Verifica ipotesi

Sunday, 14 May 2023 16:34

- Ogni qualvolta che noi facciamo dei test sui parametri, noi stiamo facendo dei test parametrici (però esistono anche test non parametrici)
    - Fatta ipotesi statistica bisogna controllare se è compatibile col campione

  T1 = ipotesi alternativa
  T0 = ipotesi nulla -> Rifiutata se risulta incompatibile con i dati del campione
  Bisogna trovare una regola, e la regola è basata sulla statistica (che deciderà)
  -> Statistica del test
  Bisogna trovare una regione critica (aka regione di rifiuto) dove rifiuto se
  I dati campionari stanno in c $(x_1 \dots x_n) \in c$
  Gli errori che possiamo avere sono di 2 tipi:
    - Prima specie: Si rifiuta H0 ed H0 è vera
    - Seconda specie: Si accetta H0 ed H0 è falsa
      Aka rifiuto H1 ed è vera

  E noi vogliamo cercare di rendere piccoli sia gli errori di prima e seconda specie
  Però, rendendo più piccolo l'errore di prima specie, diminuiamo la regione critica
  E questo vuol dire che forse avremmo più errori di seconda specie
  Ciò che si fà è creare un $\alpha = 0.1, 0.05, 0.01$
  E noi vogliamo far si che $H_0 \le \alpha$
  $P_{H_0}\big((x_1, \dots, x_n) \in c\big) \le \alpha$
- Si suppone
  $N(m, o^2)$
  $H_0: m = m_0$
  $H_1: m \neq m_0$
  E noi vogliamo che
  $|\overline{x_n} - m| > c$
  $C = \{(x_1, \dots, x_n) \in R^n: |\overline{x_n} - m_0| > c\}$
  E bisogna trovare c con
  $P_{m_0}(|\overline{x_n - m_0}| > c) = \alpha$
  Con un pò di magia
  $$P_{m_0}\left(\sqrt{n} * \frac{|x_n - m_0|}{o} > \frac{c\sqrt{n}}{o}\right) = \alpha$$
  (Così facendo abbiamo una normale standard)
  E con altre magie
  $$\left|\sqrt{n} * \frac{\overline{x_n} - m_0}{}\right| > z\alpha \to (|z| > z\alpha)$$

E quindi il nostro intervallo diventa

$$\left(\overline{x_n} \pm \frac{o}{\sqrt{n}} * z_{\frac{\alpha}{2}}\right)$$

Qui invece errore secondo tipo (credo?)

$$\bar{\alpha} = 2\left(1 - \phi\left(\left|\frac{\overline{x_n} - m_0}{\frac{o}{\sqrt{n}}}\right|\right)\right)$$

[Mi manca l'altro prof]

- Abbiamo 2 casi:
    - $H_o: m = m_0, \qquad H_1\ m > m_0$
    - $H_0: m = m_0, \qquad H_1\ m < m_0$

  In questi casi noi sappiamo che

  $$\frac{X_n - m_0}{\frac{o}{\sqrt{n}}} \sim N(0,1)$$

  Si rifiuta se

  $$z = \sqrt{n} * \frac{\overline{x_n} - m_0}{o} > z_\alpha$$

  Aka rifiuto h0 a livello di α se $m_0 \notin \left(\overline{x_n} - z_\alpha * \frac{o}{\sqrt{n}}, +\infty\right)$

  E questo è l'estremo inferiore di confidenza

  $$\bar{\alpha} = 1 - \phi\left(\frac{\overline{x_n} - m_0}{\frac{o}{\sqrt{n}}}\right)$$

  (Mi sto rompendo di sta prof, giuro non riesco a seguirla)

  Mi aspetto di rifiutare se

  L'unico fatto da ricordare è che

  $z_{1-\alpha} = -z_\alpha$
- Test per la media di una popolazione normale con varianza incognita

  $$T := \frac{\overline{x_n} - m_0}{\frac{S_n}{\sqrt{n}}} \sim t_{n-1}$$
- Test di una proporzione (grandi campioni)

  Si vuole sottoporre

  $h_0: p = p_0\ vs\ h_1: p \neq p_0$

  E supponiamo che il campione $\sim Be(p)$

  Quindi sappiamo che

  $E(X) = p$

  $X = p(1-p)$

  Formula: se

  $n \geq 30$

  $np_0 > 5$

$n(1 - p_0) > 5$
Allora possiamo supporre che $X \sim N(0,1)$
E quindi possiamo usare le regioni critiche del test Z

H_0	H_1	Statistica	Regione critica
$p = p_0$	$p \neq p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\left\lvert \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \right\rvert > z_{\alpha/2}$
$p \leq p_0$	$p > p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} > z_\alpha$
$p \geq p_0$	$p < p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} < -z_\alpha$

- Varianza campionaria combinata:

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_X + n_y - 2}$$

H_0	H_1	Statistica	Regione critica
$\mu_x = \mu_y$	$\mu_x \neq \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\left\lvert \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \right\rvert > t_{n_x+n_y-2,\alpha/2}$
$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} > t_{n_x+n_y-2,\alpha}$
$\mu_x \geq \mu_y$	$\mu_x < \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < -t_{n_x+n_y-2,\alpha}$

- Test non parametrici
  Basta mi arrendo a seguire sta prof
  Con tutto il mio impegno non riesco a comprenderla, una dei prof peggiori di quest'anno scolastico. Mi sembra di star ascoltando un muro.
  Apro gli appunti di quack e vaffanculo.
  E leggo il pdf, non riesco più ad ascoltare la voce della prof, mi genera nervoso e fa

- Esercio teorico dal pdf
  Da un comune, dopo una votazione sono usciti i seguenti risultati

Partito	A1	A2	A3	A4
%	32	26	16	25

In un altro comune,
Abbiamo avuto in totale 320 voti con i seguenti risultati

Partito	A1	A2	A3	A4
Num	118	71	62	69

Uniamo le tabelle trasformando
Le percentuali in valori $(N * x_i)$

Partito	A1	A2	A3	A4
Num	118	71	62	68

%->f_i	102.4	86.4	51.2	80

Ora per ogni valore calcoliamo lo squarto quadratico

$$Q = \sum \frac{(n_i - f_i)^2}{f_i} \sim X^2(N-1) = 8.91$$

$Scegliamo\ q = 8.91$

Si rifiuta se

$q > X^2_{n-1,\alpha} \rightarrow 8.91 > 7.815$

Quindi rifiutiamo

- Dati infiniti

  Gli incidenti seguono una legge $P_o(0,4)$

  E nelle ultime 85 settimane sono stati riletavi:

N incidenti	0	1	2	3+
N settimane	50	32	3	0

Si può affermare che il modello è ancora verificabile?

N=85

$Y \sim P_0(0.4)$

Calcoliamo ora la frequenza della poisson

$$f_0 = P(Y=0) = \frac{e^{-0.4}(0.4)^k}{k!} = e^{-0.4} = 0.67$$

Ed il nostro valore esce come

$n_0 = n * f_0 = 56.95$

$n_1 = 21.78$

$n_2 = 4.59$

$$f_3 = P(Y \geq 3) = 1 - \sum_{i=0}^{2} P(Y-i)$$

$$= 0.008$$

$n_3 = 0.68$

Mettiamo questi valori nella tabella di sopra

N incidenti	0	1	2	3+
N settimane	50	32	3	0
n_i	56.95	22.79	4.59	0.68

Notiamo che 2, 3 non seguono la regola empirica che dice che
Se il valore è inferiore di 5 allora non può essere usato.
Quindi dobbiamo accorporare 2 e 3

N incidenti	0	1	2+

N settimane	50	32	3
n_i	56.95	22.79	5.27

Ora usiamo la formula

$$Q = \sum \frac{(N_i - f_i)^2}{f_i} = 5.5576$$

$Dato\ \alpha = 5\%$

$q > X^2_{2,0.05}$

$5.5576 > 5.991 \rightarrow rifiuto\ h_o$

- Esercizio variabile continua

I tempi di vita di 100 lampadine estratte casualmente è come segue

Tempo vita	N lampadine
>1	24
1-2	16
2-3	20
3-4	14
4-5	10
5-10	16
10+	0

Si può affermare che il tempo di vita segue una legge esponenziale?

h_0: $\exp(\lambda)$

$Y \sim Exp(\lambda)$

$$E(Y) = \frac{1}{\lambda}$$

E sappiamo che $E(Y)$ è stimato con $\overline{x_n}$

E quindi

λ simato con $\frac{1}{x_n}$

Approfondiamo la tabella cercando il punto medio dei dati

Il punto medio dei dati è

Tra 0 e 1 = 0.5

Tra 1 e 2 = 1.5

Es. 1 e 3 = 2

Ora guardiamo la tabella in sè per sè

Tempo vita T_i	Punto medio $\overline{x_i}$	N lampadine
>1	0.5	24
1-2	1.5	16

2-3	2.5	20
3-4	3.5	14
4-5	4.5	10
5-10	7.5	16
10+	3	0

Il punto medio per valori verso infinito si fà

$\overline{x_{100}} = \frac{1}{n} * \sum T_i * x_i = \frac{1}{100} * (0.5 * 24 + 2.5 * 16 + 2.5 * 20 + \cdots) = \frac{1}{100} * 300 =$

Ora quindi possiamo calcolare

$\lambda = \frac{1}{\bar{x}} = \frac{1}{3} = 0.33$

(Il valore di sopra)

Ora dobbiamo calcolare le frequenze attese

$P(Y < 1) = P(y = 1) = 1 - e^{-\lambda y} = 1 - e^{-0.33*1} = 0.2811 = \pi_1$

$P(1 \leq y < 2) = e^{-0.33} - e^{-0.33*2} = 0.2021 = \pi_2$

$P(3 \leq y < 4) = \cdots = 0.1044 = \pi_3$

$P(4 \leq y < 5) = \cdots = 0.0751 = \pi_4$

$P(5 \leq Y < 10) = \cdots = 0.1552 = \pi_5$

$P(Y \geq 10) = 1 - P(Y \leq 10) = 0.0369 = \pi_6$

Quindi, rifiuto h_0 se

$$q = \sum_{i=1}^{7} \frac{(n_i - f_i)^2}{f_i} > X^2_{n-1-1,\alpha}$$

- Si chiede se il seguente segue la legge di Poisson.

Incidenti settimana	0	1	2	3+
Verificato	50	32	3	0

$H_0\colon\ Y \sim P_o(\lambda), \qquad E(Y) = \lambda$

$\lambda = \overline{x_{85}} = \frac{0 * 50 + 1 * 32 + 2 * 3}{85} = 0.447 = E(Y)$

Quindi con questo possiamo calcolare

$P(Y = 0) = e^{-0.447} = 0.64 = \pi_0 \rightarrow f_0 = 85 * 0.64 = 54.4$

$P(Y = 1) = e^{-0.447} * 0.447 = 0.29 = \pi_1 \rightarrow f_1 = 85 * 0.29 = 24.65$

$P(Y = 2) = e^{-0.447} * \left(\frac{0.447}{2}\right)^2 = 0.06 = \pi_2 \rightarrow f_2 = 85 * 0.06 = 5.1$

$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - (\pi_0 + \pi_1 + \pi_2) = 0.01 = \pi_3 \rightarrow f_3 = 0.85$

Per via della regola empirica, devo unire π_3 con π_2

Siccome $f_3 < 1$

Incidenti settimana	0	1	2+

Incidenti settimana	0	1	2+
Frequenze assolute	50	32	3
Frequenze attese	56.4	26.65	5.95

Quindi dobbiamo calcolare Q ora

$$Q = \sum^{2} \frac{(n_i - f_i)^2}{f_i} \sim X^2_{N-1-?} = X^2(1)$$

$q = \cdots$

Scelgo $\alpha = 5\% \rightarrow \alpha = 0.05$

$q > X_{1,0.05}{}^2$

Minchia ragazzi non c'ho capito na beata minchia

Esempio:

1) $N(m, 2^2), m = ?$

   Abbiamo n=10

   E $\overline{x_{10}} = 18.58$

   a. Si verifica che $H_0: m = 20$

      Con $\alpha = 5\%$

      Applichiamo il test Z

      $$Z = \sqrt{n} * \frac{\overline{x_n} - m_0}{o} \sim N(0,1)$$

      E qui rifiuto se

      $|z| > z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

      $$|z| = \left|\frac{18.58 - 20}{2}\sqrt{10}\right| = 2.24 > 1.96$$

      E rifiutiamo siccome $\alpha = 5\%$

      Se avessimo $\alpha = 10\%$

      Sappiamo per certezza che rifiutiamo siccome, aumentando α noi avremo un più alto per rifiutare h_0

      Per $\alpha = 1\%$ dobbiamo fare i calcoli

   b. Ora invece abbiamo m=21

      Siccome $21 \neq 20$

      $$P_{m=21} = \left(\left|\sqrt{10} * \frac{\overline{x_n} - 20}{2}\right| < z_{0.025}\right)$$

      Qui abbiamo il problema che, dovremmo avere

      $$\frac{\overline{x_n} - 21}{\frac{o}{\sqrt{n}}}$$

      Però noi abbiamo -20

      Allora bisogna far comparire il 21 con tanti calcoli

$$P = \left(-1.96 < \sqrt{10} * \frac{\overline{x_n} - 20}{2} < 1.96\right)$$
$$P = \left(-1.96 - \frac{\sqrt{10}}{2} < \sqrt{10} * \frac{\overline{x_n} - 21}{2} < 1.96 - \frac{\sqrt{10}}{2}\right)$$
Quindi ora $\sim N(0,1)$
$P = (-3.54 < z < 0.37) = \phi(0.37) - \phi(-3.54)$
$P = 0.6443$

2) Si scelgono casualmente 64 bambini, e dopo averli misurati
$\bar{x} = 106$
La media italiana $= 106$

Si suppone che il QI $N(m, 256)$
Quindi abbiamo $o^2 = 256$

Quindi
N=64
$\overline{x_n} = 106$
$o = 16$

- Con $\alpha = 0.025$ sono i suoi studenti più intelligenti della media?
  Quindi vuole concludere che $m > 100$
  Nella nostra relazione $N(m, 256)$
  Iniziamo a calcolare
  $$z = \frac{\overline{x_n} - m_0}{\frac{o}{\sqrt{n}}} = 3$$
  Si rifiuta se
  $$\frac{\overline{x_n} - m_0}{\frac{o}{\sqrt{n}}} > z_\alpha \to 3 > z_{0.025} \to z = 1.96 \to vero$$

- Ripetere ora supponendo o=900
  E calcolare p value
  $o^2 = 900 \to o = 30$
  $$z = \frac{106 - 100}{\frac{30}{8}} \simeq 1.6$$
  $1.6 < 1.96 \to$ non si può dire che è sopra la media
  Se $\alpha = 1\%$ allora avreia ccettato h0

  P value = $\bar{\alpha} = z_{\bar{\alpha}}$
  $z_{\bar{\alpha}} = 1.6$

E quindi
$1-\bar{\alpha}=\phi(z_{\bar{\alpha}})$
$\bar{\alpha}=1-\phi(1.6)=0.0548$

- Ipotesi di normalità è essenziale?
  No siccome il campione è numeroso:
  $n=64>30$
- Se noi sappiamo che $\alpha=1\%$
  Rifiuto h_0
  Quindi sappiamo che
  $\bar{\alpha}<0.01$
  Possiamo provare a verificare per $\alpha=0.5\%$
  E rifacendo i calcoli notiamo che accettiamo h_0
  Quindi $0.01<\bar{\alpha}<0.005$

3) Un lotto viene ritenuto inaccettabile se 8% dei pezzi su 500
Su un campione di 100 pezzi se ne trovano 9 difettosi.

$h_0: p \leq 0.08 \quad h_1: p>0.08$
Controlliamo se possiamo trasformare in N(0, 1)
$n=100>30, \qquad np_0=8>5, \qquad n(1-p_0)=92>5$
Quindi posso fare test z approssimato
$0.09=\overline{x_n}$
Usiamo la formula
$$\frac{\overline{x_n}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}=z$$
Usiamo $\alpha=5\%$
$z_\alpha=1.645$
$z=\cdots=0.3686$
$z<z\_\alpha$ quindi i dati non consentono di rifiutare h_0

4) Un gruppo di 6 pazienti per cura dimagrante
Questi risultati:
$77, 87, 104, 98, 91, 78= x_i$
$75, 88, 97, 99, 83, 70= y_i$
COn $\alpha=5\%$

a. E' la cura stata efficace?
   Iniziamo a costruire campione differenze
   $d_i=x_i-y_i$
   $-2, 1, -7, 1, -8, -8$
   SI può supporre che il peso è normalmente distribuito.
   La cura è stata efficace se $x-y<0$
   $h_0: m_\phi \geq 0 \quad h_1: m_\phi<0$
   $\overline{d_n}=3.833$

$s_n^2 = 18.967$

$$T = \frac{\overline{d_n}}{\frac{S_n}{\sqrt{n}}} = \frac{-3.833 - H_0}{\sqrt{18.967}} * \sqrt{6} = -2.156$$

$t_{5,0.05} = 2.015$

$t < -t_{5,\alpha} \rightarrow -2.156 < -2.015$

Quindi la cura è stata efficace

b. Se si suppone un decremento di almeno 5kg?

$H_0 \geq -5, \qquad H_1 < -5$

Rifiutiamo se

$$T = \frac{\bar{d} + 5}{\frac{S_6}{\sqrt{6}}} \sim t(5)$$

$t = \cdots = 0.656$

$0.656 < -t_{5,0.05} \rightarrow 0.656 < -2.015 \rightarrow falso$

5) I seguenti dati:

$X = 42, 13, 37, 54, 23, 32, 45$

$Y = 67, 70, 53, 42, 37, 74, 53, 57, 64$

Non hanno un legame ed hanno numerosità diverse

Quindi dobbiamo usare quella tabelal

Si ipetizza che $m_x = m_y \rightarrow H_0$

Calcoliamo:

Possiamo applicare la tabella se

$$\frac{1}{2} < \frac{S_x^2}{S_y^2} < 2$$

$$S_x^2 = \frac{1}{n_x - 1}(\ldots) \simeq 191.81$$

$S_y^2 = \cdots \simeq 157.78$

$$\frac{1}{2} \leq \frac{191.81}{157.79} \leq 2 \rightarrow vero$$

Quindo possiamo applicare la tabella

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_X + n_y - 2}$$

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(n_x + n_y - 2)$$

E si rifiuta se

$|t| = |T| > t_{n_x + n_y - 2, \frac{\alpha}{2}}$