Metodi statistici per le ricerche di mercato

Prof.ssa Isabella Mingo
A.A. 2016-2017

SAPIENZA Università di Roma

Facoltà di Scienze Politiche, Sociologia, Comunicazione

Corso di laurea Magistrale in «Organizzazione e marketing per la comunicazione d'impresa»

Come si determina l'ampiezza del campione

$$p - \boxed{z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}} \le P \le p + z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}$$

Nella stima ad intervallo, per popolazione non nota o infinita, l'errore che dobbiamo aggiungere e togliere alla statistica del campione per stimare il corrispondente parametro della popolazione è:

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

A partire da questa formula possiamo ricavare il valore di n

$$nE^2 = z^2{}_{\alpha/2}\, PQ$$

$$n = \frac{z^2{}_{\alpha/2}\ PQ}{E^2}$$

Se conosciamo il valore di N

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$$

A partire da questa relazione possiamo ricavare il valore di n:

$$nE^2(N-1) = z^2{}_{\propto/2}\ PQ(N-n)$$

$$nE^2(N-1) = z^2{}_{\alpha/2}PQN - z^2{}_{\alpha/2}PQn$$

$$n\left[E^2(N-1) + Z^2{}_{\alpha/2}PQ\right] = z^2{}_{\alpha/2}PQN$$

$$n = \frac{z^2{}_{\alpha/2}PQN}{E^2(N-1) + z^2{}_{\alpha/2}PQ}$$

Indicatore del grado di eterogeneità della popolazione

N. Unità della Popolazione

N. Unità del campione

$$n = \frac{z^2{}_{\alpha/2}PQN}{E^2(N-1) + z^2{}_{\alpha/2}PQ}$$

Errore della stima o precisione desiderata

Valori di Zα/2 in corrispondenza dei livelli di confidenza

1-α	sign. α	Zα/2
0,9500	0,0500	1,96
0,9545	0,0455	2,00
0,9900	0,0100	2,58
0,9973	0,0027	3,00

L'ampiezza del campione

- La dimensione del campione casuale semplice che consente di stimare un valore (parametro) percentuale o medio della popolazione ad una precisione voluta e con una data probabilità, dipende da:
  - La dimensione della popolazione (finita o infinita) e se finita dal numero **N** delle unità che la compongono.
  - Il margine di errore **e** che si vuole accettare nei risultati
    - (il valore della popolazione sarà compreso tra il valore del campione + o – l'errore)
  - Il livello di confidenza o il grado di fiducia da attribuire alla nostra stima (valore **z**).
  - La eterogeneità della popolazione .
- Si possono utilizzare apposite tavole di campionamento.

Esercizio

- Si vuole stimare la percentuale di utenti che sono d'accordo con l'apertura dei punti vendita fino alle 24.
  - Calcolare la dimensione ottima del campione casuale semplice sapendo che:
    - la popolazione di riferimento è di 4000 unità
    - ipotizzando la situazione di massima eterogeneità tra gli utenti.
    - scegliendo un errore del 5% dalla stima
    - ammettendo un livello di fiducia del 95,4%.

$$n = \frac{Z^2{}_{\alpha/2} \ PQN}{E^2(N-1) + Z^2{}_{\alpha/2} \ PQ}$$

$$n = \frac{4 \ (0{,}5)(0{,}5)4000}{0{,}05^2(3999)+4(0{,}5)(0{,}5)} = 364$$

Tavole di campionamento

Numerosità del campione e della popolazione di provenienza per stimare una percentuale

	1%	2%	3%	4%	5%
1500	1305	938	638	441	316
2000	1667	1112	714	476	333
3000	2308	1364	811	517	353
4000	2858	1539	870	541	364
5000	3334	1667	909	556	370
7000	4118	1843	959	574	378
10000	5001	2001	1000	588	385
15000	6001	2143	1034	600	390
25000	7143	2273	1064	610	394
50000	8334	2381	1087	617	397
100000	9091	2439	1099	621	398
∞	10000	2500	1111	625	400

Ipotesi: z=2 (95,4%); p=q=0,5

Dimensione campione: uso delle tavole

Determinare la dimensione del campione per stimare la percentuale di clienti che si ritengono soddisfatti del servizio di Assistenza post-vendita.
Sapendo che:
-la popolazione di riferimento è di 2000 unità
-ipotizzando la situazione di massima eterogeneità in cui i cittadini sono per metà soddisfatti e per metà insoddisfatti (p=q=0,5).
-ammettendo un errore del 2% dalla stima.
La dimensione ottima del campione è 1112 unità.

	1%	2%	3%
1500	1305	938	638
2000	1667	1112	714
3000	2308	1364	811
4000	2858	1539	870
5000	3334	1667	909
7000	4118	1843	959
10000	5001	2001	1000
15000	6001	2143	1034
25000	7143	2273	1064
50000	8334	2381	1087
100000	9091	2439	1099
¥	10000	2500	1111

Ipotesi: z=2 (95,4%); p=q=0,5

Tavola con p=0,20

Tab. 11 - Numerosità di un campione estratto da una popolazione finita per due livelli di confidenza e per assegnati valori di errore ammesso: stima *preventiva* di $P = 0{,}20$

Popolazione (N)	Margine di errore						Margine di errore					
	1%	2%	3%	4%	5%	10%	1%	2%	3%	4%	5%	10%
500	462	377	289	217	165	55	478	421	352	286	230	88
1.000	860	606	406	278	198	58	914	727	542	400	299	96
2.000	1.509	869	509	322	219	60	1.684	1.142	744	500	351	101
3.000	2.016	1.016	556	341	227	60	2.341	1.411	849	545	373	103
4.000	2.423	1.110	583	351	232	61	2.908	1.599	913	571	385	104
5.000	2.757	1.176	601	357	234	61	3.403	1.738	957	588	393	104
7.000	3.273	1.260	622	364	238	61	4.224	1.929	1.012	608	402	105
10.000	3.807	1.332	639	370	240	61	5.158	2.103	1.058	624	409	105
15.000	4.360	1.394	653	375	242	61	6.228	2.261	1.097	637	414	106
25.000	4.934	1.448	665	378	243	61	7.469	2.406	1.130	648	419	106
50.000	5.474	1.491	674	381	245	61	8.780	2.528	1.156	657	422	106
100.000	5.791	1.513	678	383	245	61	9.625	2.594	1.170	661	424	106
200.000	5.963	1.525	681	383	246	61	10.112	2.628	1.176	663	425	106
500.000	6.072	1.532	682	384	246	61	10.428	2.648	1.181	665	426	106
1.000.000	6.109	1.534	682	384	246	61	10.538	2.655	1.182	665	426	106
	Livello di confidenza del 95%						Livello di confidenza del 99%					

Dimensione campione

Determinare la dimensione del campione dell'esercizio precedente ipotizzando la situazione in cui i cittadini sono soddisfatti nel 20% dei casi.

Delineare due scenari:

1. ammettendo un errore del 2% dalla stima e un livello di confidenza del 95%;
2. ammettendo un errore del 5% e un livello di confidenza del 99%;

In quale scenario la dimensione del campione è maggiore?

Come si determina l'ampiezza del campione per stimare la media

Anche nel caso della stima del valore medio, e a partire dall'errore, possiamo ricavare il valore di n.

Nel caso di popolazioni di dimensioni non noto o infinite:

$$E = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \qquad n = \frac{z^2{}_{\alpha/2}\ \sigma^2}{E^2}$$

Nel caso di popolazioni di dimensioni note:

$$E = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\sqrt{\frac{N-n}{N-1}} \qquad n = \frac{z^2{}_{\alpha/2}\ \sigma^2 N}{E^2(N-1) + z^2{}_{\alpha/2}\ \sigma^2}$$

Varianza della Popolazione

$$n = \frac{Z^2{}_{\alpha/2}\ \sigma^2 N}{E^2(N-1)+Z^2{}_{\alpha/2}\ \sigma^2}$$

La varianza della popolazione può essere ottenuta:

- sulla base di conoscenze pregresse relative a precedenti indagini
- sulla base di una stima preliminare compiuta mediante la varianza del campione di con un sondaggio pilota
- Utilizzando il valore massimo assumibile dalla varianza:

$$\max(\sigma^2) = \frac{CV^2}{4}$$

- Dove CV è il campo di variazione del carattere ossia la differenza tra valore massimo e minimo.
- La numerosità ottenuta con questa stima è però molto pessimistica (l'ampiezza campionaria è molto alta) pertanto alcuni autori consigliano di usare :
- stima$(\sigma^2) = \frac{CV^2}{16}$

Esercizio

- Si vuole stimare l'ascolto medio televisivo degli abitanti di una città
  - Calcolare la dimensione ottima del campione casuale semplice sapendo che:
    - La popolazione di riferimento è di 20000 abitanti
    - scegliendo un errore di 0,3 ore dalla stima
    - ammettendo un livello di fiducia del 95%
    - Sapendo che la variabile assume valori compresi tra 0 e 12 ore.

$$n = \frac{z^2{}_{\alpha/2} \quad \sigma^2 N}{E^2(N-1) + z^2{}_{\alpha/2} \ \sigma^2}$$

$$n = \frac{1{,}96^2[(12)^2/16\,]\,20000}{0{,}3^2(19999)+1{,}96^2[(12)^2/16]} = 377$$

Ricalcolare la dimensione del campione utilizzando i risultati di una indagine Auditel dalla quale risulta che in quella città la varianza della visione giornaliera Tv è pari a 3,4.

Dalla dimensione campione alla precisione della stima

A partire dai criteri che utilizziamo nel determinare l'ampiezza del campione, possiamo, dopo aver trovato la statistica di interesse nel nostro campione, ricalcolare l'errore campionario e stimare il parametro della popolazione.

In particolare, riprendendo l'esempio precedente dei clienti del servizio di Assistenza post-vendita:

- Supponendo che la percentuale di soddisfatti del nostro campione di 1112 utenti, non è del 50% come avevamo ipotizzato nel calcolo della dimensione campionaria, ma è solo del 40% ,
- con una probabilità del 95,45% (livello di fiducia)
- potremmo ricalcolare il valore dell'errore E

$$\boldsymbol{E} = 2\sqrt{\frac{040 \cdot 0{,}60}{1112}} = 0{,}03$$

- la percentuale di soddisfatti nella popolazione di riferimento sarà compresa tra 37% e 43% (40%± 3% errore ammesso)

Esercizio: da n a E

- Su un campione di 500 intervistati la proporzione di soggetti soddisfatti della qualità di un nuovo cioccolatino è pari al 30%.
- Si determini, con un livello di fiducia del 95%, il margine di errore ammesso.
- Si determini la stima ad intervallo della proporzione nella popolazione.

$E = z_{\alpha/2}\sqrt{\frac{pq}{n}}$ $\quad E = 1{,}96\sqrt{\frac{0{,}30 \cdot 0{,}70}{500}} = 0{,}04$

$$0{,}30 - 0{,}04 \leq P \leq 0{,}30 + 0{,}04$$
$$0{,}26 \leq P \leq 0{,}34$$

Con un livello di fiducia del 95% la proporzione sarà compresa tra 26% e 34%

Esercizio: da n a E

- Su un campione di 380 intervistati la spesa annua per l'acquisto di abbonamenti a teatro è di 120 euro, con uno scarto quadratico medio di 9,75 euro.
- Si determini, con un livello di fiducia del 95,45%, il margine di errore ammesso.
- Si determini la stima ad intervallo della spesa media nella popolazione.

$E = z_{\alpha/2}\frac{s}{\sqrt{n-1}}$ $\quad E = 2\frac{9{,}75}{\sqrt{379}} = 1{,}00$ euro

$$120 - 1{,}00 \leq \mu \leq 120 + 1{,}00$$
$$119 \leq \mu \leq 121$$

Con un livello di fiducia del 95,45% la spesa media nella popolazione sarà compresa tra119 e 121 euro.

Decisioni in condizioni di incertezza

La statistica inferenziale utilizza opportuni metodi per individuare le caratteristiche della popolazione statistica di riferimento sulla base di informazioni ricavate da un campione estratto da tale popolazione secondo criteri probabilistici.

Le più frequenti applicazioni sono finalizzate ad ottenere:

- La stima puntuale e stima intervallare dei parametri della popolazione, che abbiamo illustrato precedentemente;

- La verifica delle ipotesi ossia una procedura finalizzata appunto ad accettare o rifiutare una ipotesi statistica:

  - Su un parametro della popolazione (media, proporzione, ecc.)

  - Sulla differenza tra due statistiche campionarie

Che cosa è una ipotesi statistica?

Nell'ambito di una indagine, l'ipotesi statistica è una supposizione sul valore di un parametro nella popolazione di interesse.

A- La spesa media dei clienti dell'azienda X per il prodotto A è di 70 euro l'anno:
$\mu = 70$
L'ipotesi sarà vera se l'uguaglianza sarà soddisfatta.

B- Si ipotizza che la percentuale di clienti insoddisfatti sia inferiore o uguali al 25%
P < 0,25
L'ipotesi sarà vera se P assumerà valori inferiori o uguali al 25%

C- Si ipotizza che le ore medie di visione Tv sia uguale nel campione di adulti del Nord e in quello di adulti del Sud
$\mu_N = \mu_S$
L'ipotesi sarà vera se le ore medie di visone saranno uguali nei due campioni.

La verifica delle ipotesi

La verifica di un'ipotesi statistica consiste nello stabilire se con una data probabilità le statistiche riscontrate in un dato campione casuale rappresentino corrispondenti caratteristiche della popolazione di riferimento, o se invece siano il frutto di fluttuazioni casuali imputabili al dato campione.
Ad esempio:
A-Si estraggono a caso 100 clienti dell'azienda X e si rileva che la loro spesa per il prodotto A è di 70,5.
Ci si chiede se la media ottenuta in questo campione sia uguale a quella ipotizzata precedentemente (70 euro) e la differenza sia solo frutto del caso, oppure se sia significativamente diversa.

B- Si estraggono a caso 150 clienti e si rileva che la percentuale di insoddisfatti è del 26%.
Ci si chiede se questo risultato ottenuto in questo campione sia solo dovuto alle oscillazioni del caso oppure se sia significativamente diverso da quella ipotizzato precedentemente (25%).

C- Si rilevano su due campioni di adulti del Nord e del Sud le ore di visione Tv. Nel primo la media è di 3,8, nel secondo di 4,0.
Si tratta di una differenza trascurabile perché dovuta al caso o significativa?

Fasi della verifica delle ipotesi

1. Formulazione dell'ipotesi nulla e della ipotesi alternativa
2. Distribuzione campionaria
3. Livello di fiducia e di significatività: individuazione del valore critico
4. Calcolo del valore della statistica test e verifica delle ipotesi

L'ipotesi nulla e l'ipotesi alternativa

- Si parte dall'ipotesi che i valori messi a confronto siano uguali e che l'eventuale differenza sia da attribuire al caso.
- ✓ Questa ipotesi ,che si sottopone a verifica, si definisce Ipotesi nulla e si indica con H_0

- Secondo l'ipotesi contraria la differenza è da considerare significativa, non dovuta al caso.
- ✓ Questa ipotesi si definisce Ipotesi alternativa e si indica con H_1

- Nei tre esempi precedenti:

A. $H_{0:}\ \mu = \mu_0$ $\quad H_1 : \mu \neq \mu_0$

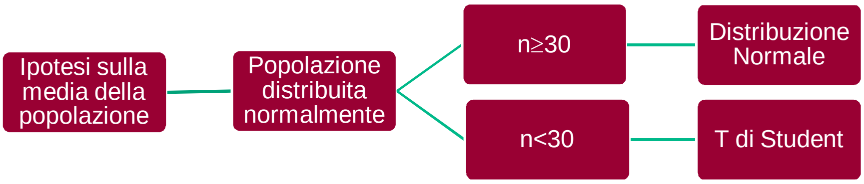
B. $H_0 : P = \mathrm{p}$ $\quad H_1 : P \neq \mathrm{p}$

C. $H_0 : \bar{X}_N = \bar{X}_s$ $\quad H_1 : \bar{X}_N \neq \bar{X}_s$

La distribuzione campionaria

- Per stabilire se le differenze sono dovuti al caso o significative si utilizzano appositi test statistici che consentono di accettare o rifiutare l'ipotesi nulla.
- I test si basano sull'uso della distribuzioni campionaria della statistica su cui verte l'ipotesi.
- Le distribuzioni campionarie (es. Distribuzione normale, T di Student) costituiscono modelli di riferimento: le caratteristiche di ciascuna distribuzione campionaria ci consentono di conoscere, dati certi requisiti, la probabilità associata ai possibili valori che quella data statistica può assumere.

- La scelta di una particolare distribuzione campionaria dipende soprattutto dal tipo di dati e dalla numerosità del campione.

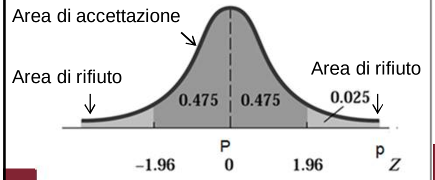


La scelta del livello di fiducia

- La procedura di controllo dell’ipotesi si basa sulla soglia di rischio di errore che si è disposti ad accettare nel pervenire ad una decisione cioè al livello di fiducia (1- α).

- Si riterrà come non accettabile un risultato campionario che cade all’esterno dei limiti corrispondenti a questo livello e dunque si rifiuterà l’ipotesi sottoposta a verifica, poiché il campione dà un risultato tanto raro da potere essere frutto del puro caso.

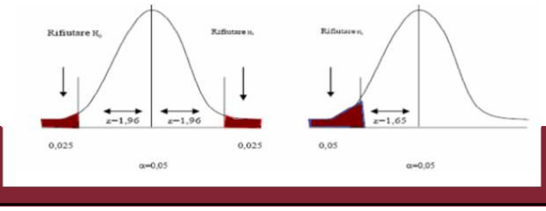
(1- α)=0,95



Test bilaterale o unilaterale

Test a due code: si individuano due valori critici che lasciano rispettivamente a sinistra e a destra una probabilità paria a $\alpha/2$: se la statistica test è all'interno di questa regione accetto H_0.
Si usa quando si cerca di dimostrare una differenza, questa può verificarsi **in entrambe le direzioni** (il test è bilaterale)

Test a una coda: si individua un valore critico che lascia in una sola coda una probabilità pari a α.
In alcuni casi è ragionevole pensare che la differenza possa verificarsi in una sola direzione (soddisfazione per un prodotto nuovo maggiore del vecchio) mentre una differenza nella direzione opposta deve essere dovuta al caso: si utilizza un test (test unilaterale)



Esercizio : test bilaterale- media

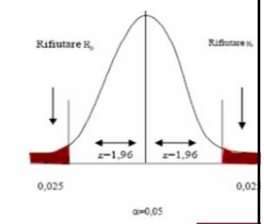
Da ricerche di mercato dello scorso anno risulta che l'età media del target di un videogame è di 28 anni, con uno scarto quadratico medio di 2,6.
Da un campione di 150 giocatori risulta invece una età media di 26 anni.
Possiamo confermare che l'età media del target sia di 28 anni o il valore trovato nel campione è significativamente diverso da quello ipotizzato?

1) $H_0: \mu = 28$ (il target ha una età media di 28 anni)
$H_1: \mu \neq 28$ (il target non ha una età media di 28 anni)

2) Distribuzione probabilistica: distribuzione normale
Test a due code (si è interessati a verificare la differenza in entrambe le direzioni)
3) Livello di confidenza $(1-\alpha)$=95%; livello di significatività : $\alpha/2$=0,025.
Valore di Z critico in corrispondenza del livello di significatività prescelto.
Z= ± 1.96

4) Calcolo del valore della statistica test (Z empirico)
$Z = \frac{26-28}{\frac{2,6}{\sqrt{150}}}$ =-9,42

Il valore Z del nostro campione è minore di -1,96 , dunque dobbiamo rifiutare l'ipotesi nulla: l'età media della popolazione da cui è estratto il campione non coincide con quella ipotizzata.
La probabilità di commettere un errore rifiutando l'ipotesi è del 5%.



Esercizio: test unilaterale - media

Se invece di un test a due code, avessimo effettuato un test a una coda, essendo interessati a verificare la differenza solo in una direzione , avremmo proceduto nel modo seguente:

1) $H_{0:}\ \mu$ = 28 (il target ha una età media di 28 anni)
$H_1 : \mu$ < 28 (il target ha una età media inferiore a 28 anni)

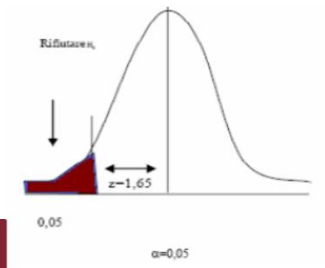
2) Distribuzione probabilistica: distribuzione normale
Test a una coda (si è interessati a verificare se la media è inferiore a quella ipotizzatala)

3) Livello di confidenza (1- α)=95%; livello di significatività : α=0,05.
Valore di Z critico in corrispondenza del livello di significatività prescelto.
Z= - 1.65

4) Calcolo del valore della statistica test (Z empirico)

$$Z = \frac{26-28}{\frac{2,6}{\sqrt{150}}} = -9,42$$

Il valore Z del nostro campione è minore di -1,65 , dunque dobbiamo rifiutare l'ipotesi nulla: l'età media della popolazione da cui è estratto il campione è minore dei 28.
La probabilità di commettere un errore rifiutando l'ipotesi è del 5%.



Esercizio

Si ritiene che in una data città il valore medio dei consumi di un prodotto X per abitante sia di 130 euro al mese con uno scarto quadratico medio di 7,8 euro. Da una indagine campionaria su un campione di n=100 individui risulta invece che il consumo per abitante è di 131,6 euro. Scegliendo un livello di significatività dell'1%, possiamo confermare che il consumo medio della popolazione sia di 130 euro o il valore trovato nel campione è significativamente maggiore di quello ipotizzato?

Esercizio

Si ritiene che in una data città il valore medio dei consumi di un prodotto X per abitante sia di 130 euro al mese con uno scarto quadratico medio di 7,8 euro. Da una indagine campionaria su un campione di n=100 individui risulta invece che il consumo per abitante è di 131,6 euro. Scegliendo un livello di significatività dell'1%, possiamo confermare che il consumo medio della popolazione sia di 130 euro o il valore trovato nel campione è significativamente maggiore di quello ipotizzato?

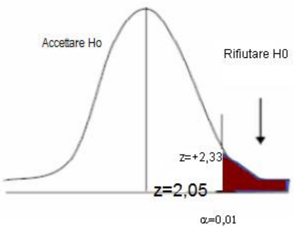
1) $H_0: \mu = 130$ $\qquad H_1: \mu > 130$

2) Distribuzione probabilistica: distribuzione normale
Test a una coda (si è interessati a verificare se la media è superiore a quella ipotizzatala)
3) Livello di confidenza (1- α)=99%; livello di significatività : α=0,01.
Valore di Z critico in corrispondenza del livello di fiducia prescelto.
Z= + 2,33

4) Calcolo del valore della statistica test (Z empirico)

$$Z = \frac{131{,}6-130}{\frac{7{,}8}{\sqrt{100}}} = 2{,}05$$

Il valore Z del nostro campione è minore di 2,33, dunque dobbiamo accettare l'ipotesi nulla: il consumo medio della popolazione da cui è estratto il campione non è maggiore di 130 euro.



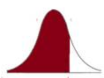


TAVOLA A

Tavole della funzione di ripartizione della variabile Normale Standardizzata:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

	Seconda cifra decimale di z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Esercizio

Utilizzando la tavola della distribuzione normale stabilire i valori di z per un test unilaterale in corrispondenza dei seguenti l7velli di fiducia:

0,95
0,9545
0,99
0,9973

Esercizio: uso del test t

Da una indagine nazionale risulta che il consumo medio di sigarette tra gli uomini di età compresa tra 18 e 35 anni sia di 9,5 al giorno. Da una ricerca effettuata su 28 uomini di quella fascia di età risulta che le sigarette consumate sono 10,04, con uno scarto quadratico medio di 3,5.
Scegliendo un livello di significatività del 5%, ed utilizzando un test bilaterale, possiamo affermare che il valore del campione è significativamente diverso da quello dell'indagine nazionale?

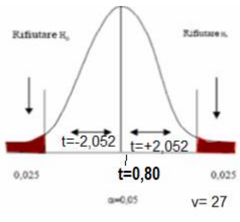
1) $H_{0:}$ μ = 9,5 H_1 : μ ≠ 9,5

2) Distribuzione probabilistica: distribuzione t di Student (n<30)
Test a due code(si è interessati a verificare se il consumo medio è diverso da quello ipotizzato)
3) Livello di confidenza (1- α)=95%; livello di significatività : α=0,05.
4) Gradi di libertà v=28-1=27
Valore di t critico in corrispondenza del livello di significatività prescelto e di v
t= ± 2,052

4) Calcolo del valore della statistica test (t empirico)

$$t = \frac{10,04-9,5}{\frac{3,5}{\sqrt{27}}} = 0,80$$

Il valore t del nostro campione cade nell'area di accettazione (è compreso tra -2,05 e + 2,05) pertanto accettiamo l'ipotesi nulla: il consumo medio di sigarette del nostro campione non è significativamente diverso di quello medio nazionale.



Esercizio: uso del test t

Se le sigarette consumate dai soggetti del nostro campione fossero state 12, con uno scarto quadratico medio di 3,5, scegliendo un livello di significatività del 5%, ed utilizzando un test ad una coda, avremmo potuto affermare che il valore del campione è significativamente maggiore di quello dell'indagine nazionale?

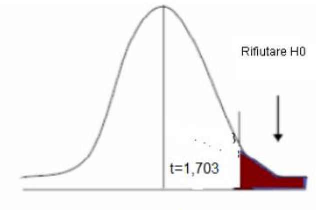
1) H_0: μ = 9,5 H_1 : μ > 9,5

2) Distribuzione probabilistica: distribuzione t di Student (n<30)
Test a una coda (si è interessati a verificare se il consumo medio è maggiore di quello ipotizzato)
3) Livello di confidenza (1- α)=95%; livello di significatività : α=0,05 (una coda v. tavole slide seguente)
4) Gradi di libertà v=28-1=27
Valore di t critico in corrispondenza del livello di significatività prescelto e di v
t= + 1,703

4) Calcolo del valore della statistica test (t empirico)

$$t = \frac{12-9,5}{\frac{3,5}{\sqrt{27}}} = 3,71$$

Il valore t del nostro campione cade nell'area di rifiuto (è maggiore di 1,703) pertanto rifiutiamo l'ipotesi nulla: il consumo medio di sigarette del nostro campione è significativamente diverso di quello medio nazionale.



α=0,05

Tavole T Student
Test a una coda

GL	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
infinity	1.282	1.645	1.960	2.326	2.576	3.091	3.291