} - \mu_0 }{s/\sqrt{n}}$, dove $s = \sqrt{\frac{Dev(X)}{n-1}}$	Z	$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
$p = p_0$	> 30	$\frac{ f - p_0 }{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Z	$f \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}$
$\mu_1 = \mu_2 = \mu$	< 30	$\frac{ \bar{x}_1 - \bar{x}_2 }{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, dove $s_p = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}$	$t_{n_1+n_2-2}$	
$\mu_1 = \mu_2 = \mu$	> 30	$\frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Z	
$p_1 = p_2 = p$	> 30	$\frac{ f_1 - f_2 }{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, dove $f = \frac{f_1 n_1 + f_2 n_2}{n}$	Z	
$\sigma_1^2 = \sigma_2^2 = \sigma^2$	qualsiasi	$\frac{s_1^2}{s_2^2}$, con $s_1^2 > s_2^2$	F_{n_1-1, n_2-1}	

CALCOLO DELLE PROBABILITÀ

1. CALCOLO COMBINATORIO

Permutazioni di n elementi distinti : $P_n = n!$

Disposizioni semplici di n elementi distinti considerati k alla volta: $D_{n,k} = \frac{n!}{(n-k)!}$

Disposizioni con ripetizione di n elementi distinti considerati k alla volta: ${}_R D_{n,k} = n^k$

Combinazioni semplici di n elementi distinti considerati k alla volta: $C_{n,k} = \binom{n}{k}$

Combinazioni con ripetizione di n elementi considerati k alla volta: ${}_R C_{n,k} = \binom{n+k-1}{k}$

2. OPERAZIONI SUGLI EVENTI

Complemento L'evento complementare \bar{E} (oppure $-E$) è vero quando l'evento E è falso, e viceversa

Unione L'evento unione $A \cup B$ è vero quando *almeno uno* dei due eventi A e B si avvera.

Intersezione L'evento intersezione $A \cap B$ è vero quando *entrambi* gli eventi A e B si avverano.

Differenza L'evento differenza $A \setminus B$ (oppure $A - B$) è vero quando A si avvera e B no.

PROPRIETÀ	UNIONE	INTERSEZIONE	DIFFERENZA
Idempotenza	$E \cup E = E$	$E \cap E = E$	_____
Commutativa	$E \cup F = F \cup E$	$E \cap F = F \cap E$	_____
Associativa	$(E \cup F) \cup G = E \cup (F \cup G)$	$(E \cap F) \cap G = E \cap (F \cap G)$	_____
Distributiva	$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$		_____
Elemento neutro	\emptyset	Ω	\emptyset
Leggi di De Morgan	$\overline{E \cup F} = \bar{E} \cap \bar{F}; \quad \overline{E \cap F} = \bar{E} \cup \bar{F}$		

Relazione duale. Data una relazione tra eventi, si chiama *relazione duale* quella che si ottiene scambiando:

\cup con \cap (e viceversa); \emptyset con Ω (e viceversa); \subset con \supset (e viceversa); E con \bar{E} (e viceversa).

Principio di dualità. Data una relazione che coinvolga operazioni con eventi, e che sia dimostrato vera, si dimostra che anche la relazione duale è vera.

3. ASSIOMI DI KOLMOGOROV

$P(E) \geq 0$; $P(\Omega) = 1$; se $E \cap F = \emptyset$ allora $P(E \cup F) = P(E) + P(F)$

4. TEOREMI	E, F compatibili ($E \cap F \neq \emptyset$)	E, F incompatibili ($E \cap F = \emptyset$)
$P(E \cup F) =$	$= P(E) + P(F) - P(E \cap F)$	$= P(E) + P(F)$
$P(E F) =$	$= P(E) - P(E \cap F)$	$= P(E)$
$P(E \cap F) =$	$= P(E) P(F E) = P(F) P(E F)$ $= P(E) P(F)$	$= 0$

5. **PROBABILITÀ CONDIZIONATA:** $P(F|E) = P(F \cap E) / P(E)$ **TEOREMA DI BAYES:** $P(H_j|E) = \frac{P(H_j)P(E|H_j)}{\sum_{i=1}^k P(H_i)P(E|H_i)}$

6. VARIABILI ALEATORIE DISCRETE

Funzione di ripartizione: $F(x_0) = P(X \leq x_0)$

Funzione di probabilità: $f(x) = P(X = x)$

Valore atteso: $E(X) = \sum_x x \cdot P(X = x)$

Varianza: $V(X) = E [X - E(X)]^2$

Proprietà lineari: $E(aX + b) = a \cdot E(X) + b$; $V(aX + b) = a^2 \cdot V(X)$

MODELLI DISCRETI (funzione di probabilità, valore atteso e varianza)

V.A. di Bernoulli: $P(X=x) = p^x \cdot (1-p)^{1-x}$, $x = 0, 1$

$E(X) = p$

$V(X) = p(1-p)$

V.A. binomiale: $P(X=x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$, $x = 0, 1, \dots, n$

$E(X) = n \cdot p$

$V(X) = n \cdot p \cdot (1-p)$

V.A. di Poisson: $P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, 2, \dots$

$E(X) = \lambda$

$V(X) = \lambda$

7. MODELLO NORMALE

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in R \quad E(Y) = \mu; \quad V(Y) = \sigma^2$$

Proprietà: 1. se $Y \sim N(\mu, \sigma^2)$, allora $Y^* = aY + b \sim N(a\mu + b, a^2\sigma^2)$

2. se $Y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$ mutuamente indipendenti, allora $Y^* = \sum a_i Y_i \sim N(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2)$

TEOREMA CENTRALE DEL LIMITE:

se $Y_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n, \dots$ sono v.a. iid,

allora, al divergere di n , la variabile somma standardizzata Z_n converge alla v.a. $N(0, 1)$