

## Esempio domande

---

Q. Dati i seguenti vettori  $b = (3 \ 2 \ 1)$ ,  $a = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$ , qual è il risultato del prodotto  $axb$ ?

a) -2

b)  $\begin{pmatrix} -3 & -2 & -2 \\ 3 & 2 & 1 \\ -3 & -2 & -1 \end{pmatrix}$

c) impossibile

Q. Qual è la traccia della matrice  $A = \begin{pmatrix} 4 & 2 \\ 8 & -2 \end{pmatrix}$

a) 0

b) 10

c) -32

Q. Per creare un dataset in R importando un database da un file csv, il comando suggerito è:

a) `import.csv()`

b) `import.data()`

c) `read.csv()`

Q. la differenza tra analisi di regressione lineare multipla e regressione lineare semplice dipende da...:

a) il numero di variabili dipendenti

b) il numero di variabili esplicative

c) il numero di equazioni di regressione

Q. immaginiamo di considerare i seguenti risultati di un'analisi di regressione multipla, dove il primo reddito da lavoro (migliaia di euro) dei laureati in Matematica è funzione del voto finale di laurea, del voto dell'esame in Analisi III e del voto nell'esame in Equazioni della Fisica Matematica ( $\alpha = 0.05$ )

	coefficiente	p-value
intercetta	1.52	0.002
Voto finale	0.08	0.018
Analisi III	0.56	0.045
Equazioni della Fisica matematica	-0.12	0.120

Quale delle seguenti affermazioni è falsa:

a) Un aumento pari a +1 nel voto finale implica una crescita nel reddito pari a 80 euro

b) Il voto nell'esame di analisi III influenza significativamente il reddito

c) Il voto in Equazioni della Fisica Matematica influenza negativamente il reddito

Q. quale delle seguenti non è una assunzione tipica sugli errori del modello di regressione lineare classico:

a) Indipendenza

b) Normalità

c) Eteroschedasticità

Q. quale delle seguenti proprietà derivano dalla standardizzazione di variabili osservate?

a) le variabili trasformate assumono valori da 0 a 1

b) le variabili trasformate assumono valori da 1 a  $+\infty$

c) le variabili trasformate hanno media nulla (0) e varianza pari a 1

Q. partendo da un database composto da 10 variabili osservate, si decide di effettuare un'Analisi per Componenti Principali. Quanti componenti vengono prodotti dall'analisi?

- a)10
- b)2
- c)non si può dire a priori

Q. Quali delle seguenti proprietà caratterizza sia la PCA che la FA?

- a) la correlazione tra le variabili originarie osservate gioca un ruolo centrale
- b)la distinzione tra varianza comune e varianza specifica (o unicità)
- c) il punto di partenza è dato dalla specificazione delle variabili risposta originarie come funzione lineare dei fattori/componenti

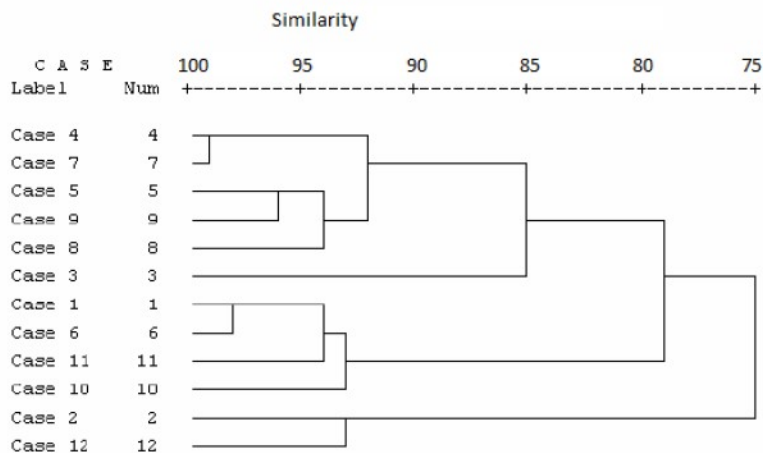
Q. Quale tra i seguenti è il comando in R per svolgere una analisi per componenti principali?

- a) factanal()
- b)princomponent()
- c)prcomp()

Q. considera il seguente output di una analisi per gruppi (o Cluster Analysis) effettuata su 12 consumatori (casi) per individuare segmenti omogenei di mercato seguendo le risposte date ad un questionario sul comportamento di acquisto.

Quali dei seguenti metodi è stato utilizzato?

Dendrogram using Ward Method



- a) gerarchico
- b) non gerarchico
- c)nessuno dei precedenti

Q.Guardando al dendrogramma riportato, qual è il numero di gruppi corrispondente ad un livello di "similarità" uguale a 80?

- a)2
- b)3
- c)4

Q.Quale tra i seguenti è un esempio di cluster naturale per l'analisi di una classe di studenti?

- a) il numero di matricola
- b) l'ordine di entrata in aula
- c) l'altezza

Q. Tra le seguenti affermazioni, quale rappresenta un obiettivo dell'analisi per gruppi (Cluster Analysis)?

- a) individuare legami di causalità tra le variabili osservate

- b) individuare relazioni di causalità fra i gruppi individuati
- c) nessuna delle precedenti

Q. dato l'indice di distanza  $d(i, i')$ , quale tra le seguenti non è una proprietà dell'indice di distanza utilizzato per l'individuazione dei gruppi nella cluster analysis?

- a)  $d(i, i') < 0$
- b)  $d(i, i') > 0$
- c)  $d(i, i') = d(i', i)$

Q. Nell'analisi per gruppi, quale tra le seguenti affermazioni è una proprietà della metrica di Minkowski

- a) nella distanza di Minkowski  $d_m(i, i')$ , se  $m=2$  si ottiene la distanza Euclidea
- b) nella distanza di Minkowski  $d_m(i, i')$ , se  $m=3$  non è possibile calcolarla
- c) nella distanza di Minkowski  $d_m(i, i')$ , se  $m=4$  si ottiene la distanza "city-block"

Q. Nel clustering gerarchico, la tecnica divisiva consiste in:

- a) divisione a tre a tre di gruppi già formati
- b) divisione casuale di gruppi già formati
- c) divisione a due a due di gruppi già formati

Q. in base al seguente modello ottenuto da un'analisi di regressione lineare semplice  $y = +3 - 0.8x$  possiamo dire che:

- a) Per ogni incremento unitario della variabile  $x$ , si stima che il valore medio di  $Y$  cresca di 0.8 unità
- b) Per ogni incremento unitario della variabile  $x$ , si stima che il valore medio di  $Y$  diminuisca di 0.8 unità
- c) Per ogni incremento unitario della variabile  $x$ , si stima che il valore medio di  $Y$  cresca di 3 unità

Q. rispetto al modello di regressione lineare semplice sopra definito, quando la variabile esplicativa osservata è pari a 0, il valore della variabile dipendente è:

- a) 0.8
- b) -0.8
- c) +3

Q. se  $R^2$  relativo al modello sopra definito è pari a 0.904, quale interpretazione possiamo dare?

- a) il 90.4% della variabilità di  $Y$  è spiegata dalla  $x$
- b) il 90.4% della variabilità di  $x$  è da attribuire a  $Y$
- c) il 90.4% della diminuzione di  $x$  è da attribuire a  $Y$

Q. riferendoci al modello sopra descritto, quanta è la variabilità non spiegata dal modello?

- a) 90.4%
- b) 9.6%
- c) non si può stabilire

Q. quale tra le seguenti è una assunzione basilare sui residui per l'analisi di regressione lineare semplice?

- a) indipendenza dei residui
- b) dipendenza dei residui
- c) linearità dei residui

Q. Si ottiene il seguente output da un'analisi di regressione lineare multivariata:

<b>Regression Statistics</b>						
<b>Multiple R</b>	<b>0.72213</b>					
<b>R Square</b>	<b>0.52148</b>					
<b>Adjusted R Square</b>	<b>0.44172</b>					
<b>Standard Error</b>	<b>47.46341</b>					
<b>Observations</b>	<b>15</b>					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
<b>Intercept</b>	<b>306.52619</b>	<b>114.25389</b>	<b>2.68285</b>	<b>0.01993</b>	<b>57.58835</b>	<b>555.46404</b>
<b>Prezzo</b>	<b>-24.97509</b>	<b>10.83213</b>	<b>-2.30565</b>	<b>0.03979</b>	<b>-48.57626</b>	<b>-1.37392</b>
<b>Pubblicità</b>	<b>74.13096</b>	<b>25.96732</b>	<b>2.85478</b>	<b>0.01449</b>	<b>17.55303</b>	<b>130.70888</b>

Indicare:

Y=

% di variabilità della Y spiegata dal modello =

Commentare i valori dei coefficienti di:

Prezzo =

Pubblicità =

Q. in un'analisi di regressione lineare multipla, il Variance Inflation Index (VIF) è maggiore di 5 indica:

- a) presenza di eteroschedasticità
- b) presenza di multicollinearità
- c) nessuna delle precedenti

Q. Il comando in R per effettuare un'analisi di regressione lineare è:

- a) reglin()
- b) regmul()
- c) lm()

Q. Il comando in R per effettuare un'analisi fattoriale è:

- a) factanal()
- b) fa()
- c) factlm()

Q. Il comando in R per effettuare un'analisi per componenti principali è:

- a) prcomp()

- b) `pca()`
- c) `pc()`

Q. Il comando in R per effettuare un'analisi per gruppi (Cluster Analysis) gerarchica è:

- a) `clust()`
- b) `gclust()`
- c) `hclust()`

Q. Il comando in R per impostare un'analisi per gruppi (Cluster Analysis) non gerarchica è:

- a) `kmeans()`
- b) `gmeans()`
- c) `hmeans()`

\*\*\*\*\*