INTERVALLI DI CONFIDENZA

Campione $X_1, X_2, \dots, X_n$	Intervallo di confidenza
$b(p)$ $n\hat{p}_n > 5 \quad n(1-\hat{p}_n) > 5$	$\hat{p}_n \pm z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}_n(1-\hat{p}_n)}{n}}$
$N(\mu, \sigma^2)$ σ^2 noto	$\bar{X}_n \pm z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
$N(\mu, \sigma^2)$ σ^2 incognito	$\bar{X}_n \pm t^{(n-1)}_{1-\alpha/2}\dfrac{\bar{S}_n}{\sqrt{n}}$ *
distr. incognita (μ, σ^2) $n > 30$ σ^2 incognito	$\bar{X}_n \pm t^{(n-1)}_{1-\alpha/2}\dfrac{\bar{S}_n}{\sqrt{n}}$ *

* se $n > 30$, si può usare $z_{1-\alpha/2}$ per il calcolo

QUIZ

Nell'intervallo di confidenza per il vero valore della media di una popolazione gaussiana, si usa la t di Student

(a) se σ^2 è nota
(b) se n è grande
(c) se σ^2 è incognita
(d) sempre

QUIZ

Si consideri un campione da una popolazione normale di media μ e varianza non nota σ^2.
L'intervallo di confidenza a livello $100(1-\alpha)\%$ per μ è dato da

(a) $\left(\bar{x} - t^{(n-1)}_{1-\alpha}\frac{\bar{s}}{\sqrt{n}},\ \bar{x} + t^{(n-1)}_{1-\alpha}\frac{\bar{s}}{\sqrt{n}}\right)$

(b) $\left(\bar{x} - t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x} + t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

(c) $\left(\bar{x} - t^{(n-1)}_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{s}^2}{n}},\ \bar{x} + t^{(n-1)}_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{s}^2}{n}}\right)$

(d) $\left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\bar{s}}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\bar{s}}{\sqrt{n}}\right)$

Esercizio 7

Nell'ambito di un'indagine sui consumi delle famiglie italiane è stato osservato un campione di n = 320 unità. È risultato che le famiglie intervistate spendono mediamente 62 euro al mese per l'acquisto di pane con varianza pari a 289 e che 297 di queste possiedono più di un'auto.

a) Si costruisca un intervallo di confidenza al 95% per la spesa media in pane delle famiglie italiane.

b) Si stimi la frequenza relativa delle famiglie che possiedono più di un'auto.

c) Si costruisca un intervallo di confidenza al 95% per il parametro di cui al punto precedente.

Esercizio 7

Nell'ambito di un'indagine sui consumi delle famiglie italiane è stato osservato un campione di n = 320 unità. È risultato che le famiglie intervistate spendono mediamente 62 euro al mese per l'acquisto di pane con varianza pari a 289 e che 297 di queste possiedono più di un'auto.

a) **Si costruisca un intervallo di confidenza al 95% per la spesa media in pane delle famiglie italiane.**

$X_1, X_2, \dots, X_{320} \sim \dots\ (\mu, \sigma^2)$ i.i.d., σ^2 **incognita**, $n = 320 > 30$

→ la media campionaria $\bar{X}_n$ ha distribuzione **t-Student** ≈ **Normale**

$$\mathrm{IC}(95\%) = \bar{x}_n \pm t^{(n-1)}_{1-\alpha/2}\frac{\bar{s}_n}{\sqrt{n}} =$$

← da scrivere all' esame !

$$= \bar{x}_n \pm z_{1-\alpha/2}\frac{\bar{s}_n}{\sqrt{n}} =$$

$$= 62 \pm 1.96\frac{\sqrt{289}}{\sqrt{320}} = (60.14,\quad 63.86)$$

$\alpha = 1 - 0.95 = 0.05 \qquad z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$

Esercizio 7

Nell'ambito di un'indagine sui consumi delle famiglie italiane è stato osservato un campione di n = 320 unità. È risultato che le famiglie intervistate spendono mediamente 62 euro al mese per l'acquisto di pane con varianza pari a 289 e che 297 di queste possiedono più di un'auto.

b) **Si stimi la frequenza relativa delle famiglie che possiedono più di un'auto.**

$$X_1, X_2, \dots, X_{320} \sim b\,(p) \text{ i.i.d.} \quad n = 320$$

p = proporzione di famiglie che possiedono più di un'auto

Proporzione campionaria

$$\hat{p}_n = \frac{297}{320} = 0.93$$

Esercizio 7

Nell'ambito di un'indagine sui consumi delle famiglie italiane è stato osservato un campione di n = 320 unità. È risultato che le famiglie intervistate spendono mediamente 62 euro al mese per l'acquisto di pane con varianza pari a 289 e che 297 di queste possiedono più di un'auto.

c) **Si costruisca un intervallo di confidenza al 95% per il parametro di cui al punto precedente.**

$$X_1, X_2, \dots, X_{320} \sim b\,(p) \text{ i.i.d.,} \quad n = 320$$

→ $\hat{p}_n = 0.93$, la proporzione campionaria ha distribuzione **Normale**

Verifico le condizioni di applicabilità :

$$n\hat{p}_n = 320 \times 0.93 = 297.6 > 5$$

$$n(1-\hat{p}_n) = 320 \times (1-0.93) = 22.4 > 5$$

$$\mathbf{IC(95\%)} = \hat{p}_n \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} =$$

$$= 0.93 \pm 1.96\sqrt{\frac{0.93(1-0.93)}{320}} = (\mathbf{0.90}, \ \ \mathbf{0.96})$$

VERIFICA DELLE IPOTESI per la media/proporzione

Campione $X_1, X_2, \dots, X_n$
$\boldsymbol{b(p)}$ $np_0 > 5 \quad n(1-p_0) > 5$
$\boldsymbol{N(\mu, \sigma^2)}$ σ^2 noto
$\boldsymbol{N(\mu, \sigma^2)}$ σ^2 incognito
distr. incognita (μ, σ^2) $n > 30$ σ^2 incognito

VERIFICA DELLE IPOTESI per la media/proporzione

Ipotesi nulla	Ipotesi alternativa	Statistica test z oppure t	Rifiuto H_0 se
$H_0: p = p_0$	$H_1: p \neq p_0$ $H_1: p > p_0$ $H_1: p < p_0$	$\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$	$\|z\| > z_{1-\alpha/2}$ $z > z_{1-\alpha}$ $z < -z_{1-\alpha}$
$H_0: \mu = \mu_0$ σ^2 nota	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$	$\|z\| > z_{1-\alpha/2}$ $z > z_{1-\alpha}$ $z < -z_{1-\alpha}$
$H_0: \mu = \mu_0$ σ^2 incognita	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\frac{\bar{X}_n - \mu_0}{\bar{s}_n/\sqrt{n}}$	$\|t\| > t_{1-\alpha/2}^{(n-1)}$ $t > t_{1-\alpha}^{(n-1)}$ * $t < -t_{1-\alpha}^{(n-1)}$

* se $n > 30$, si può usare $\boldsymbol{N(0,1)}$ per il calcolo

VERIFICA DELLE IPOTESI per la media/proporzione

Si rifiuta H_0 se p-value $< \alpha$

Ipotesi nulla	Ipotesi alternativa	Statistica test z oppure t	p-value
$H_0: p = p_0$	$H_1: p \neq p_0$ $H_1: p > p_0$ $H_1: p < p_0$	$\dfrac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$	$P(\lvert Z\rvert > \lvert z\rvert)$ $P(Z > z)$ $P(Z < z)$
$H_0: \mu = \mu_0$ σ^2 nota	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\dfrac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$	$P(\lvert Z\rvert > \lvert z\rvert)$ $P(Z > z)$ $P(Z < z)$
$H_0: \mu = \mu_0$ σ^2 incognita	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\dfrac{\overline{X}_n - \mu_0}{\bar{s}_n/\sqrt{n}}$	$P(\lvert t^{(n-1)}\rvert > \lvert t\rvert)$ $P(t^{(n-1)} > t)$ $P(t^{(n-1)} < t)$ *

*se $n > 30$, si può usare $Z \sim N(0,1)$ per il calcolo

QUIZ

Si consideri un campione di ampiezza $n > 30$ da una popolazione di media μ e deviazione standard non nota σ.
Nel test $H_0: \mu = 5$ contro $H_1: \mu < 5$ a livello di significatività α, si rifiuta l'ipotesi nulla se :

(a) $t < t_{\alpha}^{(n-1)}$ (c) $|z| < z_{1-\alpha/2}$

$\boxed{\text{(b)}\ t < -t_{1-\alpha}^{(n-1)}}$ (d) $z < -z_{\alpha}$

QUIZ

Si consideri un campione di ampiezza $n > 30$ da una popolazione di media μ e deviazione standard nota σ.
Nel test $H_0: \mu = 5$ contro $H_1: \mu < 5$ a livello di significatività α, si rifiuta l'ipotesi nulla se :

(a) $t < t_{\alpha}^{(n-1)}$ (c) $|z| < z_{1-\alpha/2}$

(b) $t < t_{1-\alpha}^{(n-1)}$ $\boxed{\text{(d)}\ z < -z_{1-\alpha}}$

QUIZ

Se in un test si rifiuta l'ipotesi nulla al livello del 2%, allora

(a) non si rifiuta a livello 5%

$\boxed{\text{(b) si rifiuta anche a livello 5\%}}$

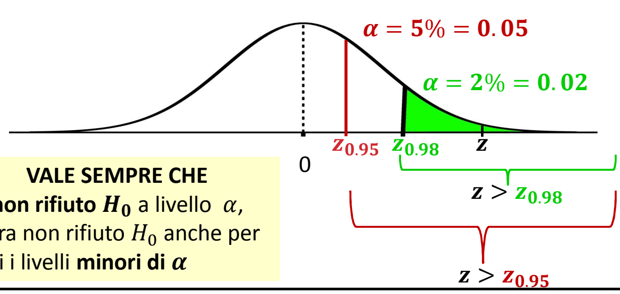
(c) p-value = 0.02

(d) nessuna delle precedenti

VALE SEMPRE CHE
Se **rifiuto H_0** a livello α, allora rifiuto H_0 anche per tutti i livelli **maggiori di α**

Esempio

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ → **Rifiuto H_0 se $z > z_{1-\alpha}$**



VALE SEMPRE CHE
Se **non rifiuto H_0** a livello α, allora non rifiuto H_0 anche per tutti i livelli **minori di α**

QUIZ

Si consideri un campione da una popolazione gaussiana di media μ e varianza σ^2 non nota.
Nel test $H_0: \mu = 5$ contro $H_1: \mu \neq 5$ a livello di significatività α, si rifiuta l'ipotesi nulla se :

(a) $t > t_{1-\alpha/2}^{(n-1)}$ $\boxed{\text{(c)}\ |t| > t_{1-\alpha/2}^{(n-1)}}$

(b) $|t| > -t_{1-\alpha/2}^{(n-1)}$ (d) $|z| > z_{1-\alpha/2}$

QUIZ

Si consideri un campione da una popolazione normale di media μ e deviazione standard σ nota.
Nel test $H_0: \mu = 5$ contro $H_1: \mu \neq 5$ a livello di significatività α, non si rifiuta l'ipotesi nulla se :

(a) $t > t_{1-\alpha/2}^{(n-1)}$ (c) $|t| > t_{1-\alpha/2}^{(n-1)}$

(b) $|t| > -t_{1-\alpha/2}^{(n-1)}$ $\boxed{\text{(d)}\ |z| > z_{1-\alpha/2}}$

QUIZ

Nel test con ipotesi nulla $H_0: p = 0.5$ contro l'alternativa $H_1: p < 0.5$ si rifiuta l'ipotesi nulla se

(a) $|z| > z_{1-\alpha/2}$ (c) $t^{(n-1)} < t_{1-\alpha}$

(b) $z < -z_{1-\alpha}$ (d) $z > z_{1-\alpha}$

QUIZ

Nella verifica di ipotesi $H_0: \mu = 2$ contro $H_1: \mu \neq 2$, la statistica test è

(a) $\mu - 2 > z_{1-\alpha}$ (c) $\frac{\bar{x}-2}{\sqrt{\bar{s}^2/n}}$

(b) $\frac{\bar{x}-2}{\bar{s}^2/\sqrt{n}}$ (d) $(\mu - 2) / \sigma$

QUIZ

Se in un test il p-value vale 0.3, allora

(a) si rifiuta H_0 a livello 1%

(b) non si rifiuta H_0 a livello 1%

(c) si rifiuta H_0 a livello 5%

(d) si rifiuta H_0 sempre

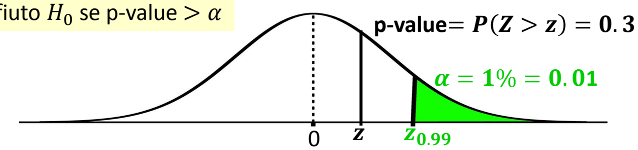
Esempio

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ $\rightarrow$ Rifiuto H_0 se $\boldsymbol{z} > z_{1-\alpha}$

$\rightarrow$ Rifiuto H_0 se p-value $< \alpha$

Rifiuto H_0 se p-value $< \alpha$

Non rifiuto H_0 se p-value $> \alpha$



QUIZ

Se si rifiuta H_0 a livello 5%, allora

(a) p-value = 0.4 (c) p-value = 0.004

(b) si rifiuta anche a livello 1% (d) nessuna delle precedenti

QUIZ

Se in un test risulta p-value = 0.7, allora

(a) si rifiuta l'ipotesi nulla (c) si rifiuta H_0 per $\alpha = 5\%$

(b) non si rifiuta l'ipotesi nulla (d) nessuna delle precedenti

QUIZ

Se in un test risulta p-value = 0.02, allora

(a) si rifiuta H_0 per $\alpha = 5\%$ ma non per $\alpha = 1\%$ (b) si rifiuta H_0 a livello $\alpha = 1\%$

(c) non si rifiuta l'ipotesi nulla (d) nessuna delle precedenti

QUIZ

La varianza della media campionaria è

(a) $\sigma^2/\sqrt{n}$ (c) $\sigma/\sqrt{n}$

(b) σ^2/n (d) nessuna delle precedenti

QUIZ

In un test statistico la probabilità di rifiutare l'ipotesi nulla quando è vera è la probabilità di

(a) non commettere un errore del primo tipo (o specie)

(b) commettere un errore del secondo tipo

(c) non commettere un errore del secondo tipo

(d) commettere un errore del primo tipo

Esercizio 1

Un docente di statistica vuole valutare la capacità di autovalutazione degli studenti che hanno appena sostenuto l'esame. Dalla correzione degli elaborati degli studenti ha ottenuto un punteggio medio di 26
Il docente interpella 8 studenti a caso che esprimono un valore di autovalutazione del proprio esame come riportato in tabella:

26	**28**	**30**	**19**	**15**	**26**	**27**	**29**

Ipotizzando una distribuzione Normale, verificare l'ipotesi nulla che gli studenti abbiano una buona capacità di autovalutazione con una probabilità di errore di primo tipo $\alpha = 0.01$.

Popolazione $N\ (\mu, \sigma^2)$ i.i.d. σ^2 incognita

Campione $n = 8$ $\quad \bar{x}_n = \ldots$ $\quad \bar{s}_n = \ldots$ $\quad \alpha = 0.01$

TEST $H_0: \mu = 26$ $\quad H_1: \mu \neq 26$ → Si rifiuta H_0 se $|t| > t_{1-\alpha/2}^{(n-1)}$

dove la statistica test è $|t| = \left|\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}}\right|$

26	**28**	**30**	**19**	**15**	**26**	**27**	**29**

Media campionaria

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{26+28+30+19+15+26+27+29}{8} = 25$$

Varianza campionaria

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

$$= \frac{(26-25)^2 + (28-25)^2 + (30-25)^2 + \cdots + (29-25)^2}{8-1} =$$

$$= \frac{(1)^2 + (3)^2 + (5)^2 + (-6)^2 + (-10)^2 + (1)^2 + (2)^2 + (4)^2}{7} =$$

$$= \frac{1+9+25+36+100+1+4+16}{7} = \frac{192}{7} = 27.4$$

Deviazione standard campionaria $\quad \bar{s}_n = \sqrt{\bar{s}_n{}^2} = \sqrt{27.4} = 5.2$

Popolazione $N\ (\mu, \sigma^2)$ i.i.d. $\boldsymbol{\sigma^2}$ **incognita**

Campione $n = 8 < \mathbf{30}$ $\quad \bar{x}_n = \mathbf{25}$ $\quad \bar{s}_n = \mathbf{5.2}$ $\quad \alpha = 0.01$

TEST $H_0: \mu = 26$ $\quad H_1: \mu \neq 26$ → Si rifiuta H_0 se $|t| > t_{1-\alpha/2}^{(n-1)}$

$$|t| = \left|\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}}\right| = \left|\frac{25-26}{5.2/\sqrt{8}}\right| = |-0.51| = 0.51$$

$$t_{1-\alpha/2}^{(n-1)} = t_{1-0.01/2}^{(8-1)} = t_{0.995}^{(7)} = 3.49948 \approx 3.50$$

Non è vero che $|t| > t_{1-\alpha/2}^{(n-1)}$

$$0.51 < 3.50$$

→ Non si rifiuta H_0 a livello 1%

→ C'è evidenza di scarsa capacità di autovalutazione degli studenti

Esercizio 2

Da un'indagine campionaria su 200 clienti di un ristorante è emerso che 18 sono insoddisfatti del servizio.

1) Calcolare la proporzione di clienti soddisfatti
2) Fornire l'intervallo di confidenza al 95% per la proporzione dei clienti insoddisfatti
3) Verificare ad un livello di significatività del 2% se la vera proporzione di clienti insoddisfatti possa ritenersi pari a 0.1 o se sia minore.

Esercizio 2
Da un'indagine campionaria su 200 clienti di un ristorante è emerso che 18 sono insoddisfatti del servizio.
1) Calcolare la proporzione di clienti soddisfatti

1)

Popolazione $Bernoulli(p)$ i.i.d. $p =$ proporzione di clienti soddisfatti

Campione $n = 200$ $x = 18$ sodisfatti

$$\hat{p}_n = \frac{n.\ soddisfatti}{n.\ clienti} = \frac{(200-18)}{200} = 0.91$$

Esercizio 2
Da un'indagine campionaria su 200 clienti di un ristorante è emerso che 18 sono insoddisfatti del servizio.
2) Fornire l'intervallo di confidenza al 90% per la proporzione dei clienti insoddisfatti

Popolazione $Bernoulli(p)$ i.i.d. $p =$ proporzione di clienti insoddisfatti

Campione $n = 200$ $\hat{p}_n = 18/200 = 0.09$

$n\hat{p}_n = 18 > 5$
$n(1-\hat{p}_n) = 182 > 5$

$\alpha = 1 - 0.90 = 0.10$ $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.10}{2}} = z_{0.95} = 1.64$

$$\begin{aligned} \mathbf{IC(90\%)} &= \hat{p}_n \pm z_{1-\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n} = \\ &= 0.09 \pm 1.64\sqrt{0.09(1-0.09)/200} = \\ &= (0.06,\quad 0.12) \end{aligned}$$

Confidiamo al 90% che la vera proporzione di clienti insoddisfatti sia compresa tra il 6% e il 12%.

Esercizio 2
Da un'indagine campionaria su 200 clienti di un ristorante è emerso che 18 sono insoddisfatti del servizio.
3) Verificare ad un livello di significatività del 2% se la vera proporzione di clienti insoddisfatti possa ritenersi pari a 0.1 o se sia minore.

Popolazione $Bernoulli(p)$ i.i.d. $p =$ proporzione di clienti insoddisfatti

Campione $n = 200$ $\hat{p}_n = 18/200 = 0.09$ $\alpha = 0.02$

TEST $H_0: p = 0.10$ $H_1: p < 0.10$ → Si rifiuta H_0 se $z < -z_{1-\alpha}$

dove la statistica test è

$$z = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Verifico le condizioni di applicabilità del test

$np_0 = 20 > 5$
$n(1-p_0) = 180 > 5$

Popolazione $Bernoulli(p)$ i.i.d. $p =$ proporzione di clienti insoddisfatti

Campione $n = 200$ $\hat{p}_n = 18/200 = 0.09$ $\alpha = 0.02$

TEST $H_0: p = 0.10$ $H_1: p < 0.10$ → Si rifiuta H_0 se $z < -z_{1-\alpha}$

$$z = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.09 - 0.10}{\sqrt{0.10(1-0.10)/200}} = -0.47$$

$$z_{1-\alpha} = z_{1-0.02} = z_{0.98} = 2.05$$

$-0.47 > -2.05$ → Non si rifiuta H_0 a livello 1%
Si può ritenere la vera proporzione di clienti insoddisfatti sia pari a 0.1 (=10%)

Esercizio 2 **Soluzione con p-value**

Da un'indagine campionaria su 200 clienti di un ristorante è emerso che 18 sono insoddisfatti del servizio.

3) Verificare ad un livello di significatività del 2% se la vera proporzione di clienti insoddisfatti possa ritenersi pari a 0.1 o se sia minore.

Popolazione $Bernoulli(p)$ i.i.d. $p =$ proporzione di clienti insoddisfatti

Campione $n = 200$ $\hat{p}_n = 18/200 = 0.09$ $\alpha = 0.02$

TEST $H_0: p = 0.10$ $H_1: p < 0.10$ → **Si rifiuta H_0 se p-value $< \alpha$**

$$z = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.09 - 0.10}{\sqrt{0.10(1-0.10)/200}} = -0.47$$

p-value $= P(Z < -0.47) = 0.31918 \approx 0.32$

$0.32 > \alpha$ per gli usuali valori di $\alpha = 0.05,\ 0.01,\ 0.001, \ldots$
per cui non si rifiuta H_0

VERIFICA DELLE IPOTESI per il confronto di medie/proporzioni

Condizioni di applicabilità del test di confronfro di medie/proporzioni di due popolazioni indipendenti:

Due campioni indipendenti
$X_1, \ldots, X_{n1} \sim b(p_1)$ $Y_1, \ldots, Y_{n2} \sim b(p_2)$ $n_1\hat{p}_1 > 5, \quad n_1(1-\hat{p}_1) > 5$ $n_2\hat{p}_2 > 5, \quad n_2(1-\hat{p}_2) > 5$
$X_1, \ldots, X_{n1} \sim \ldots\ldots (\mu_1, \sigma_1{}^2)$ $Y_1, \ldots, Y_{n2} \sim \ldots\ldots (\mu_2, \sigma_2{}^2)$ $n_1 > 30$ $n_2 > 30$ $\sigma_1{}^2 = \sigma_2{}^2$ uguale e incognita

VERIFICA DELLE IPOTESI per il confronto di medie/proporzioni

Ipotesi nulla	Ipotesi alternativa	Statistica test z oppure t	Rifiuto H_0 se
$H_0: p_1 = p_2$	$H_1: p_1 \neq p_2$ $H_1: p_1 > p_2$ $H_1: p_1 < p_2$	$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\left(\hat{p}_1 = \frac{x_1}{n_1} \quad \hat{p}_2 = \frac{x_2}{n_2} \quad \bar{p} = \frac{x_1 + x_2}{n_1 + n_2}\right)$	$\lvert z \rvert > z_{1-\alpha/2}$ $z > z_{1-\alpha}$ $z < -z_{1-\alpha}$
$H_0: \mu_1 = \mu_2$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\bar{s}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\left(\bar{s}^2 = \frac{(n_1-1)\bar{s}_1^2 + (n_2-1)\bar{s}_2^2}{n_1+n_2-2}\right)$	$\lvert t \rvert > t_{1-\alpha/2}^{(n1+n2-2)}$ $t > t_{1-\alpha}^{(n1+n2-2)}$ $t < -t_{1-\alpha}^{(n1+n2-2)}$

Esercizio 3

Si considerino due campioni, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

1) Calcolare la proporzione di fiduciosi in ciascun stato

2) Verificare ad un livello di significatività del 1% se la vera proporzione di fiduciosi dello stato A possa ritenersi pari a 0.7

3) Vi è evidenza empirica che nello stato A le persone sono più ottimiste sul futuro dell'economia?

Esercizio 3

Si considerino due campioni casuali, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

1) Calcolare la proporzione di fiduciosi in ciascun stato

Popolazione A $Bernoulli(p_A)$ i.i.d. $p_A =$ proporzione di fiduciosi in A

Popolazione B $Bernoulli(p_B)$ i.i.d. $p_B =$ proporzione di fiduciosi in B

Campione A $n_A = 900$ $\boldsymbol{\hat{p}_A = 648/900 = 0.72}$

Campione B $n_B = 800$ $\boldsymbol{\hat{p}_B = 560/800 = 0.70}$

campioni indipendenti

Esercizio 3

Si considerino due campioni casuali, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

2) Verificare ad un livello di significatività del 1% se la vera proporzione di fiduciosi nello stato A possa ritenersi pari a 0.7

Popolazione A $b(p)$ i.i.d. $p =$ proporzione di fiduciosi in A

Campione A $n = 900$ $\hat{p}_n = 648/900 = 0.72$ $\boldsymbol{\alpha = 0.01}$

TEST $H_0: p = 0.7$ $H_1: p \neq 0.7$ → **Si rifiuta H_0 se** $|z| > z_{1-\alpha/2}$

dove la statistica test è

$$|z| = \left|\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}\right|$$

Prima di tutto, verifichiamo le condizioni di applicabilità del test :

$$np_0 = 900 \times 0.7 = 630 > 5$$

$$n(1-p_0) = 900 \times (1-0.7) = 270 > 5$$

Popolazione A $b(p)$ i.i.d. $p =$ proporzione di fiduciosi in A

Campione A $n = 900$ $\hat{p}_n = 648/900 = 0.72$ $\boldsymbol{\alpha = 0.01}$

TEST $H_0: p = 0.7$ $H_1: p \neq 0.7$ → **Si rifiuta H_0 se** $|z| > z_{1-\alpha/2}$

$$|z| = \left|\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}\right| = \left|\frac{0.72 - 0.7}{\sqrt{0.7(1-0.7)/900}}\right| = 1.31$$

$$z_{1-\alpha/2} = z_{1-0.01/2} = z_{0.995} = 2.58$$

$1.31 < 2.58$ → Non si rifiuta H_0 a livello 1% → Si ritiene che la vera proporzione di fiduciosi sul futuro dell'economia nello stato A sia 70% (= 0.7)

Esercizio 3 **Soluzione con p-value**

Si considerino due campioni casuali, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

2) Verificare ad un livello di significatività del 1% se la vera proporzione di fiduciosi nello stato A possa ritenersi pari a 0.7

Popolazione A $b(p)$ i.i.d. $p =$ proporzione di fiduciosi in A

Campione A $n = 900$ $\hat{p}_n = 648/900 = 0.72$ $\boldsymbol{\alpha = 0.01}$

TEST $H_0: p = 0.7$ $H_1: p \neq 0.7$ → **Si rifiuta H_0 se p-value $< \alpha$**

$$|z| = \left|\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}\right| = \left|\frac{0.72 - 0.7}{\sqrt{0.7(1-0.7)/900}}\right| = 1.31$$

p-value $= P(|Z| > 1.31) = 2 \times P(Z < -1.31) = 2 \times 0.09510 = 0.19$

$0.19 > \alpha$ per gli usuali valori di $\alpha = 0.05,\ 0.01,\ 0.001, \dots$

dunque non si rifiuta H_0

Esercizio 3

Si considerino due campioni casuali, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

3) Vi è evidenza empirica che nello stato A le persone sono più ottimiste sul futuro dell'economia?

Popolazione A $Bernoulli(p_A)$ i.i.d. $p_A =$ proporzione di fiduciosi in A

Popolazione B $Bernoulli(p_B)$ i.i.d. $p_B =$ proporzione di fiduciosi in B

Campione A $n_A = 900$ $\hat{p}_A = 648/900 = 0.72$

Campione B $n_B = 800$ $\hat{p}_B = 560/800 = 0.70$

campioni indipendenti

TEST $H_0: p_A = p_B$ $H_1: p_A > p_B$ → **Si rifiuta H_0 se $z > z_{1-\alpha}$**

dove la statistica test è

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \qquad \text{dove} \quad \bar{p} = \frac{x_A + x_B}{n_A + n_B}$$

Campione A $n_A = 900$ $\hat{p}_A = 648/900 = 0.72$

Campione B $n_B = 800$ $\hat{p}_B = 560/800 = 0.70$

TEST $H_0: p_A = p_B$ $H_1: p_A > p_B$ → Si rifiuta H_0 se $z > z_{1-\alpha}$

Verifichiamo le condizioni di applicabilità del test :

$n_A\hat{p}_A = 900 \times 0.72 = 648 > 5$ $n_A(1-\hat{p}_A) = 900 \times 0.72 = 252 > 5$

$n_B\hat{p}_B = 800 \times 0.7 = 560 > 5$ $n_B(1-\hat{p}_B) = 800 \times 0.7 = 240 > 5$

Calcoliamo :

$$\bar{p} = \frac{x_A + x_B}{n_A + n_B} = \frac{n.\,totale\ fiduciosi}{n.\,totale\ cittadini} = \frac{648+560}{900+800} = 0.71$$

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} = \frac{0.72-0.70}{\sqrt{0.71(1-0.71)\left(\frac{1}{900}+\frac{1}{800}\right)}} = 0.91$$

Ad es. scelgo $\alpha = 0.01$ → $z_{1-\alpha} = z_{1-0.01} = z_{0.99} = 2.33$

$0.91 < 2.33$ → Non si rifiuta H_0 a livello 1% → Non c'è evidenza di maggior ottimismo sul futuro dell'economia nello stato A rispetto allo stato B

Esercizio 3 **Soluzione con p-value**

Si considerino due campioni casuali, il primo composto da 900 cittadini dello stato A e il secondo da 800 cittadini dello stato B.
Si è rilevato che il numero di cittadini fiduciosi sul futuro dell'economia è 648 nello stato A e 560 nello stato B.

3) Vi è evidenza empirica che nello stato A le persone sono più ottimiste sul futuro dell'economia?

Campione A $n_A = 900$ $\hat{p}_A = 648/900 = 0.72$

Campione B $n_B = 800$ $\hat{p}_B = 560/800 = 0.70$ $\bar{p} = 0.71$

TEST $H_0: p_A = p_B$ $H_1: p_A > p_B$ → **Si rifiuta H_0 se p-value $< \alpha$**

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} = \frac{0.72-0.70}{\sqrt{0.71(1-0.71)\left(\frac{1}{900}+\frac{1}{800}\right)}} = 0.91$$

p-value $= P(Z > 0.91) = 1 - P(Z < 0.91) = 1 - 0.81859 = 0.18$

$0.18 > \alpha$ per gli usuali valori di $\alpha = 0.05,\ 0.01,\ 0.001, \ldots$
per cui non si rifiuta H_0

Esercizio 4

Ad un esame universitario sono stati assegnati in modo casuale due compiti diversi.
Il compito A è stato dato a 102 studenti che hanno ottenuto un voto medio pari 24 con varianza 12.25; il compito B è stato svolto da 105 studenti che hanno ottenuto voto medio 24.6 con varianza 14.44
Verificare al livello di significatività 5% che i compiti A e B sono della stessa difficoltà sulla base del voto medio riportato dagli studenti nei due gruppi.

Popolazione A … $(\mu_A, \sigma_A{}^2)$ i.i.d. σ_A incognita

Popolazione B … $(\mu_B, \sigma_B{}^2)$ i.i.d. σ_B incognita

$\boldsymbol{\alpha = 0.05}$

Campione A $n_A = 102 > 30$ $\bar{x}_A = 24$ $\bar{s}_A{}^2 = 12.25$

Campione B $n_B = 105 > 30$ $\bar{x}_B = 24.6$ $\bar{s}_B{}^2 = 14.44$

campioni indipendenti

TEST $H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$ → **Si rifiuta H_0 se $|t| > t_{1-\alpha/2}^{(n_A+n_B-2)}$**

$$t = \left|\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\bar{s}^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}}\right| \qquad \bar{s}^2 = \frac{(n_A-1)\bar{s}_A{}^2 + (n_B-1)\bar{s}_B{}^2}{n_A+n_B-2}$$

Campione A $n_A = 102$ $\boldsymbol{\bar{x}_A = 24}$ $\boldsymbol{\bar{s}_A{}^2 = 12.25}$

Campione B $n_B = 105$ $\boldsymbol{\bar{x}_B = 24.6}$ $\boldsymbol{\bar{s}_B{}^2 = 14.44}$ $\boldsymbol{\alpha = 0.05}$

TEST $H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$ → **Si rifiuta H_0 se $|t| > t_{1-\alpha/2}^{(n_A+n_B-2)}$**

$$\bar{s}^2 = \frac{(n_A - 1)\bar{s}_A{}^2 + (n_B - 1)\bar{s}_B{}^2}{n_A + n_B - 2} =$$

$$= \frac{(102-1) \times 12.25 + (105-1) \times 14.44}{102 + 105 - 2} = 13.36$$

$$|t| = \left| \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\bar{s}^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} \right| = \left| \frac{24 - 24.6}{\sqrt{13.36 \left(\frac{1}{102} + \frac{1}{105} \right)}} \right| = 1.18$$

$$t_{1-\alpha/2}^{(n_A+n_B-2)} = t_{1-0.05/2}^{(102+105-2)} = t_{0.975}^{(205)} = z_{0.975} = 1.96$$

$1.18 < 1.96$ → Non si rifiuta H_0 al livello 5%.
Si ritiene che i due compiti hanno pari difficoltà

Campione A $n_A = 102$ $\boldsymbol{\bar{x}_A = 24}$ $\boldsymbol{\bar{s}_A{}^2 = 12.25}$

Campione B $n_B = 105$ $\boldsymbol{\bar{x}_B = 24.6}$ $\boldsymbol{\bar{s}_B{}^2 = 14.44}$ $\boldsymbol{\alpha = 0.05}$

TEST $H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$ → **Si rifiuta H_0 se p-value $< \alpha$**

Soluzione con p-value

$$|t| = \left| \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\bar{s}^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} \right| = 1.18$$

p-value $= P\left(|\boldsymbol{t}^{(n_A+n_B-2)}| > |t|\right) = P\left(|\boldsymbol{t}^{(205)}| > 1.18\right)$

$= 2 \times P(Z < -1.18) = 2 \times 0.11900 = 0.24$

$0.24 > \alpha$ per ogni valore solito di α, quindi non si rifiuta H_0
Si ritiene che i due compiti hanno pari difficoltà

Esercizio 5

Ad un esame universitario sono stati assegnati due compiti diversi A e B. L'esame non è stato superato dal 10% dei 150 studenti che hanno svolto il compito A e dal 9% dei 200 studenti che hanno svolto il compito B.

1) Qual è la proporzione di studenti del gruppo A che non ha superato l'esame? E il numero di studenti del gruppo A che non ha superato l'esame? Rispondere alle stesse domande anche per il gruppo B
2) Fornire l'intervallo di confidenza al 99% per la vera proporzione di insufficienze per il compito A
3) Verificare, ad un livello di significatività del 5%, se la vera proporzione di insufficienze per il compito A possa ritenersi pari all' 8% oppure maggiore
4) Verificare, al livello di significatività 0.05, se la proporzione di insufficienze nei due gruppi possa ritenersi uguale.

Esercizio 5

Ad un esame universitario sono stati assegnati due compiti diversi A e B. L'esame non è stato superato dal 10% dei 150 studenti che hanno svolto il compito A e dal 9% dei 200 studenti che hanno svolto il compito B.

1) Qual è la proporzione di studenti del gruppo A che non ha superato l'esame? E il numero di studenti del gruppo A che non ha superato l'esame? Rispondere alle stesse domande anche per il gruppo B

Popolazione A $b(p_A)$ i.i.d. p_A = proporzione di insufficienze gruppo A

Popolazione B $b(p_B)$ i.i.d. p_B = proporzione di insufficienze gruppo B

Campione A $n_A = 150$ $\boldsymbol{\hat{p}_A = 10\% = 0.10}$

Campione B $n_B = 200$ $\boldsymbol{\hat{p}_B = 9\% = 0.09}$

Il numero di insufficienze nei due gruppi :

$$\boldsymbol{x_A = 0.10 \times 150 = 15}$$

$$\boldsymbol{x_B = 0.09 \times 200 = 18}$$

Esercizio 5

Ad un esame universitario sono stati assegnati due compiti diversi A e B. L'esame non è stato superato dal 10% dei 150 studenti che hanno svolto il compito A e dal 9% dei 120 studenti che hanno svolto il compito B.

2) Fornire l'intervallo di confidenza al 99% per la vera proporzione di insufficienze per il compito A

Popolazione A $b(p)$ i.i.d. p = proporzione di insufficienze gruppo A

Campione A $n = 150$ $\hat{p}_n = 0.10$

$n\hat{p}_n = 15 > 5$

$n(1-\hat{p}_n) = 135 > 5$

$\alpha = 1 - 0.99 = 0.01$ $z_{1-\alpha/2} = z_{1-0.01/2} = z_{0.995} = 2.58$

$$\mathbf{IC(99\%)} = \hat{p}_n \pm z_{1-\alpha/2}\sqrt{\hat{p}_n(1-\hat{p}_n)/n} =$$
$$= 0.10 \pm 2.58\sqrt{0.10(1-0.10)/150} =$$
$$= (0.10 - 0.06,\quad 0.10 + 0.06) =$$
$$= (0.04,\quad 0.16) =$$

Si confida al 99% che la vera proporzione di insufficienze per il compito A sia tra 4% e 16%

Esercizio 5

Ad un esame universitario sono stati assegnati due compiti diversi A e B. L'esame non è stato superato dal 10% dei 150 studenti che hanno svolto il compito A e dal 9% dei 120 studenti che hanno svolto il compito B.

3) Verificare, ad un livello di significatività del 5%, se la vera proporzione di insufficienze per il compito A possa ritenersi pari all' 8% oppure maggiore

Popolazione A $b(p)$ i.i.d. p = proporzione di insufficienze gruppo A

Campione A $n = 150$ $\hat{p}_n = 0.10$

$n\hat{p}_0 = 12 > 5$

$n(1-\hat{p}_0) = 138 > 5$

TEST $H_0: p = 0.08$ $H_1: p > 0.08$ → **Si rifiuta H_0 se $z > z_{1-\alpha}$**

$\alpha = 0.05$

dove la statistica test è

$$z = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Popolazione A $b(p)$ i.i.d. p = proporzione di insufficienze gruppo A

Campione A $n = 150$ $\hat{p}_n = 0.10$ $\alpha = 0.05$

TEST $H_0: p = 0.08$ $H_1: p > 0.08$ → **Si rifiuta H_0 se $z > z_{1-\alpha}$**

$$z = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.10 - 0.08}{\sqrt{0.08(1-0.08)/150}} = 0.90$$

$$z_{1-\alpha} = z_{1-0.05} = z_{0.95} = 1.64$$

Essendo $0.90 < 1.64$, non si rifiuta H_0 al livello 5%
Si ritiene che la vera proporzione di insufficienze per il compito A sia 8%

Esercizio 5

Ad un esame universitario sono stati assegnati due compiti diversi A e B. L'esame non è stato superato dal 10% dei 150 studenti che hanno svolto il compito A e dal 9% dei 120 studenti che hanno svolto il compito B.

4) Verificare, al livello di significatività 0.05, se la proporzione di insufficienze nei due gruppi possa ritenersi uguale.

Popolazione A $b(p_A)$ i.i.d. p_A = proporzione di insufficienze gruppo A

Popolazione B $b(p_B)$ i.i.d. p_B = proporzione di insufficienze gruppo B

Campione A $n_A = 150$ $\hat{p}_A = 0.10$

Campione B $n_B = 200$ $\hat{p}_B = 0.09$

campioni indipendenti

$\alpha = 0.05$

TEST $H_0: p_A = p_B$ $H_1: p_A \neq p_B$ → **Si rifiuta H_0 se $|z| > z_{1-\alpha/2}$**

Verifichiamo le condizioni di applicabilità del test :

$n_A\hat{p}_A = 150 \times 0.10 = 15 > 5$ $n_A(1-\hat{p}_A) = 150 \times 0.90 = 135 > 5$

$n_B\hat{p}_B = 200 \times 0.09 = 18 > 5$ $n_B(1-\hat{p}_B) = 200 \times 0.91 = 182 > 5$

Campione A $n_A = 150$ $\hat{p}_A = 0.10$ $\boldsymbol{x_A = 15}$

Campione B $n_B = 200$ $\hat{p}_B = 0.09$ $\boldsymbol{x_B = 18}$ $\boldsymbol{\alpha = 0.05}$

TEST $H_0\colon p_A = p_B$ $H_1\colon p_A \neq p_B$ → **Si rifiuta $\boldsymbol{H_0}$ se $\boldsymbol{|z| > z_{1-\alpha/2}}$**

$$|z| = \left|\frac{\hat{p}_A - \hat{p}_B}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}}\right| = \left|\frac{0.10 - 0.09}{\sqrt{0.094(1-0.094)\left(\frac{1}{150}+\frac{1}{200}\right)}}\right| = 0.32$$

avendo prima calcolato:

$$\bar{p} = \frac{x_A + x_B}{n_A + n_B} = \frac{n.totale\ insuff.}{n.totale\ studenti} = \frac{15+18}{150+200} = 0.094$$

$$z_{1-\alpha/2} = z_{1-0.05/2} = z_{0.975} = 1.96$$

Essendo $|0.32| < 1.96$, non si rifiuta H_0 al livello 5%
Si ritiene che la proporzione di insufficienze sia uguale per i compiti A e B

Campione A $n_A = 150$ $\hat{p}_A = 0.10$ $\boldsymbol{x_A = 15}$

Campione B $n_B = 200$ $\hat{p}_B = 0.09$ $\boldsymbol{x_B = 18}$ $\boldsymbol{\alpha = 0.05}$

TEST $H_0\colon p_A = p_B$ $H_1\colon p_A \neq p_B$ → **Si rifiuta $\boldsymbol{H_0}$ se p-value $\boldsymbol{< \alpha}$**

Soluzione con p-value

Si calcola la statistica test *z* come prima

$$|z| = 0.32$$

$$\text{p-value} = P(|Z| > |z|) = P(|Z| > 0.32) = 2 \times P(Z < -0.32) =$$
$$= 2 \times 0.37448 = 0.75$$

Essendo $0.75 > 0.05$, non si rifiuta H_0 al livello 5%
Si ritiene che la proporzione di insufficienze sia uguale per i compiti A e B