

Esercizi di Statistica della 12^a settimana (Corso di Laurea in Biologia Molecolare, Università degli Studi di Padova).

Esercizio 1. Una malattia virale ha infettato il 25% della popolazione suina degli allevamenti di alcuni stati del sud degli Stati Uniti. Un test per la presenza del virus dà esito positivo l'84% delle volte in cui il maiale è effettivamente malato ed esito negativo l'80% delle volte in cui il maiale non ha la malattia. Si prende un maiale e lo si sottopone al test per la presenza del virus.

1. Se il test è positivo, qual è la probabilità che il maiale sia stato davvero infettato dal virus?
2. Se il test è negativo, qual è la probabilità che il maiale non sia stato infettato dal virus?
3. Qual è la probabilità che il test dia una diagnosi corretta?

Esercizio 2. L'azienda QuickTune offre una completa messa a punto dell'automobile a 34.98 Euro, e garantisce che la messa a punto sarà gratis se non verrà completata in non più di trenta minuti. Il tempo richiesto per la messa a punto è una variabile aleatoria X gaussiana con $\mu = 23.2$ minuti e $\sigma = 4.17$ minuti. Se ogni messa a punto costa alla QuickTune in media 24 Euro, allora:

1. Qual è la legge della variabile aleatoria che rappresenta il guadagno su una singola messa a punto?
2. Quanto guadagno si può attendere la ditta sulle prossime 200 messe a punto?
3. Che varianza si aspetta?

Esercizio 3. In un certo procedimento chimico, è di fondamentale importanza che il pH di uno dei reagenti sia esattamente 8.20. Supponiamo che 10 misurazioni indipendenti abbiano dato i seguenti valori:

8.18 8.16 8.17 8.22 8.19 8.17 8.15 8.21 8.16 8.18

1. Effettuare un test per vedere se effettivamente il valore del pH si può considerare uguale a 8.20. Riportare limitazioni al valore P .

Supponiamo ora di confrontare questo campione con un altro campione ottenuto con una procedura standard che dà in effetti un pH medio di 8.20, e di avere le richieste seguenti: se il pH medio è uguale tra i due campioni, il test deve affermarlo con probabilità 95%; d'altra parte, se il pH medio differisce tra i due campioni di almeno 0.03, tale differenza deve essere evidenziata almeno nel 95% dei casi. Supponiamo anche che la vera deviazione standard sia circa uguale a quella stimata in precedenza.

2. Quanto numerosi dovranno essere i campioni?
3. Supponiamo che si confronti il vecchio campione con un nuovo campione avente $n_Y = 15$, $\bar{Y} = 8.21$ e $s_Y = 0.025$: che conclusione si trarrebbe?

Esercizio 4. In uno studio (Senior-Mazza, The Lancet Oncology 2004) è stata studiata l'incidenza di diversi tipi di tumore nella famigerata "terra dei fuochi" campana. Le incidenze di tumore significative si sono rivelate, secondo gli studiosi, quelle al fegato, colon-retto e leucemia. I dati pubblicati, in morti su 100.000 abitanti, sono i seguenti:

tipo di tumore	Italia	Campania	ASL NA4	Distretto 73 di ASL NA4 ("terra dei fuochi")
Fegato (uomini)	14.0	15.0	38.4	35.9
Fegato (donne)	6.0	8.5	20.8	20.5
Colon-retto (uomini)	34.4	26.4	27.2	23.6
Colon-retto (donne)	31.4	26.4	21.2	29.0
Leucemia (U)	10.1	10.1	8.3	13.1
Leucemia (D)	7.6	7.5	6.7	7.8

Vogliamo ora vedere se alcune di queste incidenze sono significativamente diverse dalla media italiana. Supponiamo che nel Distretto 73 ci siano circa 125.000 uomini e 125.000 donne.

1. Quanti morti ci sono stati per tumore al fegato nel Distretto 73, sia tra gli uomini che tra le donne?
2. Calcolare un intervallo di confidenza al 95% per il tasso di incidenza del tumore al fegato sia tra gli uomini che tra le donne del Distretto 73.
3. Condurre un test di ipotesi, con $\alpha = 0.05$, per indagare se si può ritenere che il tasso di tumore al fegato nel Distretto 73 sia diverso dalla media nazionale (supponendo quest'ultimo come noto e uguale ai dati nella seconda colonna): fare il test separatamente per gli uomini e per le donne.
4. Riportare limitazioni al valore P per i test del punto precedente.

Esercizio 5. I dati seguenti rappresentano la quantità in milioni di libbre di prodotti contenenti carni bianche consumate negli Stati Uniti negli anni dal 1994 al 2002:

anno (x)	1994	1996	1998	2000	2002
quantità (y)	25.9	26.8	27.8	30.5	32.6

con le seguenti statistiche cumulative:

$$\sum X_i = 9990, \quad \sum Y_i = 143,6, \quad \sum X_i Y_i = 286947, \quad \sum X_i^2 = 19960060, \quad \sum Y_i^2 = 4154,9$$

1. Trova la retta di regressione tra gli anni ed il consumo di carni bianche.
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. Riportare limitazioni al valore P .
5. Sembra che il consumo di carni bianche stesse aumentando in quegli anni?