Esercizi

Esercizio 1

Uno studente universitario ha programmato di sostenere nella sessione estiva gli esami X e Y. Sia A l'evento "supera l'esame X" e sia E l'evento "supera l'esame Y", con P(A)=0.7, P(E)=0.5, P(A∩E)=0.4. Calcolare la probabilità che non superi nessuno dei due esami.

Risposta:

$P(\bar{A}\cap\bar{E})=1-P(A\cup E)=1-P(A)-P(E)+P(A\cap E)=$

$=1-0.7-0.5+0.4=0.2$

Esercizio 2

L'altezza di 450 studenti immatricolati all'Università di Roma Tre nel 1998 è risultata in media di 170 cm., con una ds di 7.5 cm. Nell'ipotesi che la statura si distribuisca come una Normale, quale è il numero di studenti con altezza:

a) maggiore di 180 cm.;

b) minore o uguale a 160 cm.;

c) tra 162.5 e 172.5.

Risposta:

a) $P(X>180)=P\left(Z>\frac{180-170}{7.5}\right)=P(Z>1.33)=0.092$

$n=0.092\cdot 450\cong 41$

b) $P(X\leq 160)=P\left(Z\leq\frac{160-170}{7.5}\right)=P(Z\leq -1.33)=0.092$

$n=0.092\cdot 450\cong 41$

c) $P(162.5<X<172.5)=P\left(\frac{162.5-170}{7.5}<Z<\frac{172.5-170}{7.5}\right)=$

$=P(-1<Z<0.33)=1-P(Z<-1)-P(Z>0.33)=$

$=1-0.15866-0.37070=0.47064 \quad n=0.47064\cdot 450\cong 212$

Esercizio 3

Un bimbo maschio di 4 anni ha un peso di 22.5 kg. Dalla tabella dei percentili i suoi genitori desumono che si trova al 90mo percentile. Il cugino avente la stessa età ma avente un peso di 18.2 kg si trova al 40mo percentile. Determinare μ e σ della popolazione gaussiana descrivente i pesi dei bambini di 4 anni.

Risposta: $z_{0.90}= 1.28$ $z_{0.40}= - 0.25$

$z=\frac{x-\mu}{\sigma}\Rightarrow$

$$\begin{cases}1.28=\frac{22.5-\mu}{\sigma}\\-0.25=\frac{18.2-\mu}{\sigma}\end{cases}=\begin{cases}\mu=22.5-\sigma\cdot 1.28\\\mu=18.2+\sigma\cdot 0.25\end{cases}=\begin{cases}\mu=22.5-\sigma\cdot 1.28\\22.5-\sigma\cdot 1.28=18.2+\sigma\cdot 0.25\end{cases}=$$

$$=\begin{cases}\mu=22.5-\sigma\cdot 1.28\\\sigma\cdot(1.28+0.25)=22.5-18.2\end{cases}=\begin{cases}\mu=22.5-2.81\cdot 1.28=18.90\\\sigma=4.3/1.53=2.81\end{cases}$$

Esercizio 4

A UK study of factors affecting the outcome of pregnancy among 1513 women reported that the overall incidence of pre-term births was 7.5%, SE=0.68%, 95% CI 6.1 to 8.8% (Peacock et al. 1995).

1.What is meant by SE=0.68%?

2.What is meant by 95% CI 6.1 to 8.8%?

3.What distributions are used to calculate the SE and the CI?

4.How would the CI change if 90% limits were used?

5.How would the CI change if 99% limits were used?

6.Another study conducted at about the same time in Denmark and including 51851 women, reported that the overall incidence of pre-term birth was 4.5% (95% CI 4.3 to 4.7%). Explain why this 95% CI is narrower than that reported in the UK study. Do you think that there is a real difference in pre-term birth rates between the two populations being studied?

Soluzioni

1. Standard error
2. È l'intervallo che stimiamo contenga la percentuale di nati pre-termine della popolazione generale.
3. Binomiale e gaussiana
4. Più stretto (6.4; 8.6%)
5. Più ampio (5.8; 9.2%)
6. n danese >> n UK -> più preciso e IC più stretto. Gli intervalli non si sovrappongono -> evidenza di una reale differenza

Esercizio 5

Un'urna contiene 4 palline bianche e 2 rosse, un'altra ne contiene 2 bianche e 4 rosse. Da una delle due urne scelta a caso è stata estratta una pallina rossa. Quale è la probabilità che sia stata estratta dalla prima urna?

Risposta:

$P(R|U_1)=2/6$ $P(U_1)=1/2$

$P(R|U_2)=4/6$ $P(U_2)=1/2$

$$P(U_1 \mid R) = \frac{P(U_1 \cap R)}{P(R)} = \frac{P(R \mid U_1)\cdot P(U_1)}{P(R \cap U_1) + P(R \cap U_2)} =$$

$$= \frac{P(R \mid U_1)\cdot P(U_1)}{P(R \mid U_1)\cdot P(U_1) + P(R \mid U_2)\cdot P(U_2)} = \frac{2/6 \cdot 1/2}{2/6 \cdot 1/2 + 4/6 \cdot 1/2} = \frac{1}{3}$$

Esercizio 6

L'incidenza alla nascita della sindrome genetica X è dell' 0.12% per età della madre inferiore ai 30 anni e del 0.28% per età superiore a 30 anni. La signora Y ha 10 figli di cui 3 avuti prima dei 30 anni e 7 dopo i 30 anni. Qual'è la probabilità che nessuno sia affetto da sindrome X ?

Risposta:

P(nessun figlio con sindrome di down)=$P(S_1 \cap S_2 \cap \ldots \cap S_{10})$=

Gli eventi sono tutti indipendenti, per cui

= $P(S_1) \cdot P(S_2) \cdot \ldots \cdot P(S_{10}) = (1-0.0012)^3 \cdot (1-0.0028)^7 = 0.9771$

Esercizio 7

Un individuo che prende le dovute precauzioni si ammala della malattia A con probabilità del 5%, se non le prende ha il 25% di probabilità di ammalarsi. Di un gruppo di 15 persone 10 hanno preso le precauzioni. Andrea, uno dei 15 del gruppo, non si ammala. Qual'è la probabilità che non abbia preso precauzioni?

Risposta:

A = si ammala   NA = non si ammala

P = prende precauzioni NP = non prende precauzioni

P(A|P)=0.05   P(A|NP)=0.25   P(P)=10/15   P(NP)=5/15

$$P(NP \mid NA) = \frac{P(NP \cap NA)}{P(NA)} = \frac{P(NA \mid NP)\cdot P(NP)}{P(NA \cap NP) + P(NA \cap P)} =$$

$$= \frac{P(NA \mid NP)\cdot P(NP)}{P(NA \mid NP)\cdot P(NP) + P(NA \mid P)\cdot P(P)} = \frac{(1-0.25)\cdot 5/15}{(1-0.25)\cdot 5/15 + (1-0.05)\cdot 10/15} =$$

$$= \frac{0.75/3}{0.75/3 + 0.95\cdot 2/3} = \frac{0.75}{0.75+1.9} = \frac{0.75}{2.65} = 0.28$$

Esercizio 8

Dopo anni di esperienza è nota che la distribuzione della concentrazione di rame nel sangue umano è ben descritta da una distribuzione di Gauss di parametri $\mu = 3.2 \times 10^{-5}$ e $\sigma = 2.2 \times 10^{-6}$.

All'ultimo esame del sangue trovo 9.2×10^{-5}.

Devo preoccuparmi?

$$Z = \frac{9.2\cdot 10^{-5} - 3.2\cdot 10^{-5}}{2.2\cdot 10^{-6}} = \frac{9.2-3.2}{2.2\cdot 10^{-1}} = \frac{6}{0.22} = 27$$

Il mio valore standardizzato è così elevato rispetto alla soglia di normalità pari a 1.96 che devo certamente preoccuparmi.

Esercizio 9

Un prodotto farmaceutico provoca un grave effetto collaterale a 3 pazienti su 100. Un'industria farmaceutica desidera sottoporre a prova il medicinale. Qual è la probabilità che l'effetto collaterale si verifichi al massimo in 1 soggetto in un campione casuale di 10 pazienti che prendono il medicinale? Qual è la probabilità che tutti i 10 soggetti manifestino l'effetto collaterale?

Risposte:

$P(X \le 1) = P(X=1) + P(X=0)$

$$P(X=1) = \binom{10}{1}\cdot 0.03^1 \cdot 0.97^9 = 0.2281$$

$$P(X=0) = \binom{10}{0}\cdot 0.03^0 \cdot 0.97^{10} = 0.7374$$

$$P(X \le 1) = 0.2281 + 0.7374 = 0.9655$$

$$P(X=10) = \binom{10}{10}\cdot 0.03^{10} \cdot 0.97^0 = 5.9049\cdot 10^{-16}$$

Esercizio 10

Un medico vorrebbe conoscere il valore medio della glicemia a digiuno in pazienti visti in una clinica per diabetici negli ultimi 10 anni. Determina il numero di cartelle cliniche che il medico dovrebb esaminare per ottenere un intervallo di confidenza al 90% per μ, se l'ampiezza dell'intervallo voluta è pari a 6 e un campione pilota porta ad una varianza di 60.

$$n = \frac{z^2\sigma^2}{d^2} = \frac{1.64^2 \cdot 60}{3^2} \cong 18$$

Esercizio 11

Si suppone che il peso della popolazione maschile di tesserati di una certa Federazione sportiva italiana di età >16 anni segua una legge normale di valor medio μ=75 Kg e deviazione standard σ=3.4 Kg. Qual è il peso massimo raggiunto dal 10% degli atleti più magri?

Si sa che il peso dei sedicenni italiani si distribuisce con media μ=65 e deviazione standard uguale a quella della federazione considerata. Preso un campione di 49 tesserati della federazione, si osserva un peso medio pari esattamente a 75. Gli atleti della federazione differiscono dai sedicenni della popolazione in quanto a peso?

Risposta:

$z_{10\%}$ =-1.28

$$-1.28 = \frac{x-75}{3.4} \Rightarrow x = 75 - 1.28 \cdot 3.4 = 70.65$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} = \begin{cases} H_0 : \mu = 65 \\ H_1 : \mu \neq 65 \end{cases}$$

$$z = \frac{75-65}{6.4/\sqrt{49}} = 20.59$$

Rifiuto H_0 poiché lo z osservato è di molto maggiore rispetto alla soglia pari a z=1.96. La differenza è altamente significativa

Esercizio 12

Quant'è la probabilità che su una famiglia di 5 figli, 2 siano maschi ? Che almeno 1 sia maschio? Che al massimo 4 siano femmine?

$$P(M=2) = \binom{5}{2} \cdot 0.5^2 \cdot 0.5^3 = 0.31$$

$$P(M \geq 1) = 1 - P(M=0) = 1 - \binom{5}{0} \cdot 0.5^0 \cdot 0.5^5 = 0.97$$

$$P(F \leq 4) = 1 - P(F=5) = 1 - P(M=0) = 0.97$$

Esercizio 13

Si vuole testare se un nuovo farmaco vasodilatatore per curare l'ipertensione abbia un effetto differente da quello attualmente in uso. Un precedente studio ha rilevato che un grande numero di pazienti affetti da ipertensione e sottoposti a farmaci tradizionali hanno una pressione arteriosa media pari a 123.5 mmHg con una ds pari a 17.4 mmHg. 20 soggetti ipertesi sono sottoposti a trattamento col nuovo farmaco ottenendo un valore medio di pressione arteriosa pari a 114.7. Si esegua un test statistico per valutare la significatività del risultato ottenuto (α =0.05).

1) Esprimere l'ipotesi nulla H_0 e l'ipotesi alternativa H_1.
2) Calcolare un'adeguata statistica test e trarre conclusioni
3) Scrivere l'espressione dell'intervallo di confidenza adeguato per la media e confrontarlo con il risultato del test
4) Si fissi un livello di significatività α = 2%. Si può rifiutare l'ipotesi nulla con livello di significatività α = 2%? Perchè?
5) Scrivere l'espressione dell'intervallo di confidenza al 98% per la media
6) L'intervallo di confidenza cosi ottenuto e in accordo con il risultato del test di significatività al 2%?
7) calcolare il p-value del test

Risposte:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} = \begin{cases} H_0 : \mu = 123.5 \\ H_1 : \mu \neq 123.5 \end{cases} \qquad z = \frac{114.7 - 123.5}{17.4/\sqrt{20}} = -2.26$$

Dato che sto applicando un test bilaterale le soglie della zona di accettazione sono -1.96 e 1.96. Il valore di z che osservo è inferiore rispetto al valore soglia -1.96, quindi con a=0.05 rifiuto H_0.

$$IC: \bar{x} \pm z_{\alpha/2} \cdot ES(\bar{x}) \quad \Rightarrow \quad 114.7 \pm 1.96 \cdot 17.4/\sqrt{20} \quad \Rightarrow \quad [107.08; 122.32]$$

L'intervallo di confidenza non comprende il valore del parametro di popolazione 123.5, per cui il nuovo farmaco produce un effetto diverso rispetto a quello già in uso.

Se a=0.02, le soglie della zona di accettazione diventano -2.33 e 2.33. In questo caso non posso rifiutare H_0 e dirò che il nuovo farmaco non produce effetti diversi dal farmaco vecchio.

$$IC: \bar{x} \pm z_{\alpha/2} \cdot ES(\bar{x}) \quad \Rightarrow \quad 114.7 \pm 2.33 \cdot 17.4/\sqrt{20} \quad \Rightarrow \quad [105.64; 123.76]$$

L'intervallo di confidenza concorda con il risultato del test

P-value = 2 · P(Z ≤ - 2.26) = 2 · 0.01191 = 0.02384

Esercizio 14

Rothemberg et al. hanno indagato sull'efficacia di utilizzare il hologic sahara sonometer, un dispositivo portatile che misura la densità minerale ossea (BMD) nella caviglia per prevedere una frattura. Hanno usato un valore limite biologico di densità ossea di 0.57. I risultati dell'indagine sono inseriti nella seguente tabella.

	Frattura confermata		
	Presente (D)	Assente (ND)	Totale
BMD ≤ 0.57 (T+)	214	670	884
BMD > 0.57 (T-)	73	330	403
Totale	287	1000	1287

a) Calcolare la Se dell'utilizzo del BMD con un valore di 0.57 per predire le fratture
b) Calcolare la Sp
c) Se il 10% della popolazione presentasse fratture, quale sarebbe il VPP di questo test?

Soluzione

a) 214/287=0.7456
b) 330/1000=0.33
c) VPP=0.11

Esercizio 15

Nel Framingham Study, sono stati misurati i livelli di colesterolo totale sierico in una coorte di 475 soggetti maschi di età 40-49 anni ed in buona salute. La coorte è stata poi seguita per sedici anni.

Alla fine di questo periodo, i soggetti sono stati divisi in due gruppi:

1) coloro che non avevano sviluppato malattie coronariche 354 (normali).

2) coloro che avevano sviluppato malattie coronariche 121 (malati).

La distribuzione dei livelli di colesterolo sierico nei due gruppi (all'inizio dello studio) era stata trovata approssimativamente gaussiana

- con $\mu 1$= 219 mg/dl e $\sigma 1$= 41 mg/dl nei soggetti che NON hanno poi sviluppato malattia
- con $\mu 2$ = 244 mg/dl e $\sigma 2$ = 51 mg/dl nei soggetti che hanno poi sviluppato la malattia

Quesiti

Determinare:

1.L'intervallo di confidenza al 95% del livello di colesterolo sierico per coloro che hanno sviluppato la malattia coronaria

2.L'intervallo di confidenza al 95% del livello di colesterolo sierico per coloro che NON hanno sviluppato la malattia coronaria

3.con quale probabilità un soggetto di 45 anni, che aveva un livello di colesterolo sierico totale maggiore o uguale a 250 mg/dl, svilupperà la malattia coronarica.

Soluzioni

1. C.I.=[244 ± (1.96)(51)/√121]=[234.9 ; 253.1]
2. C.I.=[219 ± (1.96)(41)/√354]=[214.8 ; 223.2]
3. p1=F[z>(250-219)/41]= 0.224
   p2=F[z>(250-244)/51]= 0.452
   p3= p2*(354/475)/[p1*(121/475)+p2*(354/475)]=
   =0.452*0.745/[0.452*0.745+0.224*0.255]=
   =0.337/[0.337+0.057]=0.855

Esercizio 16

Basandoci sui dati raccolti dal National Center for Health Statistics, considerando un campione di adulti, si osserva che il 23.53% degli adulti degli Stati Uniti è iperteso. Se scegliamo un campione casuale di 20 adulti e assumiamo che la probabilità che abbia l'ipertensione un individuo fra questi sia pari a 0.24, trovare la probabilità che il numero di soggetti ipertesi nel campione sia:

a)Esattamente 3

b)Maggiore o uguale a 3

c)Minore di 3

d)Fra 3 e sette, estremi inclusi

e)Quanti adulti ipertesi ci si aspetta di trovare in un campione di 20?

Soluzioni

a) 0.1484
b) 0.8915
c) 0.1085
d) 0.8080
e) Circa 5

Esercizio 17

I quozienti di intelligenza (QI) di individui ammessi ad una scuola statale per ritardati mentali si distribuiscono pressoché normalmente con media 60 e deviazione standard di 10.

a)Trovare la proporzione di individui con QI maggiore di 75

b) Qual è la probabilità che un individuo scelto a caso abbia un QI compreso tra 55 e 75?

c) Se campioniamo 100 soggetti che frequentano questa scuola, qual è la probabilità che il loro QI medio sia tra 55 e 75?

Soluzione

a) $P(x>75) = P[z>(75-60)/10] = P(z>1.5) = 0.06681 \rightarrow$ 6.7%
b) $P(55<x<75) = P[(55-60)/10<z<(75-60)/10] = P(-0.5<z<1.5) = 1-0.06681-0.30854=0.62465$
c) $P(55<\bar{x}<75)=P[(55-60)/(10/\sqrt{100})<z<(75-60)/(10/\sqrt{100})] =P(-5<z<15) \approx 1$

Esercizio 18

Lung function and prevalence of respiratory symptoms were compared between 42 cross-country skiers and 29 non-skiing controls (Larsson et al. 1993).

Forced expiratory volume in 1s (FEV1) was recorded as the percentage of that expected given age, sex and height.

Abnormal shortness of breath and other symptoms were recorded at an interview. Some results are shown in the table:

	N. of subjects	**FEV1**		**Abnormal breathless**	
		Mean	SD	Yes	No
Skiers	42	112	11	17	25
Non-skiers	29	105	14	4	25
		P<0.05		P<0.02	

Quesiti

- What method should be used to test the null hypotheses that skiers and non-skiers have the same mean FEV1? What conditions do the data have tu fulfil for the test to be valid?
- What method should be used to test the null hypotheses that skiers and non-skiers have the same prevalence of abnormal breathless? What conditions do the data have tu fulfil for the test to be valid?
- What conclusion can be drawn about skiing, lung function and respiratory symptoms?

Soluzione

- T-test. Distribuzione approssimativamente normale, varianze omogenee
- Chi quadro o test per proporzioni. Frequenze attese >5
- I cross-country skiers tendono ad avere migliori funzionalità polmonari ed è più verosimile che abbiano un affanno anormale rispetto ai non sciatori