Esame di Istituzioni di Matematiche II del 11 luglio 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. Siete il manager di una ditta farmaceutica, e all'inizio dell'anno avete allo studio due progetti di ricerca su nuovi farmaci, entrambi del costo di 10 miliardi. Il progetto 1 ha una probabilità di riuscita del 60%, nel qual caso guadagnerete 50 miliardi, e il progetto 2 ha una probabilità di riuscita del 40%, nel qual caso guadagnerete 80 miliardi; se un progetto non riesce, non ci si guadagna niente e rimane la perdita netta di 10 miliardi. Se alla fine dell'anno avete perso soldi invece di guadagnarne, verrete licenziati.

1. Calcolare, per entrambi i progetti, il guadagno atteso e la probabilità di essere licenziati.

2. Supponiamo ora di poter fare un "investimento misto", investendo 5 miliardi nel progetto 1 e 5 miliardi nel progetto 2, con guadagni e perdite dimezzate per entrambi i progetti. Calcolare il guadagno atteso e la probabilità di essere licenziati per questo nuovo piano di investimento, supponendo che i due progetti diano risultati indipendenti tra di loro.

3. Supponiamo ora di poter fare un investimento misto più "flessibile", investendo λ miliardi nel progetto 1 e $(10-\lambda)$ miliardi nel progetto 2, con $\lambda \in [0, 10 \text{ miliardi}]$ (chiaramente, guadagni e perdite andranno in proporzione con λ). Calcolare il guadagno atteso e la probabilità di essere licenziati per questo nuovo piano di investimento, supponendo sempre che i due progetti diano risultati indipendenti tra di loro.

Esercizio 2. Dieci uomini dai 25 ai 34 anni hanno adottato una dieta vegetariana per 1 mese. Durante la dieta, nel campione l'assunzione media giornaliera di acido linoleico è stata di 13 g, con una deviazione standard di 6 g, mentre nella popolazione generale l'assunzione media giornaliera di acido linoleico è di 15 g. Supponiamo di essere incerti sull'effetto di una dieta vegetariana sul livello di assunzione dell'acido linoleico.

1. Eseguire un test statistico sull'effetto della dieta vegetariana sul livello di acido linoleico. Riportare un valore P.

2. Supponiamo che in un campione di dieci donne dai 25 ai 34 anni con la stessa dieta vegetariana, l'assunzione media giornaliera di acido linoleico sia stata di 16 g, con una deviazione standard di 5 g. Supponendo di non conoscere la media della popolazione generale, eseguire un test statistico per verificare se il livello di acido linoleico assunto vari a seconda del sesso. Riportare un valore P.

3. Calcolare l'intervallo di confidenza al 95% per l'assunzione media di acido linoleico.

4. Che risultato avrebbe il test del punto 1 con un livello $\alpha = 0{,}05$?

Esercizio 3. Nell'ambito di un'inchiesta sul doping (La Gazzetta dello Sport, 20 gennaio 2001) è stato preso un campione di ragazzi da 11 a 13 anni, chiedendo se assumessero creatina e/o aminoacidi, con le seguenti percentuali di risposte affermative:

Età	ragazzi	ragazze
11 anni	6,13	3,77
12 anni	8,46	6,22
13 anni	9,76	8,21

Supponendo che ogni gruppo sia composto da 2000 ragazzi (per un totale di 12000 intervistati), rispondere alle seguenti domande.

1. C'è una differenza significativa nel tasso di assunzione di creatina e/o aminoacidi tra ragazzi e ragazze? Fare un test Z, riportare il valore P e rispondere alla domanda.

2. C'è una differenza significativa nei tassi di assunzione di creatina e/o aminoacidi tra i tre gruppi di età? Fare un test χ^2, riportare il valore P e rispondere alla domanda. Eventualmente, identificare i gruppi con tassi di assunzione uguali facendo test multipli.

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 madre (y), padre (x) e primogenito (t), con i seguenti risultati:

$$\sum_{i=1}^{20} x_i = 2980, \qquad \sum_{i=1}^{20} y_i = 2620, \qquad \sum_{i=1}^{20} t_i = 2030$$

$$\sum_{i=1}^{20} x_i^2 = 451350, \qquad \sum_{i=1}^{20} y_i^2 = 351350, \qquad \sum_{i=1}^{20} t_i^2 = 210850$$

$$\sum_{i=1}^{20} x_i y_i = 390825, \qquad \sum_{i=1}^{20} x_i t_i = 305700, \qquad \sum_{i=1}^{20} y_i t_i = 269550$$

1. Testare l'ipotesi che la pressione della madre e quella del padre siano (linearmente) indipendenti.
2. Trovare la retta di regressione della pressione del figlio (var. dipendente) rispetto a quella della madre (var. indipendente).
3. Eseguire un test per vedere se in effetti c'è questa relazione lineare.

(per tutti i test sopra, si usi $\alpha = 0{,}05$).

Soluzioni

Esercizio 1. Rappresentiamo i risultati dei progetti con le seguenti variabili aleatorie:

$$X_1 = \begin{cases} -10 & \text{con prob. } 0{,}4 \\ 50 & \text{con prob. } 0{,}6 \end{cases} \qquad X_2 = \begin{cases} -10 & \text{con prob. } 0{,}6 \\ 80 & \text{con prob. } 0{,}4 \end{cases}$$

L'evento "essere licenziati" viene rappresentato con $\{X < 0\}$, dove X è una variabile aleatoria che rappresenta il guadagno totale e che varia da caso a caso (1 punto).

1. Se si investe il denaro nel primo progetto si ha $X = X_1$, e:

$$\mathbb{E}[X] = \mathbb{E}[X_1] = -10 \cdot 0,4 + 50 \cdot 0,6 = 26$$

In questo caso $\mathbb{P}\{X < 0\} = \mathbb{P}\{X_1 = -10\} = 0{,}4$ (1 punto).

Se invece si investe il denaro nel secondo progetto si ha $X = X_2$, e:

$$\mathbb{E}[X] = \mathbb{E}[X_2] = -10 \cdot 0,6 + 80 \cdot 0,4 = 26$$

In questo caso $\mathbb{P}\{X < 0\} = \mathbb{P}\{X_2 = -10\} = 0{,}6$ (1 punto).

2. In questo caso, il guadagno è $X = \frac{1}{2}X_1 + \frac{1}{2}X_2$, e si ha:

$$\mathbb{E}[X] = \frac{1}{2}\mathbb{E}[X_1] + \frac{1}{2}\mathbb{E}[X_2] = \frac{1}{2} \cdot 26 + \frac{1}{2} \cdot 26 = 26$$

In questo caso abbiamo che X è minore di 0 se e solo se sia X_1 che X_2 sono minori di 0, e quindi sfruttando l'indipendenza di X_1 e X_2 si ha $\mathbb{P}\{X < 0\} = \mathbb{P}\{X_1 = -10, X_2 = -10\} = \mathbb{P}\{X_1 = -10\} \cdot \mathbb{P}\{X_2 = -10\} = 0,4 \cdot 0,6 = 0{,}24$ (2 punti).

3. Si può interpretare $\lambda/10$ come la proporzione di capitale investita nei due diversi progetti. Il guadagno è quindi $X = \frac{\lambda}{10}X_1 + (1 - \frac{\lambda}{10})X_2$, e si ha:

$$\mathbb{E}[X] = \frac{\lambda}{10}\mathbb{E}[X_1] + (1 - \frac{\lambda}{10})\mathbb{E}[X_2] = \frac{\lambda}{10} \cdot 26 + (1 - \frac{\lambda}{10}) \cdot 26 = 26$$

In questo caso X può essere minore di 0 in diversi stati a seconda di quanto vale λ. Facciamo una tabella con i possibili valori di X in dipendenza da X_1 e X_2:

X_2 \ X_1	-10	50
-10	$-10^{(1)}$	$50\frac{\lambda}{10} - 10(1 - \frac{\lambda}{10}) = 6\lambda - 10^{(2)}$
80	$-10\frac{\lambda}{10} + 80(1 - \frac{\lambda}{10}) = 80 - 9\lambda^{(3)}$	$50 + 3\lambda^{(4)}$

La casella (1) è sempre minore di zero, e la (4) è minore di 0 per ogni $\lambda \in [0, 10]$. Nella casella (2), $X \geq 0$ se e solo se $\lambda \geq \frac{10}{6} \simeq 1{,}66$, e nella casella (4), $X \geq 0$ se e solo se $\lambda \leq \frac{80}{9} \simeq 8{,}88$. Il risultato finale è che:

- se $\lambda < 1{,}66$:

$$\mathbb{P}\{X < 0\} = \mathbb{P}(\{X_1 = -10, X_2 = -10\} \cup \{X_1 = 50, X_2 = -10\}) = \mathbb{P}\{X_2 = -10\} = 0,6$$

- se $1{,}66 \leq \lambda \leq 8{,}88$:

$$\mathbb{P}\{X < 0\} = \mathbb{P}\{X_1 = -10, X_2 = -10\} = 0,6 \cdot 0,4 = 0,24$$

- se $\lambda > 8{,}88$:

$$\mathbb{P}\{X < 0\} = \mathbb{P}\{X_1 = -10\} = 0,4 \qquad \text{(3 punti)}$$

Esercizio 2.

1. Si tratta di fare un test bilatero sulla media, di ipotesi $H_0 : \mu_X = 15$, $H_1 : \mu_X \neq 15$. Abbiamo che $\bar{X} = 13$ e $s_X = 6$. Costruiamo t:
$$t = \frac{\bar{X} - 15}{\sqrt{\frac{s_X^2}{10}}} = \frac{-2}{\sqrt{3,6}} = -1,05$$
Questo valore va confrontato con i quantili in tavola per trovare P. Scopriamo che il minor quantile in tavola è $t_{0,95}(9) = 1{,}833 > |t|$; questo significa che $P > 0{,}10$ (2 punti). La differenza quindi non è da considerare significativa.

2. Facciamo un test bilatero di ipotesi $H_0 : \mu_X = \mu_Y$ e alternativa $H_1 : \mu_X \neq \mu_Y$. Abbiamo che $\bar{Y} = 16$, $s_Y^2 = 5$. Costruiamo t:
$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{10} + \frac{s_Y^2}{10}}} = -1,21$$
Questo valore va confrontato con i quantili in tavola per trovare P. Scopriamo che il minor quantile in tavola è $t_{0,95}(18) = 1{,}734 > |t|$; questo significa che $P > 0{,}10$ (2 punti). Anche stavolta la differenza non è da considerare significativa.

3. Abbiamo $s_{\bar{X}} = s_X/\sqrt{10} = 1{,}90$. L'intervallo di confidenza ha estremi $\bar{X} \pm s_{\bar{X}} t_{0,975}(9) = 13 \pm 1,90 \cdot 2,262 = 13 \pm 4,29$, e quindi $I = [8,71; 17,29]$ (1 punto).

4. Avremmo accettato H_0, sia perchè $15 \in I$, sia perchè $P > 0,1 > \alpha$ (1 punto).

Esercizio 3.

1. Prima di tutto dobbiamo ottenere le percentuali per i due gruppi maschi (M) e femmine (F). Dato che ogni campione ha taglia uguale, unificandoli si possono fare delle semplificazioni:
$$\hat{p}_M = \frac{6,13 \cdot 2000 + 8,46 \cdot 2000 + 9,76 \cdot 2000}{6000}\% = \frac{6,13 + 8,46 + 9,76}{3}\% = 0,0812$$
$$\hat{p}_F = \frac{3,77 + 6,22 + 8,21}{3}\% = 0,0607 \quad (0,5 \text{ punti})$$
Facciamo ora un test di ipotesi $H_0 : p_M = p_F$ contro l'alternativa $H_1 : p_M \neq p_F$. Abbiamo:
$$\hat{p} = \frac{8,12 \cdot 6000 + 6,07 \cdot 6000}{12000}\% = \frac{8,12 + 6,07}{2}\% = 0,0709$$
Abbiamo $n_M \hat{p} = n_F \hat{p} = 2000 \cdot 0,07 \simeq 142 \geq 5$. Siccome $\hat{p} < 1/2$, sicuramente anche $n_M(1 - \hat{p}) = n_F(1 - \hat{p})$ saranno maggiori di 5, quindi si può fare il test Z. Abbiamo:
$$Z = \frac{0,0812 - 0,0607}{\sqrt{0,07(1 - 0,07)\left(\frac{1}{6000} + \frac{1}{6000}\right)}} = 4,40$$
Il valore ottenuto non appare sulle tavole, quindi non possiamo calcolare esattamente P. Possiamo però ottenerne una stima:
$$\begin{aligned} P &= \mathbb{P}\{|Z| > 4,40\} = 2\mathbb{P}\{Z > 4,40\} = 2(1 - \mathbb{P}\{Z \leq 4,40\}) = \\ &= 2(1 - F_Z(4,40)) < 2(1 - F_Z(2,99)) = 2(1 - 0,99861) \simeq 0,00278 \end{aligned}$$
Dato che P è molto piccolo, la differenza di assunzione di creatina e/o aminoacidi tra i due sessi è da considerare significativa (4 punti).

2. Prima di tutto dobbiamo ottenere le percentuali o i numeri per i tre gruppi di età (scegliamo i numeri). Come per il punto precedente, anche qui si possono fare delle semplificazioni:
$$\begin{aligned} O_{11} &= 2000 \cdot \hat{p}_{11M} + 2000 \cdot \hat{p}_{11F} = 2000(0,0613 + 0,0377) = 190 \\ O_{12} &= 2000(0,0846 + 0,0622) = 294 \\ O_{13} &= 2000(0,0976 + 0,0821) = 359 \quad (0,5 \text{ punti}) \end{aligned}$$
Vogliamo fare un test di ipotesi $H_0 : p_{11} = p_{12} = p_{13}$ contro l'alternativa $H_1 : \exists i, j$ tali che $p_i \neq p_j$. Raccogliamo i dati osservati ed attesi nella seguente tabella:

	sì		no		totale
11	198	(284)	3802	(3716)	4000
12	294	(284)	3706	(3716)	4000
13	359	(284)	3641	(3716)	4000
	843		1157		12000

Per costruire la tabella attesa, abbiamo usato $\hat{p} = 843/12000 = 0{,}00703$. Siccome tutti i numeri della tabella attesa sono maggiori di 5, si puó usare il metodo del χ^2:

$$\begin{aligned}\chi^2 &= \frac{(198-284)^2}{284} + \frac{(3802-3716)^2}{3716} + \frac{(294-284)^2}{284} + \frac{(3706-3716)^2}{3716} + \\ &+ \frac{(359-284)^2}{284} + \frac{(3641-3716)^2}{3716} = 49,93\end{aligned}$$

Il valore critico più grande in tabella è $\chi^2_{0,999}(2) = 13{,}81$; siccome $\chi^2 > \chi^2_{0,999}(2)$, abbiamo $P < 0{,}001$. Siccome questo valore è molto piccolo, rifiutiamo H_0, e accettiamo H_1 (4 punti). Facciamo ora dei confronti multipli per vedere se ci sono gruppi omogenei.

Partiamo dai due gruppi che appaiono più vicini:

	sì		no		totale
12	294	(327)	3706	(3673)	4000
13	359	(327)	3641	(3673)	4000
	653		7347		8000

Stavolta testiamo $H_0 : p_{12} = p_{13}$ contro $H_1 : p_{12} \neq p_{13}$. Costruiamo χ^2:

$$\begin{aligned}\chi^2 &= \frac{(294-327)^2}{327} + \frac{(3706-3673)^2}{3673} + \\ &+ \frac{(359-327)^2}{327} + \frac{(3641-3673)^2}{3673} \simeq 6,86\end{aligned}$$

Abbiamo che $\chi^2_{0,99}(1) = 6{,}63 < \chi^2 < \chi^2_{0,995}(1) = 7{,}87$, quindi possiamo dire che $0{,}005 < P < 0{,}01$. Notiamo che questo valore è molto più alto di quello ottenuto per il test principale. Inoltre, se facessimo un test iniziale con $\alpha \leq 0{,}01$, allora avremmo che $\alpha/2 \leq 0{,}005$, e potremmo accettare $H_0 : p_{12} = p_{13}$, e si possono accorpare i due gruppi:

	sì		no		totale
11	198	(284)	3802	(3716)	4000
12-13	653	(562)	7347	(7438)	8000
	843		1157		12000

Si ha ora $\chi^2 = 47{,}54$. Siccome $\chi^2_{0,999}(1) = 10{,}828$ è il quantile maggiore in tavola, possiamo riportare $P < 0{,}001$. La conclusione quindi è che con $P > 0{,}01$ si può ritenere che il consumo di creatina e/o aminoacidi sia comparabile tra i 12 e i 13 anni e diverso a 11 anni (3 punti). In realtà, al di là di quanto chiesto nell'esercizio, i dati grezzi suggeriscono una distribuzione diversa in 4 gruppi (11F, 11M-12F, 12M-13F, 13M).

Esercizio 4.

1. Bisogna fare una regressione tra le variabili y (variabile dipendente) e x (variabile indipendente), facendo il 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. Partiamo calcolando le quantità:

$$\begin{aligned}
\bar{X} &= \frac{2980}{20} = 149 \quad (0,5 \text{ punti}), \qquad \bar{Y} = \frac{2620}{20} = 131 \quad (0,5 \text{ punti}),\\
s_X^2 &= \frac{1}{19}(451350 - 20 \cdot 149^2) = 385,79 \quad (0,5 \text{ punti}),\\
s_Y^2 &= \frac{1}{19}(351350 - 20 \cdot 131^2) = 427,89 \quad (0,5 \text{ punti}),\\
s_{XY} &= \frac{1}{19}(390825 - 20 \cdot 149 \cdot 131) = 23,42 \quad (0,5 \text{ punti})
\end{aligned}$$

Calcoliamo le quantità per effettuare il test:

$$\begin{aligned}
b_1 &= \frac{s_{XY}}{s_X^2} = 0,06\\
s_{Y|X} &= \sqrt{\frac{19}{18}(427,89 - 0,06^2 \cdot 385,79)} = 21,22\\
s_{b_1} &= \frac{s_{Y|X}}{\sqrt{19 \cdot s_X^2}} = 0,25
\end{aligned}$$

Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{0,06}{0,25} = 0,24$$

Abbiamo poi $\nu = 20 - 2 = 18$. Abbiamo che $t(18)$ è $t_{0,975}(18) = 2{,}100 > |t|$, quindi accettiamo H_0, e non appare una relazione lineare tra madre e padre (1,5 punti).

2. Stavolta bisogna fare una regressione tra le variabili t (variabile dipendente) rispetto a y (variabile indipendente). Partiamo calcolando le quantità che non sono ancora state calcolate:

$$\begin{aligned}
\bar{T} &= \frac{2030}{20} = 101,5 \quad (0,5 \text{ punti}), \qquad (0,5 \text{ punti}),\\
s_T^2 &= \frac{1}{19}(210850 - 20 \cdot 101,5^2) = 252,89 \quad (0,5 \text{ punti}),\\
s_{YT} &= \frac{1}{19}(269550 - 20 \cdot 101,5 \cdot 131) = 190,53 \quad (0,5 \text{ punti})
\end{aligned}$$

Calcoliamo i coefficienti della retta di regressione:

$$\begin{aligned}
b_1 &= \frac{s_{YT}}{s_Y^2} = 0,45\\
b_0 &= \bar{T} - b_1\bar{Y} = 42,55
\end{aligned}$$

La retta di regressione è quindi $t = 0,45y + 42,55$ (1 punto).

3. Come prima, 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:

$$\begin{aligned}
s_{T|Y} &= \sqrt{\frac{19}{18}(252,89 - 0,45^2 \cdot 427,89)} = 13,25\\
s_{b_1} &= \frac{s_{Y|X}}{\sqrt{19 \cdot s_X^2}} = \frac{13,25}{\sqrt{19 \cdot 427,89}} = 0,15
\end{aligned}$$

Costruiamo t:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{0,45}{0,15} = 3$$

Abbiamo ancora $\nu = 20 - 2 = 18$, e $t_{0,975}(18) = 2{,}100 < |t|$, quindi rifiuteremmo l'ipotesi, con la conclusione che c'è una relazione lineare tra le pressioni di madre e figlio (1,5 punti).

Esame di Istituzioni di Matematica II del 11 luglio 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.