INTRODUZIONE ALL'UTILIZZO DI

EXCEL

PER LA STATISTICA

morale@mat.unimi.it

Cosa è Excel ?

Excel è un'applicazione di foglio elettronico che permette di raccogliere ed elaborare i dati inseriti dall'utente.

I dati vengono raccolti in tabelle.

Tabella	Insieme di celle disposte secondo righe e colonne che costituiscono i fogli di lavoro
Cartelle di lavoro	Insieme di fogli di lavoro raccolti insieme come una rubrica telefonica e identificati da una etichetta

Il foglio di lavoro

Dr. Daniela Morale

Barra degli strumenti standard

Barra degli strumenti di formattazione

Barra della formula

Barra di stato

Schede

Intestazioni righe e colonne

Selezionare le celle

Una cella : cliccare

Un intervallo di celle contigue: cliccare e trascinare

Piu' celle non contigue: cliccare sulla prima, premere CTRL, cliccare sulle altre

Una riga (una colonna): cliccare sulla intestazione della riga (colonna)

Tutto il foglio di lavoro: cliccare il pulsante SELEZIONA TUTTO all'incrocio delle intestazioni di riga e colonna

Inserimento dati

Cliccare su una cella ed inserire con la tastiera. Se i dati immessi sono numeri (magari con la virgola) allora vengono interpretati come dati numerici, altrimenti sono interpretati com testo.

Il dato immesso compare sia nella cella sia nella barra della formula.
Si da conferma sia con INVIO sia con V.
Si cancella sia premendo ANNULLA dal menu MODIFICA, sia con X.

Trovare un dato

Dal menù MODIFICA cliccare su TROVA.

Salvaggio

Dal menù FILE cliccare SALVA o SALVA CON NOME

Dr. Daniela Morale

Gestione dei dati nei fogli di lavoro

Inserimento (cancellare) di righe / colonne

Inserimento (cancellare) celle

utilizzo copia/incolla

incolla speciale: da utilizzare se non si vogliono copiare i dati con tutte le formattazioni

Ordinamento dati

Dal menu DATI cliccare ORDINA

Inserimento funzioni

Cliccare sulla cella ed inserire un =. A questo punto si puo' scrivere una formula oppure si puo' utilizzare una formula predefinita, cliccando su INSERISCI e poi FUNZIONE oppure cliccare su F_x dalla barra degli strumenti. Vi è una lunga serie di funzioni statistiche.

Sintassi delle funzioni

OPERATORI ARITMETICI:

+ ADDIZIONE	/ DIVISIONE
- SOTTRAZIONE	% PERCENTUALE
* MOLTIPLICAZIONE	^ ELEVAMENTO A POTENZA

OPERATORI DI CONFRONTO:

= UGUALE	> MAGGIORE DI
> MINORE DI	>= MAGGIORE 0 UGUALE DI
>= MINORE O UGUALE DI	<> DIVERSO DA

OPERATORI DO CONCATENAZIONE DI TESTO:

& UNISCE O CONCATENA UNA O PIU' STRINGHE DI TESTO GENERANDO UNA SINGOLA STRINGA

OPERATORI DI RIFERIMENTO:

: OPERATORE DI INTERVALLO genera un riferimento a tutte le celle comprese tra due riferimenti, inclusi i due riferimenti stessi

. OPERATORE DI UNIONE, unisce più riferimenti generando un unico riferimento

Utilizzo dei riferimenti di cella

I riferimenti di celle inseriti in una formula possono essere espressi in tre distinti modalita':

RIFERIMENTO RELATIVO: indica al programma una cella e verra' modificato automaticamente quando la formula viene copiata in una posizione diversa da quella di creazione

RIFERIMENTO ASSOLUTO: indica al programma di utilizzare sempre la stessa cella a prescindere da dove verrà spostata la formula, si identifica la cella di riferimento assoluto con la seguente scrittura A1 per indicare la cella di colonna A riga 1

RIFERIMENTO MISTO: indica al programma un riferimento assolto solo per riga o solo per colonna con la seguente scrittura $A1 oppure A$1.

Esercizio n.1: FUNZIONI MEDIA E VARIANZA

Inserire 10 valori numerici nelle caselle comprese tra G11 e G20, far comparire la media di questi valori nella casella B8, e la loro varianza nella casella B9 usando le corrispondenti funzioni del foglio elettronico.
Inserire gli stessi dati nelle celle B13 e B14 rispettivamente usando solo le operazioni algebriche come nelle definizioni di media e varianza

Valor medio dei dati =

Varianza dei dati =

Valor medio dei dati =

Varianza dei dati =

Media	$\bar{x} = 1/n \ \Sigma \ x_i$
Varianza	$\sigma^2 = 1/n \ \Sigma_i \ (x_i - \bar{x})^2$

In EXCEL

FUNZIONI = f(x)

Funzioni statistiche:

MEDIA(num1, num2, ...) restituisce la media aritmetica degli argomenti

(numeri, riferimenti contenenti numeri)

VAR(num1, num2, ...) restituisce la varianza degli argomenti

(numeri, riferimenti contenenti numeri)

Esercizio n.2

inserimento, copia, ordinamento, calcolo di funzioni

Inserire 10 valori numerici nelle caselle comprese tra A11 e A20, e poi copiarli ordinandoli nel verso decrescente nelle caselle dalla C11 alla C20

Nella casella F11 far comparire il massimo tra i valori inseriti e alla casella G11 il minimo

Dati		Dati ordinati			Massimo	Minimo

Attenzione, una volta ordinati i valori, è facile "copiare"
il massimo e il minimo in una nuova casella.
Esistono però delle funzioni che individuano il massimo e
il minimo. Provare a utilizzarle!

Ordinamento di dati

In EXCEL

Tasto ordinamento

Funzioni

MAX(num1, num2, ...)

MIN(num1, num2, ...)

Excel
Statistica Descrittiva

ð Frequenze e funzione FREQUENZA di EXCEL

ð rappresentazione grafica dei dati : caso continuo

ð istogrammi e aereogrammi

ð utilizzo dello Strumento ANALISI DEI DATI

Tabella di distribuzione delle frequenze: CONTROLLI 1/2

N = numero di osservazioni

- FREQUENZA ASSOLUTA n_i
- La frequenza assoluta e' un numero intero compreso tra 0 e il numero totale di osservazioni
- La somma delle frequenze assolute da' il numero totale di osservazioni $\boxed{\sum_i n_i = N}$
- FREQUENZA RELATIVA $f_i = n_i / N$
- La frequenza relativa e' un numero intero compreso tra 0 e 1
- La somma delle frequenze relative da' SEMPRE 1 $\boxed{\sum_i f_i = \sum_i \frac{n_i}{N} = 1}$

Tabella di distribuzione delle frequenze: CONTROLLI 2/2

- FREQUENZA CUMULATIVA ASSOLUTA N_i
- La frequenza cumulativa assoluta e' un numero intero crescente da 0 al il numero totale di osservazioni
- ogni frequenza cum. ass. e' la somma della frequenza assoluta + la frequenza cumulativa assoluta del dato precedente

$$\boxed{N_i = N_{i-1} + n_i = \sum_{k=0}^{i} n_k}$$

- FREQUENZA CUMULATIVA RELATIVA F_i
- La frequenza cumulativa relativa e' un numero intero crescente da 0 a 1
- ogni frequenza cum. relativa e' la somma della frequenza relativa + la frequenza cumulativa relativa del dato precedente

$$\boxed{F_i = F_{i-1} + f_i = \sum_{k=0}^{i} f_k}$$

In EXCEL:

calcolo delle frequenze ---- utilizzo della funzione FREQUENZA

Calcola la frequenza RELATIVA f_i di occorrenza dei valori di un intervallo e restituisce una matrice di numeri verticale.

Sintassi:

FREQUENZA(matrice_dati; matrice_classi)

FREQUENZA viene immessa come formula matrice dopo aver selezionato un intervallo di celle adiacenti nel quale dovra' apparire il risultato.

Il numero di elementi contenuti nella matrice restituita è maggiore di una unità rispetto al numero di elementi contenuti in matrice_classi.

In EXCEL:

Formule in forma di matrice

Una formula in forma di matrice può eseguire più calcoli e restituire uno o più risultati. Le formule in forma di matrice agiscono su uno o più insiemi di valori denominati argomenti matrice. È necessario che ciascun argomento matrice sia costituito dallo stesso numero di righe e di colonne. Le formule in forma di matrice vengono create allo stesso modo delle formule a valore unico. Selezionare la cella o le celle in cui si desidera immettere la formula, creare la formula, quindi premere CTRL+MAIUSC+INVIO per immetterla.

ESEMPIO : esercizio 3 lab_1.xls

In EXCEL:

ESEMPIO : esercizio 3 lab_1.xls

Classi	Frequenze	Freq.relative	Freq.cumul.	Freq.Cum.rel.
0<x<0,2	14	0,14	14	0,14
0,2<x<0,4	23	0,23	37	0,37
0,4<x<0,6	23	0,23	60	0,6
0,6<x<0,8	19	0,19	79	0,79
0,8<x<1	21	0,21	100	1
Totali	100	1		

Istogrammi

Caso discreto:

Nel caso discreto l'istogramma e' costruito fissando sull'asse delle ascisse i valori delle classi e disegnando in corrispondenza una barra la cui altezza e' pari alla frequenza (relativa o assoluta).

Quindi l'altezza ha la stessa "unita' di misura" della probabilita' teorica.

Caso continuo:

Nel caso continuo l'istogramma e' costruito disegnando rettangoli adiacenti, le cui basi sono gli intervalli che definiscono le classi e le altezze sono date dalle frequenze (relative o assolute).

N.B.: in tal modo pero' l'altezza non ha piu' la stessa "unita' di misura" della probabilita' teorica!!

E' l'area ad avere la stessa unita' di misura della probabilita' quindi l'altezza del rettangolo deve essere scelta proporzionale al quoziente tra frequenza della classe e ampiezza dell'intervallo

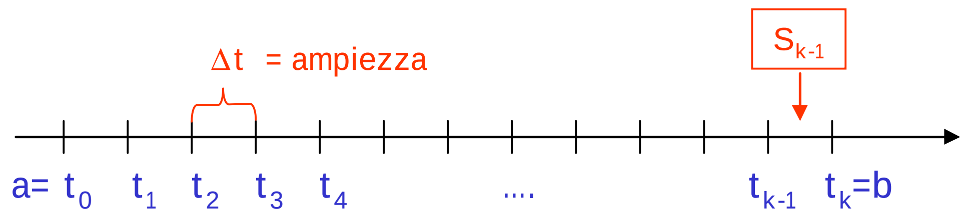
Distribuzioni e cumulative: accordo con la distribuzione teorica caso continuo

Confronto dei dati con una distribuzione teorica I 1/2

$(x_1, \dots, x_n)$: Campione estratto da un insieme (a,b),
i.e. i dati sono generati da una v.a. ass. continua X.

$a = t_0 < t_1 < \dots < t_k = b$: partizione di (a,b)

$\boxed{t_h = h\,\Delta t}$ ← Estremi delle classi



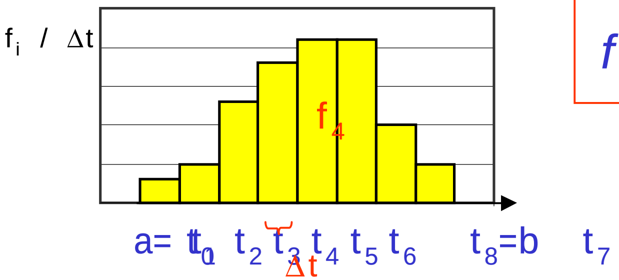
Ogni dato x_k cade in un intervallo $S_i = [t_{i-1}, t_i)$

Distribuzioni e cumulative: accordo con la distribuzione teorica caso continuo

Confronto dei dati con una distribuzione teorica I 2/2

Ogni dato cade in un intervallo $S_i = [t_{i-1}, t_i)$

n_i : frequenza assoluta dell'intervallo S_i



$$\boxed{f_i = n_i \;/\; n}$$

Frequenza relativa

Non ha piu' un significato di probabilita'

L'altezza del diagramma non rappresenta piu' la probabilita' ,

ma lo e' l'area della barra stessa. -----> l'altezza e' pari a f_i / Dt

Distribuzioni e cumulative: accordo con la distribuzione teorica caso continuo

Confronto dei dati con una distribuzione teorica II

Consideriamo la funzione

$$h(t) = \begin{cases} \dfrac{n_i}{n}\dfrac{1}{\Delta t_i} = \dfrac{f_i}{\Delta t_i} = h_i, & t \in S_i \\ 0 & \text{altrimenti} \end{cases} \qquad \Longrightarrow \qquad \int_a^b h(t)dt = \sum_{i=1}^{k} h_i \, \Delta t_i = \sum_{i=1}^{k} \frac{n_i}{n} = 1$$

Confrontiamo h_i con $\dfrac{r_i}{\Delta t}$, $\qquad r_i = P(X \in S_i)$

Se f_X e' la funzione di densita' di X, i.e. $\quad r_i = \int_{t_{i-1}}^{t_i} f_X(x)dx$

$h_i \, \Delta t = f_i$ ------> ρ_i

Distribuzioni e cumulative: accordo con la distribuzione teorica caso continuo

Confronto dei dati con una distribuzione teorica III

Se f_i sono molto piccoli $\Rightarrow$ $n_i \,/\, \Delta t$ vs. $n_i = n\,\rho_i = n f_X \,\Delta t_i$

Come si scelgono Δt_i ?

$$E[h_i] = E\left[\frac{n_i}{n}\frac{1}{\Delta t}\right] = \frac{1}{n\Delta t}E[n_i] = \frac{r_i}{\Delta t}$$

h_i ---> ρ_i

$$\Delta t \sim 0(1)$$

Δt non troppo piccolo rispetto a n

$$\mathrm{var}[h_i] = \frac{1}{n^2(\Delta t)^2}\mathrm{var}[n_i] = \frac{r_i(1-r_i)}{n(\Delta t)^2}$$

Var ---> 0

$$n(\Delta t)^2 \to \infty$$

Δt non troppo grande rispetto a n

Regola empirica

Distribuzioni e cumulative: accordo con la distribuzione teorica

caso continuo . ESEMPIO

Esperimento reale: si sparano n=96 colpi su un bersaglio e si misurano le deviazioni dei proiettili in orizzontale e in verticale. Tenendo conto dei dati in tabella disegnare un grafico e stabilire con quale distribuzione teorica si puo' descrivere il fenomeno.

ν_i/dt $N*p_i$

Orizzontale						
ti	n_i	ti*n_i	n_i* (ti - media)^2		altezza istog.freq	distrib. Teorica
-60	0	0	0		0	0,109097614
-50	3	-150	7500	media	0,3	0,31354385
-40	5	-200	8000	-11,5625	0,5	0,706693783
-30	13	-390	11700		1,3	1,249148881
-20	18	-360	7200		1,8	1,731598036
-10	21	-210	2100	dev. Standard	2,1	1,882478023
0	21	0	0	20,28443574	2,1	1,604954385
10	10	100	1000		1	1,073112914
20	5	100	2000		0,5	0,562701532
30	0	0	0		0	0,231398597
	96					

IN EXCEL: Creare un grafico

Dr. Daniela Morale

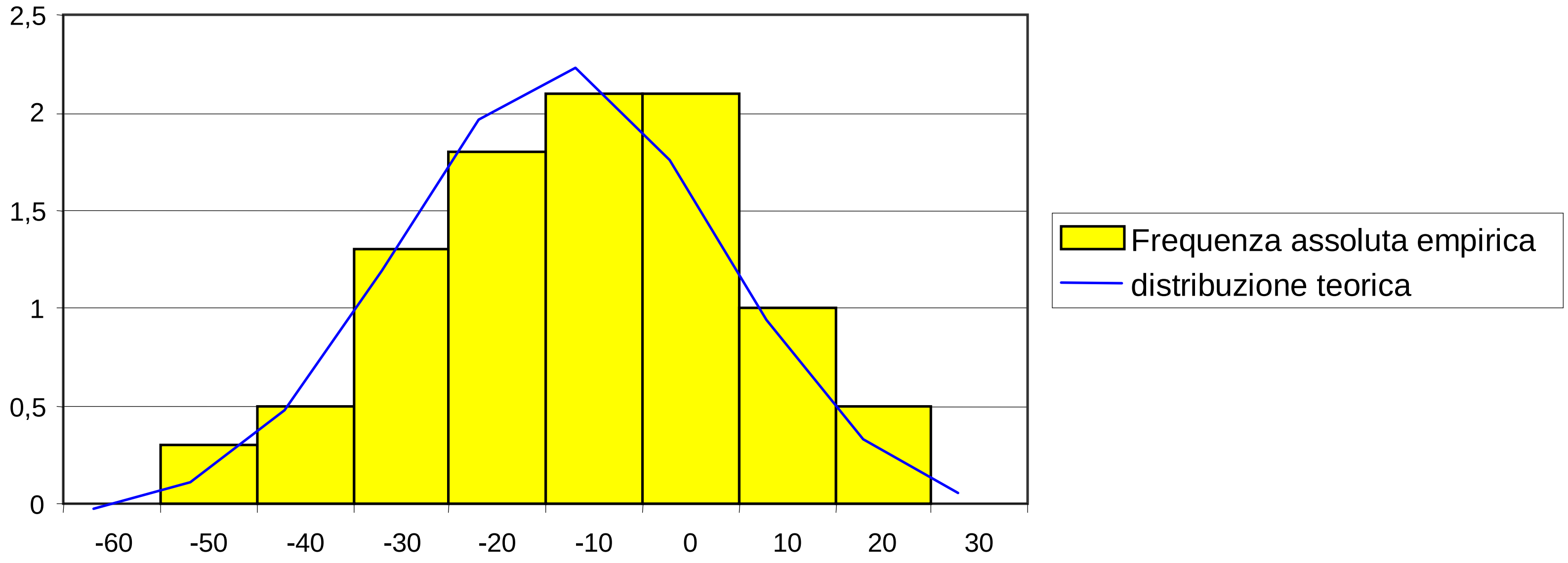
È possibile visualizzare in un grafico i dati di Excel. I grafici sono collegati ai dati del foglio di lavoro da cui sono stati creati e vengono aggiornati quando tali dati vengono modificati.

Fare clic sul pulsante Autocomposizione Grafico .

Seguire le istruzioni riportate in Autocomposizione Grafico.

Distribuzioni e cumulative: accordo con la distribuzione teorica

caso continuo . ESEMPIO



In EXCEL: ANALISI DATI

STRUMENTI DI ANALISI è un insieme di strumenti di analisi dei dati che consente di ridurre i passaggi necessari allo sviluppo di complesse analisi statistiche. Una volta forniti i dati e i parametri per ciscuna analisi, lo strumento utilizzerà le funzioni macro statistiche appropriate, visualizzando i risulatati in una tabella di output.

Alcuni strumenti generano anche dei grafici.

Come visualizzare un elenco degli strumenti di analisi disponibili:
scegliere Analisi dati dal menu Strumenti. Se tale comando non è visualizzato, eseguire il programma di installazione per installare gli Strumenti di analisi. Al termine dell'installazione l'aggiunta dovrà essere selezionata nel Gestore aggiunte (in STRUMENTI).

In EXCEL: Strumento di analisi Istogramma

Consente di calcolare le frequenze individuali e cumulative per un intervallo di celle e di classi di dati.

Opzioni della finestra di dialogo Istogramma

Intervallo di input: Immettere il riferimento di cella per l'intervallo di dati da analizzare.

Intervallo di classe (facoltativo) Immettere un intervallo di celle contenente un insieme facoltativo di valori limite che definiscano gli intervalli delle classi.

Se non si specifica l'intervallo di classe, verrà automaticamente creato un insieme di classi distribuite uniformemente tra il valore minimo e il valore massimo dei dati.

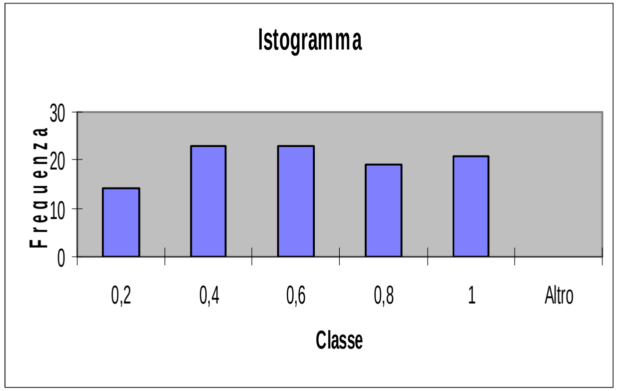
Etichette Selezionare questa casella di controllo se la prima riga dell'intervallo di input contiene etichette. In caso contrario deselezionarla, in quanto le etichette di dati appropriate per la tabella di output verranno generate automaticamente.

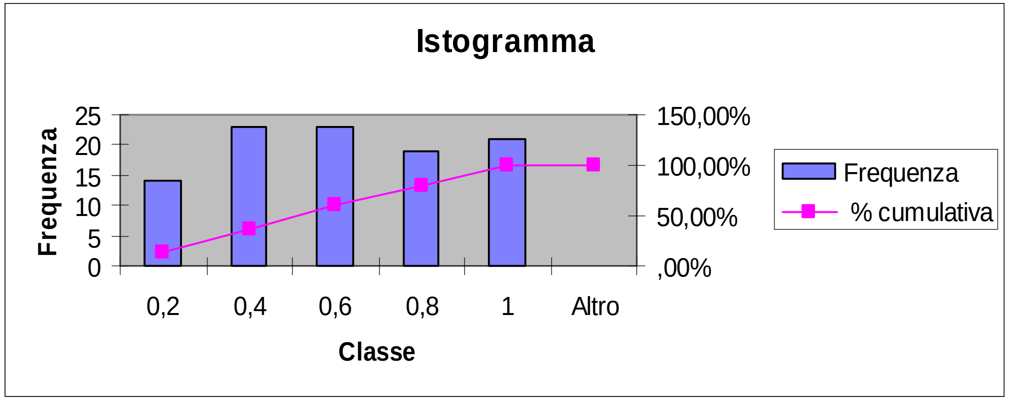
Intervallo di output Immettere il riferimento della cella superiore sinistra della tabella di output.

IN EXCEL: istogrammi

Classe	*Frequenza*
0,2	14
0,4	23
0,6	23
0,8	19
1	21
Altro	0

Classe	*Frequenza*	*% cumulativa*
0,2	14	14,00%
0,4	23	37,00%
0,6	23	60,00%
0,8	19	79,00%
1	21	100,00%
Altro	0	100,00%





Indici di posizione, di dispersione e di forma

Dr. Daniela Morale

Indici di posizione

Media $\bar{x} = 1/n \ \Sigma_i x_i$

Mediana $\begin{cases} x_{[n/2]+1} & n \text{ dispari} \\ x_{[n/2]} + x_{[n/2]+1} & n \text{ pari} \end{cases}$

Moda pto di max della distribuzione

Indici di dispersione

Varianza $\sigma^2 = 1/n \ \Sigma_i \ (x_i - \bar{x})^2$

Deviazione standard $\sigma = \sqrt{1/n \ \Sigma_i \ (x_i - x)^2}$

Range $R = x_{max} - x_{min}$

Indici di forma

Skewness (coeff. asimmetria) $1/n \ \Sigma_i \ [(x_i - x) / \sigma]^3$

> 0 ---> coda verso destra
< 0 ---> coda verso sinistra
= 0 ---> simmetrica

Curtosi $1/n \ \Sigma_i \ [(x_i - x) / \sigma]^4$

Indica se la distribuzione e' appuntita

In EXCEL: Strumento di analisi Statistica descrittiva

Genera un rapporto di statistica univariata per i dati dell'intervallo di input, fornendo informazioni sulla tendenza centrale e la variabilità dei dati.

Intervallo di input Immettere il riferimento di cella per l'intervallo di dati da analizzare che deve consistere in due o più intervalli di dati adiacenti disposti in colonne o righe.

Raggruppato in base Per indicare se i dati nell'intervallo di input sono disposti in righe o in colonne, fare clic su Righe o Colonne rispettivamente.

Livello di confidenza per media Selezionare questa opzione se si desidera includere nella tabella di output una riga per il livello di confidenza della media. Immettere quindi nella casella il livello di confidenza che si desidera utilizzare. Un valore del 95 % calcola ad esempio il livello di confidenza della media a una significatività del 5%.

In EXCEL: Strumento di analisi Statistica descrittiva

Intervallo di output Immettere il riferimento della cella superiore sinistra della tabella di output. Questo strumento genera due colonne di informazioni per ciascun insieme di dati. La colonna di sinistra contiene le etichette di statistica, mentre quella di destra contiene le statistiche. Verrà scritta una tabella di statistiche a due colonne per ciascuna colonna o riga dell'intervallo di input, a seconda dell'opzione selezionata nella casella Raggruppato in base.

Riepilogo statistiche

Selezionare questa opzione se si desidera generare un campo nella tabella di output per ciascuna delle seguenti statistiche : Media, Errore standard (della media), Mediana, Modalità, Deviazione standard, Varianza, Curtosi, Asimmetria, Intervallo, Minimo, Massimo, Somma, Conteggio, Più grande (#), Più piccolo (#) e Livello di confidenza.

ESERCIZIO : STATISTICA DESCRITTIVA

Di seguito ci sono le eta' dei pazienti affetti da tumori gliali al momento della diagnosi e il diametro del tumore : fare una analisi descrittiva dei dati

Eta'	Frequenza assoluta	Eta'	Frequenza assoluta
5	1	42	8
6	1	43	3
7	0	44	5
8	1	45	4
9	1	46	3
10	0	47	4
11	1	48	3
12	1	49	3
13	2	50	4
14	4	51	3
15	1	52	3
16	2	53	2
17	1	54	3
18	0	55	4
19	0	56	3
20	1	57	6
21	0	58	8
22	0	59	4
23	1	60	3
24	2	61	1
25	1	62	4
26	2	63	4
27	5	64	1
28	1	65	3
29	3	66	3
30	4	67	7
31	5	68	7
32	2	69	2
33	2	70	4
34	4	71	3
35	5	72	1
36	7	73	1
37	4	74	1
38	3	75	0
39	4	76	0
40	3	77	1
41	2		

DIAM	Frequenza assoluta	DIAM	Frequenza assoluta	DIAM	Frequenza assoluta
12	2	48	0	84	0
13	0	49	0	85	0
14	0	50	20	86	0
15	4	51	0	87	0
16	1	52	0	88	0
17	0	53	0	89	0
18	0	54	0	90	8
19	0	55	14	91	0
20	9	56	1	92	0
21	0	57	1	93	0
22	0	58	2	94	0
23	0	59	0	95	0
24	2	60	24	96	0
25	2	61	0	97	0
26	0	62	0	98	0
27	0	63	0	99	1
28	0	64	0		
29	0	65	2		
30	20	66	0		
31	0	67	0		
32	0	68	0		
33	0	69	0		
34	0	70	21		
35	12	71	1		
36	1	72	0		
37	1	73	0		
38	1	74	0		
39	0	75	0		
40	25	76	0		
41	0	77	0		
42	1	78	0		
43	0	79	0		
44	0	80	8		
45	5	81	0		
46	0	82	0		
47	0	83	0		

ESERCIZIO : STATISTICA DESCRITTIVA

Per fare un'analisi descrittiva ho bisogno dei dati dati grezzi quindi bisogna prima costruire la tabella dei dati, cioè fare una tabella in cui compaiono le eta' (o i diametri) tante volte quanto la frequenza

Devono essere in una sola colonna!!

Altrimenti fa uno studio per ogni colonna

ETA'				58	68
				58	68
5	30	39	48	58	68
6	30	39	48	58	68
8	30	39	48	58	68
9	30	39	49	58	68
11	31	40	49	59	68
12	31	40	49	59	69
13	31	40	50	59	69
13	31	41	50	59	70
14	31	41	50	60	70
14	32	42	50	60	70
14	32	42	51	60	70
14	33	42	51	61	71
15	33	42	51	62	71
17	34	42	52	62	71
16	34	42	52	62	72
20	34	42	52	62	73
23	34	42	53	63	74
24	35	43	53	63	77
24	35	43	54	63	
25	35	43	54	63	
26	35	44	54	64	
26	35	44	55	65	
27	36	44	55	65	
27	36	44	55	65	
27	36	44	55	66	
27	36	45	56	66	
27	36	45	56	66	
28	36	45	56	67	
29	36	45	57	67	
29	37	46	57	67	
29	37	46	57	67	
	37	46	57	67	
	37	47	57	67	
	38	47	57	67	
	38	47	58		
	38	47	58		

ESERCIZIO : Risultato strumento analisi STATISTICA DESCRITTIVA

$x = 1/n \ \Sigma \ x_i$

$x_{[n/2]} + x_{[n/2]+1}$ n pari

$\sigma = \sqrt{1/n \ \Sigma_i \ (x_i - x)^2}$

$\sigma = 1/n \ \Sigma_i \ (x_i - x)^2$

Indica se la distribuzione e' appuntita

$1/n \ \Sigma_i \ [(x_i - x) / \sigma]^4$

> 0 ---> coda verso destra
< 0 ---> coda verso sinistra
= 0 ---> simmetrica

$1/n \ \Sigma_i \ [(x_i - x) / \sigma]^3$

	ETA'
Media	45,87692308
Errore standard	1,193504384
Mediana	46
Moda	42
Deviazione standard	16,66638171
Varianza campionaria	277,7682791
Curtosi	-0,60280546
Asimmetria	-0,330795288
Intervallo	72
Minimo	5
Massimo	77
Somma	8946
Conteggio	195
Più grande(1)	77
Più piccolo(1)	5
Livello di confidenza(95,0%)	2,353908768

INTERVALLI DI CONFIDENZA

ð Intervalli di confidenza

ð Funzioni in Excel: Confidenza, DISTR.T, INV.T, INV.NORM.ST
DISTR.F, INV.F

CONFIDENZA Intervallo di confidenza per una media nel caso s nota

Restituisce l'intervallo di confidenza per una popolazione. L'intervallo di confidenza è dato da un intervallo di valori che precedono o seguono una media campione.

Sintassi

CONFIDENZA(alfa,dev_standard,dimens)

Alfa è il livello di significatività utilizzato per calcolare il livello di confidenza. Il livello di confidenza è uguale a 100*(1 - alfa)% o, in altre parole, un alfa di 0,05 indica un livello di confidenza del 95%.

Dev_standard è la deviazione standard della popolazione per l'intervallo di dati e si presuppone sia nota.

Dimens è la dimensione del campione.

Osservazioni

- Se un qualsiasi argomento non è numerico, CONFIDENZA restituirà il valore di errore #VALORE!.
- Se alfa < 0 o alfa > 1, CONFIDENZA restituirà il valore di errore #NUM!.
- Se dev_standard < 0, CONFIDENZA restituirà il valore di errore #NUM!.
- Se dimens non è un numero intero, la parte decimale verrà troncata.
- Se dimens < 1, CONFIDENZA restituirà il valore di errore #NUM!.
- Se si suppone che alfa sia uguale a 0,05, sarà necessario calcolare l'area sottostante la curva normale standard che è uguale a (1 - alfa) o al 95%. Questo valore è ± 1,96. L'intervallo di confidenza sarà quindi:

$$\bar{x} \pm 1{,}96 \frac{s}{\sqrt{n}}$$

Esempio

Si supponga che, per il campione di 50 pendolari preso in esame, la durata media del viaggio per raggiungere il posto di lavoro sia di 30 minuti con una deviazione standard della popolazione pari a 2,5. Si può avere una confidenza del 95% che la media della popolazione rientri nell'intervallo: 30 ± 0,692951

oppure:

CONFIDENZA(0,05;2,5;50) è uguale a 0,692951. In altre parole, la durata media del viaggio da casa al posto di lavoro è uguale a 30 ± 0,692951 minuti o a un valore compreso tra 29,3 e 30,7 minuti.

DISTRIB.T DISTRIBUZIONE t DI Student

Restituisce la distribuzione t di Student. La distribuzione t viene utilizzata nelle verifiche di ipotesi su piccoli insiemi di dati presi come campione. Utilizzare questa funzione al posto di una tabella di valori critici per il calcolo della distribuzione t.

Sintassi

DISTRIB.T(x;gradi_libertà;coda)

X è il valore numerico in cui calcolare la distribuzione.

Gradi_libertà è un intero che indica il numero di gradi di libertà.

Coda specifica il numero di code di distribuzione da restituire. Se coda = 1, DISTRIB.T restituirà la distribuzione ad una coda. Se coda = 2, DISTRIB.T restituirà la distribuzione a due code.

La funzione DISTRIB.T viene calcolata come DISTRIB.T = p(x<X), dove X è una variabile casuale che segue la distribuzione t.

Esempio DISTRIB.T(1,96;60;2) è uguale a 0,054645

INV.T Funzione inversa della distribuzione t di STUDENT

Restituisce l'inversa della distribuzione t di Student per i gradi di libertà specificati.

Sintassi:

INV.T(probabilità;gradi_libertà)

Probabilità è la probabilità associata alla distribuzione t di Student a due code.

Gradi_libertà è il numero di gradi di libertà che caratterizza la distribuzione.

La funzione INV.T viene calcolata come INV.T = p(t< X), dove X è una variabile casuale che segue la distribuzione t .

INV.T utilizza una tecnica iterativa per il calcolo della funzione. Dato un valore di probabilità, INV.T applica il metodo delle iterazioni fino a quando la precisione del risultato non rientra in ± 3x10^-7. Se il risultato di INV.T non converge dopo 100 iterazioni, la funzione restituirà il valore di errore #N/D.

Esempio INV.T(0,054645;60) è uguale a 1,96

Intervallo di confidenza per una media nel caso s non nota

Gli estremi di tale intervallo risultano (cf. p.81)

$$\bar{x} \pm t_{(1-a/2)} \frac{s}{\sqrt{n}}$$

dove $t_{(1-a/2)}$ e' il valor della t di Student con (n-1) gradi di liberta' che corrisponde ad una probabilita' pari a (1- α/2).

Come si determina tale valore?

Si utilizza la funzione INV.T(α; n-1)

ESERCIZIO

Si sono effettuate 40 misure del tempo di caduta (in centesimi di secondo) di un sasso da una certa altezza al suolo

63	58	74	78	70	74	75	82
76	62	72	88	65	81	79	77
86	72	79	77	60	70	65	69
72	79	65	66	70	74	84	76

Determinare un intervallo di confidenza al 99% per il tempo medio di caduta del sasso.

Si utilizza la funzione INV.T((1- α/2);39)

Valore in cui la t(39) ha probabilita' 1-0,005, ottenuta con INV.T(0,09;39)

Gradi di liberta'

Media e dev. campionaria

Estremi dell'intervallo

conf.	a	df	t(1-a/2)	media	sigma	I{-}	I{+}
0,06875	0,01	39	2,707911	73,0625	7,517453	69,84384	76,28116

Intervallo

(69.84384,76.28116)

ESERCIZIO 2 Intervalli di confidenza

Cinque persone si sono fatte misurare la capacità respiratoria prima e dopo un certo trattamento, dando luogo ai seguenti risultati:

Individuo	Prima (X)	Dopo (Y)	Variazione
A	2750	2850	100
B	2360	2380	20
C	2950	2800	-150
D	2830	2300	30
E	2250	2300	50

Si costruisca un intervallo di confidenza al 95% per mX - mY supponendo di aver campionato da popolazioni normali caratterizzate dalla stessa varianza.

ESERCIZIO 2 Intervalli di confidenza

conf.	a	df	t(1-a/2)	media	sigma	l{-}	l{+}
0,99	0,05	4	2,776451	10	94,60444	-107,467	127,4672

INTERVALLO

(-107,467;127,4672)

Intervallo di confidenza per la differenza tra due medie

nel caso s_1, s_2 note

Gli estremi di tale intervallo risultano (cf. p.84)

$$\overline{x_1} - \overline{x_2} \pm z_{(1-a/2)}\sqrt{\frac{s_{\ 1}}{n_1} + \frac{s_{\ 2}}{n_2}}$$

dove $z_{(1-a/2)}$ e' il valor della normale standard che corrisponde ad una probabilita' pari a (1- α/2).

Come si determina tale valore?

Si utilizza la funzione INV.NORM.ST(probabilità)

Intervallo di confidenza per la differenza tra due medie nel caso s_1, s_2 non note

Gli estremi di tale intervallo risultano (cf. p.84)

$$\overline{x_1} - \overline{x_2} \pm t_{(1-a/2)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_1^2}{n_1 + n_2 - 2}}$$

dove $t_{(1-a/2)}$ e' il valor della t di Student con (n_1 + n_2 -1) gradi di liberta' che corrisponde ad una probabilita' pari a (1- α/2).

Come si determina tale valore?

Si utilizza la funzione INV.T(probabilità, gradi di liberta')

ESERCIZIO 4

Date le seguenti caratteristiche di due campioni casuali estratti da due popolazioni

$n1 = 25$	$\overline{X}1 = 60.0$	$s1 = 12$
$n2 = 15$	$\overline{X}2 = 68.0$	$s2 = 10$

supponendo $\sigma 1 = \sigma 2$ si determini un intervallo di confidenza al 95% per $\mu 1 - \mu 2$

Valore in cui la t(39) ha probabilita' 1-0,005

Gradi di liberta'

Media campionaria

Estremi dell'intervallo

conf.	a	df	t(1-a/2)	media	sigma_p	rad	I(-)	I(+)
0,95	0,05	38	2,024394234	-8,00	3,356062856	0,326599	-10,2189	-5,78109

$$\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_1^2}{n_1+n_2-2}}$$

$$\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

INTERVALLO

(-10,22;-5,78)

INV.F Funzione inversa della distribuzione F di Fisher

Restituisce l'inversa della distribuzione di probabilità F. Se p = DISTRIB.F(x;...), si avrà INV.F(p;...) = x.
La distribuzione F può essere utilizzata in un test F che confronta il grado di variabilità di due insiemi di dati. È possibile ad esempio analizzare la distribuzione del reddito in Italia e in Francia per stabilire se i due paesi hanno un grado di diversità simile.

Sintassi: INV.F(probabilità;gradi_libertà1;gradi_libertà2)

Probabilità è la probabilità associata alla distribuzione cumulativa F.
Gradi_libertà1 sono i gradi di libertà al numeratore.
Gradi_libertà2 sono i gradi di libertà al denominatore.

La funzione INV.F può essere utilizzata per restituire valori critici dalla distribuzione F. Il risultato di un calcolo di ANALISI.VARIANZA spesso include dati per la statistica F, la probabilità F e il valore critico F con il livello di significatività 0,05. Per calcolare il valore critico di F, utilizzare il livello di significatività come argomento di probabilità della funzione INV.F.

Esempio INV.F(0,01;6;4) è uguale a 15,20675

Intervallo di confidenza per il rapporto di due varianze nel caso di popolazioni normali

L'intervallo di confidenza per il rapporto $\frac{s_1^2}{s_2^2}$ (cf. p.90) e'

$$\boxed{\left[\frac{s_1^2}{s_2^2}\frac{1}{F_{a/2}},\frac{s_1^2}{s_2^2}\frac{1}{F_{(1-a/2)}}\right]}$$

dove F_d e' il valor della F di Fisher con (n_1 -1, n_2 -1) gradi di liberta' che corrisponde ad una probabilita' pari a δ.

Come si determina tale valore?

Si utilizza la funzione INV.F(probabilità; gradi_libertà1;gradi_libertà2)

ESERCIZIO 5

Viene atto uno studio per valutare il tempo di risposta ad un certo stimolo. Vengono considerati due campioni il primo di taglia 21 di adulti apparentemente sani, il secondo di taglia 16 di pazienti con il morbo di Parkinson.

Le varianze campionarie risultano 1600, per il campione 1, e 1225 per il campione 2.
Per confrontarle i ricercatori vogliono un intervallo di conidenza al 95% per il rapporto.

campione	numerosita' campione	varianza campionaria
1	**21**	**1600**
2	**16**	**1225**

conf.	a	**df1**	**df2**	**F(1-a/2)**	**F(a/2)**	**s_1^2/s_2^2**	**I{+}**	**I{-}**
0,95	0,05	20	15	2,755897	0,388637	1,306122	3,36078	0,473937

TEST DI IPOTESI

- Ancora sulla distribuzione t di Student: INV.T
- TEST DI IPOTESI
- Strumenti di analisi per i test

Strumento di analisi Test z: a due campioni per media

Consente di eseguire un test z a due campioni per medie con varianze note. Questo strumento viene utilizzato per verificare le ipotesi sulla differenza tra due medie di popolazione.

Intervallo variabile 1 Immettere il riferimento di cella per il primo intervallo di dati da analizzare che deve consistere in una singola colonna o riga di dati.

Intervallo variabile 2 Immettere il riferimento di cella per il secondo intervallo di dati da analizzare che deve consistere in una singola colonna o riga di dati.

Differenza ipotizzata per le medie Immettere il numero desiderato per la variazione delle medie campione. Il valore 0 (zero) indica che si ipotizzano le stesse medie campione.

Varianza variabile 1 Immettere la varianza di popolazione nota per l'intervallo di input Variabile 1.

Varianza variabile 2 Immettere la varianza di popolazione nota per l'intervallo di input Variabile 2.

Alfa Immettere il livello di confidenza per il test che deve essere un valore compreso nell'intervallo 0...1. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

Strumento di analisi Test t: a due campioni accoppiati per medie

Consente di eseguire un test t di Student a due campioni accoppiati per determinare se le medie di un campione sono distinte. Questa forma del test t presuppone che le varianze delle due popolazioni siano uguali. È possibile utilizzare un test accoppiato quando vi è un naturale appaiamento tra le osservazioni dei campioni, come nel caso di una duplice verifica di un gruppo campione, prima e dopo un esperimento.

Intervallo variabile 1 Immettere il riferimento di cella per il primo intervallo di dati da analizzare che deve consistere in una singola colonna o riga di dati e contenere lo stesso numero di dati del secondo intervallo.

Intervallo variabile 2 Immettere il riferimento di cella per il secondo intervallo di dati da analizzare che deve consistere in un'unica colonna o riga e contenere lo stesso numero di dati del primo intervallo.

Differenza ipotizzata per le medie Immettere il numero desiderato per la variazione delle medie campione. Il valore 0 (zero) indica che si ipotizzano le stesse medie campione.

Alfa Immettere il livello di confidenza per il test che deve essere un valore compreso nell'intervallo 0...1. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

Strumento di analisi Test t: a due campioni assumendo uguale varianza

Consente di eseguire un test t di Student a due campioni. Questa forma del test t, definito test t omoschedastico, presuppone che le medie dei due insiemi di dati siano uguali.

Intervallo variabile 1 Immettere il riferimento di cella per il primo intervallo di dati da analizzare che deve consistere in una singola colonna o riga di dati e contenere lo stesso numero di dati del secondo intervallo.

Intervallo variabile 2 Immettere il riferimento di cella per il secondo intervallo di dati da analizzare che deve consistere in un'unica colonna o riga e contenere lo stesso numero di dati del primo intervallo.

Differenza ipotizzata per le medie Immettere il numero desiderato per la variazione delle medie campione. Il valore 0 (zero) indica che si ipotizzano le stesse medie campione.

Alfa Immettere il livello di confidenza per il test che deve essere un valore compreso nell'intervallo 0...1. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

Strumento di analisi Test t: a due campioni assumendo varianze diverse

Consente di eseguire un test t di Student a due campioni. Questa forma del test, definito test t eteroschedastico, presuppone che le varianze dei due intervalli di dati siano diverse. È possibile utilizzare un test t per determinare se le medie di due campioni sono uguali.

Utilizzare questo test quando i gruppi esaminati sono distinti e un test accoppiato quando un gruppo viene analizzato prima e dopo un trattamento.

Intervallo variabile 1   Immettere il riferimento di cella per il primo intervallo di dati da analizzare che deve consistere in una singola colonna o riga di dati e contenere lo stesso numero di dati del secondo intervallo.

Intervallo variabile 2   Immettere il riferimento di cella per il secondo intervallo di dati da analizzare che deve consistere in un'unica colonna o riga e contenere lo stesso numero di dati del primo intervallo.

Differenza ipotizzata per le medie   Immettere il numero desiderato per la variazione delle medie campione. Il valore 0 (zero) indica che si ipotizzano le stesse medie campione.

Alfa   Immettere il livello di confidenza per il test che deve essere un valore compreso nell'intervallo 0...1. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

Strumento di analisi Test F: a due campioni per varianze

Consente di eseguire un test F a due campioni per confrontare le varianze di due popolazioni.

Intervallo variabile 1 Immettere il riferimento per la prima colonna o riga di dati che si desidera analizzare.

Intervallo variabile 2 Immettere il riferimento per la seconda colonna o riga di dati che si desidera analizzare.

Etichette Selezionare questa casella di controllo se la prima riga o colonna dell'intervallo di input contiene etichette. In caso contrario deselezionarla, in quanto le etichette di dati appropriate per la tabella di output verranno generate automaticamente.

Alfa Immettere il livello di confidenza per il test che deve essere un valore compreso nell'intervallo 0...1. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

TEST.CHI

Restituisce il test per l'indipendenza. La funzione TEST.CHI restituisce il valore dalla distribuzione del chi quadrato (2) per un dato statistico e i gradi di libertà appropriati. È possibile utilizzare i test 2 per stabilire se i risultati ipotizzati sono confermati da un esperimento.

Sintassi TEST.CHI(int_effettivo;int_previsto)

Int_effettivo è l'intervallo di dati contenente le osservazioni da confrontare con i valori previsti.

Int_previsto è l'intervallo di dati contenente la proporzione del prodotto dei totali di riga e di colonna per il totale complessivo.

TEST.CHI restituisce la probabilità di un dato statistico c2 e i gradi di libertà gdl, dove gdl = (r - 1)(c - 1).

TEST.F

Restituisce il risultato di un test F. Un test F restituisce la probabilità a una coda che le varianze in matrice1 e matrice2 non siano sensibilmente differenti. Utilizzare questa funzione per determinare se due campioni hanno varianze diverse. Ad esempio, sulla base dei punteggi di un test da scuole pubbliche e private, è possibile verificare se la diversità di queste scuole si estende su più livelli.

Sintassi TEST.F(matrice1;matrice2)

Matrice1 è la prima matrice o il primo intervallo di dati.

Matrice2 è la seconda matrice o il secondo intervallo di dati.

TEST DI IPOTESI
E
ANALISI DELLA VARIANZA

ð Ancora sulla distribuzione t di Student: INV.T

ð TEST DI IPOTESI: TEST.Z, TEST.T

ð Analisi della Varianza

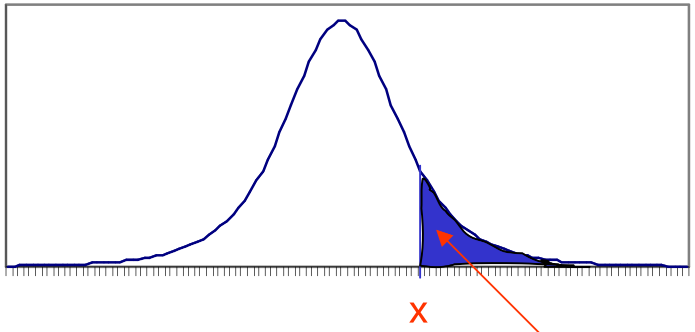
DISTRIB.T   DISTRIBUZIONE t DI Student

Sintassi :   DISTRIB.T(x;gradi_libertà;coda)

X	è il valore numerico in cui calcolare la distribuzione.
Gradi_libertà	è un intero che indica il numero di gradi di libertà.
Coda	specifica il numero di code di distribuzione da restituire.
	Se coda = 1, DISTRIB.T restituirà la distribuzione ad una coda.
	Se coda = 2, DISTRIB.T restituirà la distribuzione a due code.

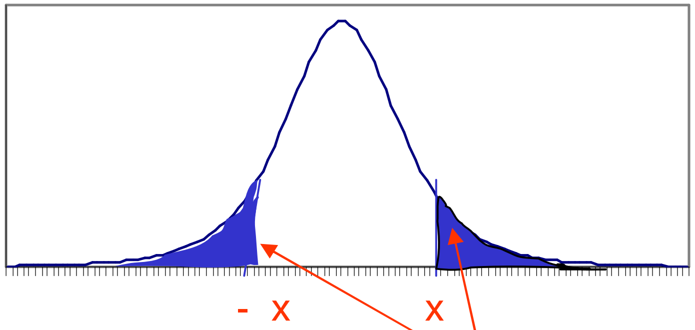
DISTRIB.T = p(x<X), .

t di Student a 4 gradi di liberta'



DISTRIB.T(x; 4 ; 1) = a = P(X > x)

t di Student a 4 gradi di liberta'



DISTRIB.T(x; 4 ; 2) = a = P(|X|> x)

INV.T Funzione inversa della distribuzione t di STUDENT

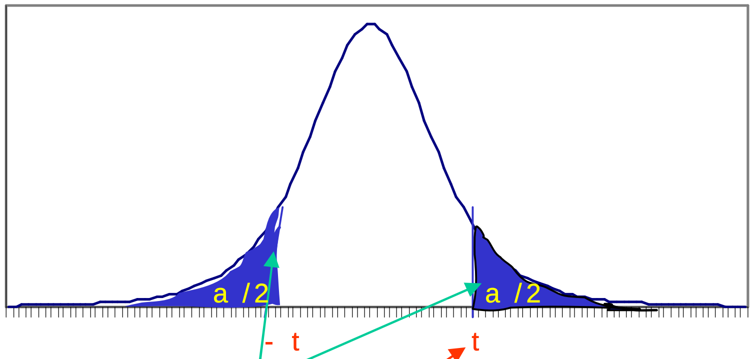
Restituisce l'inversa della distribuzione t di Student per i gradi di libertà specificati.

Sintassi: INV.T(probabilità;gradi_libertà)

Probabilità è la probabilità associata alla distribuzione t di Student a due code.
Gradi_libertà è il numero di gradi di libertà che caratterizza la distribuzione.

La funzione INV.T viene calcolata come INV.T = p(t< X), dove X è una variabile casuale che segue la distribuzione t .

t di Student a 4 gradi di liberta'



INV.T(a ; 4) = t = $t_{(1 - a/2)}$

Esempio INV.T(0,054645;60)
è uguale a 1,96

TEST.Z a due campioni per media

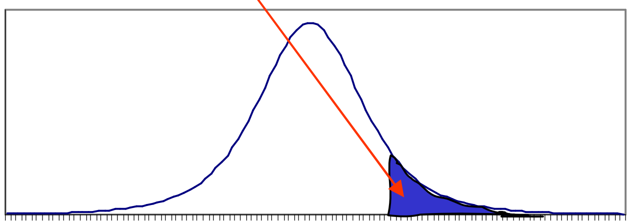
Restituisce il livello di significatività a una code per un test z.

Sintassi TEST.Z(matrice; x; sigma)

Matrice è la matrice o l'intervallo di dati con cui esaminare x.
X è il valore da esaminare.
Sigma è la deviazione standard della popolazione (nota). Se questo argomento è omesso, verrà utilizzata la deviazione standard campione.

TEST.Z(matrice;x;sigma) = 1 - DISTR.NORM.ST((media - x)/(sigma/RADQ(n)))
= P(Z > (media - x)/(sigma/RADQ(n)))

Esempio TEST.Z({3;6;7;8;6;5;4;2;1;9};4) è uguale a 0,090574



TEST a 1 coda:

1 - a = TEST.Z(matrice;x;sigma)

TEST a 2 code:

1 - a / 2 = TEST.Z(matrice;x;sigma)

ESERCIZIO 1 TEST Z

Una fabbrica di automobili ha usato finora delle guarnizioni per freni con una distanza di arresto di 90 piedi e che la ditta stia pensando di sostituirle con un altro tipo di guarnizioni, simili alle precedenti per tutti gli altri aspetti, ma che sembrano avere una distanza di arresto più breve.

In una prova le nuove guarnizioni vengono montate su 64 auto e la distanza media risulta di 87 piedi con una deviazione standard di 16 piedi.

Se il nostro compito è quello di compiere un controllo di qualità ci viene chiesto di valutare se le nuove guarnizioni siano o meno migliori delle precedenti.

TEST:			
	H 0 :	$\mu =$	90
	H 1 :	$\mu <$	90

Test unilaterale!

Dati			
	n	=	64
	X media campionaria	=	87
	S campionaria	=	16

p-value	=	0,064759

Se si sceglie di fare un test con errore di I tipo a =0.05 allora si rifiuta l'ipotesi
---> la media e' diminuita, le guarnizioni sono migliori

Se si sceglie di fare un test con errore di I tipo a =0.1 allora siaccetta l'ipotesi
---> la media non e' diminuita, le guarnizioni non sono migliori

TEST.T

Restituisce la probabilità associata ad un test t di Student. Utilizzare la funzione TEST.T per determinare se due campioni possono essere derivati dalle stesse due popolazioni aventi la stessa media.

Sintassi

TEST.T(matrice1;matrice2;coda;tipo)

Matrice1 è il primo insieme di dati.
Matrice2 è il secondo insieme di dati.
Coda specifica il numero di code di distribuzione. Se coda = 1, TEST.T utilizzerà la distribuzione ad una coda. Se coda = 2, TEST.T utilizzerà la distribuzione a due code.
Tipo è il tipo di test t da eseguire.

Se tipo è uguale a	Il test verrà eseguito
1	Accoppiato
2	Omoschedastico (varianza uguale di due campioni)
3	Eteroschedastico (varianza disuguale di due campioni)

Esempio TEST.T({3;4;5;8;9;1;2;4;5};{6;19;3;2;14;4;5;17;1};2;1) è uguale a 0,196016

ESERCIZIO 2 TEST t

Durante un esperimento su animali da laboratorio sono stati registrati i seguenti dati del flusso sanguigno nei reni sonno condizioni di controllo e durante la somministrazione di un certo anestetico.

	Flusso sanguigno ML/G/min	
animale	controllo	somministrazione di anestetico
1	2,35	2
2	2,55	1,71
3	1,95	2,22
4	2,79	2,71
5	3,21	1,83
6	2,97	2,14
7	3,44	3,72
8	2,58	2,1
9	2,66	2,58
10	2,31	1,32
11	3,43	3,7
12	2,37	1,59
13	1,82	2,07
14	2,98	2,15
15	2,53	2,05

Possiamo concludere sulla base di questi dati che l'anestetico ritarda il flusso del sangue nei reni? Si $\alpha = 0,05$. Determinare il p-value.

a =	0,05

X= controllo
Y= anestetico

TEST:

H_0 :	$\mu_X - \mu_Y =$	0
H_1 :	$\mu_X - \mu_Y <$	0

Test unilaterale! ---> 1 coda

I campioni sono accoppiati!!!! --->
utilizzo il **test t di tipo 1**

p-value	0,00564499

Quindi per al livello $\alpha = 0,05$ rifiuto l'ipotesi nulla l'anestetico ritarda il flusso….

Strumento Analisi varianza: ad un fattore

Consente di eseguire una semplice analisi della varianza per verificare l'ipotesi secondo cui i valori medi di due o più campioni, estratti da popolazioni con gli stessi valori medi, sono uguali.

Intervallo di input

Immettere il riferimento di cella per l'intervallo di dati da analizzare che deve consistere in due o più intervalli di dati adiacenti disposti in colonne o righe.

Raggruppato in base

Per indicare se i dati nell'intervallo di input sono disposti in righe o in colonne, fare clic su Righe o Colonne rispettivamente.

Alfa

Immettere il livello di valutazione dei valori critici per la statistica F. Il livello alfa è un livello di significatività correlato alla probabilità di riscontrare un errore di tipo I, ossia il rifiuto di un'ipotesi vera.

ESERCIZIO

Si misurano i battiti cardiaci per minuto di 4 gruppi di adulti: adulti al controllo annuale (gruppo A), pazienti sofferenti di angina (gruppo B), individui ipertesi (gruppo C) e individui infartuati (gruppo D). I dati raccolti forniscono evidenza sperimentale (al livello a=0.05) di una differenza di ritmo medio in questi 4 gruppi?

A	B	C	D
83	81	75	61
61	65	68	75
80	77	80	78
63	87	74	80
67	95	78	68
89	89	69	65
71	103	72	68
73	89	76	69
70	78	75	70
66	83	69	79
57	91	80	71

Analisi varianza: ad un fattore

RIEPILOGO

Gruppi	*Conteggio*	*Somma*	*Media*	*Varianza*
A	11	780	70,90909	95,49091
B	11	938	85,27273	102,8182
C	11	816	74,18182	18,36364
D	11	784	71,27273	36,81818

ANALISI VARIANZA

Origine della variazione	*SQ*	*gdl*	*MQ*	*F*	*Valore di significatività*	*F crit*
Tra gruppi	1497,727	3	499,2424	7,877875	0,000301953	2,838746127
In gruppi	2534,909	40	63,37273			
Totale	4032,636	43				

Rifiuto l'ipotesi di differenza di ritmo