

## Statistica computazionale

### Esercizi del 15 maggio 2009

#### Esercizio 1

Si consideri il file in formato csv chiamato EXAM.CSV. Si calcolino i quantili della variabile Anno e della variabile Quiz. Si stampino le prime 5 osservazioni con valori più grandi e più piccoli e si stampi accanto ad ognuna di esse il sesso. Si stampi un grafico, all'interno della PROC UNIVARIATE, con le frequenze relative al risultato nel test di laboratorio, separatamente per maschi e femmine.

#### Esercizio 2

I due file EDU1.DAT e EDU2.DAT, che si trovano in V:\CARLA\SAS\COMPITI\, contengono i dati relativi ad un campione di studenti, maschi nel primo file e femmine nel secondo, di una scuola media di un certo comune. Le variabili contenute nei file sono riportate nella seguente Tabella 1.

*Tabella 1 – Variabili archivi EDU1.DAT e EDU2.DAT*

<b>variabile</b>	<b>Codici</b>
ID	Codice identificativo dello studente
GPA	Voto medio anno precedente
IQ	Punteggio test di intelligenza
GRUPPO	Gruppo test
SESSO	Sesso (M, F)
PSICO	Punteggio test psico-attitudinale

Scrivere un programma SAS che svolge le seguenti operazioni

1. crei due SAS data set temporanei corrispondenti ai file EDU1.DAT e EDU2.DAT;

*in un unico passo di DATA:*

2. crei un data set SAS permanente a radice del disco A, unendo i due archivi precedenti, in modo tale che l'archivio risultante sia composto dalle osservazioni del primo archivio più quelle del secondo, ordinatamente per codice identificativo dello studente;
3. elimini dall'archivio tutti gli studenti del gruppo 15;
4. costruisca la nuova variabile PROFILO che assuma i seguenti valori: basso se  $GPA < 6.23$  e  $IQ < 103$ , alto se  $GPA > 8.99$  e  $IQ > 118$ , medio se GPA è compreso tra 6.23 e 8.99 (estremi inclusi) e IQ è compreso tra 103 e 118 (estremi inclusi), voto < IQ se  $GPA \leq 8.99$  e  $IQ > 103$ , voto > IQ se  $GPA > 6.23$  e  $IQ \leq 118$ ;
5. elimini dall'archivio la variabile GRUPPO;
6. assegni delle etichette di descrizione a ciascuna variabile presente nell'archivio finale.

Svolgere quindi le seguenti elaborazioni sull'archivio creato:

- (a) controllare che i codici corrispondano a quelli previsti per le variabili PROFILO e SESSO;
- (b) Stampare le variabili GPA, IQ e PSICO per le femmine con  $GPA > 8$  e  $PSICO > 55$ ;
- (c) Descrivere, anche graficamente attraverso un istogramma, la distribuzione di IQ e PSICO;
- (d) Si costruisca il BOX-PLOT relativo a GPA per sesso utilizzando la procedura BOXPLOT.
- (e) Calcolare media, deviazione standard, mediana e coefficiente di variazione di GPA, IQ e PSICO per SESSO e salvare i risultati in un archivio SAS temporaneo;
- (f) Unire all'archivio contenente i dati individuali le medie per sesso della variabile GPA appena calcolate;
- (g) Definire opportuni formati per tutte le variabili utilizzate ai punti successivi, ad esclusione di PROFILO. In particolare, per la variabile IQ definire due classi in base alla media aritmetica, ponendo nella prima classe i valori inferiori alla media, e nella seconda i valori maggiori o uguali alla media, definire 4 classi per la variabile PSICO separandole in base ai quartili della distribuzione;
- (h) Costruire la tabella di IQ (variabile di riga) per PROFILO (variabile di colonna) contenente la distribuzione di frequenza in valore assoluto percentuale e la media di PSICO.
- (i) Costruire lo scatterplot di IQ per PSICO;
- (j) Costruire un modello di regressione lineare, utilizzando PSICO come variabile dipendente e IQ come variabile esplicativa.
- (k) Costruire il grafico dei residui in base ai valori IQ: possiamo affermare che il modello lineare è corretto?
- (l) A partire dall'archivio creato al punto (e), creare un nuovo archivio che contenga un'osservazione per ogni modalità di PROFILO, la variabile PROFILO e la nuova variabile STUDENTI indicante il totale di studenti per ogni PROFILO. Stampare il contenuto dell'archivio creato per controllo (il totale di STUDENTI deve essere pari al numero di osservazioni presenti nell'archivio (g)).