COGNOME E NOME ……. MATR. ……

COMPITO B - ANALISI DEI DATI PER IL MARKETING – OTTOBRE 2010

Esercizio I

Una grande società d'informatica ha condotto un'indagine campionaria sui propri clienti, chiedendo, fra l'altro, il livello di soddisfazione globale (su scala da 1 a 10) e per aspetti specifici (quattro domande tutte su scala di Likert). SPSS ha estratto **una sola componente** principale ed ha fornito la tabella delle **comunalità** riportata di seguito.

Comunalità

	Iniziale	Estrazione
GLOBALE	1,000	,801
HARDWARE	1,000	,646
SOFTWARE	1,000	,832
SUPPORTO TECNICO	1,000	,753
SUPPORTO VENDITE	1,000	,505

- Si calcolino i coefficienti di correlazione (tutti positivi) tra ciascuna variabile e l'unica componente estratta.
- Si giudichi la validità dell'analisi con i diversi criteri.
- Si scriva l'espressione della prima componente, calcolando i valori numerici dei coefficienti di punteggio con il criterio di SPSS e si illustri il significato degli *scores*.

Esercizio II

Con riferimento a 25 modelli di *notebook* di marche differenti, il coefficiente di correlazione tra il prezzo e il peso è risultato uguale a -0.42.

- Si verifichi la significatività della correlazione al livello del 5% e dell'1% e si commenti il risultato.
- Si dica se nel caso in esame possono ritenersi soddisfatte le ipotesi sottostanti alla procedura di verifica.

Esercizio III

Si conoscono i voti di 4 studenti in 3 esami:

STUDENTE	MATEMATICA	STATISTICA	ECONOMIA
A	18	18	24
B	28	30	30
C	20	28	30
D	27	18	18

- Per i 4 studenti si calcolino la matrice delle distanze della città a blocchi e la matrice delle distanze euclidee **sui dati originari** dei voti.
- Si calcoli il coefficiente di cograduazione di Spearman tra le distanze riportate nelle due matrici suddette e se ne illustri il significato nel caso in esame.

TRACCIA DI RISOLUZIONE

Esercizio I

I coefficienti di correlazione si ottengono calcolando la radice quadrata delle comunalità e risultano i seguenti:

Matrice di componenti[a]

	Componente
	1
GLOBALE	,895
HARDWARE	,804
SOFTWARE	,912
SUPPORTO TECNICO	,868
SUPPORTO VENDITE	,710

La varianza totale spiegata dall'unica componente estratta è uguale alla somma delle comunalità e risulta uguale 3.536 (primo autovalore), cioè il 70.725%, che è inferiore alla soglia di accettabilità: 0.955 = 77,37%, per cui l'analisi non può considerarsi del tutto valida. La tabella corrispondente fornita da SPSS è riportata di seguito:

Varianza totale spiegata

Componente	Autovalori iniziali			Pesi dei fattori non ruotati		
	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumulata
1	3,536	70,725	70,725	3,536	70,725	70,725
2	,585	11,693	82,418			
3	,447	8,936	91,354			
4	,282	5,642	96,996			
5	,150	3,004	100,000			

I coefficienti di punteggio si ottengono dividendo ciascun coefficiente di correlazione per il primo auotovalore e risultano i seguenti:

Matrice dei coefficienti di punteggio dei componenti

	Componente
	1
GLOBALE	,253
HARDWARE	,227
SOFTWARE	,258
SUPPORTO TECNICO	,245
SUPPORTO VENDITE	,201

L'espressioni degli *scores* normalizzati della prima componente è:

y_1 = 0.253 z(globale) + 0.227 z(hardware) + 0.258 z(software) + 0.245 z(sup. tec.) + 0.201 z(sup. ven.)

Tale *scores* forniscono una misura del livello di soddisfazione di ciascun cliente: valori positivi indicano clienti più soddisfatti della media e valori negativi clienti meno soddisfatti della media. L'influenza maggiore sulla *customer satisfaction* è attribuibile al software e quella minore al supporto alle vendite.

Esercizio II

Si veda l'esempio a pag. 51 del testo Zani – Cerioli, 2007.
a) Si formula l'ipotesi nulla di assenza di correlazione nell'universo, contro un'ipotesi alternativa bilaterale. Per un campione di 25 elementi i gradi di libertà *g* sono uguali a 23.
Sulla tavola della variabili aleatoria T di Student per *g* = 23 si legge $t_{0.05}$ = 2.069 e $t_{0.01}$ = 2.807.
La zona di accettazione dell'ipotesi nulla al livello di significatività del 5% corrisponde all'intervallo:
-2.069; +2.048 e al livello di significatività dell'1% corrisponde all'intervallo: -2.807; 2.807.
Si calcola quindi il valore di t_r per il coefficiente di correlazione uguale a -0.42 , ottenendo:

$$t_r = \frac{-0.42}{\sqrt{1-(-0.42^2)}}\sqrt{23} = -2.219$$

che cade nella zona di rifiuto di H_0 al livello del 5%, ma nella zona di accettazione di H_0 al livello dell'1%. In base ai dati del campione di 25 elementi non si perviene ad una conclusione univoca, per cui si suggerisce di aumentare il numero di unità considerate e di ripetere l'analisi.
b) Le ipotesi sottostanti alla procedura inferenziale sono:

- le unità considerate sono un campione casuale estratto da un certo universo;
- il campione si può ritenere proveniente da un universo con distribuzione normale bivariata.

Nel caso in esame entrambe tali ipotesi appaiono poco plausibili, poiché siamo in presenza di 25 modelli diversi di notebook di varie marche, per cui essi non costituiscono un campione casuale e le differenze da modello a modello sono ragionevolmente attribuibili a fattori sistematici e non semplicemente ad errori accidentali. Pertanto, la procedura inferenziale di verifica della significatività della correlazione non risulta completamente appropriata.

ESERCIZIO III

Le matrici delle distanze sui dati originari sono le seguenti:

Matrice delle distanze

	Distanza City Block			
	1:A	2:B	3:C	4:D
1:A	,000	28,000	18,000	15,000
2:B	28,000	,000	10,000	25,000
3:C	18,000	10,000	,000	29,000
4:D	15,000	25,000	29,000	,000

Questa è una matrice di dissimilarità

Matrice delle distanze

	Distanza euclidea			
	1:A	2:B	3:C	4:D
1:A	,000	16,733	11,832	10,817
2:B	16,733	,000	8,246	17,000
3:C	11,832	8,246	,000	17,117
4:D	10,817	17,000	17,117	,000

Questa è una matrice di dissimilarità

Per verificare se esiste una relazione monotona tra le distanze nelle due matrici si calcolano i posti d'ordine delle distanze nel triangolo superiore di ciascuna di esse. Le graduatorie non risultano identiche, per cui non vi è una relazione perfettamente monotòna.

Calcolando l'indice di cograduazione tra tali graduatorie si ottiene:

Rho di Spearman = 0.943

che segnala una relazione molto forte, anche se non perfetta, tra gli ordinamenti delle distanze ottenute con le due formule.