Esame di Istituzioni di Matematica II del 5 febbraio 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. Supponiamo che il tasso annuale di mortalità per incidenti automobilistici nel 1980 sia stato di 10 ogni 100.000 abitanti nelle aree urbane degli USA. Supponiamo inoltre che il tasso annuale di mortalità per incidenti automobilistici nelle aree rurali degli USA sia stato di 100 ogni 100.000 abitanti.

1. In uno stato urbano con 2 milioni di abitanti, qual è la probabilità di non osservare più di 150 vittime automobilistiche in un dato anno?
2. In uno stato rurale con 100.000 abitanti, qual è la probabilità di non osservare più di 80 vittime automobilistiche in un dato anno?
3. Quanto deve valere x perchè la probabilità di non osservare più di x vittime automobilistiche in un dato anno in uno stato rurale con 100.000 abitanti sia del 5%?
4. Quanto deve valere x perchè la probabilità di non osservare più di x vittime automobilistiche in un dato anno in uno stato urbano con 500.000 abitanti sia del 10%?

Esercizio 2. È stato notato che c'è una relazione inversa tra consumo moderato di alcool e livello di colesterolo. Per testare questa ipotesi, è stato fatto uno studio coi seguenti risultati:

	numero	media	dev. standard
gruppo A (non bevitori)	23	205,6	25,3
gruppo B ($\leq$ 50 g/giorno)	15	182,7	21,9
gruppo C ($\geq$ 50 g/giorno)	12	199,8	30,3

1. Testare l'ipotesi che ci sia una differenza nella media dei livelli di colesterolo tra questi tre gruppi (usare $\alpha = 0{,}05$).
2. Eseguire tests di confronto multipli per capire quali delle medie sono differenti.

Esercizio 3. In uno studio del 1985, 89 donne su 283 con infertilità tubale primaria e 640 donne su 3833 di un gruppo di donne sane hanno dichiarato di aver usato un dispositivo intrauterino (IUD).

1. Qual è la stima del tasso di uso di un IUD nei gruppi di donne non fertili e sane, rispettivamente?
2. Fornire un intervallo di confidenza al 95% per le stime del punto 1.
3. Eseguire un test per verificare se il rischio di infertilità tubale primaria può dipendere dall'uso di un IUD: riportare limitazioni per il valore P.

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:

Età (x)	1	0	3	3	5	5	6	6	9	8
Pulsazioni/minuto (y)	103	125	102	86	88	78	77	78	90	75

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 più alto (supponendo un effetto lineare rispetto all'età)?
3. Eseguire il test del punto 2; riportare il valore P.

Soluzioni

Esercizio 1. Definiamo la variabile aleatoria

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esimo abitante di uno Stato è vittima di un incidente} \\ 0 & \text{altrimenti} \end{cases}$$

Allora $X_i \sim \mathcal{B}(1,p)$, con $p = 10/1.000.000 = 10^{-4}$ nelle aree urbane, e $p = 100/1.000.000 = 10^{-3}$ nelle aree rurali. Definiamo poi $S_n = \sum_{i=1}^n X_i =$ n. di incidenti in un anno in uno Stato di n abitanti. Allora $S_n \sim B(n,p)$. Fare i calcoli con questa distribuzione sarebbe molto laborioso (e anche usare una approssimazione di Poisson); siccome $np > 5$ in tutti i punti, possiamo approssimare la legge di S_n con una legge $N(np, np(1-p))$. Siccome poi il fattore $1-p$ é dell'ordine di $1-10^{-3}$, si può approssimare tranquillamente ad 1 ed approssimare quindi la legge di S_n con una legge $N(np, np)$ (2 punti).

1. Abbiamo $n = 2.000.000$, e S_n ha legge approssimabile con $N(200; 200)$. Calcoliamo:

$$\begin{aligned} \mathbb{P}\{S_n \le 150\} &\simeq \mathbb{P}\left\{\frac{S_n - 200}{\sqrt{200}} \le \frac{150-200}{\sqrt{200}}\right\} = \mathbb{P}\left\{Z \le -\frac{5}{\sqrt{2}}\right\} = \\ &= F_Z(-3,54) = 1 - F_Z(3,54) < 1 - 0,998 = 0,002 \qquad (1,5 \text{ punti}) \end{aligned}$$

2. Abbiamo $n = 100.000$, e S_n ha legge approssimabile con $N(100; 100)$. Calcoliamo:

$$\begin{aligned} \mathbb{P}\{S_n \le 80\} &\simeq \mathbb{P}\left\{\frac{S_n - 100}{\sqrt{100}} \le \frac{80-100}{\sqrt{100}}\right\} = \mathbb{P}\{Z \le -2\} = \\ &= F_Z(-2) = 1 - F_Z(2) = 1 - 0,977 = 0,023 \qquad (1,5 \text{ punti}) \end{aligned}$$

3. Abbiamo ancora $n = 100.000$, eS_n ha legge approssimabile con $N(100; 100)$. Calcoliamo:

$$0,05 = \mathbb{P}\{S_n \le x\} \simeq \mathbb{P}\left\{\frac{S_n - 100}{\sqrt{100}} \le \frac{x-100}{\sqrt{100}}\right\} = \mathbb{P}\left\{Z \le \frac{x-100}{10}\right\} = F_Z\left(\frac{x-100}{10}\right)$$

   Abbiamo quindi:

$$\frac{x-100}{10} = q_{0,05} = -q_{0,95} = -1,64$$

   e quindi $x = -16,4 + 100 = 83,6 \simeq 84$ (1,5 punti).

4. Abbiamo $n = 500.000$, eS_n ha legge approssimabile con $N(50; 50)$. Calcoliamo:

$$0,1 = \mathbb{P}\{S_n \le x\} \simeq \mathbb{P}\left\{\frac{S_n - 50}{\sqrt{50}} \le \frac{x-50}{\sqrt{50}}\right\} = \mathbb{P}\left\{Z \le \frac{x-50}{\sqrt{50}}\right\} = F_Z\left(\frac{x-50}{\sqrt{50}}\right)$$

   Abbiamo quindi:

$$\frac{x-50}{\sqrt{50}} = q_{0,1} = -q_{0,9} = -1,28$$

   e quindi $x = -1,28 \cdot \sqrt{50} + 50 = 40,95 \simeq 41$ (1,5 punti).

Esercizio 2.

1. Facciamo un test F con ipotesi $H_0 : \mu_1 = \mu_2 = \mu_3$, alternativa $H_1 : \exists i,j$ tali che $\mu_i \neq \mu_j$ e livello $\alpha =$ 0,05 (1 punto). Abbiamo:

$$\begin{aligned} s_e^2 &= \frac{1}{23+15+12-3}(22 \cdot 25,3^2 + 14 \cdot 21,9^2 + 11 \cdot 30,3^2) = 657,3 \qquad (1 \text{ punto}) \\ \bar{X} &= \frac{23 \cdot 205,6 + 15 \cdot 182,7 + 12 \cdot 199,8}{23+15+12} = 197,3 \qquad (1 \text{ punto}) \\ s_a^2 &= \frac{1}{3-1}(23(205,6-197,3)^2 + 15(182,7-197,3)^2 + 12(199,8-197,3)^2) = 2421,2 \qquad (1 \text{ punto}) \end{aligned}$$

   Abbiamo allora:

$$F = \frac{s_a^2}{s_e^2} = \frac{2421,2}{657,3} = 3,68$$

   I gradi di libertà stavolta sono $\nu_a = 3-1 = 2$, e $\nu_e = 47$, che si può approssimare con $\nu_e = 40$. Il valore critico è quindi $F_{1-\alpha}(\nu_a, \nu_e) = F_{0,95}(2,47) \simeq F_{0,95}(2,40) = 3{,}23$. Siccome $F > 3{,}23$, rifiutiamo l'ipotesi (1 punto).

2. Facciamo ora i test di Bonferroni per i diversi campioni, usando $\alpha' = \alpha/3 = 0{,}016$; approssimiamo $1 - \alpha'/2 = 0{,}992 \simeq 0{,}99$ (1 punto). Abbiamo:

   - A contro B:
   $$t = \frac{205,6 - 182,7}{\sqrt{\frac{25,3^2}{23} + \frac{21,9^2}{15}}} = 2,96$$
   da confrontare con il valore critico $t_{0,99}(36) \simeq t_{0,99}(40) = 2{,}42$. Siccome $|t| > 2{,}42$, concludiamo che $\mu_A \neq \mu_B$ (0,5 punti).
   - A contro C:
   $$t = \frac{205,6 - 199,8}{\sqrt{\frac{25,3^2}{23} + \frac{30,3^2}{12}}} = 0,56$$
   da confrontare con il valore critico $t_{0,99}(33) \simeq t_{0,99}(30) = 2{,}45$. Siccome $|t| < 2{,}45$, concludiamo che $\mu_A = \mu_C$ (0,5 punti).
   - B contro C:
   $$t = \frac{199,8 - 182,7}{\sqrt{\frac{21,9^2}{15} + \frac{30,3^2}{12}}} = 1,64$$
   da confrontare con il valore critico $t_{0,99}(25) \simeq t_{0,99}(30) = 2{,}48$. Siccome $|t| < 2{,}48$, concludiamo che $\mu_B = \mu_C$ (0,5 punti).

   Arriviamo quindi alla conclusione contraddittoria $\mu_A = \mu_C = \mu_B$ ma $\mu_A \neq \mu_C$. Ricordiamo però che il test di Bonferroni è conservativo, quindi la conclusione più plausibile è che $\mu_A = \mu_C \neq \mu_B$: in altre parole, il consumo moderato di alcool (meno di 50 g/giorno) abbassa il livello di colesterolo in modo significativo (0,5 punti).

Esercizio 3.

1. Donne non fertili: $\hat{p}_A = 89/283 = 0{,}314$; donne sane: $\hat{p}_B = 640/3833 = 0{,}167$ (1 punto)

2. Donne non fertili: l'intervallo di confidenza ha estremi
$$\hat{p}_A \pm q_{1-0,05/2}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n}} = 0,314 \pm 1,96 \cdot \sqrt{\frac{0,314 \cdot 0,686}{283}} = 0,314 \pm 0,054$$
quindi l'intervallo è $[0,260; 0,369]$ (1 punto).

   Donne sane: l'intervallo di confidenza ha estremi
$$\hat{p}_B \pm q_{1-0,05/2}\sqrt{\frac{\hat{p}_B(1-\hat{p}_B)}{n}} = 0,167 \pm 1,96 \cdot \sqrt{\frac{0,167 \cdot 0,833}{3833}} = 0,167 \pm 0,012$$
quindi l'intervallo è $[0,155; 0,179]$ (1 punto).

3. Per rispondere alla domanda, dobbiamo ordinare i dati in questo modo:

	non fertili	sane	totale
IUD	89	640	729
non IUD			
totale	283	3833	4116

   Vogliamo fare un test di ipotesi $H_0 : p_{IUD} = p_{no-IUD}$ contro l'alternativa $H_1 : p_{IUD} \neq p_{no_IUD}$, dove p_i è la probabilità che una donna del gruppo i = IUD, no-IUD, non sia fertile (in altre parole, i calcoli fatti nei punti 1 e 2 sono inutili ai fini del nostro test). Riempiamo quindi la tabella in alto e calcoliamo la tabella attesa (2,5 punti): per fare questo ci serve $\hat{p} = \frac{283}{4116} = 0{,}068$. Abbiamo

IUD	89	(50)	640	(679)	729
non IUD	194	(233)	3193	(3154)	3387
totale	283		3833		4116

Siccome tutti i numeri della tabella attesa sono maggiori di 5, si puó usare il metodo del χ^2 (0,5 punti):

$$\chi^2 = \frac{(89-50)^2}{50} + \frac{(194-233)^2}{233} + \frac{(640-679)^2}{679} + \frac{(3193-3154)^2}{3154} = 39,67 \qquad \text{(1 punto)}$$

L'ultimo valore in tavola è $\chi^2_{0,999}(1) = 10{,}828 < 39{,}67$, quindi riportiamo $P < 0{,}001$. Questo ci spinge a rifiutare l'ipotesi H_0 e ad accettare l'alternativa H_1: in altre parole, l'infertilità tubale primaria può dipendere anche dall'uso di un IUD (1 punto).

Esercizio 4. Facendo i calcoli, troviamo che

$$\sum_{i=1}^{10} X_i = 46, \qquad \sum_{i=1}^{10} Y_i = 902, \qquad \sum_{i=1}^{10} X_i^2 = 286, \qquad \sum_{i=1}^{10} Y_i^2 = 83600, \qquad \sum_{i=1}^{10} X_i Y_i = 3837$$

Calcoliamo poi le quantità:

$$\begin{aligned}
\bar{X} &= \frac{46}{10} = 4,6 \quad (0,5 \text{ punti}), \qquad \bar{Y} = \frac{902}{10} = 90,2 \quad (0,5 \text{ punti}), \\
s_X^2 &= \frac{1}{9}(286 - 10 \cdot 4,6^2) = 8,26 \quad (0,5 \text{ punti}), \\
s_Y^2 &= \frac{1}{9}(83600 - 10 \cdot 90,2^2) = 248,84 \quad (0,5 \text{ punti}), \\
s_{XY} &= \frac{1}{9}(3837 - 10 \cdot 90,2 \cdot 4,6) = 34,69 \quad (0,5 \text{ punti})
\end{aligned}$$

1. Calcoliamo i coefficienti della retta di regressione:

$$b_1 = \frac{s_{XY}}{s_X^2} = -4,19, \qquad b_0 = \bar{Y} - b_1 \bar{X} = 109,47$$

   La retta di regressione è quindi $y = -4,19x + 109,47$ (punto).

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 il metabolismo non dipende (linearmente) dall'età (1 punto).

3. Per eseguire il test, bisogna calcolare:

$$\begin{aligned}
s_{Y|X} &= \sqrt{\frac{n-1}{n-2}(s_Y^2 - b_1^2 s_X^2)} = 10,8 \quad (1 \text{ punto}) \\
s_{b_1} &= \frac{s_{Y|X}}{\sqrt{9 \cdot s_X^2}} = 1,25 \quad (1 \text{ punto})
\end{aligned}$$

   Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{-4,19}{1,25} = -3,35 \quad (0,5 \text{ punti})$$

   Nelle tavole della legge $t(8)$, t è compreso tra $t_{0,99}(8) = 2{,}89$ e $t_{0,995}(8) = 3{,}35$, quindi riportiamo $0{,}01 < P < 0{,}02$ (1 punto). La conclusione è che sembra esserci una relazione lineare tra le due quantità. In realtà, i dati sono stati presi da un esperimento più grande, in cui le età arrivavano a 21 anni, e l'effetto lineare era molto meno evidente: probabilmente la dipendenza è quindi più complessa.

Esame di Istituzioni di Matematica II del 5 febbraio 2001 (Corso di Laurea in Biotecnologie, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Sono ammessi all'orale:

Barison Erika	23,5
Boldrin Francesca	19
Carnio Silvia	23,5
Casagrande Silvia	21
Celegato Marta	18
Cuba Chiara	21,5
Faggionato Davide	18
Ferro Ambra	24,5
Fontebasso Yari	26
Gallelli Annarita	25,5
Gani Anna	22,5
Litteri Gaia	29
Lombardi Raffaele	26
Loro Emanuele	21,5
Magrini Elena	26,5
Marangoni Chiara	18
Monaco Giovanni	22,5
Nalesso Giovanna	23
Pendin Diana	21,5
Pianalto Anna	17,5
Rossato Christian	17
Sommaggio Roberta	19
Stella Roberto	26,5
Terenzio Marco	27,5
Urbanet Riccardo	24
Ventura Marcello	17
Vettori Francesca	20,5
Vivona Sandro	18
Zampetti Stefania	20,5