Lezione 29 10/12/24

TEST DI IPOTESI

Dopo aver raccolto dei dati sperimentali vogliamo verificare se sono o meno coerenti con una data ipotesi.

ESEMPIO: Lancio di una moneta : DOMANDA: È truccata?
ESTRAGGO UN CAMPIONE DI $n = 30$ lanci.

OTTENGO I SEGUENTI RISULTATI

T	C
25	5

IN GENERALE SEGUO QUESTA PROCEDURA

1) Scegliere l'ipotesi da verificare
H_0 IPOTESI NULLA → NON È TRUCCATA
H_1 IPOTESI ALTERNATIVA → È TRUCCATA

H_0 è l'ipotesi di partenza, quello che stabilisco a priori.
ESEMPIO: SE NON HO LANCIATO LA MONETA DEVO ASSUMERE CHE NON SIA TRUCCATA E POI VERIFICARE CON L'ESPERIMENTO SE È PLAUSIBILE OPPURE NO.

Calcoleremo quanto è probabile aver ottenuto i risultati che abbiamo sotto H_0 (SE FOSSE VERA L'IPOTESI NULLA).

2) Scegliere il test da effettuare a seconda dei dati, ce ne sono molti, noi ne faremo 3 →
- TEST Z
- TEST T
- TEST χ^2
  - DI ADATTAMENTO
  - DI INDIPENDENZA

(CONDIZIONI)

3) Controllare che le ipotesi del test sono verificate

4) Scegliere un livello minimo di significatività accettabile

$\alpha = 0.1$
$\alpha = 0.05$
$\alpha = 0.01$
$\alpha = 0.001$

5) Calcolare il valore del test p (p-value)

SE HAI UN CALCOLATORE

QUANTO È PROBABILE CHE SI SIA VERIFICATA LA SITUAZIONE CHE ABBIAMO OSSERVATO NEL CAMPIONE SE H_0 è VERA

6) SE $p < \alpha \rightarrow$ RIFIUTO L'IPOTESI NULLA $\rightarrow$ HO DIMOSTRATO QUALCOSA

$p > \alpha \rightarrow$ ACCETTO L'IPOTESI NULLA $\rightarrow$ NON SO SE SIA VERA O MENO H_1 MA I DATI CHE HO NON SONO SUFFICIENTI A DIMOSTRARLA.

IL TEST Z

$Z \sim N(0,1)$

- LE VARIABILI SONO NUMERICHE
- LA POPOLAZIONE SEGUE UNA LEGGE NORMALE (ALMENO APPROSSIMATIVAMENTE)

NELL' ESPERIMENTO DELLA MONETA

$$X = \begin{cases} 1 & \text{se esce teste} \\ 0 & \text{se esce croce} \end{cases}$$ ← VARIABILE BINARIA

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

MONETA NON TRUCCATA → $H_0: \rightarrow \mu = 0.5$

MONETA TRUCCATA → $H_1: \rightarrow \mu \neq 0.5$

← TEST BILATERALE

LA DISTRIBUZIONE DI X PUÒ ESSERE APPROSSIMATIVAMENTE NORMALE

se $\begin{cases} n\mu > 5 \\ n(1-\mu) > 5 \end{cases}$

$\Big[\rightarrow$ NELL'ESEMPIO FATTO PRIMA $n=30$, $\mu=\frac{1}{2}$ $\begin{cases} 30 \cdot \frac{1}{2} = 15 > 5 \\ 30\left(1-\frac{1}{2}\right) = 15 > 5 \end{cases}$ ok $\Big]$

SE $\mu = \frac{1}{500}$

$$\begin{cases} n \frac{1}{500} > 5 & n > 2500 \\ n \cdot \frac{499}{500} > 5 & n > 5 \text{ ok} \end{cases}$$

- CONOSCIAMO LA VARIANZA DELLA POPOLAZIONE $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

NEL CASO DI VARIABILE BINARIA

$$\sigma^2 = \mu(1-\mu)$$

NELL'ESEMPIO

$$\boxed{\sigma = \sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)} = \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = \sqrt{\frac{1}{4}} = \frac{1}{2}}$$

SI CALCOLA LA QUANTITÀ PIVOTALE DEL TEST

$$|Z^*| = \frac{|\bar{x} - \mu|}{\sigma} \sqrt{n} \rightarrow \left(p\text{-value}\right)$$

(SE AVESSIMO IL COMPUTER)

SI CONFRONTA $|Z^*|$ CON I DATI DELLA TABELLA

SE $|Z^*| <$ VALORE DELLA TABELLA IN 0.1 $\rightarrow$ H_0 VERA OPPURE DATI NON SUFFICIENTI

$|Z^*| >$ VALORE IN α $\rightarrow$ H_1 è VERA CON $1-\alpha$ LIVELLO DI SIGNIFICATIVITÀ

NELL' ESEMPIO

$$\bar{x} = \frac{1}{30} \sum_{i=1}^{30} x_i = \frac{1}{30}(\underbrace{1+1 \cdots 1}_{25 \text{ VOLTE}} + \underbrace{0+0+0+0+0}_{5 \text{ VOLTE}})$$

$$\bar{x} = \frac{25}{30} = \frac{5}{6}$$

$$|z^*| = \frac{|\bar{x}-\mu|}{\sigma}\sqrt{n} = \frac{\left|\frac{5}{6}-\frac{1}{2}\right|}{1/2}\sqrt{30} = 2\left|\frac{5-3}{6}\right|\sqrt{30}$$

$$= \frac{\cancel{4}^{2}}{\cancel{6}_{3}}\sqrt{30} \simeq 3.651 > 3.291$$

RIFIUTO L'IPOTESI NULLA

QUINDI CONCLUDO CHE LA MONETA È TRUCCATA AL 99.9 %

SE AVESSIMO OTTENUTO

T	C
14	16

$n = 30$

$$|z^*| = \frac{\left|\frac{14}{30}-\frac{1}{2}\right|}{1/2}\sqrt{30} = 2\frac{\left|-\frac{1}{30}\right|}{}\sqrt{30} = \frac{1}{15}\sqrt{30} \simeq 0.365$$

$0.365 < 1.645 \Rightarrow$ ACCETTO H_0

ESERCIZIO 4. DEL 20/06

$\mu = 4.0$ $\quad n=12$

$$\bar{x} = \frac{1}{12}(4+3.7+4.1+4.1+\cdots +3.6) = \frac{46.8}{12} = 3.9$$

$$|z^*| = \frac{|\mu - \bar{x}|}{\sigma}\sqrt{n} = \frac{|4-3.9|}{0.22}\sqrt{12} = 1.57 < 1.645$$

ACCETTO H_0

LA MACCHINA FUNZIONA COME DOVREBBE

IL TEST Z VA BENE SOLO SE CONOSCIAMO σ

Se non conosciamo σ LO DOBBIAMO STIMARE

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

DEVIAZIONE STANDARD CAMPIONARIA

ESEGUO UN TEST T (STESSE IPOTESI DEL TEST Z)

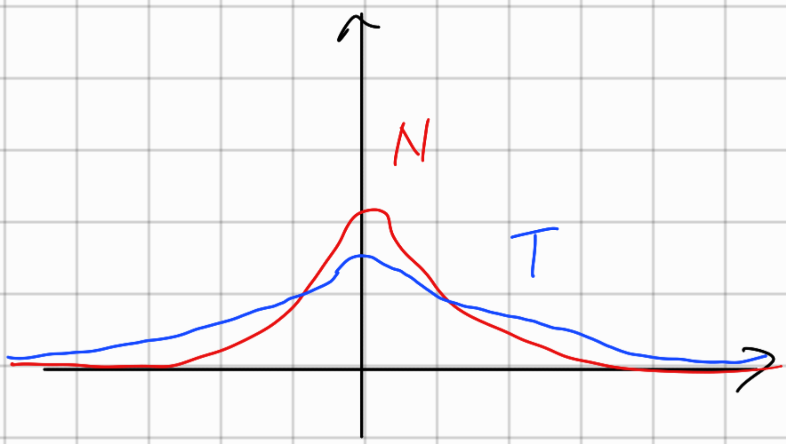
$$|T^*_{n-1}| = \frac{|\bar{x} - \mu|}{s}\sqrt{n}$$

QUANTITÀ PIVOTALE DA CALCOLARE

ESEMPIO

$\mu = 12.5$ $\qquad n = 6$

$X = (11.5,\ 11,\ 12.5,\ 13.1,\ 12.7,\ 12.4)$



NON CONOSCO LA VARIANZA $\Rightarrow$ FACCIO IL TEST T

$$\bar{x} = \frac{1}{6}(11.5 + 11 + 12.5 + 13.1 + 12.7 + 12.4) = \boxed{12.2}$$

$$(13.1-12.5)^2$$

$$S=\sqrt{\frac{1}{5}\left((11.5-12.2)^2+(11-12.2)^2+(12.5-12.2)^2+(12.7-12.2)^2+(12.4-12.2)^2\right)}$$

$$=\sqrt{\frac{1}{5}\left(0.49+1.44+0.09+0.36+0.25+0.04\right)}=0.79$$

$$|T_5^*|=\frac{|12.5-12.2|}{0.79}\sqrt{6}\simeq 0.93<2.015 \Rightarrow \text{ACCETTO } H_0$$

SUPPONIAMO DI AVERE UN ESPERIMENTO CON DUE CAMPIONI

ES. H_0 UNA MEDICINA CHE DOVREBBE ABBASSARE LA PRESSIONE E DIVIDO IL MIO CAMPIONE IN DUE: A UNA PARTE DO LA MEDICINA E ALL'ALTRA PARTE NO

OTTENGO I SEGUENTI RISULTATI $n=5$

PLACEBO $\{156, 171, 133, 102, 129\}$ X $\rightarrow \mu_x$ MEDIA DEL VALORE PRESSIONE DI CHI NON PRENDE IL FARMACO

FARMACO $\{73, 81, 103, 84, 130\}$ Y $\rightarrow \mu_y$ ME... DI CHI PRENDE IL FARMACO

$$\begin{cases} H_0 & \mu_x=\mu_y \\ H_1 & \mu_x\neq\mu_y \end{cases}$$

[IN QUESTO CASO SAREBBE PIÙ OPPORTUNO IL TEST UNILATERALE CHE NON FAREMO PER MANCANZA DI TEMPO $\begin{cases} H_0 = \mu_x \le \mu_y \\ H_1 = \mu_x > \mu_y \end{cases}$]

$$|T^*_{\underbrace{2(n-1)}_{2\cdot 4=8}}|=\frac{|\bar{x}-\bar{y}|}{\sqrt{S_x^2+S_y^2}}\sqrt{n}$$

PER CASA

$$\bar{x} = \frac{1}{5}\left(156+171+133+102+129\right) = \frac{691}{5} = 138.2$$

$$\bar{y} = \frac{1}{5}\left(73+81+103+84+130\right) = \frac{471}{5} = 94.2$$

$$S_x^2 = \frac{1}{4}\left((156-138.2)^2+(171-138.2)^2+\dots+(129-138.2)^2\right) =$$

$$= \frac{1}{4}\left((17.8)^2+(32.8)^2+(-5.2)^2+(-36.2)^2+(-9.2)^2\right) =$$

$$= \frac{1}{4}\left(316.84+1075.84+27.04+1310.04+84.64\right) = 703.7$$

$$S_y^2 = \frac{1}{4}\left((73-94.2)^2+(81-94.2)^2+(103-94.2)^2+(84-94.2)^2+(130-94.2)^2\right) =$$

$$= \dots = 521.7$$

$$|T^*_{\underset{8}{\overset{\shortparallel}{2(n-1)}}}| = \frac{|\bar{x}-\bar{y}|}{\sqrt{S_x^2+S_y^2}}\sqrt{n} = \frac{|138.2-94.2|}{\sqrt{703.7-521.7}}\sqrt{5} \simeq \frac{44}{35}\sqrt{5} \simeq 6.28$$

α	0.1	0.05	0.01	0.001
8	1.860	2.306	3.355	5.041

$$|T_8^*| > 5.041$$

Possiamo affermare al 99.9 % il farmaco funziona.