Esame di Istituzioni di Matematiche II del 20 giugno 2001 (Corso di Laurea in Biotecnologie, Universitá 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. Una ditta progetta di costruire uno stabilimento in una comunità, e c'è preoccupazione sulla salute da parete delle autorità cittadine. In particolare, si è notato che per stabilimenti analoghi c'è una incidenza insolita di intossicazione da piombo nei bambini da 0 a 3 anni. In particolare, si prevede che le intossicazioni da piombo saranno 50 su 100.000 per i bambini che vivono fino a 2 km dall'impianto, 20 su 100.000 per quelli che vivono dai 2 ai 5 km, e 5 su 100.000 per quelli che vivono a più di 5 km.

1. Se l'80% dei bambini sotto i 3 anni vive a più di 5 km, il 15% vive tra 2 e 5 km e il 5% vive a meno di 2 km, qual è la probabilità totale che un bambino di questa comunitçntragga un'intossicazione da piombo? Usare la formula
$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$
con $(A_i)_{i=1,\ldots,n}$ eventi disgiunti tali che $\mathbb{P}(A_i) > 0$ per ogni $i = 1, \ldots, n$ e $\cup_{i=1}^{n} A_i = \Omega$.

2. Dimostrare la formula usata nel punto 1.

Esercizio 2. É stato fatto uno studio sugli effetti del consumo di arachidi sul livello di colesterolo nel sangue. È stata somministrata agli stessi 18 soggetti prima una dieta a basso colesterolo HDL standard, e poi una dieta sperimentale arricchita di arachidi. Il totale di proteine, carboidrati e grassi sono stati tenuti costanti in entrambi i periodi. Alla fine, i dati per il livello di colesterolo nel sangue sono stati i seguenti:

	Dieta standard (media $\pm$ dev. st.)	Dieta con arachidi (media $\pm$ dev. st.)	Media delle diff. (arachidi $-$ standard)	Int. conf. al 95% per la diff. media
Livello di HDL	47 ± 11	45 ± 10	$-$ 2,3	$(-3,9; -0,7)$

(notiamo che la differenza delle medie e l'intervallo di confidenza sono forniti con una precisione maggiore).

1. Stimare l'errore standard della differenza delle medie di HDL tra le due diete.

2. Eseguire un test statistico per vedere se il livello di HDL si può considerare lo stesso per le due diete. Riportare un valore P il più preciso possibile.

3. Quali delle seguenti domande sono in contraddizione con la tabella e perché (non sono necessari ulteriori calcoli)?

   a) il valore P al punto 2 è minore di 0,05;

   b) l'intervallo di confidenza al 90% per la differenza media è $(-4,1; -0,5)$;

   c) l'intervallo di confidenza al 98% per la differenza media è $(-4,2; -0,4)$;

Esercizio 3. È stato fatto uno studio per stabilire quali fattori possono influenzare la probabilità che un fumatore decida di smettere, e la sua probabilità di successo (cioè se riesce ad astenersi dal fumo per almeno 1 anno). Alcuni fattori sotto esame erano il numero di sigarette fumate abitualmente in un giorno, e il livello di istruzione dei soggetti.

1. Supponiamo che su 225 fumatori che fumavano meno di 1 pacchetto al giorno, 15 abbiano tentato di smettere, e che su 150 fumatori che fumavano più di 1 pacchetto al giorno, 3 abbiano tentato di smettere. Riportare delle limitazioni per il valore P (le più precise possibili), e stabilire se la differenza è significativa o no.

2. Supponiamo poi che su 311 persone che hanno provato a smettere, si abbiano le seguenti casistiche di successo:

   - meno di un'istruzione superiore: 16 su 33;
   - istruzione superiore senza aver iniziato l'università: 47 su 76;
   - università iniziata ma non finita: 69 su 125;
   - laurea: 52 su 77.

   Questi dati mostrano un'associazione tra il livello di istruzione e il tasso di successo nello smettere (usare $\alpha = 0,05$)?

Esercizio 4. È un luogo comune che i bambini molto piccoli abbiano un metabolismo più alto che gli dà più energia rispetto agli altri bambini e agli adulti. Testiamo questa ipotesi misurando le pulsazioni cardiache al minuto di diversi bambini, con questo risultato ($x =$ età, $y =$ pressione):

$$\sum_{i=1}^{22} x_i = 233, \qquad \sum_{i=1}^{22} y_i = 1725, \qquad \sum_{i=1}^{22} x_i y_i = 16.748, \qquad \sum_{i=1}^{22} x_i^2 = 3345, \qquad \sum_{i=1}^{22} y_i^2 = 140.933$$

1. Trovare la retta di regressione per i dati in tabella.
2. Che test si può eseguire per vedere se in effeti il metabolismo è influenzato dall'età (supponendo un effetto lineare rispetto all'età)?
3. Eseguire il test del punto 2; riportare il valore P.

Soluzioni

Esercizio 1.

1. Consideriamo un generico bambino fino a 3 anni della comunità, e definiamo gli eventi

$$\begin{aligned} A_1 &= \{\text{bambino vive } < 2 \text{ km}\} \\ A_1 &= \{\text{bambino vive tra 2 e 5 km}\} \\ A_1 &= \{\text{bambino vive } > 5 \text{ km}\} \\ B &= \{\text{bambino intossicato}\} \end{aligned}$$

Sappiamo allora che:

$$\mathbb{P}(A_1) = 0,05, \quad \mathbb{P}(A_2) = 0,15, \quad \mathbb{P}(A_3) = 0,8,$$
$$\mathbb{P}(B|A_1) = \frac{50}{100000}, \quad \mathbb{P}(B|A_2) = \frac{20}{100000}, \quad \mathbb{P}(B|A_3) = \frac{5}{100000}$$

Usando la formula suggerita si ottiene:

$$\mathbb{P}(B) = \frac{50}{100000} \cdot 0,05 + \frac{20}{100000} \cdot 0,15 + \frac{5}{100000} \cdot 0,8 = 9,5 \cdot 10^{-5} = 0,000095$$

2. Abbiamo che $B = B \cap \Omega = B \cap (\cup_{i=1}^n A_i) = \cup_{i=1}^n B \cap A_i$ e l'unione è disgiunta; allora

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Esercizio 2.

1. L'intervallo di confidenza ha estremi $\bar{X} \pm s_{\bar{X}} t_{1-\alpha/2}(\nu)$, dove le X_i sono le differenze di HDL tra le diete in ogni soggetto (1 punto). L'intervallo di confidenza è quindi ampio $2s_{\bar{X}} t_{1-\alpha/2}(\nu) = 3,9 - 0,7 = 3,2$; si ha poi $\alpha = 0{,}05$, $\nu = 18 - 1 = 17$ e $t_{1-\alpha/2}(\nu) = t_{0{,}975}(17) = 2{,}109$. Allora

$$s_{\bar{X}} = \frac{3,2}{2 \cdot 2,109} = 0,76 (2 \text{ punti})$$

2. Eseguiamo un test di ipotesi $H_0 : \mu_d = 0$ e $H_1 : \mu_d \neq 0$, dove μ_d è la vera media delle differenze di HDL tra le diete. Abbiamo

$$t = \frac{\bar{X}}{s_{\bar{X}}} = -3,03$$

Abbiamo $t_{0{,}995}(17) = 2,89 < |t| < t_{0{,}9975}(17) = 3,22$, quindi $0,005 < P < 0,01$. Possiamo quindi dire che con un valore P così basso la differenza è statisticamente significativa (2 punti).

3. **a)** Abbiamo già visto che $P < 0,01$, per cui possiamo dire che il punto a) è VERO (1 punto)

   b) l'intervallo di confidenza al 90% è sempre contenuto in quello al 95%, e non viceversa come sarebbe qui: il punto b) è FALSO (1 punto)

   c) l'intervallo di confidenza al 98% contiene sempre quello al 95%, pertanto il punto c) NON È IN CONTRADDIZIONE con la tabella (1 punto)

Esercizio 3.

1. Per avere la massima precisione possibile nel valore P, usiamo il test Z. Facciamo un test di ipotesi $H_0 : p_- = p_+$ contro l'alternativa $H_1 : p_- \neq p_+$. Abbiamo:

$$\hat{p}_- = \frac{15}{225} = 0,066, \quad \hat{p}_+ = \frac{3}{150} = 0,020, \quad \hat{p} = \frac{15 + 3}{225 + 150} = 0,048$$

Abbiamo $n_-\hat{p} = 225 \cdot 0,048 = 11 \geq 5$ e $n_+\hat{p} = 150 \cdot 0,048 = 7 > 5$. Siccome $\hat{p} < 1/2$, sicuramente anche $n_-(1 - \hat{p})$ e $n_+(1 - \hat{p})$ saranno maggiori di 5, quindi si può fare il test Z. Abbiamo:

$$Z = \frac{0,066 - 0,020}{\sqrt{0,048(1 - 0,048)\left(\frac{1}{225} + \frac{1}{150}\right)}} = 1,89$$

Possiamo fare il calcolo preciso di P:

$$\begin{aligned} P &= \mathbb{P}\{|Z| > 1,89\} = 2\mathbb{P}\{Z > 1,89\} = 2(1 - \mathbb{P}\{Z \le 1,89\}) = \\ &= 2(1 - F_Z(1,89)) = 2(1 - 0,97062) \simeq 0,05876 \end{aligned}$$

Dato che $P > 0{,}05$, la differenza che si riscontra non è ancora da considerare significativa (4 punti).

2. Vogliamo fare un test di ipotesi $H_0 : p_1 = p_2 = p_3 = p_4$ contro l'alternativa $H_1 : \exists i, j$ tali che $p_i \neq p_j$, dove i p_i indicano le proporzioni degli ex-fumatori che hanno avuto successo. Raccogliamo i dati (osservati ed attesi) nella seguente tabella:

	successo		no		totale
no diploma	16	(20)	17	(13)	33
diploma	47	(45)	29	(31)	76
università	69	(74)	56	(51)	125
laurea	52	(46)	25	(31)	77
	184		127		311

Per costruire la tabella attesa, abbiamo usato $\hat{p} = 184/311 = 0{,}59$. Siccome tutti i numeri della tabella attesa sono maggiori di 5, si puó usare il metodo del χ^2:

$$\begin{aligned} \chi^2 &= \frac{(16-20)^2}{20} + \frac{(17-13)^2}{13} + \frac{(47-45)^2}{45} + \frac{(29-31)^2}{31} + \\ &\quad + \frac{(69-74)^2}{74} + \frac{(56-51)^2}{51} + \frac{(52-46)^2}{46} + \frac{(25-31)^2}{31} = 5,02 \end{aligned}$$

Il valore critico è $\chi^2_{0,95}(3) = 7{,}815$; siccome $\chi^2 < \chi^2_{0,95}(3)$, accettiamo H_0. Non appaiono quindi evidenti relazioni tra livello di istruzione e tasso di successo nello smettere di fumare (chiaramente non c'è bisogno di fare test ulteriori) (4 punti).

Esercizio 4. Partiamo calcolando le quantità:

$$\begin{aligned} \bar{X} &= \frac{233}{22} = 10,59 \quad (0,5 \text{ punti}), \qquad \bar{Y} = \frac{1725}{22} = 78,41 \quad (0,5 \text{ punti}), \\ s_X^2 &= \frac{1}{21}(3345 - 22 \cdot 10,59^2) = 41,78 \quad (0,5 \text{ punti}), \\ s_Y^2 &= \frac{1}{21}(140933 - 22 \cdot 78,41^2) = 270,35 \quad (0,5 \text{ punti}), \\ s_{XY} &= \frac{1}{21}(16748 - 22 \cdot 10,59 \cdot 78,41) = -72,44 \quad (0,5 \text{ punti}) \end{aligned}$$

1. Calcoliamo i coefficienti della retta di regressione:

$$\begin{aligned} b_1 &= \frac{s_{XY}}{s_X^2} = -1,73 \\ b_0 &= \bar{Y} - b_1\bar{X} = 96,78 \end{aligned}$$

La retta di regressione è quindi $y = -1,73x + 96,78$ (2 punti).

2. Bisogna fare un test di ipotesi $H_0 : \beta_1 = 0$ e alternativa $H_1 : \beta_1 \neq 0$. Se H_0 viene accettata, significa che non c'è dipendenza lineare (1 punto).

3. Per eseguire il test, bisogna calcolare:

$$\begin{aligned} s_{Y|X} &= \sqrt{\frac{21}{20}(270,35 - 1,73^2 \cdot 41,78)} = 12,33 \\ s_{b_1} &= \frac{s_{Y|X}}{\sqrt{21 \cdot s_X^2}} = \frac{12,33}{\sqrt{21 \cdot 41,78}} = 0,42 \end{aligned}$$

Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{-1,73}{0,42} = -4,17$$

Abbiamo poi $\nu = 22 - 2 = 20$. Il più grande quantile nelle tavole della legge $t(20)$ è $t_{0,999}(20) = 3{,}552 < |t|$, quindi riportiamo $P < 0{,}002$. Con un P così piccolo, sicuramente rifiuteremmo l'ipotesi, con la conclusione che c'è una relazione lineare tra le due quantità (2 punti).

Esame di Istituzioni di Matematica II del 20 giugno 2001 (Corso di Laurea in Biotecnologie, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Sono ammessi all'orale:

Cifani Paolo	22
Pianalto Anna	23

Visione compiti corretti: a ricevimento, oppure mercoledì 22 luglio aula D piano terra Vallisneri.