Esercitazione

Il test statistico (1)

Esercizio 1

- Si supponga di estrarre un campione casuale di 40 soggetti dalla popolazione su cui valutare il valore di uricemia. Il valore medio calcolato su tali valori è $\bar{x}$ = 5.55 mg/dl. Se σ =1.1, qual è l'intervallo di confidenza al 95% della media?

$$I.C.\,95\% = 5.55 \pm 1.96\frac{1.1}{\sqrt{40}} = [5.21, 5.89]$$

Esercizio 2

Modelliamo con una gaussiana X un indice numerico delle proprietà cromatiche di una vernice blu di prussia. Supponiamo che da sperimentazioni precedenti si possa ipotizzare σ= 2. Presi 80 barattoli ed esaminate le loro proprietà cromatiche, si trovano i valori $x_1, \ldots, x_{80}$ dell'indice X, con media = 364.

1) Stimare l'IC della media di X al 99%.

2) Quanti barattoli andrebbero esaminati per avere una precisione pari a 0.1, al 99%?

Risposte

1) Con probabilità 0.99 la media cade nell'intervallo di confidenza dove $\bar{x} \pm \delta$

$$\delta = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} = \frac{2.58 \cdot 2}{\sqrt{80}} = 0.577$$

$IC_{99\%}$: [363.42; 364.58]

2) Volendo δ = 0.1 o più piccolo, cerchiamo il più piccolo intero n tale che

$$\delta = \frac{2.58 \cdot 2}{\sqrt{n}} \le 0.1$$

ovvero

$$n \ge \left(\frac{2.58 \cdot 2}{0.1}\right)^2 = 2663$$

Esercizio 3

- L'infezione da "E.canis" è una malattia parassitaria dei cani, che talvolta viene contratta dagli uomini. Tra gli uomini infetti, la distribuzione dei globuli bianchi è approssimativamente gaussiana media μ non nota e σ=3204/mm^3. Nella popolazione generale i globuli bianchi sono in media 7250/mm^3. Si ritiene che i globuli bianchi dei soggetti infetti abbiano, in media, dimensione inferiore.

1. Quali sono le ipotesi nulla e alternativa da specificare?
2. In un campione casuale di 15 soggetti infetti, si è trovato che il numero di medio di globuli bianchi è 4767/mm3. Se si prefissa una protezione pari al 95%, che cosa si può concludere?

Risposte

1. $\begin{cases} H_0 : \mu = 7250 \\ H_1 : \mu < 7250 \end{cases}$

2. n=15 $\bar{x} = 4767 / mm^3$ α=0.05

$$z = \frac{4767 - 7250}{3204 \big/ \sqrt{15}} = \frac{-2483}{827.27} = -3.00$$

Il valore della statistica test è inferiore al valore soglia –1.64 (test ad una coda), pertanto si rifiuta l'ipotesi nulla

Esercizio 4

Il guadagno settimanale di un negozio sia descritto da una v.a. gaussiana G. Dopo un anno di esercizio, si è trovato che G è una N(100, 20). Il padrone del negozio decide di fare pubblicità sui giornali cittadini e da questa si aspetta o un aumento del guadagno o anche una diminuzione (la pubblicità ha dei costi). Presuppone invece che la varianza non cambi.

i) Dopo 10 settimane di pubblicità ha registrato i guadagni g_1, ..., g_{10}, aventi media 105. Assumendo a=0.95, possiamo affermare che la pubblicità ha avuto un effetto?

Risposte

i) Indichiamo con μ il guadagno medio in regime di pubblicità. L'ipotesi H_0 è $\mu = 100$, l'ipotesi alternativa H_1 è $\mu \neq 100$. Stiamo supponendo che il campione abbia varianza 20. Usiamo la v.a.

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2 / n}} = \frac{\bar{X} - 100}{\sqrt{20/10}}$$

che nell'ipotesi H_0 è una gaussiana N(0, 1). Il quantile gaussiano $q_{1-\alpha 2} = q_{0.975}$ vale 1.96. La regione di rifiuto è quindi {z : |z| > 1.96}. Per il nostro campione sperimentale vale

$$Z = \frac{105 - 100}{\sqrt{20/10}} = 3.54$$

quindi rigettiamo l'ipotesi H_0.

Esercizio 5

Un cardiologo vuole valutare se un nuovo agente anti-ipertensivo può essere considerato più efficace del farmaco standard.

È noto che la pressione sistolica media nei pazienti che ricevono la sostanza standard è μ_0=130 mmHg. Il cardiologo effettua lo studio somministrando a 16 pazienti il nuovo farmaco e trova che la pressione sistolica media vale 119.5 mmHg.

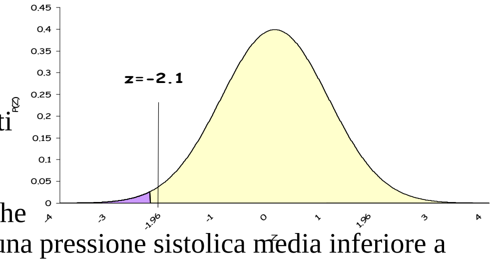
Supponendo che la deviazione standard comune per le due distribuzioni sia 20 mmHg, con un errore del primo tipo dell'uno per cento (α=0.01), quali conclusioni si posso trarre da questo risultato?

Risposte

Il sistema d'ipotesi è:
$$\begin{cases} H_0 : \mu_0 = 130\ mmHg \\ H_1 : \mu_0 < 130\ mmHg \end{cases}$$

Da cui deriva
$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{119.5 - 130}{20 / \sqrt{16}} = -2.1$$

e dalla lettura della tabella della distribuzione gaussiana sappiamo che $Z_{\alpha=0.01}$= - 2.33



Il cardiologo conclude che i risultati non offrono un'evidenza, a questo livello di significatività, per supportare l'ipotesi che i pazienti che ricevono la nuova sostanza hanno una pressione sistolica media inferiore a quella di coloro che che assumono la sostanza standard.

Esercizio 6

Il valore medio di un indicatore fisiologico, misurato nel corso di una vasta indagine sulla popolazione italiana, è risultato distribuito in modo approssimativamente gaussiano con media μ=50 e deviazione standard σ=26.3. In un campione di 25 individui residenti in una zona sospetta di inquinamento con prodotti tossici, si è osservato un valore medio pari a 60. Prefissato un livello di significatività α=0.01, ci si chiede se questo dato sia da considerarsi come scostamento casuale da 50 o se invece sia indice di una reale alterazione fisiologica attribuibile alla presenza del prodotto tossico nell'ambiente.

a. Si può affermare che nella popolazione residente in zona sospetta la media dell'indicatore è uguale a quella della popolazione di riferimento?

Risposta

Il sistema d'ipotesi è:
$$\begin{cases} H_0 : \mu_0 = 50 \\ H_1 : \mu_1 \neq 50 \end{cases}$$

Da cui deriva
$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{60 - 50}{26.3 / \sqrt{25}} = 1.90$$

e dalla lettura della tabella della distribuzione gaussiana sappiamo che $Z_{\alpha=0.01}$= 2.58

Quindi non rifiuto l'ipotesi nulla

Esercizio 7

L'indice di massa corporea, BMI, (Kg/m^2) misura il grado di sovrappeso di un soggetto. Per la popolazione di uomini di mezza età che svilupperanno diabete mellito, la distribuzione di BMI ha forma approssimativamente gaussiana con media μ non nota e deviazione standard σ=2.7 kg/m². Un campione casuale di 58 soggetti selezionati da questo gruppo ha fornito una media pari a 25kg/m².

a. Ad un livello di significatività di 0.05, saggiate se l'indice medio di BMI della popolazione di soggetti di mezza età che svilupperanno il diabete è uguale a 24 kg/m², cioè il valore medio della popolazione che non è affetta da diabete mellito.

b. Che cosa si può concludere?

Risposte

σ=2.7 kg/m²   n=58 $\overline{x}$=25 kg/m² α=0.05

a) $\begin{cases} H_0 : \mu_0 = 24 \\ H_1 : \mu_1 \neq 24 \end{cases}$ $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \dfrac{25 - 24}{2.7 / \sqrt{58}} = 2.82$

Il valore della statistica test z=2.82 è maggiore del valore soglia $z_{\alpha/2}$=1.96, pertanto si rifiuta l'ipotesi nulla.

b) Si può ritenere che il BMI dei soggetti che svilupperanno diabete mellito è mediamente maggiore di 24 kg/m²

Esercizio 8

Il calore (in calorie per grammo) emesso da un composto di cemento è (approssimativamente) normalmente distribuito di deviazione standard nota σ = 2.

Si vuole testare

$$\begin{cases} H_0 : \mu_0 = 100 \\ H_1 : \mu_1 \neq 100 \end{cases}$$

con un campione di dimensione n = 9.

1. Se la regione di accettazione fosse data dall'intervallo [98.5 – 101.5], quale sarebbe l'errore di primo tipo α?

Risposta

Sappiamo che α=P (rigettare H_0 dato che H_0 è vera) = P ($\overline{X}$ < 98.5 oppure $\overline{X}$ > 101.5| μ= 100) =

$$= P\left(\frac{\overline{X} - 100}{2/\sqrt{9}} < \frac{98.5 - 100}{2/\sqrt{9}}\right) + P\left(\frac{\overline{X} - 100}{2/\sqrt{9}} > \frac{101.5 - 100}{2/\sqrt{9}}\right) =$$

$$= P(Z < -2.25) + P(Z > 2.25) = 2 * 0.01222 = 0.02444$$

Esercizio 9

Dati tratti da: STUDIO LONGITUDINALE ITALIANO SULL'INVECCHIAMENTO ILSA (CNR)

E' noto che nella popolazione di donne in menopausa del nord Italia la distribuzione di glicemia è Normale con media=86.3 mg/dl e ds=25.6 mg/dl.
In un campione di n = 100 donne dell'Italia del sud si è ottenuta una media campionaria=96.7 mg/dl. Si assume che la ds sia ancora 25.6 mg/dl. In base a questi dati:

a) verificare se si può ritenere che le donne in menopausa provenienti dal sud abbiano valori di glicemia in media diversi da quelli delle donne del nord (α = 5%)

b) calcolare il p-value

c) costruire un intervallo di confidenza al 95% per la media della distribuzione della glicemia nelle donne del sud

Risposte (1)

- Definizione del sistema d'ipotesi

$$\begin{cases} H_0 : \mu_0 = 86.3 \\ H_1 : \mu_1 \neq 86.3 \end{cases}$$

- Calcoliamo il valore della statistica test

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{96.7 - 86.3}{25.6 / \sqrt{100}} = \frac{10.4}{2.56} = 4.06$$

- Il valore della statistica test è superiore al valore soglia pari a 1.96, pertanto rifiuto H_0 e posso affermare che le donne del sud hanno valori di glicemia diversi dalle donne del nord
- P-value=2*0.00002=0.00004

Risposte (2)

- Costruiamo l'intervallo di confidenza

$$I.C._{95\%} = 96.7 \pm 1.96 \frac{25.6}{\sqrt{100}} =$$

$$96.7 \pm 1.96 \cdot 2.56 = 96.7 \pm 5.02 =$$

$$[91.68, 101.72]$$

Esercizio 10

Un gruppo di clinici sta studiando la relazione tra trigliceridi e obesità.

Da dati CNR si ricava che negli individui con peso normale o leggero sovrappeso la distribuzione è N con media=145 mg/dl. e ds=80. mg/dl.

In un campione di n = 20 soggetti obesi si è ottenuta una media campionaria=165 mg/dl

Verificare in base a questi dati se si può ritenere che gli obesi abbiano in media valori di trigliceridi più elevati con un test ad un livello di significatività al 5%

Risposta

- Definisco il sistema d'ipotesi

$$\begin{cases} H_0 : \mu = 145 \\ H_1 : \mu > 145 \end{cases}$$

- Calcolo la statistica test

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{165 - 145}{80 / \sqrt{20}} = \frac{20}{17.89} = 1.12$$

- Z<del valore soglia pari a 1.64, pertanto non posso rifiutare H_0

Esercizio gaussiana

Per valutare l'efficacia della somministrazione di Ferro nella prevenzione dell'anemia sideropenica, si è trattato il 20% delle ragazze che frequentano la scuola secondarie con 1 compressa/die di solfato ferroso per 12 settimane consecutive.

Al termine dello studio la distribuzione dei livelli di emoglobina (Hb) aveva media **μ_T=140 g/l** e deviazione standard **σ_T=10 g/l** per le ragazze trattate, e media **μ_C=131 g/l** e deviazione standard **σ_C=12 g/l** per le ragazze non trattate.

Nell'ipotesi che la distribuzione dell'emoglobina sia gaussiana, calcola:

Domande	Risposte
La probabilità che una ragazza trattata abbia livelli di Hb minori di 125 g/l.	$z=(125-m_T)/s_T = (125-140)/10 = -1.5$ p(Hb<125) = F(-1.5) = 0.0668
La probabilità che una ragazza non trattata abbia livelli di Hb < 125 g/l.	$z=(125-m_C)/s_C=(125-131)/12= -0.5$ p(Hb<125) = F(-0.5) = 0.3085
La probabilità che una ragazza abbia livelli di Hb minori di 125 g/l.	$= f_T*p_T+f_C*p_C$ = (0.0668)(0.20) +(0.3085)(0.80) = 0.260
La probabilità che una ragazza con livelli di Hb < 125 g/l sia stata trattata.	p(trattata|Hb<125) = p(trattata&Hb<125)/p(Hb<125) = 0.0668 / 0.260 = 0.2569
La probabilità che **la media** dei livelli di Hb di 16 ragazze non trattate sia <125 g/l.	$z =(125-m_C)/(s_C/\sqrt{16})$ =(125-131)/(12/4) = -2.0 p(media(Hb)<125)= F(-2.0) = 0.0228

F(z*) è la funzione che stima l'area sotto la Gaussiana per [–inf<z<z*]