TEST DI IPOTESI

Abbiamo dei dati sperimentali e vogliamo controllare se sono o meno coerenti con un'assunzione (IPOTESI)

ESEMPIO: LANCIO DI UNA MONETA Domanda: è truccata

ESTRAGGO UN CAMPIONE DI MISURE $\{T, C, T, \dots C\}$ $n = 30$

$X = \{1, 0, 1, \dots 0\}$ $T = 25$

$C = 5$

$$\bar{X} = \frac{\#\text{TESTE}}{n} = \frac{25}{30} = \frac{5}{6}$$

μ LA MEDIA IN GENERALE

SOTTO H_0 $\mu = \frac{1}{2}$

1) Scegliere l'ipotesi da verificare
 - IPOTESI NULLA Quella che è più naturale in assenza di dati (e che vorremmo smentire) H_0

 - IPOTESI ALTERNATIVA H_1 NELL'ESEMPIO SI PUÒ TESTARE $H_1: \mu \neq \frac{1}{2}$

  $H_2 \neq \mu \underset{(>)}{>} \frac{1}{2}$

2) Scegliere il test sulla base dei dati e delle ipotesi che vogliamo verificare

3) Verificare che i REQUISITI DEL TEST siano verificati.
 - Solitamente sono ASSUNZIONI SULLA DISTRIBUZIONE DELLA POPOLAZIONE
 - O richieste che $n \geq$ valore minimo.

4) Scegliere il livello di significatività $\alpha \in \{0.1, 0.05, 0.01, 0.001\}$

Quanto sono disposto a sbagliare?
Con quale certezza voglio affermare il mio risultato?

5) Si calcola una quantità pivotale (FORMULA DIVERSA PER OGNI TEST)
Q^*

PER IL TEST UNILATERALE TOGLIERE IL VALORE ASSOLUTO

$$p = P(|Q| < Q^*)$$ P VALUE

6) Confrontiamo p con α

(RIFIUTIAMO H_0) con $1-\alpha$ LIV. DI SIGNIFICA...

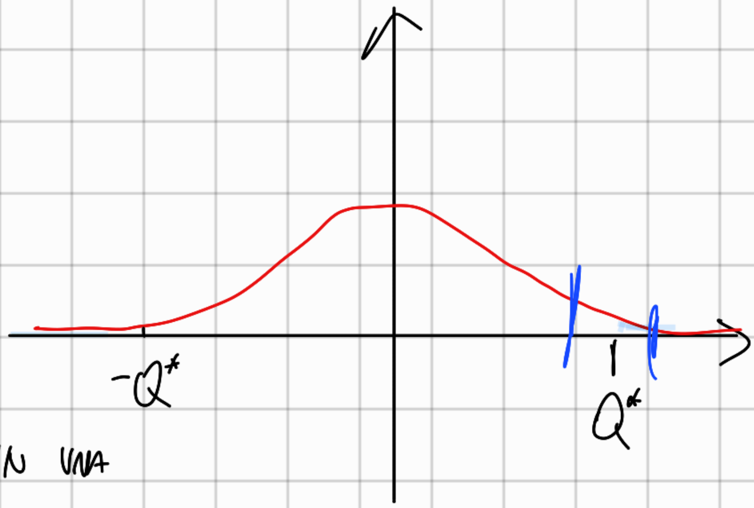
$p < \alpha \Rightarrow$ ACCETTIAMO H_1

$p > \alpha \Rightarrow$ ACCETTIAMO H_0

Se non riusciamo a calcolare p

$$p < \alpha \Leftrightarrow Q^* > K^*$$

VALORI IN UNA TABELLA



TEST Z

- Si applica quando ho valori NUMERICI e voglio sapere se LA MEDIA DELLA POPOLAZIONE È $>$ $<$ $\neq$ DA UN CERTO VALORE DATO.

- DOBBIAMO SUPPORRE DI CONOSCERE σ

NEL CASO DI TEST BINOMIALI $\{0, 1, 1, 0, 1, \dots 1\}$

ESEMPI $\{T, C\}$

$\{$GUARITO, NO$\}$

ESITO $\{$POSITIVO, NEGATIVO$\}$

NEL CASO DELLA MONETA $\mu = P(X = T) = \frac{1}{2}$

$$\sigma = \sqrt{\mu(1-\mu)} = \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2}$$

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

NELL' ESEMPIO

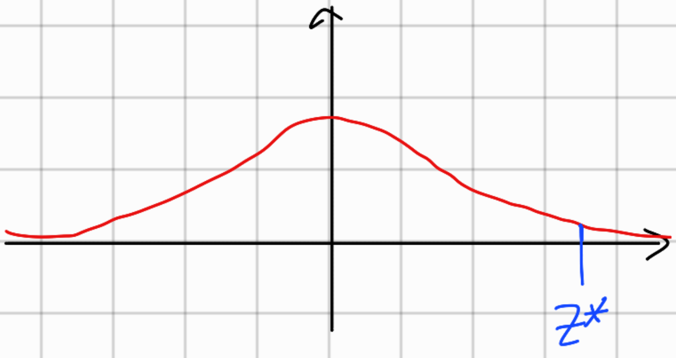
$$n\mu = 30 \cdot \frac{1}{2} = 15 > 5 \quad \text{SI} \checkmark$$
$$n(1-\mu) = 3 \cdot \left(1 - \frac{1}{2}\right) = 15 > 5 \quad \text{SI} \checkmark$$

LA QUANTITÀ PIVOTALE DA CALCOLARE È

$$Z^* = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

NELL' ESEMPIO

$$Z^* = \frac{\frac{5}{6} - \frac{1}{2}}{\frac{1}{2}} \sqrt{30} = 3.65$$



α	0.10	0.05	0.01	0.001
k^*	1.645	1.960	2.576	3.291

$|z^*| > 3.291 \Rightarrow$ COL 99.9 % DI CONFIDENZA
LA MONETA SARÀ TRUCCATA

MOLTO PROBABILMENTE LA MONETA È TRUCCATA

RIFIUTIAMO $H_0 \Rightarrow \mu \neq \frac{1}{2}$ CIOÈ

PER ESERCIZIO PROVARE ALTRI VALORI 18 TESTE
20 TESTE

SE NON CONOSCIAMO σ $\Rightarrow$ DOBBIAMO USARE LA STIMA S

SI USA IL TEST T (basato sulla distribuzione t di Student)

Solitamente per <u>V.A. CONTINUE</u>

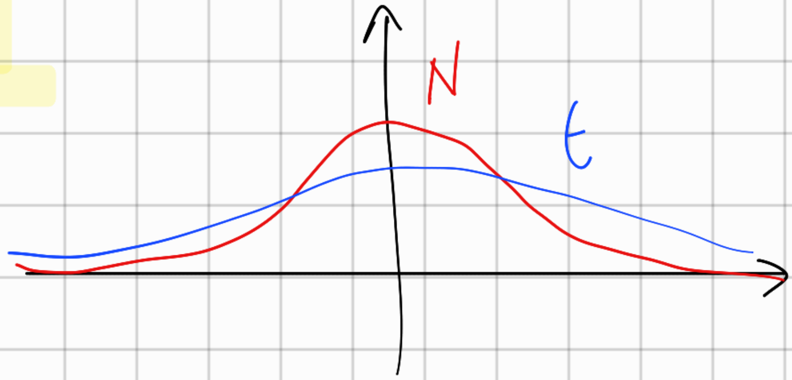
REQUISITI

- La popolazione segue una distribuzione normale
- OPPURE $n \geq 120$

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

QUANTITÀ PIVOTALE

$$T^*_{n-1} = \frac{\bar{X}-\mu}{S}\sqrt{n}$$



gradi di libertà

$\nu = n-1$

ESEMPIO

SUPPONIAMO DI AVERE UN PARAMETRO CHE DOVREBBE AVERE $\mu = 12.5$

LA POPOLAZIONE HA DISTRIBUZIONE NORMALE

ESTRAIAMO UN CAMPIONE $\boxed{n=6}$

11.5 , 11 , 12.5 , 13.1 , 12.7 , 12.4

È COERENTE CON LE ASPETTATIVE?

$$\bar{X} = \frac{1}{6}(11.5+11+12.5+\dots 12.4) = 12.2$$

$$S = \sqrt{\frac{1}{5}\left[(12.5-11.5)^2 + (12.5-11)^2 + \dots + (12.5-12.4)^2\right]} = 0.79$$

$$T^*_5 = \frac{12.2-12.5}{0.79}\sqrt{6} \simeq -0.93$$

$n \backslash \alpha$	0.1	0.05	0.01	0.001
5	2.015	2.571	4.032	6.869

$$|-0.93| = 0.93 < 2.015 \Rightarrow \text{ACCETTIAMO } H_0$$

NOTA: ESISTE UN TEST T PER IL CONFRONTO TRA LE MEDIE DI DUE CAMPIONI

HO DUE CAMPIONI:

		MEDIA	DEV. ST. C.
$X = \{x_1 \dots x_m\}$	$\rightarrow$	$\bar{x}$	S_x
$Y = \{y_1 \dots y_n\}$	$\rightarrow$	$\bar{y}$	S_y

La domanda è: è verosimile che siano stati estratti da una popolazione con la stessa media?

SI APPLICA QUANDO SI FANNO STUDI IN CUI SI MISURA UNA VARIABILE CONTINUA e si vuole verificare se questo parametro cambia tra un gruppo e l'altro

PER ESEMPIO: X HANNO RICEVUTO UN TRATTAMENTO e Y NO

oppure X MASCHI e Y FEMMINE

SE $m \neq n$ SIA $S_{xy} = \sqrt{\frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}}$

GRADI DI LIBERTÀ: $\nu = m+n-2$

LA QUANTITÀ PIVOTALE È

$$T^*_{n+m-2} = \frac{\bar{x}-\bar{y}}{S_{xy}\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

SE $n = m$ LA FORMULA SI SEMPLIFICA

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

ESEMPIO

ABBIAMO 10 INDIVIDUI a 5 DIAMO IL FARMACO a 5 NO
MISURIAMO LA PRESSIONE

PLACEBO $X = \{156, 171, 133, 102, 129\}$
FARMACO $Y = \{73, 81, 103, 84, 130\}$

Vogliamo sapere se il farmaco influenza o meno il valore della pressione

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

$$\bar{y} = \frac{1}{5}(73, 81, 103, 84, 130) = 94.2$$

$$\nu = 2(5-1) = 8$$

$n \backslash \alpha$	0.10	0.05	0.01	0.001
8	1.860	2.306	3.355	5.041
	✓	✓	✗	✗

$$S_x^2 = 703.7$$
$$S_y^2 = 521.7$$

$$T^{*}_{8} = \frac{138.2-94.2}{\sqrt{703.7+521.7}}\sqrt{5} = 2.81$$

Possiamo affermare che il farmaco riduce la pressione con un livello di significatività del 95 %
Se vogliamo livelli maggiori dobbiamo avere più dati.

ATTENZIONE!

Se x_i e y_i SONO LEGATI TRA LORO (AD ESEMPIO SONO I VALORI DELLA STESSA PERSONA PRIMA E DOPO UN TRATTAMENTO)
Si dice che i campioni sono accoppiati.

IL TEST CORRETTO DA FARE È IL SEGUENTE:
Calcolo i valori differenze $w_i = x_i - y_i$
APPLICO IL T. TEST CLASSICO PER VEDERE SE la media μ dei w_i è uguale a 0 oppure no

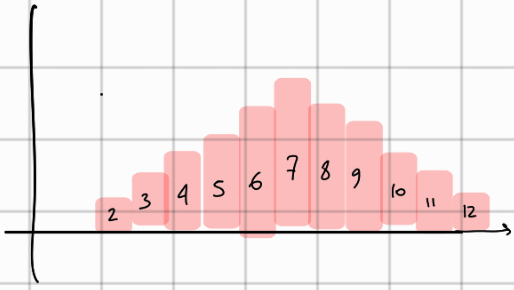
TEST χ^2 → DI ADATTAMENTO
→ DI INDIPENDENZA

Si usa per distribuzioni discrete o qualitative
(Si può applicare a distribuzioni continue se scelgo degli intervalli $[x_0,x_1)$, $[x_1,x_2)$ e così via ... $[x_{k-1},x_k]$

ESEMPIO

LANCIO DI DUE DADI n H_0 $k=11$ possibili classi

REQUISITI: OGNI CLASSE DEVE AVERE ALMENO 5 OSSERVAZIONI



ESEMPIO: Lancio di due dadi
Frequenze attese VS Frequenze ottenute

numero uscito	2	3	4	5	6	7	8	9	10	11	12
frequenza (F_i) ASSOLUTA	9	15	22	28	39	43	34	25	17	13	5

$F_i \geq 5$ ✓ ok

PROBABILITÀ ATTESE (p_i)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
FREQUENZE ATTESE (E_i)	6.94	13.89	20.83	27.78	34.72	41.67	...				

MOLTIPLICO PER $n=250$

$E_i = n p_i$

Significa che mi aspetto circa E_i volte il risultato corrispondente

QUANTITÀ PIVOTALE

$$\chi^{2*}_{k-1} = \sum_{i=1}^{n} \frac{(F_i - E_i)^2}{E_i}$$

DEVO CONFRONTARE CON I VALORI DELLA TABELLA DELLA χ^2 A $k-1$ GRADI DI LIBERTÀ

NELL' ESEMPIO

$$\chi^{2*}_{10} = 2.93$$

$n \backslash \alpha$	0.10	0.05	0.01	0.001
10	15.98	18.31	23.21	29.59

POICHÉ $2.93 < 15.98$
ACCETTO **L'IPOTESI NULLA** CHE I RISULTATI SEGUANO LA DISTRIBUZIONE ASSEGNATA

ESEMPIO: [Dati di una mutazione]

SAPPIAMO CHE IL GENOTIPO di una popolazione è così distribuito SE I GENITORI SONO Aa e Aa

$n = 800$

$$p_1 = P(AA) = 1/4 \qquad E_1 = \frac{800}{4} = 200$$
$$p_2 = P(Aa) = 1/2 \qquad E_2 = \frac{800}{2} = 400$$
$$p_3 = P(aa) = 1/4 \qquad E_3 = \frac{400}{2} = 200$$

Dopo aver inserito un gene mutato SI RISCONTRANO LE SEGUENTI FREQUENZE

(AA) $F_1 = 169$
(Aa) $F_2 = 410$
(aa) $F_3 = 221$

Verificare se è possibile affermare che la mutazione favorisca aa.

$$\chi^{2*} \quad \frac{(200-169)^2}{200} + \frac{(400-410)^2}{400} + \frac{(200-221)^2}{200} = 7.26$$

$$K-1 = 3-1 = 2$$

$n \backslash \alpha$	0.10	0.05	0.01	0.001
2	4.61	5.99	9.21	13.82
	✓	✓	✗	✗

POSSIAMO AFFERMARE CHE IL GENE MUTATO MODIFICA LA DISTRIBUZIONE (e quindi favorisce aa)

CON LIV. DI SIGNIFICATIVITÀ 0,05

CASO PARTICOLARE: $\boxed{\chi^2 \text{ DI INDIPENDENZA}}$

Supponiamo di avere un campione che ha due variabili qualitative una con k possibili categorie e l'altra con h $\begin{matrix} V_1, V_2, \dots V_k \\ U_1 U_2 \dots U_h \end{matrix}$

Vogliamo scoprire se c'è un legame tra le due oppure no

SCRIVIAMO LA TABELLA DELLE FREQUENZE OSSERVATE

	V_1	V_2	$\dots$	V_k	
U_1	n_{11}	n_{12}			$n_{1,\cdot}$
U_2	$\vdots$	$\vdots$			$n_{2,\cdot}$
$\vdots$	$\vdots$	$\vdots$			
U_h				n_{hk}	$n_{h\cdot}$
	$n_{\cdot 1}$			$n_{\cdot k}$	n

n_{ij} È IL NUMERO DI UNITÀ SIA NELLA CATEGORIA U_i che V_j

FREQUENZE MARGINALI

$n_{i\cdot}$ È IL NUMERO DI UNITÀ NELLA CATEGORIA U_i $\quad i=1\dots h$

$n_{\cdot j}$ È IL NUMERO DI UNITÀ NELLA CATEGORIA V_j $\quad j=1\dots k$

$$n_{i\cdot} = \sum_{j=1}^{k} n_{ij}$$

$$n_{\cdot\cdot} = n = \sum_{i=1}^{h} n_{i\cdot} = \sum_{j=1}^{k} n_{j\cdot}$$

$$n_{\cdot j} = \sum_{i=1}^{h} n_{ij}$$

$$\chi^2_* = \sum_{i=1}^{k}\sum_{j=1}^{h} \frac{(F_{ij} - E_{ij})^2}{E_{ij}}, \qquad \nu = k+h-2$$

Si calcolano le frequenze attese

$$E_{i,j} = \frac{m_{\cdot i} \cdot m_{\cdot j}}{m} \quad , \quad F_{ij} = m_{ij}$$

SI UTILIZZA IL TEST χ^2 a $k+h-2$ GRADI DI LIBERTÀ

ESEMPIO: TEST DI UN FARMAC $m=330$

V_1 NESSUN FARMAC
V_2 DOSE SINGOLA
V_3 DOSE DOPPIA

U_1 = paziente migliorato
U_2 = paziente non migliorato

$F_{i,j} \geq 5 \ \forall i,j$ ✓

	V_1	V_2	V_3	
migliorati	12	5	29	46 $= m_{1\cdot}$
non migliorati	114	80	90	284 $= m_{2\cdot}$
	126 $= m_{\cdot 1}$	85 $= m_{\cdot 2}$	119 $= m_{\cdot 3}$	330 $= m_{\cdot\cdot}$

Il farmaco fa effetto?

H_0: NO, I GUARITI SONO CASUALI
H_1: SI, I GUARITI SONO DI PIÙ TRA QUELLI SOTTOPOSTI AL FARMACO

Calcolo le frequenze attese

$$E_{1,1} = \frac{m_{1\cdot} \cdot m_{\cdot 1}}{m_{\cdot\cdot}} = \frac{46 \cdot 126}{330} = 17.56$$

E COSÌ VIA

	V_1	V_2	V_3	
migliorati	17.56	11.8	16.6	$46 = n_{1\cdot}$
non migliorati	108.4	73.2	102.4	$284 = n_{2\cdot}$
	$126 = n_{\cdot 1}$	$85 = n_{\cdot 2}$	$119 = n_{\cdot 3}$	$330 = n_{\cdot\cdot}$

$$\nu = (k-1)(h-1) = (3-1)(2-1) = 2$$

$$\chi^2_* = \frac{(12-17.56)^2}{17.56} + \cdots + \frac{(90-102.4)^2}{102.4} = 17.4$$

$n \backslash \alpha$	0.10	0.05	0.01	0.001
2	4.605	5.991	9.210	13.816
	✓	✓	✓	✓

IL FARMACO FUNZIONA AL 99.9 % DI CONFIDENZA

ALTRO ESEMPIO

Si somministra un farmaco a una parte di pazienti e poi si contano i guariti. La tabella delle frequenze è la seguente

FARMACO

$F_{i,j} > 5 \quad \forall i,j$ ok

GUARIGIONE

F_i	SI	NO	
SI	52	28	80
NO	8	12	20
	60	40	100

$F_{i,j}$

FARMACO

E_i	SI	NO	
SI	$\frac{80 \cdot 60}{100} = 48$	$\frac{80 \cdot 40}{100} = 32$	80
NO	$\frac{60 \cdot 20}{100} = 12$	$\frac{40 \cdot 20}{100} = 8$	20
	60	40	100

$E_{i,j}$

$$\nu = (2-1)(2-1) = 1$$

CALCOLIAMO LA QUANTITÀ PIVOTALE

$$\chi^{2*}_{1} = \frac{(52-48)^2}{48} + \frac{(58-32)^2}{32} + \frac{(8-12)^2}{12} + \frac{(12-8)^2}{8} = 4.17$$

$n \backslash \alpha$	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.828
	✓	✓	✗	✗

IL FARMACO FUNZIONA COL 95 % DI CONFIDENZA