Compendio di Teoria

Patrizio Frederic

2022-05-27

Introduzione

Il seguente materiale non è sostitutivo all'ascolto delle lezioni, qui elenco solo i punti sui quali potrò fare domande. Le domande all'esame potranno essere *dirette*, per esempio:

> Elencare le proprietà della frequenza relativa, f_j.
>
> Risposta
>
> $0 \leq f_j \leq 1, \forall j = 1, ..., K$,
>
> $\sum_{j=1}^{K} f_j = 1$.

oppure *indirette*

> Se h è uno stimatore per θ tale che $E(h) = \theta$ di quale proprietà gode h?
>
> Risposta
>
> Se h è uno stimatore tale che $E(h) = \theta$, allora h è corretto per θ

Questo materiale è diviso in due parti. Nella prima parte si riassumono i punti principali di teoria da tenere a mente. Nella seconda parte seguirà una carrellata di potenziali domande che **non** è esaustiva, ma solo esemplificativa.

RAPPRESENTAZIONI DELLA DISTRIBUZIONE UNITARIA

Quali sono i vantaggi del campionamento, rispetto al censimento?

- *Risparmio di costi* perché si deve intervistare un numero inferiore di unità statistiche sicché occorre un numero inferiore di intervistatori.
- *Risparmio di tempo* perché intervistando un numero inferiore di unità statistiche si impiega un tempo minore.
- *Maggiore quantità di informazioni* perché c'è più tempo da dedicare agli intervistati.
- *Maggiore qualità dell'informazione* perché si può dedicare più cura nella raccolta dei dati.
- *Controllo degli intervistatori* per verificare l'esecuzione corretta delle procedure.

Elencare le proprietà della frequenza relativa, f_j.

Le proprietà della frequenza relativa f_j sono:

- $0 \leq f_j \leq 1, \forall j = 1, ..., K$,
- $\sum_{j=1}^{K} f_j = 1$.

Elencare le proprietà della frequenza percentuale, $f_{\%,\,j}$.

Le proprietà della frequenza relativa (f_j) sono:

- $0 \leq f_{\%,\,j} \leq 100, \forall j = 1, ..., K$,
- $\sum_{j=1}^{K} f_{\%,\,j} = 100$.

Elencare le proprietà della frequenza assoluta, n_j.

Le proprietà della frequenza relativa (f_j) sono:

- $0 \leq n_j \leq n, \forall j = 1, ..., K$,
- $\sum_{j=1}^{K} n_j = n$.

Elencare le proprietà della media aritmetica con una descrizione sintetica.

La media aritmetica è data da

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

0. Internalità: $x_{\min} = x_{(1)} \leq \bar{x} \leq x_{(n)} = x_{\max}$
1. Invarianza della somma:
$$n\bar{x} = \sum_{i=1}^{n} x_i$$
2. Somma degli scarti dalla media è nulla: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

3. Minimizza la somma degli scarti al quadrato:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - d)^2 \quad \forall d \neq \bar{x}$$

4. Invarianza per trasformazioni lineari: se $y_i = a + bx_i$ allora $\bar{y} = a + b\bar{x}$
5. Associatività. Sia una popolazione, $\mathcal{P}$, formata da K gruppi con medie e numerosità: $(\bar{x}_1;\ n_1)$, $(\bar{x}_2;\ n_2)$, ..., $(\bar{x}_K;\ n_K)$. Allora, la media totale $\bar{x}_T$ di $\mathcal{P} =$ è data da

$$\bar{x}_T = \frac{\text{Tot}\{\mathcal{P}_1\} + \cdots + \text{Tot}\{\mathcal{P}_K\}}{n_1 + \cdots + n_K} = \frac{n_1\ \bar{x}_1 + \cdots + n_K\ \bar{x}_K}{n_1 + \cdots + n_K}$$

Definire la mediana, $x_{0.5}$, e elencare le sue proprietà, specificando anche la sua relazione con la media.

La mediana di una distribuzione, $x_{0.5}$, è quel valore della per X il quale si ha $F(x_{0.5}) = 0.5$. Le proprietà della mediana ($x_{0.5}$) sono:

- $x_{\text{min}} \leq x_{0.5} \leq x_{\text{max}}$,
- $\sum_{j=1}^{n} |x_i - x_{0.5}|$ è un minimo.
- Relazione Media-Mediana:
- Distribuzione simmetrica $\rightarrow x_{0.5} = \bar{x}$
- Distribuzione con coda lunga a destra $\rightarrow x_{0.5} < \bar{x}$
- Distribuzione con coda lunga a sinistra $\rightarrow x_{0.5} > \bar{x}$

Elencare le proprietà della varianza (σ^2).

Le proprietà della varianza (σ^2) sono:

- $\sigma^2(X) \geq 0$
- $\sigma^2(\text{costante}) = 0$
- se $y_i = a + bx_i$ allora allora $\sigma_Y^2 = b^2\sigma_X^2$

Elencare le proprietà della deviazione standard (σ).

Le proprietà della deviazione standard (σ) sono:

- $\sigma \geq 0$
- $\sigma = 0$
- se $y_i = a + bx_i$ allora allora $\sigma_Y = |b|\sigma_X$

Si definisca il percentile p-esimo

Il p-esimo percentile, x_p, è quel valore tale che $F(x_p) = p$.

PROBABILITÀ

Sia dato uno spazio di probabilità Ω. Sia A un evento in Ω. Indicare le proprietà della funzione di probabilità, $P(A)$.

- $0 \leq P(A) \leq 1$.
- $P(\emptyset) = 0 \qquad , \qquad P(\Omega) = 1$.

Enunciare il postulato empirico del caso.

In un gruppo di prove ripetute più volte *nelle stesse condizioni*, ciascuno degli eventi possibili si presenta con una frequenza relativa che tende alla probabilità all'aumentare del numero di prove; ossia

$$P(A) = \frac{n_A}{n} + \epsilon_n \quad \text{dove} \quad \epsilon_n \to 0 \quad \text{per} \quad n \to \infty.$$

Elencare i postulati del calcolo delle probabilità.

Sia dato un spazio di probabilità $\{\Omega, \mathscr{A}, P\}$

- $P(A) \geq 0 \qquad \forall A \in \mathscr{A}$,
- $P(\Omega) = 1$,
- $\forall A, B \in \mathscr{A} : A \cap B = \emptyset$ allora $P(A \cup B) = P(A) + P(B)$

Definire le principali proprietà della probabilità

- $0 \leq P(A) \leq 1 \;\; \forall A$,
- $P(\emptyset) = 0$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\overline{A}) = 1 - P(A)$

Quando due eventi si dicono incompatibili?

Due eventi A e B si dicono incompatibili se e solo se

$$A \cap B = \emptyset$$

Definire la probabilità condiziona di A dato B

la probabilità condizionata di A dato B è data da

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definire il concetto di indipendenza tra due eventi

Due eventi A e B si dicono indipendenti se

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

ne consegue che se A e B sono indipendenti, allora

$$P(A \cap B) = P(A)P(B)$$

Scrivere la *chain rule* (o regola del prodotto) per due eventi

Dati due eventi A e B la *chain rule* è

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

---

scrivere la *chain rule* (o regola del prodotto) per tre eventi

Dati tre eventi A, B e C la *chain rule* è

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Enunciare il teorema delle probabilità totali per due eventi

Dati due eventi A e B il teorema delle probabilità totali afferma che

$$P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B})$$

VARIABILI CASUALI

Definire in generale il supporto di una variabile casuale X

- Il supporto S_X è l'insieme dei possibili valori che può assumere la VC X.

Definire in generale il valore atteso di una variabile casuale X

In generale il valore atteso è dato da

$$E(X) = \sum_{\forall x \in S_X} xP(X = x)$$

Elencare le proprietà del valore atteso, $E(X)$.

Le proprietà del valore atteso, $E(X)$ sono:

- $x_{\min} \leq E(X) \leq x_{\max}, \quad x_{\min},\ x_{\max} \in S_X,$
- $E\Big(X - E(X)\Big) = 0,$
- $E\Big(X - E(X)\Big)^2 < E(X - d)^2 \quad \forall\ d \neq E(X),$
- $E(a + bX) = a + b\ E(X).$

Definire in generale la varianza di una variabile casuale X

La varianza è data da

$$V(X) = E\Big(\big(X - E(X)\big)^2\Big) = \sum_{\forall x \in S_X} (x - E(x))^2 P(X = x)$$

Elencare le proprietà della varianza, $V(X)$.

Le proprietà della varianza, $V(X)$, sono:

- $V(X) \geq 0,$
- $V(X) = 0$ se e solo se $P(X = x) = 1$
- Se X e Y sono **indipendenti**, allora

$$V(aX + bY) = V(aX - bY) = a^2V(X) + b^2V(Y), \forall\ a, b \in \mathbb{R}$$

nota: se $a = 1$ e $b = 1$ allora

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

Elencare le proprietà della funzione di probabilità, $f(x)$.

Le proprietà della funzione di probabilità, $f(x)$ sono:

- $0 \leq f(x) \leq 1, \forall x \in S_X,$
- $\sum_{\forall\, x \in S_X} f(x) = 1.$

Elencare le proprietà della funzione di ripartizione

La funzione di ripartizione F di una VC X è, per definizione:

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

F gode delle seguenti proprietà:

- Non decrescente, ossia $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
- $\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1.$
- Continua a destra, ossia $\lim_{x\to x_0^+} F(x) = F(x_0)$.
- $P(a < X \leq b) = F(b) - F(a)$.

Definire la variabile casuale X di Bernoulli, riportando anche la funzione di probabilità e *tutte* le sue proprietà.

Sia $X \sim \text{Ber}(\pi)$, allora X può assumere solo i valori 1 in caso di successo e 0 in caso di insuccesso.

$$\begin{aligned} S_X &= \{0,\, 1\} \\ f(x) &= \pi^x\ (1-\pi)^{1-x} \\ \Theta &= \pi \in [0,\, 1] \subset \mathbb{R}^+ \\ E(X) &= \pi \\ V(X) &= \pi\,(1-\pi)\,. \end{aligned}$$

Definire la variabile casuale Binomiale, riportando *tutte* le sue proprietà.

La variabile binomiale conta il numero di successi in n esperimenti e può essere interpretata come al somma di n Bernoulli IID di parametro π. La funzione di probabilità, il valore atteso e la varianza sono:

$$\begin{aligned} S_X &= \{0,\, 1,\, 2,\, \ldots,\, n\} \\ f(x) &= \binom{n}{x}\ \pi^x\ (1-\pi)^{n-x} \\ \Theta &= \pi \in [0,\, 1] \subset \mathbb{R}^+ \quad , \quad n \in \{1, 2, 3, \ldots\} \\ E(X) &= n\ \pi \\ V(X) &= n\ \pi\ (1-\pi) \end{aligned}$$

Gode della proprietà riproduttiva: siano $X_1, \ldots, X_K$, n variabili casuali indipendenti distribuite secondo una $\text{Bin}(n_i,\, \pi)$, per $i = 1, \ldots, K$. Allora, la loro somma, $S_K = X_1 + \cdots + X_K$, si distribuisce secondo una $\text{Bin}(n_1 + ... + n_K, \pi)$.

Definire la variabile casuale X distribuita secondo una Poisson(λ), riportando *tutte* le sue proprietà.

Le seguenti risposte valgono in tutto 1.

$$
\begin{aligned}
S_X &= \{0,\,1,\,2,\,\ldots\} \\
P(X=x) &= \frac{\lambda^x\ e^{-\lambda}}{x!} \\
\Theta &= \lambda\ \in\ (0,\infty) \equiv \mathbb{R}^+ \\
E(X) &= \lambda \\
V(X) &= \lambda\,.
\end{aligned}
$$

Gode della proprietà riproduttiva: siano $X_1, \ldots, X_K$ K variabili casuali indipendenti distribuite secondo una Poisson(λ_k), per $k = 1, \ldots, K$. Allora, la loro somma, $S_K = X_1 + \cdots + X_K$, si distribuisce secondo una Poisson($\lambda_1 + \cdots + \lambda_K$).

Definire la variabile casuale X distribuita riportando *tutte* le sue proprietà.

Sia $X\ \sim\ N(\mu,\ \sigma^2)$ un variabile casuale, allora la distribuzione è simmetrica rispetto al parametro μ,

$-\infty < X < \infty$ oppure $x \in \mathbb{R}$

una descrizione verbale completa della forma

$\Theta\ :\ \mu\ \in\ (-\infty,\infty)\quad,\quad \sigma^2\ \in\ (0,\infty)$

$E(X) = \mu$

$V(X) = \sigma^2\,.$

Sia $X\ \sim\ N(\mu_X,\ \sigma^2_X)$ un variabile casuale normale, allora

- Ogni trasformazione lineare della X, sia $Y = a + bX$, si distribuisce secondo una normale: $Y\ \sim\ N(a + b\mu_X,\ b^2\sigma^2_X)$
- La somma di due variabili casuali normali, siano $X\ \sim\ N(\mu_X,\ \sigma^2_X)$ e $Y\ \sim\ N(\mu_Y,\ \sigma^2_Y)$ si distribuisce secondo una normale $X + Y\ \sim\ N(\mu_X + \mu_Y,\ \sigma^2_X + \sigma^2_Y)$.

Definire la variabile casuale chi-quadrata, χ^2,

Siano $Z_1, ..., Z_n$, n variabili casuali IID, tc $Z_i \sim N(0,1)$; allora, la variabile casuale chi-quadrata, χ^2, è definita come

$$\chi^2_n = \sum_{i=1}^{n} Z_i^2$$

- Supporto: χ^2_n assume valori tra 0 e ∞;ossia, $0 \leq \chi^2_n < \infty$.
- La funzione di densità ha una coda lunga a destra e dipende da n.
- $\Theta\ :\ n \in \{1,\ 2,\ \ldots,\ \}$; i gdl non sono oggetto di stima.
- $E(\chi^2_n) = n$
- $V(\chi^2_n) = 2\,n$.

Definire la variabile casuale t di Student

Sia $Z \sim N(0,1)$ e sia $Y \sim \chi^2_n$. Allora, la t di Student è data dal rapporto

- $T_n = \frac{Z}{\sqrt{Y/n}}$
- Supporto: t_n assume valori tra $-\infty$ e ∞; ossia, $-\infty < t_n < \infty$.
- La funzione di densità è simmetrica rispetto allo zero, le aree esterne sono più alte di quelle della normale, le aree interne sono più basse di quelle della normale, e l'andamento di ogni curva dipende da n.
- $\Theta : n \in \{1, 2, \dots, \}$; i gdl non sono oggetto di stima.
- $E(t_n) = 0$
- $V(t_n) = \frac{n}{n-2} \qquad \text{per } n \geq 3$

TEOREMA CENTRALE DEL LIMITE

Enunciare il Teorema centrale del limite per la Somma

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1, ..., n$. Posto

$$S_n = X_1 + ... + X_n,$$

allora

$$S_n \underset{a}{\sim} N\left(n\mu, n\sigma^2\right)$$

Enunciare il Teorema centrale del limite per la Media

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1, ..., n$. Posto

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\bar{X} \underset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Enunciare il Teorema centrale del limite per la Proporzione

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $X_i \sim \text{Ber}(\pi)$, $\forall i = 1, ..., n$. Posto

$$\hat{\pi} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\hat{\pi} \underset{a}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

INFERENZA

Definire le proprietà desiderabili degli stimatori per una dimensione campionaria fissata (n fissato).

- La *correttezza*:

Lo stimatore h si dice **corretto** se

$$E(h) = \theta \qquad \forall\theta \in \Theta$$

- l'*efficienza*: L'efficienza di uno stimatore è misurata con il **Errore Quadratico Medio** (*Mean Squared Error*):
$$MSE(h) = E((h - \theta)^2) = V(h) + B^2(h)$$
dove
$$B(h) = |E(h) - \theta|$$
  - se h è corretto allora
  $$MSE(h) = V(h)$$

se h_1 e h_2 sono due stimatori per θ, si dice che h_1 è **più efficiente** di h_2 se e solo se

$$MSE(h_1) < MSE(h_2)$$

h è detto lo stimatore *più efficiente* se e solo se

$$MSE(h) \leq MSE(h^*), \forall h^* \neq h$$

Definire le proprietà desiderabili degli stimatori per la dimensione campionaria che diverge ($n \to \infty$).

- Lo stimatore h si dice **asintoticamente corretto** se

$$\lim_{n\to\infty} E(h(X_1, ..., X_n)) = E(h) = \theta$$

- Lo stimatore h si dice **consistente** (in media quadratica) se e solo se

$$\lim_{n\to\infty} MSE(h(X_1, ..., X_n)) = \lim_{n\to\infty} MSE(h) = 0$$

- Essendo
$$MSE(h) = V(h) + B^2(h)$$
allora
$$\lim_{n\to\infty} MSE(h) = 0, \text{ se e solo se } \lim_{n\to\infty} V(h) = 0 \text{ e } \lim_{n\to\infty} B^2(h) = 0$$

Descrivere la differenza tra *standard deviation* di popolazione, *standard error* e *standard deviation stimata*

- La *standard deviation* (SD) σ, rappresenta la dispersione degli individui dalla media, è un indicatore di *variabilità* della *popolazione*, per esempio in una popolazione finita di N individui:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

  - La *deviazione standard* σ è la radice della varianza della popolazione σ^2.
- Lo *standard error* $SE(h)$ di uno stimatore h per θ è un indicatore della *variabilità* dello stimatore nello *spazio dei parametri*

$$SE(h) = \sqrt{V(h)}$$

  - Lo *standard error* $SE(h)$ di uno stimatore h per θ è la radice della varianza della VC h.
- La *standard deviation stimata* $\hat{\sigma}$, rappresenta la dispersione degli individui *del campione* dalla media *del campione*, è un indicatore di *variabilità* del *campione*:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

  - La *deviazione standard stimata* $\hat{\sigma}$ è la radice della varianza del campione $\hat{\sigma}^2$.

Definire la funzione di verosimiglianza, di log-verosimiglianza e lo stimatore di massima verosimiglianza in generale

- Siano $x_1, ..., x_n$ n osservazioni di $X \sim \mathscr{L}(\theta)$, $\theta \in \Theta$, si definisce la verosimiglianza L di θ la funzione:

$$L(\theta; x_1, ..., x_n) = L(\theta) \propto P(X_1 = x_1, ..., X_n = x_n; \theta)$$

- Se $x_1, .., x_n$ sono osservazioni IID otteniamo

$$\begin{aligned} L(\theta) &\propto P(X_1 = x_1; \theta) \cdot ... \cdot P(X_n = x_n; \theta) \\ &\propto \prod_{i=1}^{n} f(x_i; \theta) \end{aligned}$$

- Si definisce $\ell(\theta)$ la log-verosimiglianza

$$\ell(\theta) = \log L(\theta)$$

- Lo stimatore di *massima verosimiglianza* per θ è quel valore $\hat{\theta}$ che rende massima la verosimiglianza:

$$\begin{aligned} \hat{\theta} &= \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta) \\ &= \underset{\theta \in \Theta}{\operatorname{argmax}}\, \ell(\theta) \end{aligned}$$

Scrivere la funzione di verosimiglianza per una Bernoulli, la sua log-verosimiglianza, individuare gli stimatori di massima verosimiglianza e le loro proprietà

- La verosimiglianza è

$$\begin{aligned} L(\pi) &\propto \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} \\ &= \pi^{s_n}(1-\pi)^{n-s_n}, \qquad s_n = \sum_{i=1}^{n} x_i \end{aligned}$$

- La log-verosimiglianza è

$$\begin{aligned}\ell(\pi) &= \log L(\pi) \\ &= s_n \log \pi + (n - s_n) \log(1 - \pi)\end{aligned}$$

- Lo stimatore di massima verosimiglianza

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- $\hat{\pi}$ è corretto per π, infatti

$$E(\hat{\pi}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{\pi + ... + \pi}{n} = \frac{n}{n}\pi = \pi$$

- Mean Squared Error:

$$MSE(\hat{\pi}) = V(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$$

- Lo stimatore $\hat{\pi}$ per π è *consistente*

$$\lim_{n \to +\infty} MSE(\hat{\pi}) = \lim_{n \to +\infty} \frac{\pi(1-\pi)}{n} = 0$$

Scrivere la funzione di verosimiglianza per una Poisson, la sua log-verosimiglianza, individuare gli stimatori di massima verosimiglianza e le loro proprietà

- La verosimiglianza è

$$\begin{aligned}L(\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\ &\propto \lambda^{s_n} e^{-n\lambda}, \qquad s_n = \sum_{i=1}^{n} x_i\end{aligned}$$

- La log-verosimiglianza è

$$\begin{aligned}\ell(\lambda) &= \log \lambda^{s_n} e^{-n\lambda} \\ &= \log \lambda^{s_n} + \log e^{-n\lambda} \\ &= s_n \log \lambda - n\lambda, \qquad \text{in quanto } \log e^a = a\end{aligned}$$

- Lo stimatore di massima verosimiglianza

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Correttezza:

$$E(\hat{\lambda}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \lambda = \lambda$$

- Mean Squared Error:

$$MSE(\hat{\lambda}) = V(\hat{\lambda}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n}{n^2}\lambda = \frac{\lambda}{n}$$

- Consistenza:

$$\lim_{n\to+\infty} MSE(\hat{\lambda}) = \lim_{n\to+\infty} \frac{\lambda}{n} = 0$$

Scrivere la funzione di verosimiglianza per una Normale,individuare gli stimatori di massima verosimiglianza e le loro proprietà

- Siano $X_1, ..., X_n$ n VC IID, replicazioni della stessa $X \sim N(\mu, \sigma^2)$, e dunque con funzione di probabilità:

$$f(x_i; \mu, \sigma^2)$$

- Gli stimatori sono

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2 \end{aligned}$$

- Correttezza per μ:

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

- Mean Squared Error per μ:

$$MSE(\hat{\mu}) = V(\hat{\mu}) = \frac{\sigma^2}{n}$$

- Consistenza per μ:

$$\lim_{n\to+\infty} MSE(\hat{\mu}) = \lim_{n\to+\infty} \frac{\sigma^2}{n} = 0$$

- Correttezza per $\hat{\sigma}^2$:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

  $\hat{\sigma}^2$ non è stimatore corretto per σ^2.

- Correzione di $\hat{\sigma}^2$

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

Indicare le proprietà degli stimatori di massima verosimiglianza

- Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per per θ, allora
- $\hat{\theta}$ non è sempre stimatore corretto ma è sempre corretto asintoticamente:

$$E(\hat{\theta}) \xrightarrow{n\to\infty} \theta$$

- $\hat{\theta}$ non è sempre stimatore a *massima efficienza* ma lo è sempre asintoticamente:

$$V(\hat{\theta}) \xrightarrow{n\to\infty} I^{-1}(\theta)$$

  - $I^{-1}(\theta)$ è un risultato teorico ed un limite sotto al quale nessuno stimatore può scendere.
  - Se esiste lo stimatore più efficiente allora è quello di *massima verosimiglianza*.

- $\hat{\theta}$ è asintoticamente distribuito normalmente

$$\hat{\theta} \underset{a}{\sim} N(\theta, I^{-1}(\theta))$$

- Lo stimatore di massima verosimiglianza è invariante alle trasformazioni monotone invertibili g: se $\psi = g(\theta)$, allora $\hat{\psi} = g(\hat{\theta})$.

Definire un intervallo di confidenza

- Un *intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è costruito su quella coppia di statistiche L_1 e L_2 tali che

$$P(L_1 < \theta < L_2) = 1 - \alpha$$

- Un *intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è l'intervallo $[L_1, L_2]$ calcolato sui dati del campione.

Definire il livello di confidenza

- Un *intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è costruito in modo tale che l'intervallo copre il vero θ con probabilità $(1-\alpha)$, quindi il livello di confidenza è la probabilità di coprire, nel long run, il vero parametro.

Cos'è un test statistico?

Un *Test Statistico* è la scelta tra due ipotesi diverse su θ alla luce dei dati che osserveremo:

$$\begin{cases} H_0 : \theta \in \Theta_0, & \Theta_0 \subset \Theta \\ H_1 : \theta \in \Theta_1, & \Theta_1 \subset \Theta \end{cases}$$

- Se $\Theta_0 = \{\theta_0\}$ è un solo punto si dice che H_0 è un'ipotesi **semplice**, altrimenti è **composta** - Se $\Theta_1 = \{\theta_1\}$ è un solo punto si dice che H_1 è un'ipotesi **semplice**, altrimenti è **composta**

Descrivere la tavola della verità

	Decisione	
Stato di natura	Decido H_0	Decido H_1
è vera H_0	corretta	errore di I tipo
	$1-\alpha$	α
è vera H_1	errore di II tipo	corretta
	β	$1-\beta$

- Dove

$$\alpha = P(\text{Errore I tipo}) = P(\text{Decidere } H_1; H_0)$$

  α è il livello di **significatività** del test

  - α è la probabilità di scegliere H_1 quando invece è vera H_0

- e

$$\beta = P(\text{Errore II tipo}) = P(\text{Decidere } H_0; H_1)$$

- Infine

$$1-\beta = P(\text{Decidere } H_1; H_1)$$

  $1-\beta$ è la **potenza del test**

- $1-\beta$ è la probabilità di scegliere H_1 quando H_1 è vera.

Definire la probabilità di significatività osservata p_{value}

La probabilità di significatività p_{value} è

$$p_{\text{value}} = P(|T| > |t_{\text{obs}}|; H_0)$$

La probabilità di significatività osservata p_{value} esprime la probabilità, *se fosse vera* H_0, di trovare un campione ancora più in favore di H_1 di quello disponile

---

REGRESSIONE

Elencare gli assunti del modello di regressione

0. Dati $(x_1, y_1), ..., (x_n, y_n)$, n coppie di punti, si assume che

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \forall i$$

1. Il valore atteso dell'errore è nullo

$$E(\varepsilon_i) = 0 \quad \forall i$$

2. Omoschedasticità

$$V(\varepsilon_i) = \sigma^2_\varepsilon, \qquad \text{costante } \forall i$$

3. Indipendenza dei residui

$$\varepsilon_i \text{ è indipendente da } \varepsilon_j \quad \forall i \neq j$$

4. Indipendenza tra i residui e la X

$$X_i \text{ è indipendente da } \varepsilon_i \quad \forall i$$

5. *Esogeneità* della X: la distribuzione su X non è oggetto di inferenza

6. Normalità dei residui

$$\varepsilon_i \sim N(0, \sigma^2_\varepsilon) \quad \forall i$$

   - Normalità delle Y

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2_\varepsilon) \quad \forall i$$

Proprietà della covarianza

- se la covarianza è positiva, c'è associazione lineare diretta tra X ed Y
- se la covarianza è negativa, c'è associazione lineare inversa tra X ed Y
- se la covarianza è prossima a zero, non c'è associazione lineare tra X ed Y
- Simmetria

$$\text{cov}(x, y) = \text{cov}(y, x)$$

- Campo di variazione

$$-\hat{\sigma}_X \hat{\sigma}_Y \leq \text{cov}(x, y) \leq +\hat{\sigma}_X \hat{\sigma}_Y$$

Definire la previsione

- La previsione è

$$\hat{Y}_{(X=x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Se $\min\{x\} \leq x \leq \max\{x\}$ si parla di interpolazione.
- Se $x < \min\{x\}$ oppure $x > \max\{x\}$ si parla di estrapolazione.
- L'errore di previsione cresce al crescere di $(x - \bar{x})^2$.

Definire le previsioni osservate e i residui osservati

- Le previsioni, sulle x osservate
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
- Le stime degli errori
$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

Elencare le proprietà della retta dei minimi quadrati

- Valgono le seguenti proprietà:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}, \qquad \text{la retta passa nel punto delle medie} \tag{1}$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i, \qquad \text{la media delle previsioni coincide con quella degli } y \tag{2}$$
$$0 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i, \qquad \text{la media dei residui osservati è zero} \tag{3}$$

Proprietà del coefficiente di correlazione

1. $-1 \leq r \leq 1$

   1.1 Il SEGNO indica la direzione della relazione

   1.2 $r > 0$, al crescere di X, *in media*, cresce Y

   1.3 $r < 0$, al crescere di X, *in media*, decresce Y

   1.4 $r = 1$, associazione perfetta diretta

   1.5 $r = -1$, associazione perfetta indiretta

2. r è un numero puro, ovvero è privo di unità di misura
3. è simmetrico: $r_{XY} = r_{YX} = r$
4. è invariante per cambiamenti di scala:

$$\text{se } W = a + bY, \text{allora } r_{X,W} = \text{sign}(b) r_{XY}, \text{ dove la funzione } \text{sign}(b) = \begin{cases} +1, & \text{se } b > 0 \\ -1, & \text{se } b < 0 \end{cases}$$

5. r misura l'associazione lineare:

   5.1 r misura come i punti si addensano intorno alla retta.

   5.2 $f(x)$ **non lineare** r è parzialmente inutile

   5.3 il valore di r, da solo, non è in grado di descrivere tutte le possibili relazioni che si possono realizzare tra due variabili.

6. r è più elevato se i dati sono aggregati in medie o percentuali

---

Riportare la scomposizione della varianza

- variabilità totale di y è scomponibile nella somma di due parti

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

$$TSS = ESS + RSS$$

  - TSS Total Sum of Squares
  - ESS Explained Sum of Squares
  - RSS Residual Sum of Squares
- La variabilità totale di y è scomponibile nella somma di due parti

$$\left\{ \begin{array}{c} \text{varibilità di } y \\ \text{intorno alla sua media} \end{array} \right\} = \left\{ \begin{array}{c} \text{varibilità della retta} \\ \text{intorno alla media} \end{array} \right\} + \left\{ \begin{array}{c} \text{varibilità delle } y \\ \text{intorno alla retta} \end{array} \right\}$$

Definire il coefficiente di determinazione lineare R^2

- l'indice di determinazione lineare è il quadrato dell'indice di correlazione

$$R^2 = \left(\frac{ESS}{TSS}\right) = r^2 = \left(\frac{\text{cov}(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}\right)^2$$

rappresenta la quota di varianza spiegata dal modello

- $$0 \leq R^2 \leq 1$$
  - Se $R^2 = 1$, allora $r = -1$ oppure $r = +1$, associazione lineare perfetta, 100% della variabilità spiegata
  - Se $R^2 = 0$, allora $r = 0$ associazione lineare nulla, 0% della variabilità spiegata
  - Se $R^2 > 0.75$, allora considereremo l'associazione lineare *soddisfacente*.

Enunciare il Teorema di Gauss-Markov

1. Gli stimatori dei minimi quadrati sono corretti

$$E(\hat{\beta}_1) = \beta_1, \qquad E(\hat{\beta}_0) = \beta_0$$

2. Gli stimatori $\hat{\beta}_0$ e $\hat{\beta}_1$ di β_0 e β_1 sono, tra tutti gli stimatori lineari corretti per β_0 e β_1, *BLUE* (*Best Linear Unbiased Estimators* Best: i più efficienti; Unbiased: corretti; Linear Estimators: stimatori lineari).

Spiegare brevemente cos'è l'analisi dei residui

- La **analisi dei residui** è una serie di procedure diagnostiche per controllare che gli assunti del modello di regressione siano rispettati.
- Le procedure consistono nel produrre statistiche e grafici sui residui osservati $\hat{\varepsilon}_i$.

Definire il diagramma dei residui e la retta dei residui

- il diagramma dei residui consiste nel mettere in ordinata le x_i e in ascissa i residui $\hat{\varepsilon}_i$
- la **retta dei residui**, che è la retta di regressione tra x e $\hat{\varepsilon}$ è parallela all'asse delle x e coincide con esso.

Definire il Normal QQ plot

- Si tratta di un grafico che mette sull'asse delle x i *quantile* (percentili in inglese) teorici della normale e in ordinata i *quantile* osservati dei residui sul campione.
- Se i percentili teorici e quelli osservati giacciono su una retta, allora gli errori si possono assumere normali, tanto più i punti si allontanano tanto più l'ipotesi è violata.

Definire i Punti di leva, Outliers e punti influenti

- Outlier: osservazione con residuo anomale (sulle y)
- Leverage: (punto di leva), valore anomalo (sulle x)
- Influence Points: (punti influenti) osservazioni con comportamento anomalo che influenzano notevolmente i risultati

---

Definire i residui Studentizzati

- I residui studentizzati sono dati da:

$$\tilde{\varepsilon}_i = \frac{\hat{\varepsilon}_i}{S_\varepsilon\sqrt{1-h_i}} \sim t_{n-2}$$

- Si preferiscono i residui studentizzati perché incorporano le leve e sono più confrontabili.
- Tanto più alto è $|\tilde{\varepsilon}_i|$ tanto più il punto i è influente

---

Scrivere la relazione tra gli α_1, β_1 ed r

- $$\begin{aligned} \hat{\beta}_1 &= \frac{\text{cov}(x,y)}{\hat{\sigma}_X^2} \\ \hat{\beta}_1 &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r \end{aligned}$$

- $$\begin{aligned} \hat{\alpha}_1 &= \frac{\text{cov}(x,y)}{\hat{\sigma}_Y^2} \\ \hat{\alpha}_1 &= \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} r \end{aligned}$$

- Quindi

$$\hat{\beta}_1 = \hat{\alpha}_1 = r, \text{ se e solo se } \hat{\sigma}_X = \hat{\sigma}_Y$$

Scrivere i coefficienti della regressione tra variabili standardizzate

- I coefficienti sono dati da

$$\hat{\beta}_{1Z} = r \qquad \hat{\alpha}_{1Z} = r \tag{4}$$

$$\hat{\beta}_{0Z} = 0 \qquad \hat{\alpha}_{0Z} = 0 \tag{5}$$

$$\tag{6}$$

Possibili Domande

RAPPRESENTAZIONI DELLA DISTRIBUZIONE UNITARIA

---

1. qual è il campo di variazione della media aritmetica?

$x_{(1)} \leq \bar{x} \leq x_{(n)}$, la media varia tra il minimo e il massimo dei dati

---

2.qual è la proprietà di internalità della media aritmetica?

$x_{(1)} \leq \bar{x} \leq x_{(n)}$, la media varia tra il minimo e il massimo dei dati

---

3. $\bar{x}$ può essere maggiore di $x_{(n)}$?

No, $x_{(1)} \leq \bar{x} \leq x_{(n)}$, la media varia tra il minimo e il massimo dei dati

---

4. $\bar{x}$ può essere minore di $x_{(1)}$?

No, $x_{(1)} \leq \bar{x} \leq x_{(n)}$, la media varia tra il minimo e il massimo dei dati

---

5. a quanto sommano gli scarti dalla media aritmetica?

la somma degli scarti dalla media è nulla: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

---

6. cosa vuol dire che la somma degli scarti dalla medie è nulla?

vuol dire che: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

---

7. qual è la media degli scarti dalla media?

la somma degli scarti dalla media è nulla, e quindi anche la sua media

---

8. la media aritmetica minimizza la somma degli scarti al quadrato

sì,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - d)^2 \quad \forall d \neq \bar{x}$$

---

9. la media aritmetica minimizza la somma del valore assoluto degli scarti

NO, la media aritmetica minimizza la somma degli scarti al quadrato ,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - d)^2 \quad \forall d \neq \bar{x}$$

---

10. la media aritmetica massimizza la somma degli scarti al quadrato

NO, la media aritmetica minimizza la somma degli scarti al quadrato ,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - d)^2 \quad \forall d \neq \bar{x}$$

---

11. cosa vuol dire che la media è invariante alle trasformazioni lineari?

che se $y_i = a + bx_i$ allora $\bar{y} = a + b\bar{x}$

---

12. se $y_i = a - bx_i$ allora $\bar{y} = b^2\bar{x}$

no, la media è invariante alle trasformazioni lineari $y_i = a - bx_i$ allora $\bar{y} = a + (-b)\bar{x}$

---

13. se $y_i = a - bx_i$ allora $\bar{y} = a + b\bar{x}$

no, la media è invariante alle trasformazioni lineari $y_i = a - bx_i$ allora $\bar{y} = a + (-b)\bar{x}$

---

14. se $y_i = a - bx_i$ allora $\bar{y} = a - b\bar{x}$

sì, la media è invariante alle trasformazioni lineari $y_i = a - bx_i$ allora $\bar{y} = a + (-b)\bar{x}$

---

15. se $y_i = x_i^2$ allora $\bar{y} = \bar{x}^2$

no, la media è invariante solo alle trasformazioni lineari

---

16. Cos'è la proprietà di associatività della media?

Associatività. Sia una popolazione, $\mathcal{P}$, formata da K gruppi con medie e numerosità: $(\bar{x}_1;\ n_1)$, $(\bar{x}_2;\ n_2)$, ..., $(\bar{x}_K;\ n_K)$. Allora, la media totale $\bar{x}_T$ di $\mathcal{P} =$ è data da

$$\bar{x}_T = \frac{\mathrm{Tot}\{\mathcal{P}_1\} + \cdots + \mathrm{Tot}\{\mathcal{P}_K\}}{n_1 + \cdots + n_K} = \frac{n_1\ \bar{x}_1 + \cdots + n_K\ \bar{x}_K}{n_1 + \cdots + n_K}$$

---

17. Una popolazione di n gruppi e divisa in due gruppi, uno con media $\bar{x}_1$ di numerosità n_1, l'altro di media x_2 e numerosità n_2. Qual è la media della popolazione?

la media totale di popolazione $\bar{x}_T$ è data da:

$$\bar{x}_T = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

---

18. Definire la mediana, $x_{0.5}$,

La mediana di una distribuzione, $x_{0.5}$, è quel valore della X per il quale si ha $F(x_{0.5}) = 0.5$.

---

19. qual è il campo di variazione della mediana?

la mediana varia tra il minimo e il massimo dei dati $x_{\min} \leq x_{0.5} \leq x_{\max}$,

---

20. qual è la proprietà di internalità della mediana?

la mediana varia tra il minimo e il massimo dei dati $x_{\min} \leq x_{0.5} \leq x_{\max}$,

---

21. $x_{0.5}$ può essere maggiore di $x_{(n)}$?

no, la mediana varia tra il minimo e il massimo dei dati $x_{\min} \leq x_{0.5} \leq x_{\max}$,

---

22. $x_{0.5}$ può essere minore di $x_{(1)}$?

no, la mediana varia tra il minimo e il massimo dei dati $x_{\min} \leq x_{0.5} \leq x_{\max}$,

---

23. la mediana minimizza la somma degli scarti al quadrato

no, minimizza il valore assoluto degli scarti $\sum_{j=1}^{n} |x_i - x_{0.5}|$ è un minimo.

---

24. la mediana minimizza la somma del valore assoluto degli scarti

sì, minimizza il valore assoluto degli scarti $\sum_{j=1}^{n} |x_i - x_{0.5}|$ è un minimo.

---

25. la mediana massimizza la somma degli scarti al quadrato

no, minimizza il valore assoluto degli scarti $\sum_{j=1}^{n} |x_i - x_{0.5}|$ è un minimo.

---

26. $\bar{x}$ può essere maggiore di $x_{0.5}$?

Sì, se la distribuzione ha una coda lunga a destra

---

27. $\bar{x}$ può essere minore di $x_{0.5}$?

Sì, se la distribuzione ha una coda lunga a sinistra

---

28. $\bar{x}$ può essere uguale a $x_{0.5}$?

Sì, se la distribuzione è simmetrica

---

29. la varianza può essere maggiore di zero?

Sì, la varianza è sempre maggiore o uguale zero $\sigma_X^2 \geq 0$

---

30. la varianza può essere minore di zero?

No, la varianza è sempre maggiore o uguale zero $\sigma_X^2 \geq 0$

---

31. qual è il campo di variazione della varianza?

La varianza è sempre maggiore o uguale zero $\sigma_X^2 \geq 0$

---

32. la varianza può essere uguale a zero?

Sì, la varianza è sempre maggiore o uguale zero $\sigma_X^2 \geq 0$

è uguale a zero solo se $x_i = \text{costante}, \forall i$

---

33. se $y_i = a + bx_i,\ \forall i$, qual è la varianza di Y?

la varianza di Y è $\sigma_Y^2 = b^2\ \sigma_X^2$

---

34. se $y_i = a - bx_i,\ \forall i$, qual è la varianza di Y?

la varianza di Y è $\sigma_Y^2 = b^2\ \sigma_X^2$

---

35. se $y_i = a + bx_i,\ \forall i$, la varianza di Y è $\sigma_Y^2 = a + b\sigma_X^2$

no, la varianza di Y è $\sigma_Y^2 = b^2\ \sigma_X^2$

---

36. se $y_i = a - bx_i,\ \forall i$, la varianza di Y è $\sigma_Y^2 = -(b^2)\sigma_X^2$

no, la varianza di Y è $\sigma_Y^2 = b^2\ \sigma_X^2$

---

37. se $y_i = a + bx_i,\ \forall i$, la varianza di Y è $\sigma_Y^2 = (a + b)^2\sigma_X^2$

no, la varianza di Y è $\sigma_Y^2 = b^2\ \sigma_X^2$

---

38. la deviazione standard può essere maggiore di zero?

la deviazione standard è sempre maggiore o uguale zero $\sigma_X \geq 0$

---

39. la deviazione standard può essere minore di zero?

no, la deviazione standard è sempre maggiore o uguale zero $\sigma_X \geq 0$

---

40. qual è il campo di variazione della deviazione standard

la deviazione standard è sempre maggiore o uguale zero $\sigma_X \geq 0$

---

41. la deviazione standard può essere uguale a zero?

la deviazione standard è sempre maggiore o uguale zero $\sigma_X \geq 0$

è uguale a zero solo se i dati sono costanti

---

42. se $y_i = a + bx_i,\ \forall i$ qual è la deviazione standard di Y?

la sd di Y è

$$\sigma_Y = |b|\sigma_X$$

---

43. se $y_i = a - bx_i,\ \forall i$ qual è la deviazione standard di Y?

la sd di Y è

$$\sigma_Y = |b|\sigma_X$$

---

44. se $y_i = a + bx_i, \ \forall i$ la deviazione standard di Y è $\sigma_Y = a + b\sigma_X$.

No, la sd di Y è

$$\sigma_Y = |b|\sigma_X$$

---

45. se $y_i = a - bx_i, \ \forall i$ la deviazione standard di Y è $\sigma_Y = -b\sigma_X$

No, la sd di Y è

$$\sigma_Y = |b|\sigma_X$$

---

46. se $y_i = a + bx_i, \ \forall i$ la deviazione standard di Y è $\sigma_Y = (a + b)\sigma_X$

No, la sd di Y è

$$\sigma_Y = |b|\sigma_X$$

PROBABILITÀ

---

47. la probabilità varia tra 0 e 100.

No,

$$0 \leq P(A) \leq 1 \quad \forall A$$

---

48. la probabilità varia tra 0 e 1.

Sì,

$$0 \leq P(A) \leq 1 \quad \forall A$$

---

49. la probabilità varia tra -1 e 1

No,

$$0 \leq P(A) \leq 1 \quad \forall A$$

---

50. qual è la probabilità dell'evento nullo?

La probabilità dell'evento nullo è zero

$$P(\emptyset) = 0$$

---

51. qual è la probabilità dell'evento impossibile?

La probabilità dell'evento nullo è zero

$$P(\emptyset) = 0$$

---

52. la probabilità può essere zero?

Sì, se $A = \emptyset$, allora

$$P(A) = P(\emptyset) = 0$$

---

53. Sia $\{\Omega, \mathscr{A}, P\}$ uno spazio di probabilità, quando è vero che

$$P(A \cup B) = P(A) + P(B) \ ?$$

$$P(A \cup B) = P(A) + P(B)$$

se e solo se A e B sono incompatibili: $A \cap B = \emptyset$

---

54. Quando è vero che $P(\overline{A}) = 1 - P(A)$?

è sempre vero

---

55. cosa vuol dire che A è B sono due eventi incompatibili?

vuol dire che A e B hanno intersezione nulla: $A \cap B = \emptyset$

---

56. Qual è la probabilità dell'intersezione di due eventi incompatibili?

se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

---

57. Se A e B sono due eventi incompatibili allora $P(A \cap B) = P(A)P(B)$

No, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

---

58. Se A e B sono due eventi incompatibili allora $P(A \cup B) = P(A) + P(B)$

sì, A e B sono due eventi incompatibili, allora

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

essendo $P(A \cap B) = 0$

---

59. Se A e B sono due eventi incompatibili allora $P(A \cap B) = 1$

No, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

---

60. Se A e B sono due eventi incompatibili allora $P(A \cap B) = 0$

Sì, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

---

61. se A e B sono eventi incompatibili allora $P(A|B) = P(A)$

No, se A e B sono due eventi incompatibili, allora

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0$$

---

62. se A e B sono eventi incompatibili allora $P(A|B) = 1$

No, se A e B sono due eventi incompatibili, allora

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0$$

---

63. se A e B sono eventi incompatibili allora $P(A|B) = 0$

Sì, se A e B sono due eventi incompatibili, allora

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0$$

---

64. se A e B sono eventi incompatibili allora $P(A|B) = P(B)$

No, se A e B sono due eventi incompatibili, allora

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0$$

---

65. se A e B sono due eventi incompatibili allora sono indipendenti

No, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

Se A e B sono indipendenti

$$P(A \cap B) = P(A)P(B)$$

---

66. se A e B sono due eventi incompatibili allora possono essere indipendenti

No, mai, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

Se A e B sono indipendenti

$$P(A \cap B) = P(A)P(B)$$

---

67. se A e B sono due eventi incompatibili allora non possono essere indipendenti

Sì, se A e B sono due eventi incompatibili, allora

$$P(A \cap B) = P(\emptyset) = 0$$

Se A e B sono indipendenti

$$P(A \cap B) = P(A)P(B)$$

---

68. se A e B sono eventi indipendenti allora $P(A|B) = P(A)$

Sì, se A e B sono indipendenti, allora

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

---

69. se A e B sono eventi indipendenti allora $P(A|B) = 1$

No, se A e B sono indipendenti, allora

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

---

70. se A e B sono eventi indipendenti allora $P(A|B) = 0$

No, se A e B sono indipendenti, allora

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

---

71. se A e B sono eventi indipendenti allora $P(A|B) = P(B)$

No, se A e B sono indipendenti, allora

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

---

72. Dati due eventi A e B tali che $P(A \cap B) = P(A)P(B)$, in che relazione sono A e B?

A e B sono indipendenti.

---

73. dati due eventi A e B quando $P(A \cap B) = P(A)P(B)$?

$P(A \cap B) = P(A)P(B)$ se e solo se A e B sono indipendenti.

---

74. dati due eventi A e B allora $P(A \cap B) = P(A|B)P(B|A)$

No, se A e B sono indipendenti, allora

$$P(A|B) = P(A), \qquad \text{e} \qquad P(B|A) = P(B)$$

---

75. dati tre eventi A, B e C quando $P(A \cap B \cap C) = P(A)P(B)P(C)$?

$P(A \cap B \cap C) = P(A)P(B)P(C)$ se e solo se A, B e C sono indipendenti.

Variabili Casuali

---

76. Una VC X è tale che

$$P(X = 0) = 1/3, \quad P(X = 1) = 1/3, \quad P(X = 2) = 1/3$$

quale è il supporto di X?

Il supporto S_X è l'insieme dei possibili valori che può assumere la VC X. In questo caso

$$S_X = \{0, 1, 2\}$$

---

77. Una VC X è tale che

$$P(X = 0) = 1/3, \quad P(X = 1) = 1/3, \quad P(X = 2) = 1/3$$

quale è valore atteso di X?

In generale il valore atteso è dato da

$$E(X) = \sum_{\forall x \in S_X} xP(X = x)$$

, in questo caso

$$E(X) = 0\frac{1}{3} + 1\frac{1}{3} + 2\frac{1}{3} = 1$$

---

78. Per il valore atteso stesse domande che per la media aritmetica

---

79. Per la varianza di una VC stesse domande della varianza descrittiva

---

80. Qual è la varianza di $2X - Y$?

$$V(2X - Y) = 2^2V(X) + (-1)^2V(Y) = 4V(X) + V(Y)$$

---

81. Qual è la varianza di $-X$?

$$V(-X) = (-1)^2V(X) = V(X)$$

---

82. Se X e Y sono due VC qualunque è vero che $V(aX + bY) = a^2V(X) + b^2V(Y)$?

No, è vero solo se X e Y sono indipendenti

---

83. Se X e Y sono due VC indipendenti è vero che $V(aX + bY) = aV(X) + bV(Y)$?

No, se X è Y sono indipendenti

$$V(aX + bY) = a^2V(X) + b^2V(Y)$$

---

84. Se X e Y sono due VC indipendenti è vero che $V(aX - bY) = a^2V(X) - b^2V(Y)$?

No, se X è Y sono indipendenti

$$V(aX - bY) = a^2V(X) + (-b)^2V(Y) = a^2V(X) + b^2V(Y)$$

---

85. Se X e Y sono due VC indipendenti è vero che $V(X - Y) = V(X) - V(Y)$?

No, se X è Y sono indipendenti

$$V(X - Y) = 1^2V(X) + (-1)^2V(Y) = V(X) + V(Y)$$

---

86. Se X e Y sono due VC qualunque è vero che $V(X - Y) = V(X) - V(Y)$?

No, se X è Y sono indipendenti

$$V(X - Y) = 1^2V(X) + (-1)^2V(Y) = V(X) + V(Y)$$

---

87. Se X e Y sono due VC indipendenti è vero che $V(X-Y) = V(X) + V(Y)$?

Sì, vale solo se X è Y sono indipendenti

$$V(X-Y) = 1^2 V(X) + (-1)^2 V(Y) = V(X) + V(Y)$$

---

88. Qual è il sopporto della Bernoulli?

La Bernoulli assume due soli valori 0 e 1. Il suo supporto è

$$S_X = \{0, 1\}$$

---

89. Qual è il valore atteso della Bernoulli?

Se $X \sim \text{Ber}(\pi)$, allora

$$E(X) = \pi$$

---

90. Qual è la varianza della Bernoulli?

Se $X \sim \text{Ber}(\pi)$, allora

$$V(X) = \pi(1-\pi)$$

---

91. Qual è il parametro della Bernoulli e quale lo spazio del parametro?

Se $X \sim \text{Ber}(\pi)$, allora *pi* è il parametro e

$$\pi \in [0, 1]$$

---

92. Qual è la VC che conta il numero di successi in n esperimenti IID?

È la variabile binomiale $X \sim \text{Bin}(n; \pi)$ dove n è il numero di prove e π è la probabilità di successo nella singola prova

---

93. Qual è il sopporto della Binomiale?

la VC Binomiale, che conta il numero di successi in n esperimenti IID, può assumere i valori da 0 ad n:

$$S_X = \{0, 1, ..., n\}$$

---

94. Qual è il valore atteso della Binomiale?

Se $X \sim \text{Bin}(n; \pi)$, allora

$$E(X) = n\pi$$

---

95. Qual è la varianza della Binomiale?

Se $X \sim \text{Bin}(n; \pi)$, allora

$$V(X) = n\pi(1-\pi)$$

---

96. Quali sono i parametri della Binomiale e quale lo spazio dei parametri?

È la variabile binomiale $X \sim \text{Bin}(n; \pi)$ dove n è il numero di prove e π è la probabilità di successo nella singola prova

$$\pi \in [0, 1]$$

e

$$n \in \{1, 2, 3, ...\}$$

---

97. Se $X \sim \text{Bin}(5, 0.4)$ e $Y \sim \text{Bin}(7, 0.4)$ come si distribuisce $X + Y$?

Se X e Y sono indipendenti allora

$$X + Y \sim \text{Bin}(7 + 7, 0.4)$$

se non sono indipendenti non sappiamo dirlo

---

98. Se $X \sim \text{Bin}(7, 0.4)$ e $Y \sim \text{Bin}(7, 0.5)$ è vero che $X + Y \sim \text{Bin}(7 + 7, 0.5 + 0.4)$?

No, hanno due π diversi, non sappiamo scrivere la distribuzione

---

99. Se $X_1 \sim \text{Ber}(0.5)$, $X_2 \sim \text{Ber}(0.5)$, $X_3 \sim \text{Ber}(0.5)$, indipendenti, come si distribuisce $X_1 + X_2 + X_3$?

La somma di Bernoulli IID è binomiale e quindi

$$X_1 + X_2 + X_3 = \text{Bin}(3; 0.5)$$

---

100. Se $X_1 \sim \text{Bin}(n_1, \pi)$, $X_2 \sim \text{Bin}(n_2, \pi)$, indipendenti, come si distribuisce $X_1 + X_2$?

Se X e Y sono indipendenti allora

$$X + Y \sim \text{Bin}(n_1 + n_2, \pi)$$

se non sono indipendenti non sappiamo dirlo

---

101. Qual è il sopporto della Poisson?

La Poisson assume valori su tutti i numeri naturali

$$S_X = \{0, 1, 2, ...\} = \mathbb{N}$$

---

102. Qual lo spazio dei parametri della Poisson?

Se $X \sim \text{Pois}(\lambda)$, allora

$$\lambda > 0, \qquad \text{ovvero} \qquad \lambda \in \mathbb{R}^+$$

---

103. Qual è il valore atteso della Poisson?

Se $X \sim \text{Pois}(\lambda)$, allora

$$E(X) = \lambda$$

---

104. Qual è la varianza della Poisson?

Se $X \sim \text{Pois}(\lambda)$, allora

$$V(X) = \lambda$$

---

105. Qual è la proprietà di riproduttività della Poisson?

siano $X_1, \ldots, X_K$, K variabili casuali indipendenti distribuite secondo una Poisson(λ_k), per $k = 1, \ldots, K$. Allora, la loro somma, $S_K = X_1 + \cdots + X_K$, si distribuisce secondo una Poisson($\lambda_1 + \cdots + \lambda_K$).

---

106. Se $X_1 \sim \text{Pois}(\lambda_1)$ e $X_2 \sim \text{Pois}(\lambda_2)$ è vero che $X_1 + X_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$?

Sì, ma solo se X_1 e X_2 sono indipendenti

---

107. Se $X \sim \text{Bin}(n; \pi)$, $n \to +\infty$ e $\pi \to 0$, tc $n\pi = \lambda$, come si distribuisce X?

Se n tende a divergere si crea la Poisson

$$X \to \text{Pois}(\lambda)$$

---

108. Se $X \sim \text{Pois}(\lambda)$ e $Y \sim \text{Pois}(\lambda)$, come si distribuisce $X + Y$?

Se x E y sono indipendenti, allora

$$X + Y \sim \text{Pois}(\lambda + \lambda)$$

---

109. Sia $X \sim N(\mu, \sigma^2)$, descrivere la distribuzione di X.

- X ha una distribuzione simmetrica rispetto a μ,
- è a forma campanulare, il suo massimo è in μ
- presenta due flessi, uno in $\mu - \sigma$, l'altro in $\mu + \sigma$.

---

110. Qual è il supporto della normale?

La normale assume valori sull'intera retta reale, quindi

$$S_X = \{-\infty < X < +\infty\} = \mathbb{R}$$

---

111. Quali sono il valore atteso e la varianza della normale standard?

Se $Z \sim N(0, 1)$, allora

$$E(Z) = 0 \qquad V(Z) = 1$$

---

112. Qual è lo spazio dei parametri della normale?

Se $X \sim N(\mu, \sigma^2)$, parametri μ e σ^2 sono tc

$$\mu \in \mathbb{R},\ (-\infty < \mu < +\infty) \qquad \sigma^2 \in \mathbb{R}^+,\ (\sigma^2 > 0)$$

---

113. Qual è la proprietà di linearità della normale?

Se $X \sim N(\mu, \sigma^2)$, allora

$$a + bX \sim N(a + b\mu, b^2\sigma^2)$$

---

114. Se $X_1 \sim N(\mu, \sigma^2)$ e $X_2 \sim N(\mu, \sigma^2)$, come si distribuisce

$$\frac{X_1 + X_2}{2} \quad ?$$

Se $X_1 \sim N(\mu, \sigma^2)$ e $X_2 \sim N(\mu, \sigma^2)$, allora

$$\frac{X_1 + X_2}{2} \sim N\Big(\frac{\mu + \mu}{2}, \frac{1}{4}(\sigma^2 + \sigma^2)\Big)$$

$$\frac{X_1 + X_2}{2} \sim N\Big(\mu, \frac{\sigma^2}{2}\Big)$$

---

115. Qual è la proprietà di riproduttività della normale?

Se $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ sono indipendenti, allora $X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$.

---

116. Se $X_1 \sim N(\mu, \sigma^2)$ e $X_2 \sim N(\mu, \sigma^2)$, è vero che $X_1 \cdot X_2 \sim N(\mu_1 \cdot \mu_2, \sigma_1^2 \cdot \sigma_2^2)$?

No, la normale è chiusa rispetto alla somma, non al prodotto

---

117. Siano $Z_1, ..., Z_n$, n variabili casuali IID, tc $Z_i \sim N(0,1)$, come si distribuisce $Z_1^2 + ...Z_n^2$?

La somma di normali standard indipendenti al quadrato è un chi quadro con n gradi di libertà

$$Z_1^2 + ...Z_n^2 \sim \chi_n^2$$

---

118. Se $X_1 \sim N(0,1)$ e $X_2 \sim N(0,1)$, è vero che $X_1^2 + X_2^2 \sim \chi_2^2$?

Sì, la somma di normali standard indipendenti al quadrato è un chi quadro con n gradi di libertà, ma solo se X_1 e X_2 sono indipendenti.

---

119. Sia $Y \sim \chi_n^2$, descrivere la distribuzione di Y.

La funzione di densità ha una coda lunga a destra e dipende da n.

---

120. Qual è il supporto della VC χ_n^2?

La VC chi-quadro è la somma di variabili al quadrato e quindi assume solo valori positivi

$$S_{\chi^2} = \mathbb{R}^+, \qquad (\chi^2 > 0)$$

---

121. Sia $T \sim t_n$, descrivere la distribuzione di T.

La funzione di densità è simmetrica rispetto allo zero, le aree esterne sono più alte di quelle della normale, le aree interne sono più basse di quelle della normale, e l'andamento di ogni curva dipende da n.

122. Qual è il supporto della VC t_n?

La VC t può assumere ogni valore reale

$$S_T = \mathbb{R}, \qquad (-\infty < T < +\infty)$$

123. Sia $Z \sim N(0,1)$ e $Y \sim \chi^2_n$, è vero che

$$\frac{Z}{\sqrt{Y/n}} \sim t_n \quad ?$$

Sì, ma solo se Z e Y sono indipendenti.

124. Sia $T \sim t_n$ come si distribuisce T per $n \to +\infty$?

La t_n per n che diverge tende alla normale standard

$$t_n \to N(0,1), \text{ per } n \to \infty$$

125. Enunciare il Teorema centrale del limite per la Somma

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1, ..., n$. Posto

$$S_n = X_1 + ... + X_n,$$

allora

$$S_n \underset{a}{\sim} N\left(n\mu, n\sigma^2\right)$$

126. Enunciare il Teorema centrale del limite per la Media

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1, ..., n$. Posto

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\bar{X} \underset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

127. Enunciare il Teorema centrale del limite per la Proporzione

Siano $X_1, ..., X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $X_i \sim \text{Ber}(\pi)$, $\forall i = 1, ..., n$. Posto

$$\hat{\pi} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\hat{\pi} \underset{a}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

INFERENZA

---

128. Quando un stimatore si dice corretto?

Lo stimatore h si dice **corretto** se

$$E(h) = \theta \qquad \forall \theta \in \Theta$$

---

129. Sia h uno stimatore per θ, tale che:
$$E(h) = \theta$$
di quale proprietà gode h?

Lo stimatore h è **corretto**

$$E(h) = \theta \qquad \forall \theta \in \Theta$$

---

130. Come si misura l'efficienza di uno stimatore?

L'efficienza di uno stimatore è misurata con il **Errore Quadratico Medio** (*Mean Squared Error*):
$$MSE(h) = E((h-\theta)^2) = V(h) + B^2(h)$$
dove
$$B(h) = |E(h) - \theta|$$
+ se h è corretto allora
$$MSE(h) = V(h)$$

---

131. Definire il Mean Squared Error di uno stimatore

L'efficienza di uno stimatore è misurata con il **Errore Quadratico Medio** (*Mean Squared Error*):
$$MSE(h) = E((h-\theta)^2) = V(h) + B^2(h)$$
dove
$$B(h) = |E(h) - \theta|$$
+ se h è corretto allora
$$MSE(h) = V(h)$$

---

132. Se h_1 e h_2 sono due stimatori per θ, cosa significa dire che h_1 è più efficiente di h_2?

se h_1 e h_2 sono due stimatori per θ, si dice che h_1 è **più efficiente** di h_2 se e solo se
$$MSE(h_1) < MSE(h_2)$$

---

133. Se h_1 e h_2 sono due stimatori corretti per θ, cosa significa dire che h_1 è più efficiente di h_2?

se h_1 e h_2 sono due stimatori per θ, si dice che h_1 è **più efficiente** di h_2 se e solo se
$$MSE(h_1) < MSE(h_2)$$

essendo h_1 e h_2 corretti, allora
$$V(h_1) < V(h_2)$$

134. Sia h uno stimatore per θ, definire la correttezza asintotica di h.

Lo stimatore h si dice **asintoticamente corretto** se

$$\lim_{n\to\infty} E(h(X_1, ..., X_n)) = E(h) = \theta$$

135. Sia h uno stimatore per θ, definire la consistenza di h.

- Lo stimatore h si dice **consistente** (in media quadratica) se e solo se

$$\lim_{n\to\infty} MSE(h(X_1, ..., X_n)) = \lim_{n\to\infty} MSE(h) = 0$$

- Essendo

$$MSE(h) = V(h) + B^2(h)$$

allora

$$\lim_{n\to\infty} MSE(h) = 0, \text{ se e solo se} \lim_{n\to\infty} V(h) = 0 \text{ e } \lim_{n\to\infty} B^2(h) = 0$$

136. Sia h uno stimatore per θ, tc

$$\lim_{n\to\infty} MSE(h(X_1, ..., X_n)) = \lim_{n\to\infty} MSE(h) = 0,$$

h è asintoticamente corretto?

Sì, essendo h **consistente** (in media quadratica) , il suo MSE va a zero all'aumentare di n, ed essendo:

$$MSE(h) = V(h) + B^2(h)$$

allora

$$\lim_{n\to\infty} MSE(h) = 0, \text{ se e solo se} \lim_{n\to\infty} V(h) = 0 \text{ e } \lim_{n\to\infty} B^2(h) = 0$$

137. Descrivere la differenza tra *standard deviation* di popolazione, *standard error* e *standard deviation stimata*

- La *standard deviation* (SD) σ, rappresenta la dispersione degli individui dalla media, è un indicatore di *variabilità* della *popolazione*, per esempio in una popolazione finita di N individui:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

  - La *deviazione standard* σ è la radice della varianza della popolazione σ^2.
- Lo *standard error* $SE(h)$ di uno stimatore h per θ, è un indicatore della *variabilità* dello stimatore nello *spazio dei parametri*

$$SE(h) = \sqrt{V(h)}$$

  - Lo *standard error* $SE(h)$ di uno stimatore h per θ, è la radice della varianza della VC h.
- La *standard deviation stimata* $\hat{\sigma}$, rappresenta la dispersione degli individui *del campione* dalla media *del campione*, è un indicatore di *variabilità* del *campione*:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

– La *deviazione standard stimata* $\hat{\sigma}$ è la radice della varianza del campione $\hat{\sigma}^2$.

---

138. Definire la funzione di verosimiglianza in generale

- Siano $x_1, ..., x_n$ n osservazioni di $X \sim \mathscr{L}(\theta)$, $\theta \in \Theta$, si definisce la verosimiglianza L di θ la funzione:

$$L(\theta; x_1, ..., x_n) = L(\theta) \propto P(X_1 = x_1, ..., X_n = x_n; \theta)$$

- Se $x_1, .., x_n$ sono osservazioni IID otteniamo

$$\begin{aligned} L(\theta) &\propto P(X_1 = x_1; \theta) \cdot ... \cdot P(X_n = x_n; \theta) \\ &\propto \prod_{i=1}^{n} f(x_i; \theta) \end{aligned}$$

---

139. Definire la funzione log-verosimiglianza in generale

Si L la funzione di verosimiglianza per θ, si definisce $\ell(\theta)$ la log-verosimiglianza

$$\ell(\theta) = \log L(\theta)$$

---

140. Lo stimatore di massima verosimiglianza in generale

Si L la funzione di verosimiglianza per θ, e $\ell(\theta)$ la log-verosimiglianza

Lo stimatore di *massima verosimiglianza* per θ, è quel valore $\hat{\theta}$ che rende massima la verosimiglianza:

$$\begin{aligned} \hat{\theta} &= \operatorname*{argmax}_{\theta \in \Theta} L(\theta) \\ &= \operatorname*{argmax}_{\theta \in \Theta} \ell(\theta) \end{aligned}$$

---

141. Scrivere la funzione di verosimiglianza per una Bernoulli, la sua log-verosimiglianza, individuare gli stimatori di massima verosimiglianza e le loro proprietà

- La verosimiglianza è

$$\begin{aligned} L(\pi) &\propto \prod_{i=1}^{n} \pi^{x_i} (1-\pi)^{1-x_i} \\ &= \pi^{s_n} (1-\pi)^{n-s_n}, \qquad s_n = \sum_{i=1}^{n} x_i \end{aligned}$$

142. Scrivere la funzione di verosimiglianza per una Bernoulli, la sua log-verosimiglianza, individuare gli stimatori di massima verosimiglianza e le loro proprietà

La log-verosimiglianza è

$$\begin{aligned} \ell(\pi) &= \log L(\pi) \\ &= s_n \log \pi + (n - s_n) \log(1-\pi) \end{aligned}$$

---

143. Individuare lo stimatore di massima verosimiglianza e le sue proprietà per una Bernoulli,

Lo stimatore di massima verosimiglianza

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- $\hat{\pi}$ è corretto per π, infatti

$$E(\hat{\pi}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{\pi + ... + \pi}{n} = \frac{n}{n}\pi = \pi$$

- Mean Squared Error:

$$MSE(\hat{\pi}) = V(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$$

- Lo stimatore $\hat{\pi}$ per π è *consistente*

$$\lim_{n\to+\infty} MSE(\hat{\pi}) = \lim_{n\to+\infty} \frac{\pi(1-\pi)}{n} = 0$$

---

144. Possion: vedi Bernoulli

---

145. Normale: vedi Bernoulli

---

146. Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per θ, $\hat{\theta}$ è corretto?

- $\hat{\theta}$ non è sempre stimatore corretto ma è sempre corretto asintoticamente:

$$E(\hat{\theta}) \xrightarrow{n\to\infty} \theta$$

---

147. Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per θ, $\hat{\theta}$ è a massima efficienza?

- $\hat{\theta}$ non è sempre stimatore a *massima efficienza* ma lo è sempre asintoticamente:

$$V(\hat{\theta}) \xrightarrow{n\to\infty} I^{-1}(\theta)$$

  - $I^{-1}(\theta)$ è un risultato teorico ed un limite sotto al quale nessuno stimatore può scendere.
  - Se esiste lo stimatore più efficiente allora è quello di *massima verosimiglianza*.

---

148. Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per θ, com'è distribuito $\hat{\theta}$?

- $\hat{\theta}$ è asintoticamente distribuito normalmente

$$\hat{\theta} \underset{a}{\sim} N(\theta, I^{-1}(\theta))$$

---

149. Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per θ, cosa vuol dire che è invariante alle trasformazioni monotone invertibili g?

- Lo stimatore di massima verosimiglianza è invariante alle trasformazioni monotone invertibili g: se $\psi = g(\theta)$, allora $\hat{\psi} = g(\hat{\theta})$.

---

150. Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per θ, qual è la proprietà di invarianza alle trasformazioni lineari?

- Lo stimatore di massima verosimiglianza è invariante alle trasformazioni monotone invertibili g: se $\psi = g(\theta)$, allora $\hat{\psi} = g(\hat{\theta})$.

---

151. Definire un intervallo di confidenza

- Un *intervallo di confidenza* per θ, al livello $(1-\alpha) \times 100\%$ è costruito su quella coppia di statistiche L_1 e L_2 tali che
$$P(L_1 < \theta < L_2) = 1 - \alpha$$
- Un *intervallo di confidenza* per θ, al livello $(1-\alpha) \times 100\%$ è l'intervallo $[L_1, L_2]$ calcolato sui dati del campione.

---

152. Definire il livello di confidenza

Un *intervallo di confidenza* per θ, al livello $(1-\alpha) \times 100\%$ è costruito in modo tale che l'intervallo copre il vero θ con probabilità $(1-\alpha)$, quindi il livello di confidenza è la probabilità di coprire, nel long run, il vero parametro.

---

153. Cos'è un test statistico?

Un *Test Statistico* è la scelta tra due ipotesi diverse su θ alla luce dei dati che osserveremo:

$$\begin{cases} H_0 : \theta \in \Theta_0, & \Theta_0 \subset \Theta \\ H_1 : \theta \in \Theta_1, & \Theta_1 \subset \Theta \end{cases}$$

- Se $\Theta_0 = \{\theta_0\}$ è un solo punto si dice che H_0 è un'ipotesi **semplice**, altrimenti è **composta** - Se $\Theta_1 = \{\theta_1\}$ è un solo punto si dice che H_1 è un'ipotesi **semplice**, altrimenti è **composta**

---

154. Cos'è l'errore di primo tipo

È l'errore che si commette scegliendo H_1 quando è vera H_0

---

155. Cos'è l'errore di secondo tipo

È l'errore che si commette scegliendo H_0 quando è vera H_1

---

156. Definire la probabilità di significatività α

$$\alpha = P(\text{Errore I tipo}) = P(\text{Decidere } H_1; H_0)$$

α è il livello di **significatività** del test α è la probabilità di scegliere H_1 quando invece è vera H_0 (errore I tipo)

---

157. Definire la probabilità di significatività di errore del secondo tipo

$$\beta = P(\text{Errore II tipo}) = P(\text{Decidere } H_0; H_1)$$

---

158. Definire la potenza di un test

$$1 - \beta = P(\text{Decidere } H_1; H_1)$$

$1 - \beta$ è la **potenza del test**

$1 - \beta$ è la probabilità di scegliere H_1 quando H_1 è vera.

---

159. Definire la probabilità di significatività osservata p_{value}

La probabilità di significatività p_{value} è

$$p_{\text{value}} = P(|T| > |t_{\text{obs}}|; H_0)$$

La probabilità di significatività osservata p_{value} esprime la probabilità, *se fosse vera* H_0, di trovare un campione ancora più in favore di H_1 di quello disponile

---

160. Se in un test chi-quadro sull'indipendenza tra due VC $p_{\text{value}} = 0.0341$, cosa possiamo concludere?

Il test è significativo al 5% ma non all'1%. C'è una buona evidenza campionaria che respinge H_0 (le due variabili non sono indipendenti) anche se non fortissima

---

161. Se in un test chi-quadro sull'indipendenza tra due VC $p_{\text{value}} = 0.00002871$, cosa possiamo concludere?

Il test è significativo all'1% e oltre. C'è una fortissima evidenza campionaria che respinge H_0 (le due variabili non sono indipendenti).

---

162. Se in un test sull'uguaglianza tra due medie

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

$p_{\text{value}} = 0.00002871$, cosa possiamo concludere?

Il test è significativo all'1% e oltre. C'è una fortissima evidenza campionaria che respinge H_0 (le due due medie sono diverse).

REGRESSIONE

---

163. Scrivere l'assunto di base del modello di regressione

Dati $(x_1, y_1), ..., (x_n, y_n)$, n coppie di punti, si assume che

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \forall i$$

---

164. Elencare gli assunti sul valore atteso e la varianza dell'errore nel modello di regressione

Il valore atteso dell'errore è nullo

$$E(\varepsilon_i) = 0 \quad \forall i$$

Omoschedasticità: la varianza dell'errore è costante

$$V(\varepsilon_i) = \sigma^2_\varepsilon, \qquad \text{costante } \forall i$$

---

165. In che relazione stanno gli errori tra di loro e rispetto alle x?

Indipendenza dei residui

$$\varepsilon_i \text{ è indipendente da } \varepsilon_j \quad \forall i \neq j$$

Indipendenza tra i residui e la X

$$X_i \text{ è indipendente da } \varepsilon_i \quad \forall i$$

---

166. Cosa vuole dire che X è esogena?

Esogeneità della X: la distribuzione su X non è oggetto di inferenza

---

167. Come si assume si distribuiscano gli errori (o residui) del modello di regressione?

Si assume la normalità dei residui

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \quad \forall i$$

---

168. Come si assume si distribuiscano le Y del modello di regressione?

Si assume la normalità delle Y

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_\varepsilon^2) \quad \forall i$$

---

169. se la covarianza è positiva, che tipo di associazione mi aspetto tra X e Y

se la covarianza è positiva allora c'è associazione lineare diretta tra X ed Y

---

170. se la covarianza è negativa, che tipo di associazione mi aspetto tra X e Y

se la covarianza è positiva allora c'è associazione lineare inversa tra X ed Y

---

171. se la covarianza è prossima a zero, che tipo di associazione mi aspetto tra X e Y

se la covarianza è prossima a zero, non c'è associazione lineare tra X ed Y

---

172. Cosa vuol dire che la covarianza è simmetricaˆ

Significa che

$$\text{cov}(x, y) = \text{cov}(y, x)$$

---

173. Qual è il campo di variazione della covarianza?

- Campo di variazione

$$-\hat{\sigma}_X \hat{\sigma}_Y \leq \text{cov}(x, y) \leq +\hat{\sigma}_X \hat{\sigma}_Y$$

---

174. Definire la previsione

- La previsione è

$$\hat{Y}_{(X=x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

175. Definire l'interpolazione e l'estrapolazione

in un modello di regressione stimato

$$\hat{y}_{(X=x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Se $\min\{x\} \leq x \leq \max\{x\}$ si parla di interpolazione.
- Se $x < \min\{x\}$ oppure $x > \max\{x\}$ si parla di estrapolazione.

---

176. In che modo la scelta di x incide sull'errore di previsione

- L'errore di previsione cresce al crescere di $(x - \bar{x})^2$.

---

177. Definire le previsioni osservate e i residui osservati

- Le previsioni, sulle x osservate

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Le stime degli errori

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

---

178. Sia dato in modello di regressione stimato

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

quanto vale

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{x} \ ?$$

- $$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}, \text{la retta passa nel punto delle medie}$$

---

179. Sia dato in modello di regressione stimato

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

qual è la media degli $\hat{y}_i$?

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i, \quad \text{la media delle previsioni coincide con quella degli } y$$

---

180. Siano dati i residui stimati di un modello di regressione

$$\hat{\varepsilon}_i = y - \hat{y}_i$$

quanto vale

$$\sum_{i=1}^{n} \varepsilon_i \ ?$$

$$0 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i, \quad \text{la media dei residui osservati è zero}$$

---

181. Qual è il campo di variazione del coefficiente di correlazione?

il coefficiente di correlazione r varia tra meno uno e uno

$$-1 \leq r \leq 1$$

---

182. Il coefficiente di correlazione può essere minore di zero?

Sì, il coefficiente di correlazione r varia tra meno uno e uno

$$-1 \leq r \leq 1$$

---

183. Il coefficiente di correlazione può essere maggiore di zero?

Sì, il coefficiente di correlazione r varia tra meno uno e uno

$$-1 \leq r \leq 1$$

---

184. Il coefficiente di correlazione può essere maggiore uguale a zero?

Sì, il coefficiente di correlazione r varia tra meno uno e uno

$$-1 \leq r \leq 1$$

---

185. Il coefficiente di correlazione può essere maggiore di uno?

No, il coefficiente di correlazione r varia tra meno uno e uno

$$-1 \leq r \leq 1$$

---

186. stesse domande su segno e associazione e simmetria della covarianza

---

187. cosa vuol dire che r è è invariante per cambiamenti di scala?

Significa che:

$$\text{se } W = a + bY, \text{allora } r_{X,W} = \text{sign}(b) r_{XY}, \text{ dove la funzione sign}(b) = \begin{cases} +1, & \text{se } b > 0 \\ -1, & \text{se } b < 0 \end{cases}$$

---

188. cosa misura r?

r misura l'associazione lineare: - r misura come i punti si addensano intorno alla retta. - $f(x)$ **non lineare** r è parzialmente inutile - il valore di r, da solo, non è in grado di descrivere tutte le possibili relazioni che si possono realizzare tra due variabili.

---

189. cosa succede se r viene misurato se i dati sono aggregati in medie o percentuali??

r è più elevato se i dati sono aggregati in medie o percentuali

---

190. Riportare la scomposizione della varianza in un modello di regressione

- variabilità totale di y è scomponibile nella somma di due parti

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

$$TSS = ESS + RSS$$

  - TSS Total Sum of Squares
  - ESS Explained Sum of Squares
  - RSS Residual Sum of Squares

---

191. Interpretare la scomposizione della varianza in un modello di regressione

- La variabilità totale di y è scomponibile nella somma di due parti

$$\left\{\begin{array}{c}\text{varibilità di } y\\ \text{intorno alla sua media}\end{array}\right\} = \left\{\begin{array}{c}\text{varibilità della retta}\\ \text{intorno alla media}\end{array}\right\} + \left\{\begin{array}{c}\text{varibilità delle } y\\ \text{intorno alla retta}\end{array}\right\}$$

---

192. Definire il coefficiente di determinazione lineare R^2

- l'indice di determinazione lineare è il quadrato dell'indice di correlazione

$$R^2 = \left(\frac{ESS}{TSS}\right) = r^2 = \left(\frac{\mathrm{cov}(x,y)}{\hat{\sigma}_x \hat{\sigma}_y}\right)^2$$

---

193. Interpretare il coefficiente di determinazione lineare R^2

R^2 rappresenta la quota di varianza spiegata dal modello

- $$0 \le R^2 \le 1$$
  - Se $R^2 = 1$, allora $r = -1$ oppure $r = +1$, associazione lineare perfetta, 100% della variabilità spiegata
  - Se $R^2 = 0$, allora $r = 0$ associazione lineare nulla, 0% della variabilità spiegata
  - Se $R^2 > 0.75$, allora considereremo l'associazione lineare *soddisfacente*.

---

194. Gli stimatori dei minimi quadrati sono corretti?

Sì, in base al Teorema di Gauss-Markov, gli stimatori dei minimi quadrati sono corretti

$$E(\hat{\beta}_1) = \beta_1, \qquad E(\hat{\beta}_0) = \beta_0$$

---

195. Cosa vuol dire che gli stimatori dei minimi quadrati sono BLUE?

Significa che gli stimatori $\hat{\beta}_0$ e $\hat{\beta}_1$ di β_0 e β_1 sono, tra tutti gli stimatori lineari corretti per β_0 e β_1, *BLUE* (*Best Linear Unbiased Estimators* Best: i più efficienti; Unbiased: corretti; Linear Estimators: stimatori lineari).

---

196. Cos'è l'analisi dei residui

- La **analisi dei residui** è una serie di procedure diagnostiche per controllare che gli assunti del modello di regressione siano rispettati.

- Le procedure consistono nel produrre statistiche e grafici sui residui osservati $\hat{\varepsilon}_i$.

---

197. Definire il diagramma dei residui e la retta dei residui

- il diagramma dei residui consiste nel mettere in ordinata le x_i e in ascissa i residui $\hat{\varepsilon}_i$

---

198. Definire la retta dei residui

- la **retta dei residui**, che è la retta di regressione tra x e $\hat{\varepsilon}$ è parallela all'asse delle x e coincide con esso.

---

199. Definire il Normal QQ plot

- Si tratta di un grafico che mette sull'asse delle x i *quantile* (percentili in inglese) teorici della normale e in ordinata i *quantile* osservati dei residui sul campione.

---

200. Come si interpreta il Normal QQ plot?

- Se, in un QQ plot, i percentili teorici e quelli osservatati giacciono su una retta, allora gli errori si possono assumere normali, tanto più i punti si allontanano tanto più l'ipotesi è violata.

---

201. Definire i Punti di leva, Outliers e punti influenti

- Outlier: osservazione con residuo anomale (sulle y)
- Leverage: (punto di leva), valore anomalo (sulle x)
- Influence Points: (punti influenti) osservazioni con comportamento anomalo che influenzano notevolmente i risultati

---

202. Definire i residui Studentizzati

- I residui studentizzati sono dati da:

$$\tilde{\varepsilon}_i = \frac{\hat{\varepsilon}_i}{S_\varepsilon\sqrt{1-h_i}} \sim t_{n-2}$$

---

203. Interpretare i residui Studentizzati

- Si preferiscono i residui studentizzati perché incorporano le leve e sono più confrontabili.
- Tanto più alto è $|\tilde{\varepsilon}_i|$ tanto più il punto i è influente

---

204. Si considerino i due modelli di regressione stimati

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad \hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 y_i$$

Dove si incrociano le due rette $y = \hat{\beta}_0 + \hat{\beta}_1 x$ e $x = \hat{\alpha}_0 + \hat{\alpha}_1 y$?

- Le due rette si incrociano nel luogo delle medie $(\bar{x}, \bar{y})$

$$\begin{aligned} \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \bar{x} &= \hat{\alpha}_0 + \hat{\alpha}_1 \bar{y} \end{aligned}$$

205. Si considerino i due modelli di regressione stimati

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad \hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 y_i$$

Se

$$\bar{x} = 0, \text{ e } \bar{y} = 0$$

Dove si incrociano le due rette $y = \hat{\beta}_0 + \hat{\beta}_1 x$ e $x = \hat{\alpha}_0 + \hat{\alpha}_1 y$?

- Le due rette si incrociano nel luogo delle medie $(0, 0)$

$$\begin{aligned} 0 &= 0 + \hat{\beta}_1 0 \\ 0 &= 0 + \hat{\alpha}_1 0 \end{aligned}$$

206. Si considerino i due modelli di regressione stimati

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad \hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 y_i$$

Se x e y sono standardizzate, quanto valgono $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0, \hat{\beta}_1$?

se x e y sono standardizzate allora

$$\hat{\alpha}_0 = \hat{\beta}_0 = 0$$

e

$$\hat{\alpha}_1 = \hat{\beta}_1 = r$$