

Corso di Statistica Multivariata
Docente Mini Valentina
2018/2019

**DI SEGUITO VENGONO INDICATE ALCUNE DOMANDE PER AGEVOLARE LO STUDENTE AD ESERCITARSI
UTILIZZANDO I CONCETTI, LE NOZIONI E LE APPLICAZIONI VISTE DURANTE IL CORSO**

Q1. Quale tra i seguenti è il comando corretto per effettuare un'analisi fattoriale in R?

- a. *factanal()*
- b. *factor()*
- c. *fa()*

Q2. Considerate i seguenti risultati di una regressione lineare multipla, dove il profitto netto di un'azienda Y (in milioni di euro) è funzione degli investimenti in R&S (X_1), investimenti in marketing (X_2) e il prezzo di vendita medio (X_3), (con $\alpha = 0.01$):

	Coefficienti	
Intercetta	2.45	0.009
X_1	0.89	0.018
X_2	0.25	0.101
X_3	-0.52	0.002

Quale delle seguenti affermazioni è falsa?

- a. L'intercetta è significativa
- b. Un aumento nel prezzo di vendita medio ha un effetto negativo significativo sul profitto
- c. L'investimento in marketing è l'unica variabile esplicativa che non ha effetto significativo sul profitto

Q3. Quale tra i seguenti è il comando corretto per effettuare un'analisi per componenti principali in R?

- a. *pca()*
- b. *pcanal()*
- c. *prcomp()*

Q4. Quale tra i seguenti è il comando corretto per effettuare un'analisi per gruppi (Cluster Analysis) gerarchica?

- a. *clust()*
- b. *gclust()*
- c. *hclust()*

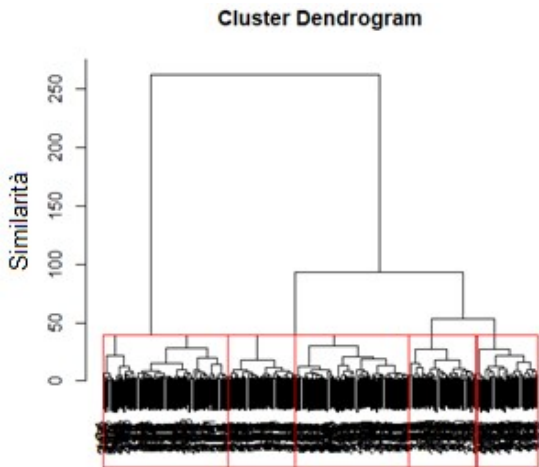
Q5. Si consideri WD quale devianza nei gruppi, TD devianza totale e g il numero di gruppi in una analisi per gruppi (cluster Analysis). Tra le seguenti affermazioni riguardanti l'indice $R^2 = 1 - WD/TD$ è vera?

- a. R^2 diminuisce al diminuire di g
- b. R^2 cresce al diminuire di g
- c. R^2 non è una funzione monotona di g

Q6. 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()*

Q7. Si osservi il dendrogramma sotto riportato relativo ad una analisi per gruppi (Cluster Analysis) condotta su un campione di 386 rispondenti ad un questionario sulle preferenze relative alla tipologia di vino consumato.



Quali dei seguenti metodi è stato utilizzato?

- a. Gerarchico
- b. Non gerarchico
- c. Nessuno dei precedenti

Q8. Guardando l'output di analisi riportato nella domanda precedente, qual è il numero di gruppi corrispondente ad un livello di "similarità" uguale a 75?

- a. 2
- b. 3
- c. 4

Q9. Si osservi l'output di una analisi per componenti principali sotto riportato.

```
> summary(PC)
Importance of components%s:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7
Standard deviation  2.6281 1.28078 1.23787 1.0049 0.91494 0.88279 0.7941
Proportion of Variance 0.4063 0.09649 0.09014 0.0594 0.04924 0.04584 0.0371
Cumulative Proportion 0.4063 0.50277 0.59291 0.6523 0.70155 0.74740 0.7845
      PC8      PC9      PC10      PC11      PC12      PC13      PC14
Standard deviation  0.76531 0.7177 0.66280 0.63848 0.63425 0.58348 0.54650
Proportion of Variance 0.03445 0.0303 0.02584 0.02398 0.02366 0.02003 0.01757
Cumulative Proportion 0.81895 0.8492 0.87509 0.89907 0.92273 0.94276 0.96033
      PC15      PC16      PC17
Standard deviation  0.50860 0.47053 0.44086
Proportion of Variance 0.01522 0.01302 0.01143
Cumulative Proportion 0.97554 0.98857 1.00000
```

Quante sono le variabili presenti nel database originario?

- a. Più di 17
- b. 17
- c. Guardando l'output riportato non è possibile stabilirlo

Q10. Seguendo l'approccio della varianza cumulata, quanti sono i componenti che devono essere estratti?

- a. 5
- b. 17
- c. 4

Q11. Da un'analisi di regressione lineare multipla si ottiene il seguente risultato:

```

Call:
lm(formula = LIKE_PAS ~ LIKE_AROMA + LIKE_SWEET + LIKE_TASTE)

Residuals:
    Min       1Q   Median       3Q      Max
-4.1751 -0.4982  0.1616  0.6278  2.3758

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.12803    0.17372   0.737 0.461592
LIKE_AROMA   0.42853    0.05601   7.652 1.63e-13 ***
LIKE_SWEET   0.19714    0.05441   3.623 0.000331 ***
LIKE_TASTE   0.24836    0.05409   4.591 5.99e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.042 on 382 degrees of freedom
Multiple R-squared:  0.5919,    Adjusted R-squared:  0.5887
F-statistic: 184.7 on 3 and 382 DF,  p-value: < 2.2e-16

```

Utilizzando i valori ottenuti, scrivere il modello stimato:

Y=.....

Q12. Osservando l'output riportato in Q.11, quale percentuale di variabilità di LIKE_PAS è spiegata dal modello?

- a. circa il 10%
- b. circa il 59%
- c. circa il 62%

Q13. Osservando l'output riportato in Q.11, interpretare il valore stimato del coefficiente relativo a LIKE_AROMA:

.....

Q14. Osservando l'output riportato in Q.11, quando LIKE_AROMA è pari a 2, LIKE_SWEET assume valore 0 e LIKE_TASTE è pari a 3, QUANTO è il valore di LIKE_PAS?

.....

Q15. Se nell'analisi che ha prodotto l'output riportato in Q.11, il VIF è maggiore di 5 allora possiamo dire che:

- a. vi è presenza di eteroschedasticità
- b. vi presenza di multicollinearità
- c. nessuna delle precedenti