

Esercizi di Statistica della 8^a settimana (Corso di Laurea in Biotecnologie, Università degli Studi di Padova).

Esercizio 1. In uno studio compiuto di recente, si è vista la variazione del livello di colesterolo (in mg/dl) dopo aver adottato una dieta vegetariana, con i risultati nella seguente tabella.

variazione	< 10	10–30	> 30
numero casi	6	11	7

Vogliamo ora testare l'ipotesi che questi dati provengano da una distribuzione $N(19.8; 16.8^2)$.

1. Se $X \sim N(19.8; 16.8^2)$, calcolare la probabilità che X appartenga ad ogni intervallo della tabella sopra.
2. Testare l'ipotesi che i dati sopra provengano da una distribuzione $N(19.8; 16.8^2)$, riportando limitazioni al valore P .
3. Si può concludere che i dati nella tabella provengano da una distribuzione normale?

Esercizio 2. In una certa zona, i dati in possesso delle assicurazioni dicono che in un anno il numero di incidenti per automobilista segue una legge di Poisson. In particolare, su un campione di 440 automobilisti, nell'ultimo anno 366 non hanno avuto incidenti, 68 ne hanno avuto uno, 5 ne hanno avuti 2 e 1 ne ha avuti 3.

1. Stimare la media λ di incidenti pro-capite del campione.
2. Chiamato $\hat{\lambda}$ lo stimatore del punto 1., calcolare la probabilità che un automobilista medio abbia 0, 1, 2 o almeno 3 incidenti in un anno se il numero di incidenti ha distribuzione di Poisson con parametro uguale a $\hat{\lambda}$.
3. Possiamo concludere che il campione abbia distribuzione di Poisson? Fare un test χ^2 di adattamento usando $\alpha = 0.05$.
4. Riportare limitazioni al valore P per il punto 3.

Esercizio 3. I dati nella tabella seguente sono misure che mettono in relazione un particolare parametro di funzionalità epatica (SGOT) con il livello di colesterolo HDL nel sangue.

SGOT	9.5	11.0	13.5	15.5	17.5	19.5	20.5
HDL (mg/dL)	40.0	41.2	42.3	42.8	43.8	43.6	46.5

Abbiamo quindi i seguenti dati aggregati ($x = \text{SGOT}$, $y = \text{HDL}$):

$$\sum x_i = 107, \quad \sum x_i^2 = 1740.5, \quad \sum x_i y_i = 4637.6$$

$$\sum y_i = 300.2, \quad \sum y_i^2 = 12900.2.$$

1. Trovare la retta di regressione tra il livello di SGOT (var. indipendente) e il livello di HDL (var. dipendente).

2. Fare un test di livello $\alpha = 0.05$ per vedere se c'è una relazione lineare tra le due quantità di cui sopra.
3. Supponiamo che nella misurazione dei dati sia stato fatto un errore sistematico per cui i livelli di SGOT sono da alzare tutti di 0.5, e che invece i livelli di HDL siano giusti. Che effetto ha questo errore sulle risposte date ai punti 1. e 2.? (suggerimento: non sono necessari ulteriori calcoli)

Esercizio 4. I seguenti dati si riferiscono all'area di contaminazione dell'acqua col passare del tempo, dovuta alla fuga di un agente chimico tossico in un punto (l'area Y è misurata in acri, il tempo X in anni):

anni (x)	1.3	2.4	4.4	0.5	3.6
acri (y)	4.8	5.3	19.7	1.5	10.1

1. Trovare la retta di regressione tra il tempo di esposizione all'agente tossico e l'area inquinata.
2. Che test bisogna fare per stabilire se c'è una relazione lineare significativa tra le due quantità?
3. Eseguire il test del punto 2, riportando il valore P .

Esercizio 5. In un elettrocardiogramma sotto sforzo vengono misurate i battiti cardiaci al minuto (bpm) in funzione della potenza erogata. Per un certo individuo sono stati misurati i seguenti dati:

potenza (Watt)	bpm
0	61
30	77
60	85
90	95
120	111
150	125
180	140
210	151
240	167

Si ha quindi

$$\sum x_i = 1080, \quad \sum x_i^2 = 183600, \quad \sum x_i y_i = 145020$$

$$\sum y_i = 1012, \quad \sum y_i^2 = 124136,$$

1. Trovare la retta di regressione tra la potenza erogata e i battiti cardiaci.
2. Calcolare l'intervallo di confidenza al 95% sul coefficiente angolare.

Soluzioni su <http://www.math.unipd.it/~vargiolu/Statistica/>