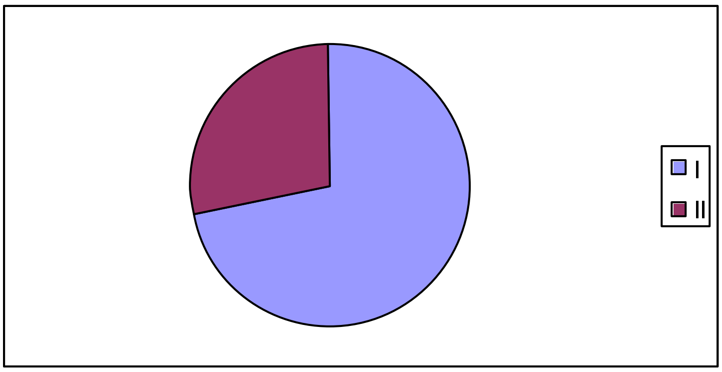
PROVA SCRITTA DI STATISTICA
CLEA-CLEFIN-CLELI (COD. 5047-4038)
6 febbraio 2003 -- A
SOLUZIONI

ESERCIZIO 1 **(8 punti)**
Uno studio legale ha 7 uffici. Per ciascuno di questi uffici, vengono rilevati i caratteri:
X: numero di personal computer presenti nell'ufficio;
Y: numero di impiegati che lavorano nell'ufficio;
Z: superficie (in metri quadrati) dell'ufficio;
W: piano in cui l'ufficio si trova.
I dati rilevati sono presentati nella tabella che segue:

UFFICIO	X	Y	Z	W
I	3	3	16	I
II	2	3	22	II
III	2	2	13	I
IV	1	1	18	I
V	3	6	34	II
VI	2	3	20	I
VII	1	2	12	I

a) Si fornisca una opportuna rappresentazione grafica del carattere W.
b) Si dica, sulla base del valore di un opportuno indice numerico, se il carattere X può ritenersi molto, poco o mediamente concentrato.
c) Si determini la tabella a doppia entrata relativa ai caratteri X e Y.
d) Si determini la funzione di regressione di Y su X.

a) Il grafico a torta di W è il seguente:



b) Il rapporto di concentrazione di Gini è R=0.2381, per cui X può ritenersi poco concentrato.

c)

X\Y	1	2	3	6
1	1/7	1/7	0	0
2	0	1/7	2/7	0
3	0	0	1/7	1/7

d) $m_2(x) = \begin{cases} 3/2 & x=1 \\ 8/3 & x=2 \\ 9/2 & x=3 \end{cases}$

<u>ESERCIZIO 2</u> (6 punti)

Si lanciano due dadi regolari a 4 facce e si considerano le variabili aleatorie X:"somma dei risultati dei due dadi" ed Y:"minimo tra i risultati dei due dadi".

a) Si calcoli il valore atteso di X.
b) Si calcoli P(X<5).
c) Si dica se X e Y sono indipendenti.

a) La distribuzione di X è:

$$p(x)\begin{cases} 1/16 & x=2 \\ 2/16 & x=3 \\ 3/16 & x=4 \\ 4/16 & x=5 \\ 3/16 & x=6 \\ 2/16 & x=7 \\ 1/16 & x=8 \end{cases}$$

Quindi E(X)=80/16=5.

b) P(X<5)=6/16=0.375.
c) X e Y non sono indipendenti. Infatti, ad esempio, P(X=8,Y=1)=0, il che esclude l'indipendenza tra X e Y.

<u>ESERCIZIO 3</u> (4 punti)

Un quiz si compone di due parti. Il tempo (in minuti) necessario ad un individuo per risolvere la I parte è una variabile aleatoria con distribuzione normale con media 12 e varianza 9, mentre il tempo necessario alla risoluzione della seconda parte è una variabile aleatoria con distribuzione normale di media 9 e varianza 7. Si possono assumere indipendenti i tempi necessari alla risoluzione delle due parti.

a) Si calcoli la probabilità che l'individuo risolva la prima parte del quiz in più di 13 minuti.

b) Si calcoli la probabilità che l'individuo risolva l'intero quiz in meno di 17 minuti.

a) Indichiamo con X il tempo necessario alla risoluzione della prima parte e con Y il tempo necessario alla risoluzione della seconda parte.

Si ha: $P(X > 13) = P(\frac{X-12}{3} > \frac{13-12}{3}) = P(Z > 0.33) = 0.3707$.

b) Indicata con W la variabile aleatoria X+Y, W ha distribuzione normale con media 21 e varianza 16. Quindi $P(W < 17) = P(\frac{W-21}{4} < \frac{17-21}{4}) = P(Z < -1) = 0.1587$.

ESERCIZIO 4 (4 punti)

Si vuole verificare se la durata media m delle comunicazioni effettuate con telefono cellulare in una determinata fascia oraria è inferiore a 100 secondi. A questo scopo, si rilevano le durate $x_1, x_2, ..., x_{12}$ in un campione di 12 comunicazioni, ottenendo $\bar{x} = 98$ e $s^2 = \frac{1}{12}\sum(x_i - \bar{x})^2 = 10$. Si assuma, per la durata di una telefonata, una distribuzione normale.

a) Si fornisca una stima, ottenuta da uno stimatore non distorto, per la varianza s^2 della durata delle telefonate.
b) Si determini una regione di rifiuto, relativa ad un test di dimensione 0.1, per l' ipotesi nulla $H_0 : m \geq 100$ contro l'alternativa $H_1 : m < 100$.
c) Si decida, sulla base dei dati del campione, se rifiutare oppure no l'ipotesi nulla utilizzando il test di cui al punto precedente.

a) Una stima per la varianza, ottenuta da uno stimatore non distorto è

$$s_c^2 = \frac{12}{11}s^2 = 10.91.$$

b) $R = \left\{(x_1, ..., x_n) : \frac{\bar{x} - 100}{s_c / \sqrt{12}} \leq -1.363\right\}$.

c) Poiché $\frac{\bar{x} - 100}{s_c / \sqrt{12}} = -2.0975$, l'ipotesi nulla è da rifiutare.

ESERCIZIO 5 (5 punti)

Si consideri un modello lineare $Y = b_1 x + b_0 + e$ per i dati $(x_1, y_1), ..., (x_n, y_n)$, con

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 2, \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 1, \bar{x} = 3, \bar{y} = 2.$$

a) Si determinino stime, ottenute da stimatori non distorti, per i parametri b_0 e b_1 del modello.
b) Sapendo che (0.32,0.68) è un intervallo di confidenza per b_1 con coefficiente di confidenza 0.95, si decida se rifiutare oppure no, al livello 0.05, l'ipotesi nulla $b_1 = 0$ contro l'alternativa $b_1 \neq 0$. Motivare la risposta.

c) Si dica quale significato assumerebbe, con riferimento al modello lineare, il mancato rifiuto dell'ipotesi nulla $b_1 = 0$.

a) Le stime dei minimi quadrati di b_0 e b_1 sono:

$$\hat{b}_0 = \overline{y} - \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\overline{x} = 0.5 \quad \text{e} \quad \hat{b}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0.5.$$

b) Per decidere se rifiutare l'ipotesi nulla al livello 0.05 si può rilevare se il valore 0 appartiene oppure no all'intervallo di confidenza di coefficiente 0.95; poiché l'intervallo (0.32,0.68) non contiene 0, rifiutiamo l'ipotesi nulla.

c) Il mancato rifiuto dell'ipotesi nulla indicherebbe inadeguatezza del modello lineare a spiegare la relazione tra le variabili X e Y.