STANDARDIZZAZIONE

X
variabile aleatoria
di media μ
e varianza σ^2

$\longrightarrow$

$Z = \dfrac{X-\mu}{\sigma}$
variabile standardizzata
ha sempre media 0
e varianza 1

Le probabilità per $\boldsymbol{X}$ si possono calcolare dalle corrispettive per $\mathbf{Z}$

$$P(X < x) = P\left(Z < \frac{x-\mu}{\sigma}\right)$$

NOTA BENE

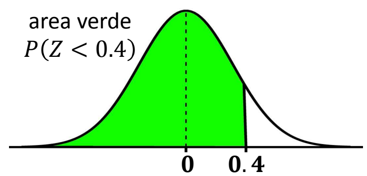
$$X \sim \boldsymbol{N}(\mu, \sigma^2) \longrightarrow Z = \frac{X-\mu}{\sigma} \sim \boldsymbol{N}(0,1)$$

$X \sim N(-10, 25)$

$$P(X < -8) = \mathrm{P}\left(Z < \frac{-8+10}{\sqrt{25}}\right) =$$
$$= P(Z < 0.4) = 0.65542$$

Nota Bene
Le tavole statistiche forniscono $\boldsymbol{P(Z < z)}$ $\rightarrow$ guardate la figura in alto

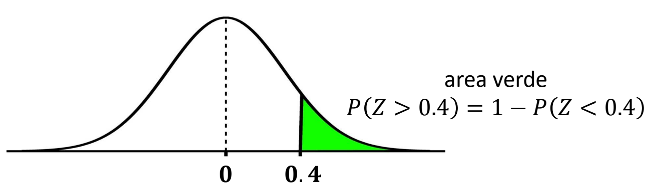
$z = 0.4$ è positivo $\rightarrow$ Tavola Normale pag. 219
$\rightarrow$ $z = 0.4 = 0.4 + 0.00$
$\rightarrow$ incrocio riga di 0.4 e colonna di 0.00
$\rightarrow$ trovo 0.65542



$X \sim N(-10, 25)$

$$P(X \geq -8) = \mathrm{P}\left(Z \geq \frac{-8+10}{\sqrt{25}}\right) =$$
$$= P(Z \geq 0.4) =$$
$$= 1 - P(Z \geq 0.4) =$$
$$= 1 - 0.65542 = 0.34458$$

Nota Bene
Le tavole statistiche forniscono $\boldsymbol{P(Z < z)}$
Tavola Normale pag. 219
$P(Z < 0.4) = 0.65542$

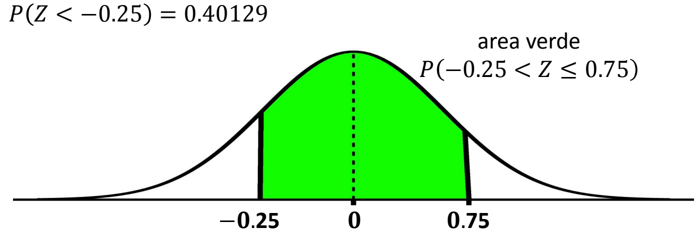


$X \sim N(3, 16)$

$$P(2 < X \leq 6) = \mathrm{P}\left(\frac{2-3}{\sqrt{16}} < Z \leq \frac{6-3}{\sqrt{16}}\right) =$$
$$= P(-0.25 < Z \leq 0.75) =$$
$$= P(Z < 0.75) - P(Z < -0.25) =$$
$$= 0.77337 - 0.40129 = 0.37208$$

Nota Bene
Le tavole statistiche forniscono $\boldsymbol{P(Z < z)}$
Tavola Normale pag. 218-219

$P(Z < 0.75) = 0.77337$
$P(Z < -0.25) = 0.40129$



$Z \sim N(0,1)$ $\qquad$ $P(-1.8 < Z \leq 1.8) = 1 - 2 \times P(Z < -1.80) =$

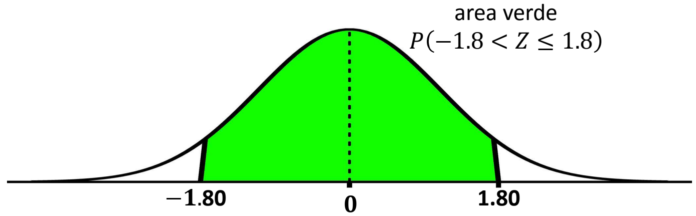
$= 1 - 2 \times 0.03593 =$

$= 0.92814$

Per simmetria
le due code hanno uguale probabilità

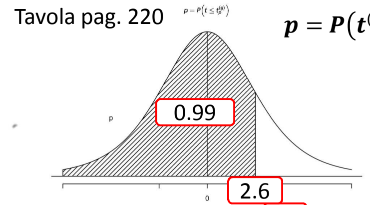
Tavola Normale pag. 218

$P(Z < -1.80) = P(Z > 1.80) = 0.03593$



t-Student

Tavola pag. 220



$\boldsymbol{n}$ → prima colonna

$\boldsymbol{p = P\left(t^{(n)} \leq t\right) = P\left(t^{(n)} < t\right)}$ → prima riga

$\boldsymbol{t}$ **quantile** → dentro la tabella

p / n	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	636.61925
2	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	31.59905
3	0.76489	1.63775	2.35338	3.18245	4.54070	5.84091	12.92398
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60410	8.61030
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03216	6.86883
6	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.95882
7	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	5.40790
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	5.04131
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.78091
10	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.58689
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.43698
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	4.31779
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	4.22083
14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	4.14045
15	0.69120	1.34061	1.75305	2.13145	2.60248	2.94671	4.07277
16	0.69013	1.33676	1.74588	2.11991	2.58349	2.92078	4.01500
17	0.68920	1.33338	1.73961	2.10982	2.56693	2.89823	3.96513
18	0.68836	1.33039	1.73406	2.10092	2.55238	2.87844	3.92165
19	0.68762	1.32773	1.72913	2.09302	2.53948	2.86093	3.88341
20	0.68695	1.32534	1.72472	2.08596	2.52798	2.84534	3.84952
21	0.68635	1.32319	1.72074	2.07961	2.51765	2.83136	3.81928
22	0.68581	1.32124	1.71714	2.07387	2.50832	2.81876	3.79213
23	0.68531	1.31946	1.71387	2.06866	2.49987	2.80734	3.76763
24	0.68485	1.31784	1.71088	2.06390	2.49216	2.79694	3.74540
25	0.68443	1.31635	1.70814	2.05954	2.48511	2.78744	3.72514
26	0.68404	1.31497	1.70562	2.05553	2.47863	2.77871	3.70661
27	0.68368	1.31370	1.70329	2.05183	2.47266	2.77068	3.68959
28	0.68335	1.31253	1.70113	2.04841	2.46714	2.76326	3.67391
29	0.68304	1.31143	1.69913	2.04523	2.46202	2.75639	3.65941
30	0.68276	1.31042	1.69726	2.04227	2.45726	2.75000	3.64596
40	0.68067	1.30308	1.68385	2.02108	2.42326	2.70446	3.55097
60	0.67860	1.29582	1.67065	2.00030	2.39012	2.66028	3.46020
120	0.67654	1.28865	1.65765	1.97993	2.35782	2.61742	3.37345
∞	0.67449	1.28155	1.64485	1.95996	2.32635	2.57583	3.29053

Tabella A3. - Tavola della t di Student. La tavola restituisce i valori di $t_p^{(g)}$ dove g sono i gradi di libertà. Si tenga sempre conto della relazione $t_p^{(g)} = -t_{1-p}^{(g)}$.

Nota Bene
Le tavole statistiche
forniscono $\boldsymbol{P\left(t^{(n)} < t\right)}$

Trovare il quantile t con probabilità $P\left(t^{(15)} \leq t\right) = 0.99$

Si incrociano riga di $n = 15$

e colonna di $p = 0.99$

per trovare

$t = 2.60248 \approx 2.6$

POPOLAZIONE $X \sim$ **distribuzione** (μ, σ^2)

CAMPIONE $X_1, X_2, X_3, \ldots, X_n \sim$ **distribuzione** (μ, σ^2)

MEDIA CAMPIONARIA

$$\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

VARIANZA CAMPIONARIA

$$\mathrm{S}_n{}^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1} = \frac{(X_1 - \overline{X}_n)^2 + (X_1 - \overline{X}_n)^2 + \cdots + (X_n - \overline{X}_n)^2}{n-1}$$

$\overline{X}_n$ è uno stimatore della media μ perché $E(\overline{X}_n) = \mu$. Inoltre :

$V(\overline{X}_n) = \sigma^2/n$ $\qquad$ se σ^2 è nota

$V(\overline{X}_n) \approx \mathrm{S}_n{}^2/n$ $\qquad$ se σ^2 è incognita

per cui $\overline{X}_n$ approssima μ tanto meglio quanto più il campione è grande essendo che la sua varianza diminuisce al crescere di n

POPOLAZIONE $X \sim$ **bernoulli** (p)

$$X = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p \end{cases} \qquad \text{media} = p \qquad \text{varianza} = p(1-p)$$

CAMPIONE $X_1, X_2, X_3, \ldots, X_n \sim$ **bernoulli** (p)

PROPORZIONE CAMPIONARIA = media campionaria

= proporzione di successi su n prove

$$\hat{p}_n = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\textit{numero successi}}{\textit{numero prove}}$$

$\hat{p}_n$ è uno stimatore della media p perché $E(\hat{p}_n) = p$. Inoltre :

$$V(\hat{p}_n) = \frac{p(1-p)}{n}$$

per cui $\hat{p}_n$ approssima p tanto meglio quanto più il campione è grande

essendo che la sua varianza diminuisce al crescere di n

Esercizio 1

Calcolare media, varianza e deviazione standard del seguente campione di ampiezza n = 5 :

1	1	3	4	6

Media campionaria

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+1+3+4+6}{5} = 3$$

Varianza campionaria

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

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

$$= \frac{(-2)^2+(-2)^2+(0)^2+(1)^2+(3)^2}{4} =$$

$$= \frac{4+4+0+1+9}{4} = \frac{18}{4} = 4.5$$

Deviazione standard campionaria $\bar{s}_n = \sqrt{\bar{s}_n{}^2} = \sqrt{4.5} = 2.1$

Esercizio 2

L'Ufficio del Turismo è interessato a sapere quanti italiani hanno intenzione di andare in vacanza quest'estate.
A tal fine sono state intervistate 12 persone a caso con il seguente esito:

Si	Si	No	No	Si	Si	Si	Si	No	Si	No	Si

In base al campione, stimare proporzione e varianza attesa di italiani che hanno intenzione di andare in vacanza quest'estate.

Proporzione campionaria (= media campionaria di campioni bernoulliani !)

$$\hat{p}_n = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+1+0+0+1+1+1+1+0+1+0+1}{12} = \frac{8}{12} = 0.67$$

Varianza campionaria

$$\bar{s}_n{}^2 = \frac{\hat{p}_n(1-\hat{p}_n)}{n} = \frac{0.67 \times (1-0.67)}{12} = 0.018$$

Deviazione standard campionaria $\bar{s}_n = \sqrt{\bar{s}_n{}^2} = \sqrt{0.018} = 0.13$

INTERVALLI DI CONFIDENZA

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

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

Esercizio 3

Sono state rilevate le velocità di 7 automobili in una strada dove vige il limite di 50 Km/h :

79	73	68	77	86	71	69

Assumendo che la variabile velocità sia distribuita normalmente, determinare l'intervallo di confidenza al 95% per la velocità media

$X_1, X_2, \dots, X_7 \sim N(\mu, \sigma^2)$ i.i.d., σ^2 **incognita**, $n = 7 \leq 30$

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

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

Bisogna calcolare:

- $\bar{x}_n$ media campionaria
- $\bar{s}_n$ deviazione standard campionaria
- $t_{1-\alpha/2}^{(n-1)}$ quantile dalla distribuzione t-Student

79	73	68	77	86	71	69

Media campionaria, varianza campionaria e deviazione standard :

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{79+73+68+77+86+71+69}{7} = 74.7$$

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

$$= \frac{(79-74.7)^2 + (73-74.7)^2 + (68-74.7)^2 + \cdots + (69-74.7)^2}{7-1} =$$

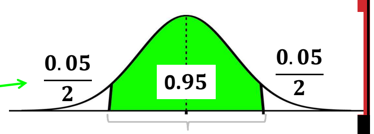
$$= 40.9$$

$$\bar{s}_n = \sqrt{\bar{s}_n{}^2} = \sqrt{40.9} = 6.4$$

Intervallo di confidenza al 95%

$\alpha = 1 - 0.95 = 0.05$

probabilità da ripartire in modo simmetrico sulle code della densità t-Student



Quantile :

$$t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{1-\frac{0.05}{2}}^{(7-1)} = t_{0.975}^{(6)} = 2.44691$$

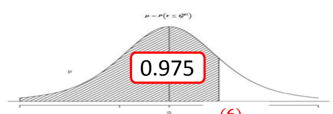


Tavola t-Student pag. 220

n \ p	0.75	0.90	0.95	0.975	0.99
1	1.00000	3.07768	6.31375	12.70620	31.82052
2	0.81650	1.88562	2.91999	4.30265	6.96456
3	0.76489	1.63775	2.35338	3.18245	4.54070
4	0.74070	1.53321	2.13185	2.77645	3.74695
5	0.72669	1.47588	2.01505	2.57058	3.36493
6	0.71756	1.43976	1.94318	2.44691	3.14267
7	0.71114	1.41492	1.89458	2.36462	2.99795
8	0.70639	1.39682	1.85955	2.30600	2.89646
9	0.70272	1.38303	1.83311	2.26216	2.82144
10	0.69981	1.37218	1.81246	2.22814	2.76377

$n = 7$

$\bar{x}_n = 74.7$

$\bar{s}_n = 6.4$

$t_{1-\frac{\alpha}{2}}^{(n-1)} = 2.44691$

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

$$= 74.7 \pm 2.44691 \frac{6.4}{\sqrt{7}} =$$

$$= (74.7 - 5.9, \quad 74.7 + 5.9) =$$

$$= (\mathbf{68.8}, \quad \mathbf{80.6})$$

Possiamo considerare il limite di velocità di 50 Km/h rispettato?

Non proprio. Confidiamo al 95% che la vera velocità media delle automobili sul tratto di strada in esame sia tra 68.8 e 80.6 Km/h

Esercizio 4

In un sondaggio è stato chiesto ai residenti di una città quanti minuti al giorno dedicano all'attività fisica. Da un campione di 50 persone risulta un tempo medio di 30 minuti con deviazione standard 4.2 minuti.
Costruire un intervallo di confidenza del 90% per il tempo medio dedicato all'esercizio fisico da parte degli abitanti di quella città.

$X_1, X_2, \dots, X_{50} \sim$ distr. incognita (μ, σ^2) i.i.d., σ^2 incognita, $n = 50 > 30$

→ $\bar{X}_n$ ha distribuzione **t-Student** ≈ **Normale** per TLC

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

→ da scrivere all' esame !

Bisogna calcolare:

- $\bar{x}_n$ media campionaria → $\bar{x}_n = 30$
- $\bar{s}_n$ deviazione standard campionaria → $\bar{s}_n = 4.2$
- $z_{1-\alpha/2}$ quantile dalla distribuzione Normale → $\alpha = 1 - 0.90 = 0.10$

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

$n = 50$

$\bar{x}_n = 30$

$\bar{s}_n = 4.2$

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

$$\mathbf{IC(90\%)} = \bar{x}_n \pm z_{1-\alpha/2} \frac{\bar{s}_n}{\sqrt{n}} = 30 \pm 1.64 \frac{4.2}{\sqrt{50}} = (30 - 0.97,\ 30 + 0.97) = (\mathbf{29.03},\ \mathbf{30.97})$$

Confidiamo al 90% che tutti i residenti della città in esame dedichino all'esercizio fisico in media tra 29.03 e 30.97 minuti al giorno.

QUIZ

Nell'esercizio 3, l'intervallo di confidenza del 95% risulta :

a. (15.2, 20.4)
b. (32.5, 40.2)
c. (28.8, 31.2)
d. (28.0, 31.0)

a. e b. si escludono perché non contengono la media campionaria 30.

d. si esclude perché non è simmetrico rispetto alla media campionaria 30.

Esercizio 5

In un campione casuale di 95 aziende , 67 hanno dichiarato di aver ottenuto la certificazione ISO negli ultimi cinque anni.
Determinare l'intervallo di confidenza a livello 99% per la proporzione di ditte che sono state certificate negli ultimi cinque anni.

$X_1, X_2, \dots, X_{95} \sim$ Bernoulli (p) i.i.d., $n = 95 > 30$

$$X_j = \begin{cases} 1 = \text{certificazione Si} & \text{con prob. p} \\ 0 = \text{certificazione No} & \text{con prob. } 1 - p \end{cases}$$

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

$$\mathbf{IC(99\%)} = \hat{p}_n \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

Bisogna calcolare:

- $\hat{p}_n$ proporzione campionaria
- $z_{1-\alpha/2}$ quantile dalla distribuzione Normale

Esercizio 5

In un campione casuale di 95 aziende , 67 hanno dichiarato di aver ottenuto la certificazione ISO negli ultimi cinque anni.
Determinare l'intervallo di confidenza, a livello 99%, per la proporzione di ditte che sono state certificate negli ultimi cinque anni.

Ampiezza campionaria

$$n = 95$$

Proporzione campionaria

$$\hat{p}_n = \frac{67}{95} = 0.7$$

Quantile dalla distribuzione normale

$$\alpha = 1 - 0.99 = 0.01$$

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

$n = 95$

$\bar{x}_n = 0.7$

$z_{1-\frac{\alpha}{2}} = 2.58$

$$\mathbf{IC(99\%)} = \hat{p}_n \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} =$$

$$= 0.7 \pm 2.58\sqrt{\frac{0.7(1-0.7)}{95}} =$$

$$= 0.7 \pm 2.58 \times 0.047 =$$

$$= (0.7 - 0.12, \quad 0.7 + 0.12) =$$

$$= (\mathbf{0.58}, \quad \mathbf{0.82})$$

Confidiamo al 99% che la vera percentuale di aziende che hanno ottenuto la certificazione ISO negli ultimi cinque anni sia tra il 58% e l'82%.

Esercizio 6

Un processo di produzione fabbrica un tipo particolare di mattone il cui peso è distribuito normalmente con deviazione standard 0.12 Kg.
Si estrae un campione casuale di 16 mattoni dalla produzione e si rileva un peso medio per mattone di 4.07 Kg.
Determinare un intervallo di confidenza a livello 95% per il peso medio di un mattone.

$X_1, X_2, \dots, X_7 \sim N(\mu, 0.122)$ i.i.d., σ^2 **nota**, $n = 16 \leq 30$

→ la media campionaria $\bar{X}_n$ ha distribuzione **Normale**

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

Bisogna calcolare:

- $\bar{x}_n$ media campionaria → $\bar{x}_n = 4.07$
- $z_{1-\alpha/2}$ quantile dalla distribuzione Normale → $\alpha = 1 - 0.95 = 0.05$

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

$n = 16$

$\bar{x}_n = 4.07$

$\sigma = 0.12$

$z_{1-\frac{\alpha}{2}} = 1.96$

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

$$= 4.07 \pm 1.96\frac{0.12}{\sqrt{16}} =$$

$$= 4.07 \pm 1.96 \times 0.03 =$$

$$= (4.07 - 0.59, \quad 4.07 + 0.59) =$$

$$= (\mathbf{3.48}, \quad \mathbf{4.66})$$

Confidiamo al 95% che il peso medio di un mattone in produzione sia tra 3.48 Kg e 4.66 Kg.

Esercizio 6

Un processo di produzione fabbrica un tipo particolare di mattone il cui peso è distribuito normalmente con deviazione standard 0.12 Kg.
Si estrae un campione casuale di 16 mattoni dalla produzione e si rileva un peso medio per mattone di 4.07 Kg.
Determinare un intervallo di confidenza a livello 95% per il peso medio di un mattone.

Come si risolverebbe il problema nel caso in cui non fosse specificato che il peso ha distribuzione Normale?

Si risolverebbe nello stesso modo ipotizzando che la popolazione ha distribuzione normale.
In questo caso l'ipotesi di normalità è plausibile in quanto le variazioni sul peso dei mattoni si possono attribuire ad errore casuale nella procedura di misurazione.

E se, oltre a non conoscere la distribuzione del peso, il campione fosse grande (ad esempio 100 mattoni anziché 16) ?

Si risolverebbe nello stesso modo in virtù del Teorema del Limite Centrale (ampiezza campionaria *n=100 > 30*).

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.**

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

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

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

da scrivere all' esame !

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

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

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

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.**

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.**

$X_1, X_2, \dots, X_{320} \sim \boldsymbol{Bernoulli}\ (p)$ i.i.d., $\boldsymbol{n = 320 > 30}$

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

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

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

Esercizio 8
La lunghezza dei pezzi prodotti da un macchinario ha distribuzione normale con media μ incognita.
Si estraggono 200 pezzi che risultano lunghi mediamente 52 mm con deviazione standard 0.35 mm.
Determinare l'intervallo di confidenza per μ a livello 92%.

$X_1, X_2, \dots, X_{320} \sim N(\mu, \sigma^2)$ i.i.d., σ^2 **incognita**, $n = 200 > 30$

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

$$\mathbf{IC(92\%)} = \bar{x}_n \pm t_{1-\alpha/2}^{(n-1)} \frac{\bar{s}_n}{\sqrt{n}} =$$

da scrivere all' esame !

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

$$= 52 \pm 1.75 \frac{0.35}{\sqrt{200}} = (51.96, \quad 52.04)$$

$\alpha = 1 - 0.92 = 0.08$ $\qquad z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.08}{2}} = z_{0.96} = 1.75$

VERIFICA DELLE IPOTESI

Verifica d'ipotesi: μ

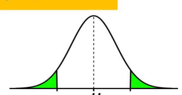
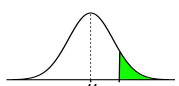
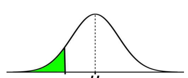
TLC

t-test

$(X_1, \dots, X_n)$ campione aleatorio $X_i \sim ???$

$n > 30$

σ^2 **non** nota ⇒ $s_n{}^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$

H_0	H_1		
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} \right\rvert > t(n-1)_{1-\alpha/2}$	
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} > t(n-1)_{1-\alpha}$	
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} < -t(n-1)_{1-\alpha}$	

Si rifiuta H_0 se $\bar{x}_n$ è lontano da μ_0 nella zona verde di rifiuto

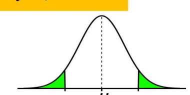
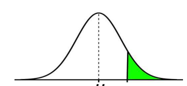
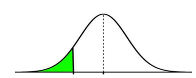
Verifica d'ipotesi: μ

t-test

$(X_1, \dots, X_n)$ campione aleatorio $N(\mu, \sigma^2)$

σ^2 **non** nota ⇒ $s_n{}^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$

$$\frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t(n-1)$$

H_0	H_1		
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right\rvert > z_{1-\alpha/2}$ $\left\lvert \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} \right\rvert > t(n-1)_{1-\alpha/2}$	
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}$ $\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} > t(n-1)_{1-\alpha}$	
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}$ $\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} < -t(n-1)_{1-\alpha}$	

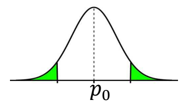
Si rifiuta H_0 se $\bar{x}_n$ è lontano da μ_0 nella zona verde di rifiuto

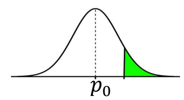
TLC

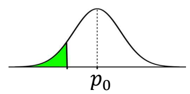
Verifica d'ipotesi: p

$(X_1, \dots, X_n)$ i.i.d $X_i \sim b(p)$, $\quad \boldsymbol{np_0 \geq 5}$ & $\boldsymbol{n(1-p_0) \geq 5}$

H_0	H_1	si rifiuta H_0 se:
$p = p_0$	$p \neq p_0$	$\left\lvert \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \right\rvert > z_{1-\frac{\alpha}{2}}$
$p = p_0$	$p > p_0$	$\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{1-\alpha}$
$p = p_0$	$p < p_0$	$\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} < -z_{1-\alpha}$







Si rifiuta H_0 se $\hat{p}_n$ è lontano da p_0 nella zona verde di rifiuto

Campione $X_1, X_2, \dots, X_n$	Statistica test	distribuzione Statistica test
distr. incognita (μ, σ^2) $n > 30$ σ^2 incognita	$\frac{\bar{X}_n - \mu_0}{\bar{s}_n/\sqrt{n}}$	* $\boldsymbol{t(n-1)}$
$\boldsymbol{b(p)}$ $n > 30$ $np > 5 \quad n(1-p) > 5$	$\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$	$\boldsymbol{N(0,1)}$
$\boldsymbol{N(\mu, \sigma^2)}$ σ^2 nota	$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$	$\boldsymbol{N(0,1)}$
$\boldsymbol{N(\mu, \sigma^2)}$ σ^2 incognita	$\frac{\bar{X}_n - \mu_0}{\bar{s}_n/\sqrt{n}}$	* $\boldsymbol{t(n-1)}$

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

Esercizio 9

Un impianto confeziona scatole di biscotti del peso di 250g con varianza 400. Durante un controllo periodico dell'impianto emerge che il peso medio di 16 confezioni di biscotti prese a campione è pari a 255g. Verificare se l'impianto funziona correttamente a livello 1%.

La distribuzione non è nota e il campione è piccolo.
Tuttavia è plausibile che il peso abbia distribuzione Normale:

Popolazione $\boldsymbol{N}\ (\mu, \boldsymbol{\sigma^2 = 400})$ i.i.d. σ^2 nota

Campione $n = 16 \quad \bar{x}_n = 255 \quad\quad \alpha = 0.01$

TEST $H_0: \mu = 250 \quad H_1: \mu \neq 250$ → Si rifiuta H_0 se $\left\lvert \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right\rvert > z_{1-\alpha/2}$

$$\left\lvert \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right\rvert = \left\lvert \frac{255 - 250}{\sqrt{400}/\sqrt{16}} \right\rvert = 1$$

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

Non è vero che $\left\lvert \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right\rvert > z_{1-\alpha/2}$

$1 < 2.58$

→ Non si rifiuta H_0 a livello 1%

→ Si può ritenere che l'impianto funzioni bene

Esercizio 10

Un partito politico ha ottenuto il 13% dei voti alle ultime elezioni.
Dopo qualche tempo il partito commissiona un sondaggio per verificare se la propria quota di elettori è aumentata. Su un campione casuale di 500 elettori, 86 manifestano l'intenzione di votare per quel partito.
Verificare con livello di significatività dell'1% se la quota di elettori del partito è effettivamente aumentata.

Popolazione $Bernoulli(p)$ i.i.d. $\quad p =$ proporzione di elettori del partito

Campione $n = 500 \quad \hat{p}_n = 86/500 = 0.17 \quad\quad \alpha = 0.01$

TEST $H_0: p = 0.13 \quad H_1: p > 0.13$ → Si rifiuta H_0 se $\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{1-\alpha}$

$$\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.17 - 0.13}{\sqrt{0.13(1-0.13)/500}} = 2.66$$

$$z_{1-\alpha} = z_{1-0.01} = z_{0.99} = 2.33$$

È vero che $\frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{1-\alpha}$

$2.66 > 2.33$

Si rifiuta H_0 a livello 1% → La quota di elettori si può ritenere aumentata

Esercizio 11

Verificare la seguente ipotesi per la media di una popolazione Normale

$$H_0: \mu = 100 \qquad H_1: \mu < 100$$

sulla base di un campione casuale di ampiezza $n = 25$ che fornisce le seguenti statistiche campionarie $\bar{x}_n = 95$, $\bar{s}_n = 10$, ed una probabilità di errore di primo tipo $\alpha = 0.05$

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

Campione $n = 25$ $\quad \bar{x}_n = 95$ $\quad \bar{s}_n = 10$ $\quad \alpha = 0.05$

TEST $H_0: \mu = 100$ $\quad H_1: \mu < 100$ → Si rifiuta H_0 se $\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}} < -\boldsymbol{t}_{\boldsymbol{1-\alpha}}^{(\boldsymbol{n-1})}$

$$\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}} = \frac{95 - 100}{10/\sqrt{25}} = -2.5$$

$$\boldsymbol{t}_{\boldsymbol{1-\alpha}}^{(\boldsymbol{n-1})} = \boldsymbol{t}_{\boldsymbol{1-0.05}}^{(\boldsymbol{25-1})} = \boldsymbol{t}_{\boldsymbol{0.95}}^{(\boldsymbol{24})} = 1.71088$$

È vero che $\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}} < -t_{1-\alpha}^{(n-1)}$

$-2.5 < -1.71088$

Si rifiuta H_0 a livello 5%

Esercizio 11 bis

Verificare la seguente ipotesi per la media di una popolazione Normale

$$H_0: \mu = 100 \qquad H_1: \mu < 100$$

sulla base di un campione casuale di ampiezza $n = 25$ che fornisce le seguenti statistiche campionarie $\bar{x}_n = \mathbf{106}$, $\bar{s}_n = 10$, ed una probabilità di errore di primo tipo $\alpha = 0.05$

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

Campione $n = 25$ $\quad \bar{x}_n = \mathbf{106}$ $\quad \bar{s}_n = 10$

TEST $H_0: \mu = 100$ $\quad H_1: \mu < 100$

Si rifiuta H_0 quando la media campionaria è molto più piccola di 100.

In questo caso la media campionaria 106 è più grande di 100,

quindi **sicuramente non si rifiuta $\boldsymbol{H_0}$** a livello 5%

Esercizio 12

Da un campione casuale di 150 bottiglie di birra di una certa marca risulta un contenuto medio di 0.99 litri con deviazione standard 0.05.
Verificare, con un livello di significatività 0.01, se la quantità di birra contenuta è 1 litro come indicato sull'etichetta.

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

Campione $\boldsymbol{n = 150}$ $\quad \bar{x}_n = 0.99$ $\quad \bar{s}_n = 0.05$ $\quad \alpha = 0.01$

TEST $H_0: \mu = 1$ $\quad H_1: \mu \neq 1$ → Si rifiuta H_0 se $\left|\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}}\right| > \boldsymbol{t}_{\boldsymbol{1-\alpha/2}}^{(\boldsymbol{n-1})}$

→ Si rifiuta H_0 se $\left|\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}}\right| > \boldsymbol{z}_{\boldsymbol{1-\alpha/2}}$

$$\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}} = \frac{0.99 - 1}{0.05/\sqrt{150}} = -2.45$$

$$\boldsymbol{z}_{\boldsymbol{1-\alpha/2}} = \boldsymbol{z}_{\boldsymbol{1-0.01/2}} = \boldsymbol{z}_{\boldsymbol{0.995}} = 2.58$$

Non è vero che $\left|\frac{\bar{x}_n - \mu_0}{\bar{s}_n/\sqrt{n}}\right| > \boldsymbol{z}_{\boldsymbol{1-\alpha/2}}$

$2.45 < 2.58$

Non si rifiuta H_0 a livello 1%

L'etichetta sulla bottiglia corrisponde al contenuto