

## I COMPITO (3 crediti)

### ESERCIZIO 1

E' stata effettuata una ricerca per studiare se il raccolto di una particolare specie di cereale Y (t/ha) è correlato linearmente con la quantità di fertilizzante applicato X (kg/ha). Sono state applicate diverse quantità di fertilizzante (da 0 a 4kg/ha) su diversi appezzamenti e per ognuno è stata misurata la quantità di raccolto.

Si è ottenuta la retta di regressione:

$$y = 4.85 + 0.05x$$

i) Si tratta di uno studio di osservazione

**V F**

ii) Quale delle seguenti affermazioni è corretta?

**a b c**

- (a) La variabile esplicativa è il raccolto di cereale
- (b) Se si applicano 30kg/ha di fertilizzante il raccolto stimato è pari a 6.35t/ha.
- (c) Per ottenere un raccolto stimato pari a 5.05t/ha occorre applicare 4kg/ha di fertilizzante.

iii) Il coefficiente di correlazione tra le due variabili è pari a 0.68. Calcolate il coefficiente di determinazione e spiegate cosa rappresenta in questo contesto.

0.46

iv) Qual è l'incremento di raccolto che ci si può attendere utilizzando un kg/ha aggiuntivo di fertilizzante?

0.05 t/ha

v) Qual è l'unità di misura del coefficiente angolare in questo contesto?

nessuna

vi) Spiegate se si può dire che esiste una relazione di causa effetto tra le due variabili.

### ESERCIZIO 2

In una classe il 15% degli studenti ha ottenuto un voto superiore a 7. Assumendo che la distribuzione dei voti sia approssimativamente normale, con una deviazione standard pari a 1.5, calcolare il voto medio degli studenti.

$$\mu = 7 - 1.5 * (\text{inv.norm.st}(0.85)) = 5.45$$

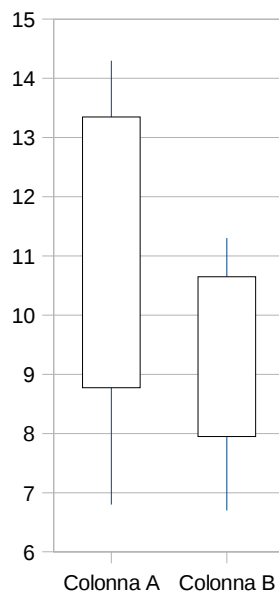
### ESERCIZIO 3

I dati che seguono sono le misure, ottenute con un particolare strumento, che misura come cambia la resistenza della superficie di una foglia. Qualunque trattamento che influenzi la chiusura degli stomi della foglia fa aumentare la resistenza della foglia e viceversa.

Sono stati considerati 2 trattamenti I e II ciascuno con sei osservazioni prese nell'ordine in cui compaiono. Le osservazioni si riferiscono alle misure di resistenza delle superfici delle foglie **QUI MANCA IL PUNTO** Si vogliono confrontare i 2 trattamenti.

<u>I trattamento</u>	<u>II trattamento</u>
8,2	7,4
10,5	9,6
6,8	6,7
14,3	10,9
13,4	9,9
13,2	11,3

i) Per confrontare graficamente i due trattamenti riportate qui sotto i box-plots dei due campioni di



osservazioni **affiancandoli**.

ii) Usate tutti gli indici sintetici ( di posizione e di variabilità) opportuni per effettuare, anche, un confronto numerico relativo ai 2 trattamenti. Calcolate anche il coefficiente di variazione. Riportate i risultati ottenuti.

--

iii) Commentate sulla base dei risultati ottenuti, sia grafici, sia numerici, le differenze tra i 2 trattamenti.

--

v) Accorpate le osservazioni relative ai 2 trattamenti.e per le 12 osservazioni ottenute costruite il normal plot senza riportarlo sul foglio. Si può dire che i dati seguono una distribuzione approssimativamente normale? Spiegate perché.

<p>Si', buona linearita' del normal plot.</p>
---

vi) Supponendo che le 12 osservazioni siano distribuite approssimativamente secondo una normale:  
a) calcolate media e scarto standard di questo campione di dati, b) calcolate a quale valore corrisponde il 60-mo percentile.

<p>Mu=10.66, sigma=2.58  P(X&lt;=11.31)=60%</p>
---

**ESERCIZIO 4**

In uno studio per analizzare gli effetti, sul raccolto, della bruciatura di stoppie, sono state misurate quattro variabili. Per ciascuna specificate di che tipo di variabile si tratta:

- 1) pH del suolo (misurato riportando una cifra decimale)
- 2) tipo di coltivazione (0=grano, 1=orzo, 2=avena, 3=altro)
- 3) quantità di stoppia bruciata (0=piccola, 1=media, 2=grande)
- 4) data finale del raccolto (ad es., 10 Set. 2006)

<ol style="list-style-type: none"> <li>1)</li> <li>2)</li> <li>3)</li> <li>4)</li> </ol>	
--	--

## ESERCIZIO 5

In un sondaggio sulle assunzioni di alcolici è stata condotta un'inchiesta tra gli studenti di un college scegliendo a caso 10 classi differenti e intervistando tutti gli studenti delle classi selezionate.

Si tratta di un campionamento casuale

**V**      **F**

## COMPITO 2 ( 3 crediti)

### ESERCIZIO 1

Ad alcuni topi, sono state somministrate, in modo casuale, diverse dosi di Alar. Si tratta di un composto chimico spruzzato sugli alberi di mele per evitare che le mele cadano prima di giungere a maturazione. Dopo un determinato tempo si è studiato il numero di tumori che avevano colpito i topi.

a)

Sono state misurate quattro variabili per ogni topo. Per ciascuna specificate di che tipo di variabile si tratta:

- 1) sesso (0=femmina, 1=maschio)
- 2) peso (g)
- 3) dose di Alar ( nessuna, piccola, grande)
- 4) numero di tumori (il numero di tumori possibili è circa 7)

1)	
2)	
3)	
4)	

b)

Si tratta di uno studio di osservazione o di un esperimento? Spiegate

--

### ESERCIZIO 2

1. Si vuole studiare se esiste una relazione tra le misurazioni ottenute con lo strumento A e quelle ottenute rispettivamente con gli strumenti B e C nelle stesse condizioni. Calcolate il coefficiente di

correlazione, il coefficiente angolare e la retta di regressione (forse **coeff.angolare e intercetta della retta di regressione?**) per i 2 insiemi di dati accoppiati: X1, Y1; X2, Y2. Riportate qui sotto i risultati ottenuti. Cosa potete osservare? Commentate.

X1(A)	Y1(B)	X2(A)	Y2(C)
10	8,04	10	9,14
8	6,95	8	8,14
13	7,58	13	8,74
9	8,81	9	8,77
11	8,33	11	9,26
14	9,96	14	8,10
6	7,24	6	6,13
4	4,26	4	3,10
12	10,84	12	9,13
7	4,82	7	7,26
5	5,68	5	4,74

1)  
corr = 0.8164  
pendenza=0.5001  
intercetta=3.0001

2)  
0.8162  
0.5000  
3.0009

Commenti

2. Disegnate un diagramma di dispersione con la relativa retta di regressione per ciascuno dei 2 insiemi di dati (senza riportarli sul foglio). In quale dei 2 casi usereste la retta di regressione per prevedere il valore di y dato x=14? Riportate tale valore.

Userei il primo caso (curiosita': **il secondo come e' stato ottenuto? :**)  
Valore=10,001

3. Disegnate i grafici dei residui relativi ad entrambe le rette (senza riportarli sul foglio). Dal confronto di tali grafici, vi sembra che la retta di regressione rappresenti un buon modello per i dati in entrambi i casi? Spiegate perché si o perché no.

1

**ESERCIZIO 3**

i) Accorpate in un'unica colonna i due campioni Y1 e Y2. Per i dati accorpati disegnatel'istogramma, considerando sull'asse delle ordinate le **frequenze relative percentuali**. Riportatelo sul foglio qui di seguito. Dividete i dati nelle classi che seguono:

3.10—4.30

4.30—5.50

5.50—6.70

6.70—7.90

7.90—9.10

9.10—10.30

10.30—11.50

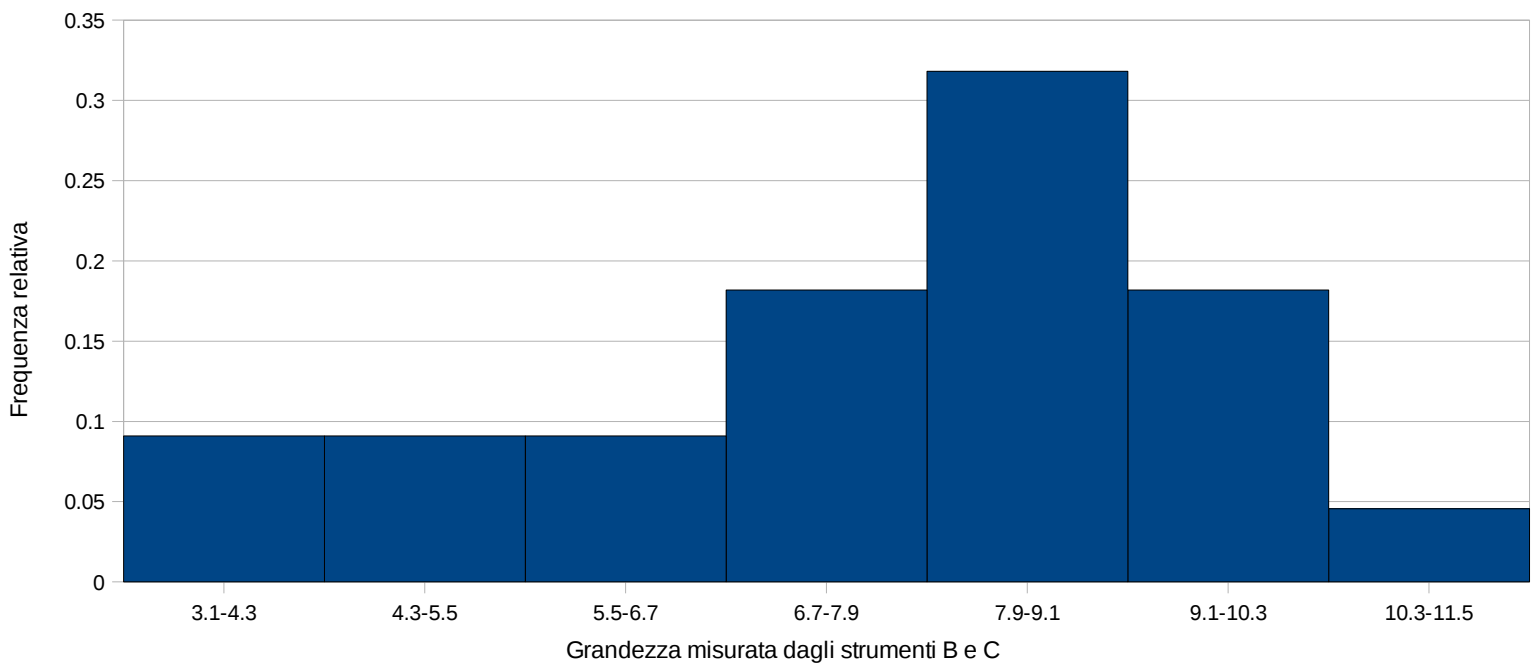
ii) Il grafico mostra che i dati sono approssimativamente normali. a) Calcolate la media e lo scarto standard di questo campione di dati. b) Calcolate a quale valore corrisponde il 70-mo percentile

.

a)media=7.5 ; dev.st=1.98

b)valore (normale) corrispondente al 70mo percentile= 8.54

Istogramma



#### ESERCIZIO 4

Per una distribuzione normale standard calcolare la percentuale di valori che sono distanti più di 2 deviazioni standard dalla media.

4.6%

#### ESERCIZIO 5

Un botanico vuole ricavare dei dati riguardo alle piante cresciute in appartamento. Il campione viene ottenuto telefonando alle prime 250 persone registrate nell'elenco telefonico. Si tratta di un campione casuale.

V

F

#### ESERCIZIO 6

In un campione di 50 osservazioni la frequenza di una data classe 25—30 è pari a 10. Inoltre, la frequenza cumulata di tutte le classi superiori a questa classe è pari a 40. La frequenza cumulata per questa classe è:

- a) 10
- b) 50
- c) 40
- d) nessuna delle precedenti

Risposta a)

In una scuola agli esami di maturità si sono presentati 200 studenti. 50 di questi hanno preso voti tra 70 ed un numero  $X$  da determinare. Possiamo assumere che la distribuzione dei voti sia ben approssimata da una distribuzione normale con media pari a 75 e deviazione standard pari a 15. Determinare il numero  $X$ , approssimato all'intero più vicino.

Un'analisi di due liste di dati ha prodotto una retta di regressione di equazione

$$Y = 14 + b \cdot X.$$

Calcolare  $b$  sapendo che il residuo di  $Y$  quando  $X = 5$  e  $Y = 21$  risulta essere  $-2$ .

Possiamo affermare che tale modello lineare ben interpreta i dati?



Si vuole risalire all'equazione della retta di regressione di due insiemi di dati

$$Y = b \cdot X + a$$

sapendo che la variabile  $X$  ha media 25 e dev. standard 4,

la variabile  $Y$  ha media 10 e dev. standard 1.5

il coefficiente di correlazione vale  $r=0.8$

Calcolare i parametri  $a$  e  $b$ .

Determinare il rapporto tra la variabilità del peso  $Y$  spiegata dalla retta di regressione e la variabilità totale.

$$X = N(m=5) \quad P(X < 0) = 10\%; \quad s? \quad \square \quad z = -1,28 \quad s = 3,91$$