Verifica delle ipotesi

Monday, 5 June 2023 13:27

(Sto rifacendo l'esercitazione siccome non avevo capito una beata minchia)

1) Si nota che il diametro di anelli è approssimatamente Normalmente distribuito
   Ed ha una deviazione standard $\sigma = 0.001mm$
   Dato un campione $n = 15$
   E' nota $\bar{x} = 74.036mm$
   a. Si testi che la media del diametro degli anelli = 74.035mm ad un livello di signific
      $\alpha = 5\% = 0.05$
      E si calcoli il $pvalue$

      Dobbiamo prima di tutto trovare
      H0 = ipotesi nulla
      Qui noi vogliamo che la media del diametro = 74.035
      $H_0: m = 74.035$

      H1 = ipotesi alternativa (inversa)
      $H_1: m \neq 74.035$

      Leggendo dal formulario

H_0	H_1	Statistica	Regione critica
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n}$	$\left\lvert\frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n}\right\rvert > z_{\alpha/2}$

$RC = \left|\frac{\overline{x_n} - m_0}{o}\sqrt{n}\right| > z_{\frac{\alpha}{2}}$

$n = 15 \rightarrow 15\ anelli$

$\overline{x_n} = media = 74.036$

$\sigma = 0.001$

$\alpha = 5\% = 0.05$

$m = 74.035$

Quindi abbiamo tutti i valori da sostituire (Yey!)

$z_{\frac{\alpha}{2}} = z_{0.025}$

Noi abbiamo una gaussiana tale che

$\phi(z) = 1 - 0.025$

Quindi

$\phi(z) = 0.975$

Nodiamo che è nell'incrocio tra 1.9 e 0.06

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750

$z = 1.96$

Quindi

$\phi(1.96) = 0.975$

Quindi se

$$\left|\frac{\overline{x_n} - m_0}{o}\sqrt{n}\right| > 1.96$$

$$= \left|\frac{74.036 - 74.035}{0.001} * \sqrt{15}\right| >^? 1.96$$

$$= \cdots = \sqrt{15} > 1.96 \to 3.87 > 1.96$$

Quindi è vero, e quindi rifiuto h_0 con $\alpha = 5\%$

Questo è un errore di prima specie

$P(Errore\ 1\ specie) = 5\%$

$P(Errore\ 1\ specie) =$ rifiutare h_0 quando h_0 è vera

Ora dobbiamo calcolare pvalue $= \bar{\alpha}$

E cos'è il pvalue? Praticamente il valore massimo che α può avere per poter acc
h_0

Il pvalue ci dice quanto i valori sono in contraddizione tra di loro.

$$\bar{\alpha} \to \left|\frac{\overline{x_n} - m_0}{o}\sqrt{n}\right| = z_{\frac{\bar{\alpha}}{2}}$$

$3.87 = z_{\frac{\bar{\alpha}}{2}}$

$\bar{\alpha} = 2\big(1 - \phi(3.87)\big)$

(Formula da ricordare a memoria)

Dobbiamo trovare

$\phi(3.87) \simeq 0.0001$

$\bar{\alpha} = 0.0002$

b. Assumendo $\alpha = 5\%$ l'ipotesi che la media del diametro $= 74.035mm$ non veng
rifiutata quando il vero valore della media $= 74.034$

$m = 74.034$

Ora vogliamo trovare che non rifiuta

Formula che rifiuta:

$$\left|\frac{\overline{x_n} - m_0}{o}\sqrt{n}\right| > z_{\frac{\alpha}{2}}$$

Formula che non rifiuta:

$$-z_{\frac{\alpha}{2}} \le \left|\frac{\overline{x_n} - m_0}{o}\sqrt{n}\right| \le z_{\frac{\alpha}{2}}$$

$$P_{m=74.034}\left(-1.96 \le \left|\frac{\overline{x_n} - 74.035}{0.001}\sqrt{15}\right| \le 1.96\right)$$

Noi qui stiamo cercando

La probabilità di non rifiutare h0 quando h0 è falsa = errore seconda specie

Noi dobbiamo ricondurre il tutto ad una gaussiana standard

$N(74.034, (0.001)^2)$

Il problema nostro è questo: 74.035 dovrebbe essere 74.034

Per farlo dobbiamo sommare 0.001 sopra

Però per farlo dobbiamo sommare 0.001 sia a sinistra che a destra

Quindi

$$P_{m=74.034}\left(-1.96 + \frac{0.001}{0.001}\sqrt{15} \le \frac{\overline{x_{15}} - 74.035 + 0.001}{0.001}\sqrt{15} \le 1.96 + \frac{0.001}{0.001}\right.$$

$$= P\left(1.91 \le \frac{\overline{x_{15}} - 74.034}{0.001} \le 5.83\right) \sim N(0,1)$$

Quindi ora possiamo applicare le nostre formule carine e coccolose

$= \phi(5.83) - \phi(1.91) = 1 - 0.9719 \simeq 0.03$

2) Dicono che la media è pari a 14 ore.

Su un campione di 10 batterie, questi casi:

18	17	14	16	15	12	13	15	17	1

a. Si può approvare con $\alpha = 5\%$

Per la ipotesi, a noi ci importa la media

h_0: $m \le 14$

h_1: $m > 14$

Quindi controlliamo in maniera pessimista

$$RC = \frac{\overline{x_n} - m}{S_n}\sqrt{n} > t_{n-1,\alpha}$$

$m = 14$

$n = 10$

$\alpha = 0.05$

$$\overline{x_n} = \frac{1}{n}\sum x_i = \frac{1}{10}(18 + 17 + \cdots) = 15$$

$$S_n^2 = \frac{1}{n-1}\sum(\overline{x_n} * x_i)^2 = \frac{1}{9}((18 * 15)^2 + (17 - 15)^2 + \cdots)) = 4$$

$S_n = 2$

$t_{9,0.05} = 1.833$

Quindi ora possiamo sostituire

$$\frac{15-14}{2}\sqrt{10} >^? 1.833 \rightarrow 1.58 > 1.833$$

Quindi accetto h_0 a livello di significavità 5%

Quindi statisticamente non è vero per forza che la media sarà 14

b. Calcolare $\bar{\alpha}$

$$\frac{\overline{x_n} - m}{S_n}\sqrt{n} = t_{n-1,\bar{\alpha}}$$

Noi sappiamo che

$1.58 = t_{9,\bar{\alpha}}$

Quindi dobbiamo usare la tavola ora all'inverso

Guardando la tavola notiamo che

$0.05 < \bar{\alpha} < 0.10$

3) Una azienda dice il 90% delle persone che soffrono vengono guarite.
Dai test su 100 individui, 88 dichiarano che è stato efficace.
Supponiamo p la vera proporzione.

Noi dobbiamo mostrare che è vero

h_0: $p \geq 0.9$

h_1: $p < 0.9$

Questo è un campione numeroso estratto da una popolazione bernoulliana (aka 1
con media e varianza incognita

$np > 5 \rightarrow si$

$n(1-p) > 5 \rightarrow si$

Quindi possiamo utilizzare la cosa di h_0, h_1

$p \geq p_0$	$p < p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} < -z_\alpha$

$\alpha = 5\% = 0.05$

$z_\alpha = z_{0.05}$

E noi dobbiamo trovare quel valore tale per cui

$\phi(x) = 1 - 0.05 = 0.95$

$x = 1.645$

Quindi ora possiamo fare il test

$$\frac{\overline{x_n} - p}{\sqrt{p - (1-p)}}\sqrt{n} <^{?} -1.645$$

$\simeq -0.667 < -1.645 \rightarrow no$

Quindi siccome non è minore

Accetto h_0

$\bar{\alpha} > 0.05$

$\bar{\alpha} \rightarrow 0.667 = -z_{\bar{\alpha}} \rightarrow z_{\bar{\alpha}} = 0.667$

$$\bar{\alpha} = 1 - \phi(0.667)$$

4) I dati seguenti mettono in relazione frequenza cardiaca di 12 individui prima e dop masticato tabacco.

Individuo	Freq prima	Freq dopo
1	73	77
2	67	69
3	68	73
4	60	70
5	76	74
6	80	88
7	73	74
8	77	82
9	62	69
10	58	61
11	82	84
12	78	80

a. Verificare al livello 5% che masticare tabacco non aumenti la frequenza card (in h0 vera caso contrario)

h_0: $m_d \leq 0$

h_1: $m_d \geq 0$

(affermazione forte)

$m_d = m_x - m_y$

$m_x =$ vera media dopo masticato tabacco

$m_y =$ vera media prima masticato tabacco

$$RC = \frac{\overline{x_d}}{S_d}\sqrt{m_d} > t_{n-1,\alpha}$$

1

$$\overline{x_d} = \frac{}{12}\left((77-73)+(69-67)+(73-68)+\cdots\right) = 3.75$$
$$S_d^2 = \frac{1}{12-1}\left((4-3.75)^2+(2-3.75)^2+\cdots\right) \simeq 9.477$$
$$S_d = \sqrt{S_d^2} \simeq 3.078$$
$$RC = \frac{3.75}{3.078}\sqrt{12} >^{?} t_{11,0.05}$$
$$\simeq 4.22 > 1.796$$
Quindi rifiuto h_0 al livello 5%

b. $\bar{\alpha}$ del test
Siccome l'abbiamo rifiutato sicuramente
$\bar{\alpha} < \alpha$
E per trovarlo
$$\frac{\overline{x_d}}{S_d}\sqrt{n} = t_{n-1,\bar{\alpha}} \to \bar{\alpha} = 0.0014$$
Quindi i dati sono in contraddizione significativa con h_0

5) ...
$\alpha = 95\% \to (-0.61, 0.78)$
$H_0: m = 1, H_1: m \neq 1$
Da notare che $m = 1$
E nel nostro intercallo non abbiamo $m = 1$
Quindi rifiutiamo al tesi
E quindi α è l'opposto
$\alpha = 0.05 \to \alpha < 0.05$

6) $n = 25$
Primi $= 86.1$
Secondi $= 92.2$
$\sigma_1 = 2.09$
$\sigma_2 = 2.49$

a. $$S_P^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2} = \frac{24 * 2.09^2 + 24 * 2.49^2}{50-2} = 5.28$$

b. $\alpha = 5\%$
$\alpha = 1\%$
$H_0: m_x = m_y$
$H_1: m_x \neq m_y$
$$\left|\frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}\right| > t_{n_x+n_y-2,\frac{\alpha}{2}}$$
$$\left|\frac{86.1-92.2}{5.28\sqrt{\frac{2}{25}}}\right| > t_{48,\frac{\alpha}{2}}$$

$t_{48,0.025} = 2.011$

$\left|\frac{-6.1}{1.49}\right|$

$4.09 > 2.01$

Rifiuto h_0

$4.09 > t_{48,0.005} = 4.09 > 2.682$

Rifiuto h_0

7) $H_0: lol = lo_0$
Viene rifiutato a livello $\alpha = 3\%$
Ed allora il p-value sarà più grande
Quindi il p-value del test è minore di 0.03
Alpha > pvalue = rifiuto
Alpha < pvalue = accetto

8) $N(m, \sigma^2)$
$H_0: m < 5$
$H_1: m \geq 5$
$\bar{\alpha} = 0.06$

Più α è grande, e più l'intervallo di confidenza si restringe
Più α è piccolo, e più l'intervallo di confidenza è largo
Noi dobbiamo scegliere un alpha che include $\bar{\alpha}$
Siccome stiamo lavorando sui sottoinsiemi
Quindi $\alpha = 5\%$

9) $\sigma = 0.5$
$n = 100$
$\bar{x}_n = 15.1$

a. Media intervallo $\alpha = 1 - 0.9 = 0.10$

$\left(\bar{x}_n \pm z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$

$\phi(z_{1-0.05}) = \phi(z_{0.95}) = 1.645$

$\left(15.1 \pm 1.645 * \frac{0.5}{10}\right) = (15.017, 15.182)$

b. $h_0: m = 15$

$h_1: m \neq 15$

c. $\left|\frac{\overline{x_n} - m}{\sigma}\sqrt{n}\right| > z_{\frac{\alpha}{2}}$

$\left|\frac{15.1 - 15}{0.5} * 10\right| > 1.645$

$2 > 1.645$

Quindi il contenuto dei sacchi è diverso da 15kg

d. $\phi(2) = 1 - \frac{\bar{\alpha}}{2}$
$(\phi(2) - 1) * 2) = -\bar{\alpha}$
$-(0.9772 - 1) * 2 = \bar{\alpha}$
$\ldots = \cdots = 0.0456$

10) $\sigma^2 = 16 \rightarrow \sigma = 4$
$n = 15$
$\bar{x} = 31.33$

a. $\alpha = 0.05$
$z_{0.025} = \phi(z_{1-0.025)=\phi(0.975)} = 1.96$
$\left(31.33 \pm 1.96 * \frac{4}{\sqrt{15}}\right) = 31.33 \pm 2.024 = (29.306, 33.354)$

b. $h_0: m = 32$
$h_1: m \neq 32$
$\left|\frac{\bar{x}_n - m}{\sigma}\sqrt{n}\right| > z_{\frac{\alpha}{2}}$
$\left|\frac{31.33 - 32}{4}\sqrt{15}\right| > 1.96$
$0.64 > 1.96$
Accetto h_0
Notare che questo sarebbe stato possibile farlo senza calcoli
Guardare l'intervallo di confidenza

$\alpha = 1\%$ si accetta sicuramente

c. N affinchè ampiezza inferiore 1cm
$1.96 * \frac{4}{\sqrt{n}} = 0.5$
$\left(\frac{1.96 * 4}{0.5}\right)^2 = n = 245.86$
$n \geq 245.86$

11) N=70
$P = \frac{37}{70}$
Sappiamo che è bernulliana siccome o sei miopa oppure no

a. $\left(\overline{x_n} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x_n}(1 - \overline{x_n}}{n}}\right)$
$\sqrt{\frac{37}{}\left(1 - \frac{37}{}\right)}$

$$\frac{37}{70} \pm 1.96\sqrt{\frac{\frac{37}{70}\left(1-\frac{37}{70}\right)}{70}} = \frac{37}{70} \pm 0.11 = (0.4116, 0.6455)$$
$\alpha = 0.05$
$\phi\left(z_{1-\frac{0.05}{2}}\right) = 1.96$

b. Sappaimo che $\bar{x} = \frac{45}{100}$
E' più diffuso per $\alpha = .05$
$h_0: p = 0.45$
$h_1: p \neq 0.45$

$$\left|\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}\right| > z_{\frac{\alpha}{2}}$$
$$\left|\frac{\frac{37}{70} - \frac{45}{100}}{\sqrt{\frac{45}{100} * \frac{55}{100}}}\sqrt{70}\right| > 1.96$$
$1.19 > 1.96$
Falso quindi accetto h_0
Quindi non posso concludere

c. F
$$\frac{196}{100}\sqrt{\frac{\frac{37}{70}\left(1-\frac{37}{70}\right)}{n}} = \frac{2}{100}$$
$$\sqrt{\frac{\frac{37}{70}\left(1-\frac{37}{70}\right)}{n}} = \frac{2}{196}$$
$$\frac{\frac{37}{70}\left(1-\frac{37}{70}\right)}{n} = \left(\frac{2}{196}\right)^2$$
$$\left(\frac{196}{2}\right)^2 * \left(\frac{37}{70}\right)\left(\frac{33}{70}\right) = n$$
$n = 2393$

12) $\bar{x} = ?$
$\sigma^2 = 0.0256$
$n = 25$
$\bar{x}_n = 0.68$

a. Costruire livello confidenza
$$\overline{x_n} \pm z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

$$0.68 \pm 1.96 * \frac{\sqrt{0.0256}}{\sqrt{25}} = (0.917, 1.04)$$

b. 1 è dentro l'intervallo, quindi accetto h_0
c. Hai rotto i coglioni

13) Regressione lineare FINALMENTE

X	Y	X^2	Y^2	XY
7	25	49	625	175
17	0	289	0	0
8	11	64	121	88
36	-26	1296	676	936
23	-2	529	4	46
19	-18	361	324	342
17	8	289	64	136
2	26	4	676	52
5	35	25	1225	105

$n = 9$
$\bar{x} = 14.89$
$\bar{y} = 6.56$
$y = \alpha + \beta \bar{x}$
$\alpha = \bar{y} - \beta x$
$\beta = \frac{S_{xy}}{S_{xx}}$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 2906 - 9 * 14.89^2 = 910$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 1000.89 \rightarrow errore = -1576.44$$

$\beta = 1.09 \rightarrow erore \rightarrow -1.73$
$\alpha = 14.89 - 1.09 * 14.89 = -0.8 \rightarrow errore \rightarrow 32.32$

$$S_{yy} = \sum y_i^2 - n\bar{y}^2 = 3327.7$$

$$S_{rr} = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}} = 596.05$$

(Ho sbagliato qualche calcolo)

$$R^2 = 1 - \frac{SS_r}{S_{yy}} = 0.82$$

Quindi ora posso rispondere alla domande della prof

$Y = 32.32 - 1.73x$
Appare adeguato siccome
SS_r è sopra a 0.7