Esame di Statistica del 1 settembre 2004 (Corso di Laurea 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. Supponiamo di inoculare il vaccino dell'influenza ad una popolazione particolarmente sensibile di 500 unità tra i 60 e i 64 anni, e che il tasso di mortalità dopo questa operazione sia $p = 0{,}0028$. Utilizzando l'approssimazione di Poisson, calcolare:

1. la probabilità che muoiano esattamente 4 persone;
2. la probabilità che muoiano almeno 4 persone;
3. il numero atteso di morti;
4. la deviazione standard del numero di morti.

Esercizio 2. In un carcere è stato condotto uno studio su come il livello di testosterone possa influenzare il comportamento. Si sono presi tre gruppi di prigionieri: aggressivi (in carcere per reati violenti), socialmente dominanti (in carcere per reati non violenti, ma che avevano una posizione alta nella gerarchia del carcere) e non violenti (in carcere per reati non violenti e che non avevano una posizione alta nella gerarchia del carcere). Ogni gruppo è formato da 12 persone. Le analisi del sangue hanno rilevato i seguenti livelli di testosterone:

gruppo	media	dev. standard
A (aggressivi)	10,10	2,29
D (dominanti)	8,46	2,36
N (non aggressivi)	5,99	1,20

1. Stabilire se il livello di testosterone è differente tra i tre gruppi (usare $\alpha = 0{,}01$).
2. Stabilire quali sono i gruppi che si possono considerare con lo stesso livello medio di testosterone.

Esercizio 3. In un collegio c'è una epidemia di gastroenterite: esattamente metà degli occupanti si sono ammalati. In questo collegio, gli occupanti sono divisi in stanze come segue: ci sono 39 stanze doppie, 12 triple e 45 quadruple. Una domanda naturale è se ci possano essere state infezioni tra le persone di stanze uguali.

1. In una generica stanza da $n = 2, 3, 4$ posti, che distribuzione avrà la variabile aleatoria X_n = "numero di malati" assumendo che non ci siano infezioni tra gli individui della stessa stanza (e che quindi la gastroenterite sia provocata solo da cause esterne)?
2. Sotto l'ipotesi del punto precedente, calcola $\mathbb{P}\{X_n < 2\}$ e $\mathbb{P}\{X_n \geq 2\}$ per $n = 2, 3, 4$.
3. Per rispondere alla domanda iniziale, una procedura standard richiede di calcolare il numero atteso di stanze con meno di due malati e il numero atteso di stanze con due o più malati nelle stanze da 2, 3 e 4 posti rispettivamente. Eseguire il calcolo (precisazione: devono venire 6 numeri).

La distribuzione osservata di malati nelle stanze è la seguente:

	stanze con < 2 malati			stanze con ≥ 2 malati		
n. letti per stanza	2	3	4	2	3	4
	27	9	18	12	3	27

4. Eseguire un test χ^2, usando i numeri calcolati al punto 3, per verificare se non ci sono state infezioni. Riportare limitazioni per il valore P.

Esercizio 4. Vogliamo testare fino a che punto l'ipertensione è un fenomeno genetico. A questo scopo, sono state esaminate 20 famiglie, prendendo la pressione arteriosa di padre (x) e primogenito (t), con i seguenti risultati:

$$\sum_{i=1}^{20} x_i = 2980, \qquad \sum_{i=1}^{20} t_i = 2030$$

$$\sum_{i=1}^{20} x_i^2 = 451350, \qquad \sum_{i=1}^{20} t_i^2 = 210850, \qquad \sum_{i=1}^{20} x_i t_i = 305700$$

1. Trovare la retta di regressione della pressione del figlio (var. dipendente) rispetto a quella del padre (var. indipendente).
2. Eseguire un test per vedere se in effetti c'è questa relazione lineare. Si usi $\alpha = 0{,}05$.

Soluzioni

Esercizio 1. Se chiamiamo X il numero di morti, abbiamo che $X \sim B(n,p)$, con $n = 500$ e $p = 0{,}0028$. Dato che $n > 100$ e $p < 0{,}01$, si può usare l'approssimazione di Poisson, quindi X ha legge approssimabile con $Po(\lambda)$, con $\lambda = 500 \cdot 0,0028 = 1{,}4$.

1. Dobbiamo calcolare $\mathbb{P}\{X = 4\} \simeq e^{-1{,}4}\frac{1{,}4^4}{4!} = 0{,}0395$

2. Dobbiamo calcolare $\mathbb{P}\{X \geq 4\} = 1 - \mathbb{P}\{X \leq 3\} = 1 - \sum_{k=0}^{3} \mathbb{P}\{X = k\}$. Usando la regola ricorsiva per la legge di Poisson si ha:
$$\begin{aligned}
\mathbb{P}\{X=0\} &= e^{-1,4} = 0,2466 \\
\mathbb{P}\{X=1\} &= \frac{1,4}{1} \cdot 0,2466 = 0,3452 \\
\mathbb{P}\{X=2\} &= \frac{1,4}{2} \cdot 0,3452 = 0,2417 \\
\mathbb{P}\{X=3\} &= \frac{1,4}{3} \cdot 0,2417 = 0,1128
\end{aligned}$$
Abbiamo dunque $\mathbb{P}\{X \geq 4\} = 1 - 0,2466 - 0,3452 - 0,2417 - 0,1128 = 0,0537$.

3. Abbiamo $\mathbb{E}[X] = \lambda = 500 \cdot 0,0028 = 1,4$.

4. Abbiamo $\mathrm{Var}\,[X] = \lambda = 1,4$, e quindi la deviazione standard è uguale a $\sqrt{1,4} = 1{,}18$.

Esercizio 2.

1. Facciamo una analisi della varianza di ipotesi $H_0 : \mu_A = \mu_D = \mu_N$ e alternativa $H_1 : \exists i,j$ tali che $\mu_i \neq \mu_j$. La varianza entro i campioni è
$$s_e^2 = \frac{2,29^2 + 2,36^2 + 1,20^2}{3} = 4,08$$
La media tra i campioni è
$$\bar{X} = \frac{10,10 + 8,46 + 5,99}{3} = 8,18$$
La varianza tra i campioni è
$$s_a^2 = \frac{12}{2}((10,10 - 8,18)^2 + 16(8,46 - 8,18)^2 + 14(5,99 - 8,18)^2) = 51,37$$
La F vale quindi:
$$F = \frac{s_a^2}{s_e^2} = \frac{51,37}{4,08} = 12,58$$
I gradi di libertà sono $\nu_a = 3 - 1 = 2$ e $\nu_e = 12 + 12 + 12 - 3 = 33$. Bisogna quindi confrontare F con il quantile $F_{0{,}99}(2,33)$, che non è sulle tavole. Notiamo però che $F_{0{,}99}(2,33) < F_{0{,}99}(2,30) = 5{,}39$. Siccome $12{,}58 > 5{,}39$, l'ipotesi è rifiutata.

2. Facciamo ora i test di Bonferroni per i diversi campioni; il comune denominatore è $\sqrt{s_e^2(\frac{1}{12} + \frac{1}{12})} = \sqrt{4,08(\frac{1}{12} + \frac{1}{12})} = 0{,}83$. Abbiamo:
   - A contro D:
   $$t_{\mathrm{AD}} = \frac{10,10 - 8,46}{0,83} = 1,99$$
   - D contro N:
   $$t_{\mathrm{DN}} = \frac{8,46 - 5,99}{0,83} = 2,99$$
   - A contro N:
   $$t_{\mathrm{AN}} = \frac{10,10 - 5,99}{0,83} = 4,98$$

   Questi valori vanno confrontati con il quantile $t_{1-\alpha'/2}(33)$, che non è sulle tavole; infatti $\alpha' = \alpha/3 = 0,0033$, quindi $1 - \alpha'/2 = 0{,}9983$. Siccome $2{,}97 = t_{0{,}9975}(40) < t_{0{,}9983}(33) < t_{0{,}999}(30) = 3{,}38$, abbiamo che gli ultimi due test danno come risultato una differenza di medie, mentre il primo no. Possiamo quindi dire che i dati ci suggeriscono $\mu_A = \mu_D \neq \mu_N$.

Esercizio 3.

1. Per ogni generica camera, definiamo le variabili aleatorie

$$Y_i = \begin{cases} 1 & \text{se l'}i\text{-esimo occupante è malato} \\ 0 & \text{altrimenti} \end{cases}$$

Allora $Y_i \sim Be(p)$, con $p = 49/98 = 1/2$. Allora $X_n = \sum_{i=1}^{n} Y_i$ è il numero di occupanti malati in una stanza da n posti. Se non ci sono infezioni tra gli occupanti della stessa stanza, è ragionevole supporre che le Y_i siano indipendenti, e quindi $X_n \sim B(n,p)$.

2. Per $n = 2$ abbiamo:

$$\begin{aligned} \mathbb{P}\{X_2 \geq 2\} &= \mathbb{P}\{X_2 = 2\} = \binom{2}{2}\left(\frac{1}{2}\right)^2 = \frac{1}{4} = 0,25 \\ \mathbb{P}\{X_2 < 2\} &= 1 - 0,25 = 0,75 \end{aligned}$$

Per $n = 3$ abbiamo:

$$\begin{aligned} \mathbb{P}\{X_3 < 2\} &= \mathbb{P}\{X_3 = 0\} + \mathbb{P}\{X_3 = 1\} = \binom{3}{0}\left(1-\frac{1}{2}\right)^3 + \binom{3}{1}\frac{1}{2}\left(1-\frac{1}{2}\right)^2 = \frac{1}{8} + \frac{3}{8} = 0,5 \\ \mathbb{P}\{X_3 \geq 2\} &= 1 - \mathbb{P}\{X_3 < 2\} = 0,5 \end{aligned}$$

Infine, per $n = 4$ abbiamo:

$$\begin{aligned} \mathbb{P}\{X_4 < 2\} &= \mathbb{P}\{X_4 = 0\} + \mathbb{P}\{X_4 = 1\} = \binom{4}{0}\left(1-\frac{1}{2}\right)^4 + \binom{4}{1}\frac{1}{2}\left(1-\frac{1}{2}\right)^3 = \frac{1}{16} + \frac{4}{16} = \frac{5}{16} = 0,3125 \\ \mathbb{P}\{X_4 \geq 2\} &= 1 - \mathbb{P}\{X_4 < 2\} = 0,6875 \end{aligned}$$

3. Possiamo raccogliere i risultati nella seguente tabella:

	stanze con <2 malati	stanze con ≥ 2 malati	totale
stanze doppie	$39 \times 0,75 = 29,25 \simeq 29$	$39 \times 0,25 = 9,75 \simeq 10$	39
stanze triple	$12 \times 0,5 = 6$	$12 \times 0,5 = 6$	12
stanze quadruple	$45 \times 0,3125 = 14,0625 \simeq 14$	$45 \times 0,6875 = 30,94 \simeq 31$	45

4. Vogliamo fare un test di ipotesi $H_0 : X_n \sim B(n,1/2)$ e alternativa $H_1 : X_n \nsim B(n,1/2)$. Raccogliamo i dati (osservati ed attesi) nella seguente tabella:

	stanze con <2 malati		stanze con ≥ 2 malati		totale
stanze doppie	27	(29)	12	(10)	39
stanze triple	9	(6)	3	(6)	12
stanze quadruple	18	(14)	27	(31)	45

Per costruire la tabella attesa, abbiamo usato i risultati del punto 3. Siccome tutti i numeri della tabella attesa sono maggiori di 5, si puó usare il metodo del χ^2:

$$\chi^2 = \frac{(27-29)^2}{29} + \frac{(12-10)^2}{10} + \frac{(9-6)^2}{6} + \frac{(3-6)^2}{6} + \frac{(18-14)^2}{14} + \frac{(27-31)^2}{31} = 5,3$$

Il valore più vicino della tabella è $\chi^2_{0,95}(2) = 5{,}99$; siccome $\chi^2 < \chi^2_{0,95}(2)$, accettiamo H_0. La conclusione è che possiamo accettare l'ipotesi di indipendenza con $P > 0{,}05$. Siccome l'indipendenza era il modo per modellizzare matematicamente l'assenza di infezioni tra gli occupanti della stessa stanza, dai dati non sembra che la gastroenterite sia infettiva.

Esercizio 4.

1. Per calcolare la retta di regressione, servono le quantità:

$$\begin{aligned} \bar{X} &= \frac{2980}{20} = 149 \qquad \bar{T} = \frac{2030}{20} = 101,5 \\ s_X^2 &= \frac{1}{19}(451350 - 20 \cdot 149^2) = 385,7 \\ s_T^2 &= \frac{1}{19}(210850 - 20 \cdot 101,5^2) = 252,9 \\ s_{XT} &= \frac{1}{19}(305700 - 20 \cdot 149 \cdot 101,5) = 170 \end{aligned}$$

   Possiamo ora calcolare i coefficienti:

$$\begin{aligned} b_1 &= \frac{s_{XY}}{s_X^2} = 0,441 \\ b_0 &= \bar{Y} - b_1\bar{X} = 35,8 \end{aligned}$$

   La retta di regressione è quindi $y = 0,441x + 35,8$.

2. Bisogna fare un test di ipotesi $H_0 : \beta_1 = 0$ e alternativa $H_1 : \beta_1 \neq 0$. Per questo bisogna calcolare:

$$\begin{aligned} s_{Y|X} &= \sqrt{\frac{19}{18}(252,9 - 0,441^2 \cdot 385,7)} = 13,707 \\ s_{b_1} &= \frac{s_{Y|X}}{\sqrt{19 s_X^2}} = \frac{13,707}{\sqrt{19 \cdot 385,7}} = 0,160 \end{aligned}$$

   Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{0,441}{0,160} = 2,76$$

   Il valore critico è $t_{0,975}(18) = 2{,}1$. Abbiamo che $|t| > 2{,}1$, quindi rifiutiamo H_0 e accettiamo H_1: si può quindi affermare che c'è una relazione lineare tra i dati.