

Esercizi di Statistica della 11^a settimana (Corso di Laurea in Biologia Molecolare, Università degli Studi di Padova).

Esercizio 1. 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 = SGOT, y = 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), e stimare la varianza del rumore.
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 2. Il referendum del 2011 ha evidenziato che sembra esserci una relazione, in un referendum svolto in 2 giornate, tra l'affluenza parziale alla fine del primo giorno e la percentuale finale di votanti, cruciale per il quorum. Di seguito ci sono i dati degli ultimi referendum in cui si è votato in due giornate:

anno (argomento)	affluenza parziale	affluenza finale
1990 (caccia e pesticidi):	31,5%	43,4%
1991 (preferenza unica alla Camera):	45,7%	62,2%
1993 (droghe leggere, 2 ministeri, maggioritario, ecc.):	57,2%	77,0%
2003 (licenziamento senza giusta causa):	17,5%	25,7%

Vogliamo vedere se ci può essere una relazione lineare tra i due dati.

1. Calcolare i coefficienti della retta di regressione.
2. Eseguire un test con $\alpha = 5\%$ per vedere se la dipendenza lineare è significativa.
3. Calcolare l'intervallo di previsione al 95% per una affluenza parziale uguale al 41%.
4. Con una tale affluenza parziale era ragionevole affermare che il quorum del 50% sarebbe stato raggiunto?

Esercizio 3. Il termine “regressione” fu usato per la prima volta da Francis Galton a proposito delle leggi sull’ereditarietà. Galton pensava che tali leggi prescrivessero per la discendenza degli estremi di una popolazione una “regressione verso la media”, intendendo con questo che i figli di individui con caratteristiche eccezionalmente alte o basse tendono ad essere più nella media dei genitori.

Lo statistico Karl Pearson, per dimostrare la tesi di Galton, prese le stature (in pollici) di 10 figli e le confrontò con quelle dei loro padri:

Padre	60	62	64	65	66	67	68	70	72	74
Figlio	64	65	66	65	67	67	67	68	70	70

con le seguenti statistiche cumulate:

$$\sum_{i=1}^{10} X_i = 668, \quad \sum_{i=1}^{10} X_i^2 = 44794, \quad \sum_{i=1}^{10} X_i Y_i = 44766$$

$$\sum_{i=1}^{10} Y_i = 669, \quad \sum_{i=1}^{10} Y_i^2 = 44793,$$

1. Calcolare la retta di regressione per questi dati.
2. Calcolare l’errore standard del coefficiente angolare.

Vogliamo ora controllare se i dati evidenziano in modo significativo un coefficiente angolare minore di 1. Usiamo sempre $\alpha = 0.05$.

3. Fare un test unilatero con $H_0 : \beta_1 = 1$ e $H_1 : \beta_1 < 1$. Galton aveva ragione?
4. Fare un test bilatero con $H_0 : \beta_1 = 0.5$ e $H_1 : \beta_1 \neq 0.5$. Perché un test proprio per $\beta_1 = 0.5$?

Esercizio 4. I dati seguenti provengono da un esperimento di Francis Galton, e mettono in relazione il diametro di un seme discendente con quello del seme genitore, in mm.

diámetro del seme genitore (x)	15	16	17	18	19	20	21
diámetro del seme discendente (y)	15.3	16.4	15.5	16.2	16.0	17.4	17.5

1. Trova la retta di regressione tra i diametri dei semi genitori e quelli dei semi discendenti.
2. Trovare l’intervallo di confidenza al 95% del coefficiente angolare.
3. La relazione lineare è significativa? Fare un test con $\alpha = 0.05$.
4. Fare un test sulla “regressione alla media”, ponendo $H_0 : \beta_1 \geq 1$ e $H_1 : \beta_1 < 1$ e fornendo limitazioni al valore P . Quali sono le conclusioni?