Esame di Istituzioni di Matematica II del 18 gennaio 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. Il tasso di infarti cardiaci in 5 anni in donne sane tra i 50 e i 59 anni è 10 su 1000. Supponiamo che tra 1000 donne sane tra i 50 e i 59 anni che abbiano assunto ormoni post-menopausa, si siano osservati 3 infarti in 5 anni.

1. Calcolare la probabilità di osservare un numero di infarti minore o uguale a 3 usando la distribuzione binomiale.
2. Rispondere alla domanda 1 usando l'approssimazione di Poisson.
3. Confrontate le quantità delle domande 1 e 2.

Esercizio 2. É stato fatto uno studio sulle ore settimanali di attività fisica moderata in gruppi di uomini di diverse età, con i seguenti risultati:

età	n	media ore	dev. st. ore
20-34	487	8,1	1,4
35-49	233	9,7	1,2
50-64	191	7,9	1,2

1. Fare una analisi della varianza per confrontare le medie dei 3 gruppi, usando il livello $\alpha = 0{,}01$.
2. Se necessario, fare una procedura di confronto multipla per identificare quali specifiche medie sono differenti.

Esercizio 3. Supponiamo di voler studiare l'aggregazione familiare di malattie respiratorie sulla base di malattie specifiche. Sono identificate 100 famiglie in cui uno dei due genitori ha l'asma (famiglie di tipo A) e 200 famiglie in cui nessuno dei genitori ha l'asma (famiglie di tipo B). Supponiamo che nelle famiglie del tipo A, in 15 il primogenito abbia l'asma e in 9 il primogenito abbia un altro tipo di malattia respiratoria. Inoltre, supponiamo che nelle famiglie del tipo B, in 4 il primogenito abbia l'asma e in 6 il primogenito abbia un altro tipo di malattia respiratoria.

1. Dire se il tasso di primogeniti malati di asma nei due tipi di famiglie può ritenersi uguale: fare un test con $\alpha = 0{,}01$ e fornire il valore P (esatto o con delle limitazioni).
2. Dire se il tasso di primogeniti con malattie respiratorie non asmatiche nei due tipi di famiglie può ritenersi uguale: fare un test con $\alpha = 0{,}01$ e fornire il valore P (esatto o con delle limitazioni).

Esercizio 4. I dati nella tabella seguente sono misure da un gruppo di 11 uomini con una malattia cardiaca prese all'autopsia:

$$\sum x_i = 611,7, \quad \sum x_i^2 = 35.350,49, \quad \sum x_i y_i = 280.031,5$$

$$\sum y_i = 4950, \quad \sum y_i^2 = 2.421.650,$$

dove abbiamo indicato con x_i il peso corporeo totale e con y_i il peso del cuore dell'i-esimo individuo.

1. Costruire un intervallo di confidenza al 90% per il peso del cuore.
2. Trovare la retta di regressione tra il peso corporeo totale e il peso del cuore.
3. Fare un test per vedere se c'é una relazione lineare tra le due quantità di cui sopra; riportare il valore P.

Soluzioni

Esercizio 1. Definiamo la variabile aleatoria

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esima donna ha un infarto in 5 anni} \\ 0 & \text{altrimenti} \end{cases}$$

Allora $X_i \sim \mathcal{B}(1,p)$, con $p = 10/1000 = 0{,}01$. Definiamo poi $S_{1000} = \sum_{i=1}^{1000} X_i =$ n. di donne inizialmente sane che hanno un infarto in 5 anni (1 punto).

1. Calcoliamo:

$$\begin{aligned} \mathbb{P}\{S_{1000} \le 3\} &= \sum_{k=0}^{3} \mathbb{P}\{S_{1000} = k\} = \\ &= \binom{1000}{0}(1-p)^{1000} + \binom{1000}{1}p(1-p)^{999} + \binom{1000}{2}p^2(1-p)^{998} + \binom{1000}{3}p^3(1-p)^{997} = \\ &= 0,99^{1000} + \frac{1000}{1}\cdot 0,01 \cdot 0,99^{999} + \frac{1000\cdot 999}{1\cdot 2}\cdot 0,01^2 \cdot 0,99^{998} + \frac{1000\cdot 999\cdot 998}{1\cdot 2\cdot 3}\cdot 0,01^3\cdot 0,99^{997} = \\ &= 0,00004 + 0,00043 + 0,00220 + 0,00739 = 0,01006 \quad \text{(3 punti)} \end{aligned}$$

2. Abbiamo $np = 1000\cdot\ 0{,}01 = 10 > 5$; approssimiamo quindi $S_{1000} \sim \mathbb{P}(10)$:

$$\begin{aligned} \mathbb{P}\{S_{1000} \le 3\} &= \sum_{k=0}^{3} \mathbb{P}\{S_{1000} = k\} \simeq \sum_{k=0}^{3} e^{-10}\frac{10^k}{k!} = \\ &= e^{-10}\left(1 + \frac{10}{1} + \frac{100}{1\cdot 2} + \frac{1000}{1\cdot 2\cdot 3}\right) = 0,0000454 \cdot 227,6 = 0,01033 \quad \text{(3 punti)} \end{aligned}$$

3. I risultati sono uguali fino alla quinta cifra decimale; ciò non stupisce, poiché per "n grande" e "p piccolo" una variabilie aleatoria binomiale può essere approssimata da una variabile di Poisson (1 punto).

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 $\alpha = 0{,}01$. Abbiamo:

$$\begin{aligned} s_e^2 &= \frac{1}{487+233+191-3}(486\cdot 1,4^2 + 232\cdot 1,2^2 + 190\cdot 1,2^2) = 1,718 \\ \bar{X} &= \frac{487\cdot 8,1 + 233\cdot 9,7 + 191\cdot 7,9}{457+233+191} = 8,46 \\ s_a^2 &= \frac{1}{3-1}(487(8,1-8,46)^2 + 233(9,7-8,46)^2 + 191(7,9-8,46)^2) = 240,61 \end{aligned}$$

Abbiamo allora:

$$F = \frac{s_a^2}{s_e^2} = \frac{240,61}{1,718} = 140$$

I gradi di libertà stavolta sono $\nu_a = 3-1 = 2$, e $\nu_e = 908$, che si può approssimare con $\nu_e = \infty$. Il valore critico è quindi $F_{1-\alpha}(\nu_a, \nu_e) = F_{0,99}(2,908) \simeq F_{0,99}(2,\infty) = 4{,}60$. Siccome $F > 4{,}60$, rifiutiamo l'ipotesi (5 punti).

2. Facciamo ora i test di Bonferroni per i diversi campioni; abbiamo:

   - 20-34 contro 35-49:

$$t = \frac{8,1-9,7}{\sqrt{\frac{1,4^2}{487}+\frac{1,2^2}{233}}} = -15,8$$

   - 35-49 contro 50-64:

$$t = \frac{9,7-7,9}{\sqrt{\frac{1,2^2}{233}+\frac{1,2^2}{191}}} = 15,36$$

- 20-34 contro 50-64:
$$t = \frac{8,1-7,9}{\sqrt{\frac{1,4^2}{487}+\frac{1,2^2}{191}}} = 1,85$$

  Questi valori vanno confrontati con i quantili $t_{1-\alpha'/2}(\nu)$, che non sono sulle tavole: infatti $\nu = 487 + 233 - 2$, $233 + 191 - 2$ o $487 + 191 - 2$, e tutti e tre questi numeri sono approssimabili con ∞; inoltre si ha che $\alpha' = \alpha/3 = 0,0033$, quindi $1 - \alpha'/2 = 0{,}9983$. Siccome la distribuzione $t(\nu)$ si puó approssimare con una $N(0,1)$ per ν abbastanza grande (cioè approssimabile con ∞), otteniamo che $t_{0,9983}(\nu) \simeq q_{0,9983} = 2,94$. In alternativa, si possono approssimare i quantili con $t_{0,9975}(\infty) = 2{,}807$ o $t_{0,999}(\infty) = 3{,}090$. Qualunque quantile si usi, si vede che 20-34 contro 50-64 non hanno una differenza significativa ($|1,85| < t_{0,9983}(\nu)$), mentre 20-34 contro 35-49 e 35-49 contro 50-64 sì ($|15,36| > t_{0,9983}(\nu)$ e $|-15,8| > t_{0,9983}(\nu)$). Il risultato finale é quindi che si può ritenere che $\mu_1 = \mu_3 \neq \mu_2$ (3 punti).

Esercizio 3. Raccogliamo i dati nella tabella:

	asma	altro	sani	totale
A	15	9	76	100
B	4	6	190	200
	19	15	266	300

In entrambe le domande, bisognerà confrontare due campioni ognuno con due caratteristiche; si potrà quindi usare indifferentemente il test Z o il test χ^2. Per mostrare entrambi gli approcci, risolviamo il primo punto usando il χ^2 e il secondo usando il test Z.

1. Vogliamo fare un test di ipotesi $H_0 : p_1 = p_2$ contro l'alternativa $H_1 : p_1 \neq p_2$, dove p_1 indica la proporzione dei primogeniti asmatici nel gruppo A e p_2 indica la proporzione dei primogeniti asmatici nel gruppo B. Raccogliamo i dati (osservati ed attesi) nella seguente tabella:

	asma		non asma		totale
A	15	(6,3)	85	(93,7)	100
B	4	(12,6)	196	(187,4)	200
	19		281		300

   Per costruire la tabella attesa, abbiamo usato $\hat{p} = 19/300 = 0{,}063$. Siccome tutti i numeri della tabella attesa sono maggiori di 5, si puó usare il metodo del χ^2:
$$\chi^2 = \frac{(15-6,3)^2}{6,3} + \frac{(4-12,6)^2}{12,6} + \frac{(85-93,7)^2}{93,7} + \frac{(196-187,4)^2}{187,4} = 19,06$$

   Il valore critico è $\chi^2_{0,99}(1) = 6{,}63$; siccome $\chi^2 > \chi^2_{0,99}(1)$, rifiutiamo H_0 e accettiamo H_1. Siccome dalle tavole risulta $\chi^2_{0,999}(1) = 10{,}828$, riportiamo $P < 0{,}001$ (4 punti). Se invece avessimo fatto un test Z, saremmo riusciti a concludere solo che $P < 0{,}00278$.

2. Vogliamo ora fare un test di ipotesi $H_0 : p_3 = p_4$ contro l'alternativa $H_1 : p_3 \neq p_4$, dove p_3 indica la proporzione dei primogeniti con malattie non asmatiche nel gruppo A e p_4 indica la proporzione dei primogeniti con malattie non asmatiche nel gruppo B. Abbiamo:
$$\hat{p}_3 = \frac{9}{100} = 0,09, \quad \hat{p}_4 = \frac{6}{200} = 0,03, \quad \hat{p} = \frac{9+6}{100+200} = 0,05$$

   Abbiamo $n_A\hat{p} = 100 \cdot 0,05 = 5 \geq 5$ e $n_B\hat{p} = 200 \cdot 0,05 = 10 > 5$. Siccome $\hat{p} < 1/2$, sicuramente anche $n_A(1-\hat{p})$ e $n_B(1-\hat{p})$ saranno maggiori di 5, e quindi si può fare il test Z. Abbiamo:
$$Z = \frac{0,09-0,03}{\sqrt{0,05(1-0,05)\left(\frac{1}{100}+\frac{1}{200}\right)}} = 2,25$$

   Il valore critico è $q_{1-\alpha/2} = q_{0,995} = 2{,}58$; siccome $|Z| < 2{,}58$, accettiamo H_0. Stavolta possiamo fare il calcolo preciso di P:
$$\begin{aligned} P &= \mathbb{P}\{|Z| > 2,25\} = 2\mathbb{P}\{Z > 2,25\} = 2(1 - \mathbb{P}\{Z \leq 2,25\}) = \\ &= 2(1 - F_Z(2,25)) = 2(1 - 0,98778) \simeq 0,0244 \quad \text{(4 punti)} \end{aligned}$$

   Se avessimo usato il test χ^2 avremmo potuto concludere solo che $0{,}01 < P < 0{,}025$.

Esercizio 4. Partiamo calcolando le quantità:

$$\begin{aligned}
\bar{X} &= \frac{611,7}{11} = 55,6 \quad (0,5 \text{ punti}), \qquad \bar{Y} = \frac{4950}{11} = 450 \quad (0,5 \text{ punti}),\\
s_X^2 &= \frac{1}{10}(35350,49 - 11 \cdot 55,6^2) = 133,44 \quad (0,5 \text{ punti}),\\
s_Y^2 &= \frac{1}{10}(2421650 - 11 \cdot 450^2) = 19415 \quad (0,5 \text{ punti}),\\
s_{XY} &= \frac{1}{10}(280031,5 - 11 \cdot 450 \cdot 55,6) = 481,15 \quad (0,5 \text{ punti})
\end{aligned}$$

1. L'errore standard della media del peso del cuore è $s_{\bar{Y}} = \sqrt{s_Y^2/11} = \sqrt{19415/11} = 42{,}011$, e gli estremi dell'intervallo sono quindi $\bar{Y} \pm t_{1-\alpha/2}(\nu) = 450 \pm t_{0,95}(10) \cdot 42,011 = 450 \pm 1,812 \cdot 42,011 = 450 \pm 76,12$, e l'intervallo è $I = [373,87; 526,12]$ (1,5 punti).

2. Calcoliamo i coefficienti della retta di regressione:

$$\begin{aligned}
b_1 &= \frac{s_{XY}}{s_X^2} = 3,60\\
b_0 &= \bar{Y} - b_1\bar{X} = 249,52
\end{aligned}$$

   La retta di regressione è quindi $y = 3,60x + 249,52$ (2 punti). Notiamo che in realtà la variabile y è espressa in grammi, e la variabile x in chilogrammi; volendo quindi tenere le stesse unità per entrambe le variabili, la retta di regressione sarebbe (ad esempio in g) $y = 0,0036x + 249,52$ g.

3. 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. Per eseguire il test, bisogna calcolare:

$$\begin{aligned}
s_{Y|X} &= \sqrt{\frac{10}{9}(19415 - 3,60^2 \cdot 133,44)} = 140,18\\
s_{b_1} &= \frac{s_{Y|X}}{\sqrt{11 \cdot s_X^2}} = \frac{140,18}{\sqrt{11 \cdot 133,44}} = 3,66
\end{aligned}$$

   Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{3,60}{3,66} = 0,984$$

   Il più piccolo quantile nelle tavole della legge $t(9)$ è $t_{0,95}(9) = 1{,}833 > |t|$, quindi riportiamo $P > 0{,}10$. Siccome gli α comunemente usati sono tutti minori di 0,10, sicuramente accetteremmo un test, con la conclusione che non c'è una relazione lineare tra le due quantità (2 punti).

Esame di Istituzioni di Matematica II del 18 gennaio 2001 (Corso di Laurea in Biotecnologie, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Sono ammessi all'orale:

Condini Chiara	27
Damesin Alessandro	20,5
Fiocco Francesca	28
Fogal Stefano	27,5
Manfrin Giampaolo	24
Meneghello Giulia	29
Minetto Silvia	22,5
Nalesso Giovanna	18
Pinato Odra	22,5
Rossi Alessandro	23,5
Solito Samantha	20,5
Sommaggio Roberta	17
Turato Cristian	18

Visione compiti corretti: a ricevimento, oppure mercoledì 25 gennaio aula D piano terra Vallisneri.