Esame di Statistica del 12 febbraio 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. Un'azienda produce occhiali utilizzando tre diversi macchinari. Il primo macchinario produce mediamente un paio di occhiali difettosi ogni 100, il secondo ogni 200, il terzo ogni 300. Gli occhiali vengono imballati in scatole identiche, contenenti 100 paia. Ogni scatola contiene occhiali scelti a caso tra quelli prodotti da una sola delle tre macchine. Si supponga che il primo macchinario abbia una produzione doppia rispetto agli altri due, cioè una scatola scelta a caso ha probabilità $1/2$ di essere prodotta dal primo macchinario, $1/4$ da secondo e $1/4$ dal terzo. Un ottico riceve una di queste scatole.

1. Qual è la probabilità che trovi almeno un paio di occhiali difettoso?
2. Se l'ottico trova esattamente due paia difettose, qual è la probabilità che gli occhiali siano stati prodotti dal primo macchinario?

Esercizio 2. In un esperimento del 1943, 10 ratti albini furono usati per studiare l'efficacia del tetracloruro di carbonio nel trattamento dei vermi. Le cavie furono infettate con larve e dopo 10 giorni divise a caso in due gruppi da 5, trattati con dosi diverse della sostanza. Due giorni dopo i ratti furono soppressi e i vermi contati, coi seguenti risultati:

0.032 cc	421	462	400	378	413
0.063 cc	207	17	412	74	116

1. Calcolare gli intervalli di confidenza al 95% per entrambi i campioni.
2. Supponendo che la varianza tra le popolazioni sia la stessa, dire se questi dati dimostrano che il dosaggio superiore è stato più efficace. Riportare limitazioni al valore P.

Esercizio 3. Un campione aleatorio di 500 famiglie degli Stati Uniti è stato classificato per regione e reddito (in migliaia di dollari), ottenendo i risultati seguenti:

reddito	sud	nord
0–10	42	53
10–30	102	178
30 o più	36	89

1. Ci sono differenze tra le diverse classi di reddito nel fatto di appartenere agli Stati del nord o agli Stati del sud? Effettuare il test con un livello $\alpha = 0.05$ e riportare un valore P.
2. Stabilire quali sono le classi di reddito in cui il tasso di appartenenza geografica si può ritenere uguale.

Esercizio 4. Le cifre che seguono sono le medie annuali dei prezzi di tutti i libri recensiti dalla rivista *Science* dal 1990 al 1996.

anno (-1900) (x)	90	91	92	93	94	95	96
dollari (y)	54.43	54.08	57.58	51.21	59.96	60.52	62.13

con le seguenti statistiche aggregate:

$$\sum_{i=1}^{7} x_i = 651, \qquad \sum_{i=1}^{7} y_i = 399.91, \sum_{i=1}^{7} x_i^2 = 60571, \qquad \sum_{i=1}^{7} y_i^2 = 22943.20, \qquad \sum_{i=1}^{7} x_i y_i = 37229.99$$

1. Trovare la retta di regressione del prezzo medio dei libri rispetto all'anno.
2. Eseguire un test per vedere se in effetti c'è una relazione lineare. Riportare limitazioni al valore P.
3. Fornire un intervallo di confidenza al 95% della media dei prezzi di tutti i libri recensiti nel 1997.

Soluzioni

Esercizio 1. Definiamo gli eventi $A_j := \{$ la scatola proviene dal j-esimo macchinario$\}$, $j = 1, 2, 3$. Allora $\Omega = A_1 \cup A_2 \cup A_3$, e l'unione è disgiunta, e si ha $\mathbb{P}(A_1) = 1/2$, $\mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/4$. Modellizziamo inoltre i difetti degli occhiali con delle variabili aleatorie di Bernoulli definite da

$$X_i := \begin{cases} 1 & \text{se l'}i\text{-esimo occhiale è difettoso} \\ 0 & \text{altrimenti} \end{cases}$$

Allora le $(X_i)_i$ sono i.i.d. con legge $Be(p_i)$ sotto ciascuna delle $\mathbb{P}(\cdot|A_i)$, dove $p_1 = 1/100$, $p_2 = 1/200$ e $p_3 = 1/300$. Consideriamo poi $S_{100} := \sum_{i=1}^{100} X_i \sim B(100, p_i)$ sotto $\mathbb{P}(\cdot|A_i)$.

1. Applicando la formula della probabilità totale, abbiamo:

$$\begin{aligned}
\mathbb{P}\{S_{100} > 0\} &= 1 - \mathbb{P}\{S_{100} = 0\} = 1 - \sum_{j=1}^{3} \mathbb{P}(\{S_{100} = 0\}|A_j)\mathbb{P}(A_j) = \\
&= 1 - \sum_{j=1}^{3} \binom{100}{0} (1 - p_j)^{100} \mathbb{P}(A_j) = \\
&= 1 - \left(\left(1 - \frac{1}{100}\right)^{100} \cdot \frac{1}{2} + \left(1 - \frac{1}{200}\right)^{100} \cdot \frac{1}{4} + \left(1 - \frac{1}{300}\right)^{100} \cdot \frac{1}{4} \right) = \\
&= 1 - 0.513492 = 0.486508
\end{aligned}$$

2. Applicando la formula di Bayes, bisogna calcolare

$$\begin{aligned}
\mathbb{P}(A_1|\{S_{100} = 2\}) &= \frac{\mathbb{P}(\{S_{100} = 2\}|A_1)\mathbb{P}(A_1)}{\sum_{j=1}^{3} \mathbb{P}(\{S_{100} = 2\}|A_j)\mathbb{P}(A_j)} = \frac{\binom{100}{2} p_1^2 (1 - p_1)^{98} \frac{1}{2}}{\sum_{j=1}^{3} \binom{100}{2} p_j^2 (1 - p_j)^{98} \mathbb{P}(A_j)} = \\
&= \frac{\binom{100}{2} \left(\frac{1}{100}\right)^2 \left(1 - \frac{1}{100}\right)^{98} \cdot \frac{1}{2}}{\binom{100}{2} \left(\left(\frac{1}{100}\right)^2 \left(1 - \frac{1}{100}\right)^{98} \cdot \frac{1}{2} + \left(\frac{1}{200}\right)^2 \left(1 - \frac{1}{200}\right)^{98} \cdot \frac{1}{4} + \left(\frac{1}{300}\right)^2 \left(1 - \frac{1}{300}\right)^{98} \cdot \frac{1}{4} \right)} = \\
&= \frac{0.092432}{0.121275} = 0.762172
\end{aligned}$$

dove per il denominatore abbiamo applicato la formula della probabilità totale.

Esercizio 2.

1. Calcolando media e varianza campionarie dei due gruppi otteniamo per il primo $\bar{X} = 414.8$ e $s_X^2 = 960.70$ e per il secondo $\bar{Y} = 165.2$ e $s_Y^2 = 23839.70$, quindi $s_{\bar{X}} = \sqrt{s_X^2/5} = 13.86$ e $s_{\bar{Y}} = \sqrt{s_Y^2/5} = 69.05$. Abbiamo poi per entrambi i campioni $t_{1-\alpha/2}(\nu) = t_{0.975}(4) = 2.776$. Allora il primo intervallo di confidenza ha estremi $\bar{X} \pm t_{1-\alpha/2}(\nu) s_{\bar{X}} = 414.8 \pm 2.776 \cdot 13.86$, e risulta uguale a $[376; 453]$; il secondo invece ha estremi $\bar{Y} \pm t_{1-\alpha/2}(\nu) s_{\bar{Y}} = 165.2 \pm 2.776 \cdot 69.05$, e risulta uguale a $[-26; 357]$

2. Facciamo un test t sulla differenza delle medie con ipotesi $H_0 : \mu_X = \mu_Y$ e alternativa $H_1 : \mu_X > \mu_Y$. Supponendo che la varianza delle popolazioni iniziali sia la stessa, possiamo usare un test di Student. Abbiamo

$$t = \frac{\bar{X} - \bar{Y}}{s_{\bar{X}-\bar{Y}}} = \frac{414.8 - 165.2}{\sqrt{\frac{960.70}{5} + \frac{23839.70}{5}}} = 3.54$$

I gradi di libertà sono $\nu = 5 + 5 - 2 = 8$. La regione critica è della forma $[t_{1-\alpha}(\nu), +\infty)$. Siccome $t_{0.995}(8) = 3.355 < t < t_{0.9975}(8) = 3.833$, possiamo riportare $0.0025 < P < 0.005$. Questo significa che possiamo rifiutare H_0 e accettare H_1: il dosaggio maggiore sembra più efficace.

Esercizio 3.

1. Vogliamo fare un test di ipotesi $H_0 : p_1 = p_2 = p_3$ contro l'alternativa $H_1 : \exists i, j$ tali che $p_i \neq p_j$. Raccogliamo i dati osservati ed attesi nella seguente tabella:

	sud		nord		totale
1	42	(34)	53	(61)	95
2	102	(101)	178	(179)	280
3	36	(45)	89	(80)	125
	180		320		500

   Per costruire la tabella attesa, abbiamo usato

$$\hat{p} = \frac{42 + 102 + 36}{95 + 280 + 125} = 0.36$$

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

$$\begin{aligned}\chi^2 &= \frac{(42-34)^2}{34} + \frac{(53-61)^2}{61} + \frac{(102-101)^2}{101} + \frac{(178-179)^2}{179} + \\ &+ \frac{(36-45)^2}{45} + \frac{(89-80)^2}{80} = 5.61\end{aligned}$$

   Il valore critico è $\chi^2_{0.95}(2) = 5.991$; siccome $\chi^2 < \chi^2_{0.95}(2)$, accettiamo H_0: non sembra che l'appartenenza regionale sia legata all'appartenenza alle diverse classi di reddito. Inoltre possiamo dire che $P > 0.05$ (6 punti).

2. Poichè abbiamo già accettato H_0, i tassi si possono ritenere tutti uguali fra di loro senza fare ulteriori tests (2 punti).

Esercizio 4. Partiamo calcolando le quantità:

$$\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i = 93 \quad (0.5 \text{ punti}), \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 57.13 \quad (0.5 \text{ punti}),\\
s_X^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = 4.66 \quad (0.5 \text{ punti}),\\
s_Y^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right) = 16.05 \quad (0.5 \text{ punti}),\\
s_{XY} &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y}\right) = 6.39 \quad (0.5 \text{ punti})
\end{aligned}$$

1. Calcoliamo i coefficienti della retta di regressione:

$$\begin{aligned}
b_1 &= \frac{s_{XY}}{s_X^2} = 1.37\\
b_0 &= \bar{Y} - b_1\bar{X} = -70.28
\end{aligned}$$

   La retta di regressione è quindi $y = 1.37x - 70.28$.

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

$$\begin{aligned}
s_{Y|X} &= \sqrt{\frac{n-1}{n-2}(s_T^2 - b_1^2 s_X^2)} = 2.96\\
s_{b_1} &= \frac{s_{T|X}}{\sqrt{(n-1)\cdot s_X^2}} = 0.56
\end{aligned}$$

   Abbiamo allora

$$t = \frac{b_1 - 0}{s_{b_1}} = 2.44$$

   Bisogna confrontare $|t|$ con una legge di Student a $\nu = 7 - 2 = 5$ gradi di libertà. Abbiamo che $t_{0.95}(5) = 2.015 < |t| = 2.44 < t_{0.975}(5) = 2.570$, quindi $0.05 < P < 0.1$. Con una P così alta siamo portati ad accettare H_0, il che significa che non sembra esserci una relazione lineare della media dei prezzi dei libri recensiti rispetto all'anno.

3. La media che ci attendiamo per i prezzi del 1997 (quindi con $x = 97$) è $y = b_1x + b_0 = 62.61$. L'errore standard della stima della y è

$$s := s_{Y|X}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{X})^2}{(n-1)s_x^2}} = 3.87$$

   e $t_{0.975}(5) = 2.570$, quindi l'intervallo di confidenza cercato ha estremi $y \pm t_{1-\alpha/2}(\nu)s = 62.61 \pm 2.570 \cdot 3.87$, ed è quindi $[52.66; 72.56]$

Esame di Statistica del 12 febbraio 2007 (Corso di Laurea in Biotecnologie, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Hanno superato la prova:

Agnolon Valentina	$21 + 3^-$
Anselmi Giulia	$24 + 3^+$
Bisin Marco	$18 + 3^-$
Bordin Fulvio	$20.5 + 3^-$
Cantarelli Irene Xochilt	$20.5 + 3^+$
Casano Alessandra Maria	$21.5 + 3^-$
Civettini Michele	$15 + 3^+$
De Prà Valentina	17
Ferro Chiara	20
Filograna Roberta	$25.5 + 3^+$
Foletto Mattia	18.5
Fortin Andrea	$15.5 + 3^+$
Gardiman Elisa	$16.5 + 3^-$
Gottardo Lisa	$23.5 + 3^-$
Loschi Luca	$16 + 3^+$
Lunardon Alice	$22.5 + 3^+$
Massarotto Luca	$16.5 + 3^-$
Nigris Sebastiano	$19 + 3^-$
Orlandi Veronica	$22 + 3^+$
Pianca Nicola	19
Pinto Marcella	$26 + 3^+$
Rigato Annafrancesca	25
Romoli Ottavia	$24 + 3^-$
Salvagnin Umberto	20
Stasi Fabio	$22.5 + 3^-$
Szekely Serena	$24.5 + 3^+$
Tebaldi Marta	$20 + 3^-$
Zilio Federica	$18 + 3^-$

Visione compiti corretti, registrazione voto e/o orali: giovedì 15 febbraio ore 14.30 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.