

## LABORATORIO DI MATEMATICA

## LA STIMA E LA VERIFICA DELLE IPOTESI CON EXCEL

La funzione	determina
INV.T( $\alpha$ ; $\nu$ )	$z_{1-\frac{\alpha}{2}}$ , il limite di confidenza, dove $\alpha$ è il livello di significatività e $\nu$ sono i gradi di libertà.
DISTRIB.T( $z$ ; $\nu$ ; coda)	$\alpha$ , il livello di significatività, dove $z$ è il limite di confidenza, $\nu$ sono i gradi di libertà e coda vale 0 o 2.

## ESERCITAZIONE GUIDATA

Costruiamo un foglio elettronico che permetta di rappresentare la carta di controllo statistico relativa a una macchina confezionatrice di barattoli, sapendo che la distribuzione è normale ed è nota la varianza  $\sigma^2$ .

Il foglio deve leggere:

- $\mu$ , il peso medio in grammi di un barattolo;
- $\sigma$ , lo scarto quadratico medio in grammi;
- $r$ , il livello di confidenza ( $1 - \alpha$ );
- $c$ , il numero dei campioni da esaminare;
- $n$ , il numero dei barattoli di un campione;
- i pesi medi rilevati degli  $n$  campioni;

quindi determinare il numero dei campioni che sono fuori controllo e rappresentare graficamente la carta di controllo.

Supponiamo che i dati siano  $\mu = 500$  g,  $\sigma = 15$  g,  $n = 20$ ,  $c = 15$ ,  $r = 99,74\%$  e che le medie dei campioni siano risultate le seguenti:

campione:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
media:	512	490	506	501	511	491	501	496	508	498	508	491	492	508	497

## I Dati

- Entriamo in ambiente Excel, scriviamo le didascalie e mettiamo dei bordi alle celle B3, B4, B5, B6 e B7 per indicare dove inserire i dati come vediamo in figura 1.
- Immettiamo in B3, B4, B5, B6 e B7 i dati proposti dal problema.

	A	B	C	D	E
1	Sul controllo statistico della qualità				
2					
3	$\mu =$	500		la media dichiarata	
4	$\sigma =$	15		lo scarto quadratico medio	
5	$1 - \alpha =$	99,74%		il livello di confidenza	
6	$n =$	20		la numerosità di un campione	
7	$c =$	15		il numero dei campioni	

▲ Figura 1

## I limiti di controllo

- Calcoliamo  $\alpha$ , il livello di significatività, con  $= 1 - B5$  in B9 (figura 2).
- Troviamo  $z_{1-\frac{\alpha}{2}}$ , il limite di confidenza, con  $= INV.T(B9; 1000000)$  in D10.

- Determiniamo i limiti di controllo, traducendo le formule  $LCL = \mu - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$  e  $UCL = \mu + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$  con  $=B3-D10*B4/RADQ(B6)$  e  $=B3+D10*B4/RADQ(B6)$  rispettivamente in B11 e in E11.

	A	B	C	D	E
9	$\alpha =$	0,26%		il livello di significatività	
10	$z$ di $(1 - \alpha/2) =$			3,0114461	
11	LCL =	489,8993		UCL =	510,1007

▲ Figura 2

**La tabella per la carta di controllo**

- Prepariamo una tabella con gli indici dei campioni nella colonna A, scrivendo 1 in A14, 2 in A15 e copiando la zona A14:A15 sino alla A28 (figura 3).
- Registriamo nella colonna B dalla B14 alla B28 le quindici medie dei campioni.

	A	B	C	D	E
13	Campione	Media		Controllo 1	Controllo 2
14	1	512		fuori	sopra
15	2	490		dentro	sotto
16	3	506		dentro	sopra
17	4	501		dentro	sopra
18	5	511		fuori	sopra
19	6	491		dentro	sotto
20	7	501		dentro	sopra
21	8	496		dentro	sotto
22	9	508		dentro	sopra
23	10	498		dentro	sotto
24	11	508		dentro	sopra
25	12	491		dentro	sotto
26	13	492		dentro	sotto
27	14	508		dentro	sopra
28	15	497		dentro	sotto

▲ Figura 3

- Per controllare se ognuna delle medie si trova all'interno della striscia compresa fra *LCL* e *UCL*, scriviamo  $=SE(E(B14 > \$B\$11; B14 < \$E\$11); "dentro"; "fuori")$  in D14 e la copiamo sino alla D28.
- Effettuiamo, inoltre, un controllo qualitativo sulle medie, testando se ognuna di esse è sopra o sotto media. Scriviamo  $=SE(B14 > \$B\$3; "sopra"; SE(B14 < \$B\$3; "sotto"; "pari"))$  in E14 e la copiamo sino alla E28.

**Il conteggio dei campioni**

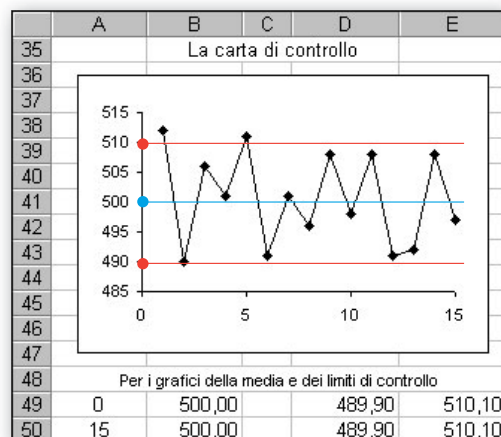
- Contiamo i campioni che hanno media fuori controllo digitando  $=CONTA.SE(D14:D28; "fuori")$  in C31 (figura 4).
- Con  $=CONTA.SE(E14:E28; "sopra")$  in C32 e  $=CONTA.SE(E14:E28; "sotto")$  in C33 valutiamo il numero dei campioni che hanno la media rispettivamente sopra e sotto a quella dichiarata.
- Calcoliamo le corrispondenti percentuali, scrivendo  $=C31/B7$  in D31,  $=C32/B7$  in D32 e  $=C33/B7$  in D33 e dichiarando in formato percentuale la zona D31:D33.

	A	B	C	D	E
30	Il numero dei valori			La percentuale	
31	fuori controllo		2	13,33%	
32	sopra media		8	53,33%	
33	sotto media		7	46,67%	

▲ Figura 4

**La carta di controllo**

- Per tracciare la carta di controllo, evidenziamo la zona A14:B28 e diamo *Inserisci\_grafico*.
- Nella prima finestra di dialogo selezioniamo *Dispers. (XY)* e scegliamo il tipo *Dispersione con coordinate unite da linee*.
- Nelle altre tre confermiamo le scelte di Excel.
- Per sistemare il grafico facciamo clic con il tasto destro del mouse sull'area del disegno e selezioniamo le opportune opzioni per operare le seguenti variazioni: togliamo il colore allo sfondo, cambiamo il colore delle linee e degli indicatori, leviamo la griglia, stabiliamo da 0 a 15 la variazione di *x*.
- Per aggiungere nel grafico la retta della media scriviamo nelle celle A49 e A50 rispettivamente 0 e 15 (scelte come ascisse di due punti della retta) e nelle celle B49 e B50 il riferimento all'ordinata (contenuta in B3). Facciamo clic con il tasto destro del mouse sull'area del disegno, selezioniamo *Dati di origine*, poi *Serie*, facciamo clic su *Aggiungi*, sulla riga *Valori X*, nella zona del foglio A49:A50, nella riga *Valori Y*, nella zona del foglio B49:B50, e infine su *OK*.
- Operiamo in modo simile per le rette dei limiti e al termine vediamo la carta di controllo come in figura 5.



▲ Figura 5

## Esercitazioni

### Sugli stimatori

- 1** Per verificare che la media campionaria è uno stimatore corretto e consistente, costruisci un foglio elettronico che legga i consumi  $A, B, C, D, E$  ed  $F$  di una popolazione e che determini la media e la varianza dei campioni con i valori di  $n$  che vanno da 2 a 5 sia nel caso di campionamento bernoulliano sia di campionamento in blocco.

Per verificare ognuna delle seguenti affermazioni costruisci un foglio elettronico che permetta di inserire i cinque consumi  $A, B, C, D$  ed  $E$  di una popolazione, contenga lo spazio dei campioni di numerosità  $n = 3$ , sia con ripetizione sia senza ripetizione, e determini le grandezze necessarie.

- 2** La media campionaria è uno stimatore più efficiente dello stimatore mediana.
- 3** Lo stimatore mediana campionaria è uno stimatore corretto.
- 4** La varianza campionaria non è uno stimatore corretto, mentre lo è la varianza campionaria corretta.
- 5** I due stimatori della media della popolazione:

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}, \quad X^* = \frac{h \cdot X_1 + k \cdot X_2 + j \cdot X_3}{h + k + j}, \quad \text{con } h, k \text{ e } j \text{ interi e positivi,}$$

se il campionamento avviene con ripetizione, sono corretti e il primo, se  $h, k$  e  $j$  non sono tutti e tre uguali a 1, è più efficiente del secondo.

### Sulla stima puntuale

Costruisci per ognuno dei seguenti casi un foglio elettronico che permetta di assegnare i valori alle grandezze (esprese con le lettere usuali) del primo elenco e che determini, sia in un campionamento bernoulliano sia in uno in blocco, i valori di quelle del secondo elenco (dove  $a$  indica il numero degli elementi di un campione che godono di una certa proprietà).

Prova i fogli con i dati suggeriti.

- 6** Dati  $N, n, \bar{x}, s, a$ , trova  $s_{\bar{x}}, s_F$  e l'errore totale. Prova con  $N = 4000, n = 200, \bar{x} = 15, s = 4, a = 50$ .  
[0,2835, 3,06%, 1134; 0,2763, 2,98%, 1105]
- 7** Dati  $N, \bar{x}, s, s_{\bar{x}}$ , trova  $n$ . Prova con  $N = 4000, \bar{x} = 15, s = 4, s_{\bar{x}} = 0,2$ .  
(Suggerimento. Puoi usare la funzione di Excel *Ricerca obiettivo*.) [401; 365]
- 8** Dati  $N, a, s_F$ , trova  $n$ . Prova con  $N = 10\,000, a = 100, s_F = 1\%$ . [946; 897]

### Sulla stima per intervallo

Per ognuno dei seguenti problemi costruisci un foglio elettronico che possa leggere i dati e trovi quanto richiesto. Oltre alle lettere che indicano le grandezze note, considera che la  $r$  esprime in percentuale il livello di confidenza  $1 - \alpha$ , la  $I$  la stima dell'intervallo di confidenza e  $var$  la varianza di un campione. Supponi, se non diversamente espresso, un campionamento bernoulliano.

Applica poi il foglio con i valori suggeriti.

- 9** Dati  $n, \sigma, s$  e  $I$ , trova, sia con  $\sigma$  sia con  $s$ , il livello di confidenza  $r$ .  
Poni  $n = 12, \sigma = 3, s = 3, I = 3,6$ . (Suggerimento. Usa la funzione di Excel DISTRIB.T.) [96,23%; 92,8%]

- 10** Dati  $N, n, \bar{x}, \sigma, s$  e  $r$ , stima sia in un campionamento bernoulliano, sia in uno in blocco, prima con  $\sigma$  poi con  $s$  gli intervalli della media.

$N = 10000, n = 400, \bar{x} = 88, \sigma = 10, s = 16, r = 95\%$ .

[bernoulliano: ]87,02; 88,98[; ]87,02; 88,98[; in blocco: ]87,03; 88,96[; ]87,03; 88,96[

- 11** Dati  $n, \bar{x}, var$  di un campione, con  $var$  variabile da  $v_1$  a  $v_2$  con passo  $dv$ , segnala, a un livello di fiducia dell' $r\%$ , quando il valore della varianza campionaria rende l'ampiezza dell'intervallo maggiore di un valore assegnato  $I$ . Comunica inoltre l'intervallo di confidenza per ogni valore della varianza.  
Poni  $r = 95, n = 10, \bar{x} = 60, var = 15, I = 5, v_1 = 8, v_2 = 15, dv = 1$ .

[11; ]57,5; 62,5[

- 12** Dati  $r$  e i dieci valori di un campione, determina la stima dell'intervallo di confidenza della media. Prova con  $r = 95\%$  e i seguenti valori: 240 270 264 288 294 312 288 282 252 270.

[260,86; 291,14[

- 13** Dati  $f, \alpha$  ed  $n$  variabile da  $n_1$  a  $n_2$  con passo  $dn$ , indica quando la stima, valutata a un livello di fiducia dell' $r\%$ , è inferiore a *soglia*.  
Poni  $f = 45\%, r = 90, n_1 = 1350, n_2 = 1800, dn = 50$  e *soglia* = 2%.

[[1350; 1650]: non minore; [1700; 1800]: minore]

## La verifica delle ipotesi

Stabilito che:

- un alimento è prodotto in confezioni dal peso medio  $\mu$  e con lo scarto quadratico medio  $\sigma$ ;
- se  $\sigma$  non è noto, compare  $s$ , lo scarto quadratico medio del campione estratto;
- il peso medio in un campione di numerosità  $n$  è  $\bar{x}$ ;
- il peso delle confezioni è una grandezza che si distribuisce normalmente;
- la produzione è verificata al livello di significatività dell' $r\%$  con un test bilaterale a due code;
- i pesi sono espressi in grammi;

per ognuno dei seguenti casi costruisci un foglio elettronico che legga i valori delle grandezze indicate, faccia assumere sedici valori equidistanti alla grandezza variabile nell'intervallo assegnato e dica quando la produzione è sotto controllo.

- 14** Dati:  $\mu, \sigma, n, \bar{x}$  e  $[r_1; r_2]$ . Poni  $\mu = 16, \sigma = 1,6, n = 50, \bar{x} = 16,50, r_1 = 90$  ed  $r_2 = 99$ .

[[90%; 95,40%]:  $H_1$ ; [96%; 99%]:  $H_0$ ]

- 15** Dati:  $\mu, s, \bar{x}, r$  e  $[n_1; n_2]$ . Poni  $\mu = 2300, s = 80, n = 50, \bar{x} = 2350, r = 99\%, n_1 = 10$  ed  $n_2 = 25$ .

[[10; 21]:  $H_0$ ; [22; 25]:  $H_1$ ]

- 16** Dati:  $\mu, n, \bar{x}$  e  $[s_1; s_2]$ . Poni  $\mu = 180, n = 40, \bar{x} = 192, s_1 = 30$  ed  $s_2 = 90$ .

[[30; 58]:  $H_1$ ; [62; 90]:  $H_0$ ]

- 17** Dati:  $n_1, s_1$  e  $\mu_{\bar{X}_1}$  di un primo campione,  $n_2, s_2$  e  $\mu_{\bar{X}_2}$ , variabile in  $[\mu_1; \mu_2]$ , di un secondo campione ed  $r$ .  
Poni  $n_1 = 12, s_1 = 0,3, \mu_{\bar{X}_1} = 6,5, s_2 = 0,32, \mu_1 = 6,2, \mu_2 = 6,8$  ed  $r = 95$ .

[[6,2; 6,24] e [6,76; 6,8]:  $H_1$ ; [6,28; 6,72]:  $H_0$ ]