Esame di Statistica del 16 aprile 2007 (Corso di Laurea Triennale in Biotecnologie, Università degli Studi di Padova).

Cognome Nome Matricola

Es. 1	Es. 2	Es. 3	Es. 4	Somma	Voto finale

Attenzione: si consegnano SOLO i fogli di questo fascicolo.

Esercizio 1. Il 46% degli elettori di un comune si ritiene di centro, il 30% di sinistra e il 24% di destra. Alle ultime elezioni sono andati a votare il 35% degli elettori di centro, il 62% di quelli di sinistra e il 58% di quelli di destra. Un elettore è scelto a caso. Sapendo che l'elettore ha votato alle ultime elezioni, qual è la probabilità che si tratti di un elettore:

1. del centro;
2. di sinistra;
3. di destra.
4. Quale percentuale di elettori ha partecipato alla scorsa elezione?

Esercizio 2. La produzione di grossi trasformatori e condensatori elettrici richiede l'impiego di sostanze tossiche (PCB), molto pericolose quando vengono disperse nell'ambiente. Si vogliono confrontare due metodi per misurare il livello di PCB nel pesce di un lago nelle cui prossimità vi è un impianto di grosse dimensioni. I risultati delle misurazioni sono riportati sotto.

Lago A	11.5	10.8	11.6	9.4	12.4	11.4	12.2	11.0	10.6	10.8
Lago B	11.8	12.6	12.2	12.5	11.7	12.1	10.4	12.6		

1. Calcolare l'intervallo di confidenza al 95% per la differenza delle medie tra i campioni.
2. Supponendo che la varianza tra le popolazioni sia la stessa, dire se da questi dati si può supporre che i due metodi producano misurazioni confrontabili. Riportare limitazioni al valore P.

Esercizio 3. In un celebre esperimento per determinare l'efficacia dell'acido acetilsalicilico nella prevenzione degli infarti, 22000 uomini di mezza età vennero divisi casualmente in due gruppi, e fu loro somministrata una dose giornaliera del farmaco o di un placebo. Alla conclusione dell'esperimento, erano stati colpiti da infarto 104 uomini nel gruppo trattato e 189 in quello di controllo.

1. In base a questi dati si può dire che l'assunzione quotidiana di acido acetilsalicilico modifichi la probabilità di subire un infarto? Usare $\alpha = 0.05$.
2. Calcolare l'intervallo di confidenza al 95% delle differenze tra i due tassi di cui sopra.
3. Durante lo stesso periodo, 119 uomini del gruppo trattato e 98 del gruppo di controllo subirono un ictus. Si può dire che l'assunzione quotidiana di acido acetilsalicilico modifichi la probabilità di subire un ictus? Usare $\alpha = 0.01$.
4. Calcolare l'intervallo di confidenza al 95% delle differenze tra i due tassi di cui sopra.

Esercizio 4. I dati seguenti illustrano la relazione esistente tra il prezzo unitario di un certo bene in 6 luoghi differenti e il numero di unità dello stesso bene che sono state ordinate.

Prezzo	50	40	35	30	20	15
n. pezzi	88	112	123	136	158	172

con le seguenti statistiche cumulative:

$$\sum_{i=1}^{6} x_i = 190, \qquad \sum_{i=1}^{6} y_i = 789, \qquad \sum_{i=1}^{6} x_i^2 = 6850, \qquad \sum_{i=1}^{6} y_i^2 = 108461, \qquad \sum_{i=1}^{6} x_i y_i = 23005$$

1. Calcolare la retta di regressione del numero di pezzi ordinati rispetto al prezzo.
2. Eseguire un test per vedere se in effetti c'è una relazione lineare. Riportare limitazioni al valore P.
3. Stimare il numero di unità ordinate se il prezzo fosse 25.
4. Fornire un intervallo di confidenza al 95% per la quantità del punto 3.

Soluzioni

Esercizio 1. Definiamo gli eventi

$$\begin{aligned} A_1 &:= \{\text{l'elettore è di sinistra}\}, & A_2 &:= \{\text{l'elettore è di centro}\}, \\ A_3 &:= \{\text{l'elettore è di destra}\}, & V &:= \{\text{l'elettore ha votato alle ultime elezioni}\}, \end{aligned}$$

allora i dati si possono riscrivere così:

$$\begin{aligned} \mathbb{P}(A_1) = 0.3, \qquad \mathbb{P}(A_2) &= 0.46, \qquad \mathbb{P}(A_3) = 0.24, \\ \mathbb{P}(V|A_1) = 0.62, \qquad \mathbb{P}(V|A_2) &= 0.35, \qquad \mathbb{P}(V|A_3) = 0.58. \end{aligned}$$

Ovviamente per fare i primi 3 punti è indispensabile fare prima il punto 4.

4) Dato che $\cup_{i=1}^{3} A_i = \Omega$ e l'unione è disgiunta, usando la formula della probabilità totale calcoliamo

$$\mathbb{P}(V) = \sum_{i=1}^{3} \mathbb{P}(V|A_i)\mathbb{P}(A_i) = 0.62 \cdot 0.3 + 0.35 \cdot 0.46 + 0.58 \cdot 0.24 = 0.4862$$

1,2,3) Usando la formula di Bayes, le probabilità richieste sono rispettivamente:

$$\begin{aligned} \mathbb{P}(A_2|V) &= \frac{\mathbb{P}(V|A_2)\mathbb{P}(A_2)}{\mathbb{P}(V)} = \frac{0.35 \cdot 0.46}{0.4862} = 0.331139, \\ \mathbb{P}(A_1|V) &= \frac{\mathbb{P}(V|A_1)\mathbb{P}(A_1)}{\mathbb{P}(V)} = \frac{0.62 \cdot 0.3}{0.4862} = 0.382559, \\ \mathbb{P}(A_2|V) &= \frac{\mathbb{P}(V|A_3)\mathbb{P}(A_3)}{\mathbb{P}(V)} = \frac{0.58 \cdot 0.24}{0.4862} = 0.286302 \end{aligned}$$

Esercizio 2.

1. Calcolando media e varianza campionarie dei due gruppi otteniamo per il primo $\bar{X} = 11.17$ e $s_X^2 = 0.7423$ e per il secondo $\bar{Y} = 11.98$ e $s_Y^2 = 0.5298$, quindi $s_{\bar{X}-\bar{Y}} = \sqrt{s_X^2/10 + s_Y^2/8} = 0.3904$. Il numero di gradi di libertà è $\nu = 10 + 8 - 2 = 16$, e quindi il quantile da usare è $t_{1-\alpha/2}(\nu) = t_{0.975}(16) = 2.119$. Allora l'intervallo di confidenza ha estremi $\bar{X} - \bar{Y} \pm s_{\bar{X}-\bar{Y}} t_{1-\alpha/2}(\nu) = -0.81 \pm 0.3904 \cdot 2.119$ e risulta uguale a $[-1.637; 0.017]$.

2. Facciamo un test t sulla differenza delle medie con ipotesi $H_0 : \mu_X = \mu_Y$ e alternativa $H_1 : \mu_X \neq \mu_Y$. Supponendo che la varianza delle popolazioni iniziali sia la stessa, possiamo usare un test di Student. Abbiamo
$$t = \frac{\bar{X} - \bar{Y}}{s_{\bar{X}-\bar{Y}}} = \frac{11.17 - 11.98}{0.3904} = -2.074$$
I gradi di libertà sono $\nu = 14$. Siccome $1.761 = t_{0.95}(14) < |t| < t_{0.975}(14) = 2.144$, possiamo riportare $0.05 < P < 0.1$. Siccome il valore P non è abbastanza basso, non abbiamo elementi sufficienti a concludere che i due metodi producano risultati differenti.

Esercizio 3.

1. Vogliamo fare un test di ipotesi $H_0 : p_1 = p_2$ contro l'alternativa $H_1 : p_1 \neq p_2$. Per effettuare il test possiamo utilizzare il test Z oppure il test χ^2, che producono gli stessi risultati. Utilizzando il metodo del test Z, calcoliamo le seguenti quantità:

$$\hat{p}_1 = \frac{104}{11000} = 0.00945, \quad \hat{p}_2 = \frac{189}{11000} = 0.01718, \quad \hat{p} = \frac{104 + 189}{11000 + 11000} = 0.01331$$

   Siccome $n_1 \cdot \hat{p} = n_2 \cdot \hat{p} = 11000 \cdot 0.01331 = 146.5 > 5$, si puó usare il metodo del test Z:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}}} = -4.999$$

   Il valore critico è $q_{0.975} = 1.96$; siccome $|Z| > q_{0.975}$, rifiutiamo H_0 in favore di H_1: sembra che l'assunzione quotidiana di acido acetilsalicilico modifichi la probabilità di subire un infarto.

2. L'intervallo di confidenza cercato è $\hat{p}_1 - \hat{p}_2 \pm q_{0.975} s_{\hat{p}} = -0.00773 \pm 1.96 \cdot 0.00154$, dove

$$s_{\hat{p}} = \sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}} = 0.00154$$

   quindi l'intervallo di confidenza è $[-0.0107; -0.0047]$.

3. Anche in questo caso vogliamo fare un test di ipotesi $H_0 : p_1 = p_2$ contro l'alternativa $H_1 : p_1 \neq p_2$. Utilizzando il metodo del test Z, calcoliamo le seguenti quantità:

$$\hat{p}_1 = \frac{119}{11000} = 0.01082, \quad \hat{p}_2 = \frac{98}{11000} = 0.00891, \quad \hat{p} = \frac{119 + 98}{11000 + 11000} = 0.00986$$

   Siccome $n_1 \cdot \hat{p} = n_2 \cdot \hat{p} = 11000 \cdot 0.00986 = 108.5 > 5$, si puó usare il metodo del test Z:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}}} = 1.432$$

   Il valore critico stavolta è $q_{0.995} = 2.57$; siccome $|Z| < q_{0.995}$, accettiamo H_0: sembra che l'assunzione quotidiana di acido acetilsalicilico non modifichi la probabilità di subire un ictus.

4. L'intervallo di confidenza cercato è $\hat{p}_1 - \hat{p}_2 \pm q_{0.975} s_{\hat{p}} = 0.00191 \pm 1.96 \cdot 0.00274$, dove

$$s_{\hat{p}} = \sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}} = 0.00274$$

   quindi l'intervallo di confidenza è $[-0.0035; 0.0073]$.

Esercizio 4. Partiamo calcolando le quantità:

$$\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i = 31.666 \quad (0.5 \text{ punti}), \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 131.5 \quad (0.5 \text{ punti}),\\
s_X^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = 166.666 \quad (0.5 \text{ punti}),\\
s_Y^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right) = 941.5 \quad (0.5 \text{ punti}),\\
s_{XY} &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y}\right) = -396 \quad (0.5 \text{ punti})
\end{aligned}$$

1. Calcoliamo i coefficienti della retta di regressione:

$$\begin{aligned}
b_1 &= \frac{s_{XY}}{s_X^2} = -2.376\\
b_0 &= \bar{Y} - b_1\bar{X} = 206.740
\end{aligned}$$

   La retta di regressione è quindi $y = -2.376x + 206.74$.

2. Bisogna eseguire un test di ipotesi $H_0 : \beta_1 = 0$ e alternativa $H_1 : \beta_1 \neq 0$. Bisogna allora calcolare:

$$\begin{aligned}
s_{Y|X} &= \sqrt{\frac{n-1}{n-2}(s_Y^2 - b_1^2 s_X^2)} = 0.869\\
s_{b_1} &= \frac{s_{Y|X}}{\sqrt{(n-1)\cdot s_X^2}} = 0.030
\end{aligned}$$

   Abbiamo allora

$$t = \frac{b_1 - 0}{s_{b_1}} = -78.93$$

   Bisogna confrontare $|t|$ con una legge di Student a $\nu = 7 - 2 = 5$ gradi di libertà. Abbiamo che $t_{0.999}(4) = 7.173 < |t|$, quindi $P < 0.002$. Con una P così bassa siamo portati a rifiutare H_0 in favore di H_1, il che significa che sembra esserci una relazione lineare.

3. La media che ci attendiamo per un prezzo unitario di 25 è $y = b_1x + b_0 = 147.34$.

4. L'errore standard della stima della y è

$$s := s_{Y|X}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{X})^2}{(n-1)s_x^2}} = 0.959$$

   e $t_{0.975}(4) = 2.776$, quindi l'intervallo di confidenza cercato ha estremi $y \pm t_{1-\alpha/2}(\nu)s = 147.34 \pm 2.776 \cdot 0.959$, ed è quindi $[144.677; 150.002]$

Esame di Statistica del 16 aprile 2007 (Corso di Laurea in Biotecnologie, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Hanno superato la prova:

Barzaghi Andrea	17.5
Brazzale Daniele	26.5
Cappellotto Matteo	20
Cecchin Stefano	22
Fusato Ivan	17.5
Ghelli Filippo	17.5
Podda Rachele	18.5
Torcellan Tommaso	20

Visione compiti corretti, registrazione voto e/o orali: giovedì 18 aprile ore 10.00 aula Aula 1BC/50 torre Archimede.
Verrà data precedenza alla registrazione voti a chi accetta il voto dello scritto e ha il bonus di + 3.