

ESERCITAZIONE N° 4

4A

Es. 3.21 PAG. 67

y = % di seggi occupata da donne (variabile quantitativa)

NAZIONE	y
N	38
I	33
A	28
IR	14
S	45
C	24
G	11
USA	15
SV	25
PB	34

$$n = 10$$

$$\bar{y} = ?$$

$$s = ?$$

$$\bar{y} = \frac{\sum x_i}{n} = \frac{(38+33+\dots+34)}{10} = \frac{267}{10} = \underline{26.7}$$

$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{1112.01}{9}} = \sqrt{123.56} = \underline{11.11}$$

(*)	38	$1^2 = 11,3^2 = 127,69$
	33	$1^2 = 6,3^2 = 39,69$
	28	$1^2 = 1,3^2 = 1,69$
	14	$1^2 = -12,7^2 = 161,69$
	45	$-26,7^2 = 18,3^2 = 334,89$
	24	$1^2 = -2,7^2 = 7,29$
	11	$1^2 = -15,7^2 = 246,49$
	15	$1^2 = -11,7^2 = 136,89$
	25	$1^2 = -1,7^2 = 2,89$
	34	$1^2 = 7,3^2 = 53,29$
		<u>1112.01</u>

Es. 3.23 PAG. 68

$\mu = 50000$ distr. b. campanulone
 $\sigma = 10000$

a) 68% $\rightarrow \mu - 1\sigma$ $\mu + 1\sigma$
 $50000 - 10000$ $50000 + 10000$
 40000 60000

95% $\rightarrow 50000 - 2(10000)$ $50000 + 2(10000)$
 30000 70000

TOT. $\rightarrow 50000 - 3(10000)$ $50000 + 3(10000)$
 20000 80000

b) 100000 outlier? Sì, ricade a più di 3 deviazioni standard dalla media.

↳ 5 dev. st sopra la media $\rightarrow$ insolito se distrib. campanulone

$$Z = \frac{100000 - 50000}{10000} = 5$$

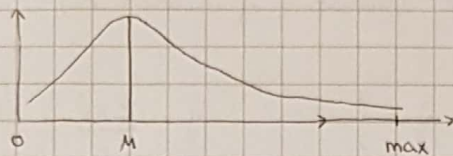
Es. 3.29 PAG. 68

$$\mu = 10449$$

$$\sigma = 7489$$

$$\max = 336240$$

a) Forma distrib? ASIMMETRICA POSITIVA (maggiore coda dx)



b) Max e' outlier perche' ricade a 43.5 dev.st. sopra la media.

$$z = \frac{336240 - 10449}{7489} = 43.5$$

$$\text{osservazione} \leftarrow \underbrace{\text{max} - \mu}_{\sigma}$$

Es. 3.33 PAG. 69

DATI SUL LIBRO! =>

Se escludo outlier (valore 44) :

- media diminuisce
 - dev. st. diminuisce (variabilita' dati diminuisce)
- L7 e anche range diminuisce

N.B. Mediana e altri indici di posizione rimangono uguali (sono meno influenzati da outlier)

Es. 3.45 PAG. 70

$$\mu = 8.3$$

$$\sigma = 3.3$$

a) $z = \frac{x_i - \mu}{\sigma} = \frac{21.1 - 8.3}{3.3} = 3.878 \Rightarrow$ il valore 21.1 e' 3.878 dev. st. sopra la media

b) $z = \frac{5.8 - 8.3}{3.3} = -0.76 \Rightarrow$ il valore 5.8 e' 0.76 dev. st. sotto la media

c) $x_1 = 16.5 \quad z = \frac{16.5 - 8.3}{3.3} = 2.48 \Rightarrow$ sopra la media, oltre 95% ma compreso nella tot.

$x_2 = 20.1 \quad z = \frac{20.1 - 8.3}{3.3} = 3.58 \Rightarrow$ sopra la media (non compreso nella distrib.)

↓
NB Problema non dice che distrib. e' campanulare

$$\mu = 4998$$

$$\sigma = 1786$$

a) $x_i = 4222$

$$z = \frac{4222 - 4998}{1786} = -0.43 \Rightarrow 0.43 \text{ dev. st. sotto la media}$$

b) $x_i = 11067$

$$z = \frac{11067 - 4998}{1786} = 3.40 \Rightarrow \text{sopra la media}$$

c) Se distribuzione campanulone valore 11067 sarebbe outlier perché totalità distrib. sarebbe compresa nell'intervallo $\mu - 3\sigma$ e $\mu + 3\sigma$ e 11067 sarebbe 3.4 dev. st. sopra la media

Es. 4.21 PAG 104

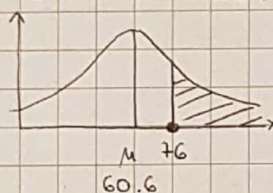
$\mu = 60.6$

$\sigma = 18.9$

distrib. normale

a) proporzione uso > 76 ?

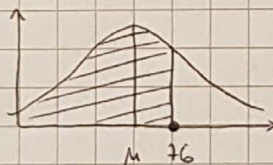
Quale area sotto la curva devo trovare?



$$z = \frac{76 - 60.6}{18.9} = 0.81$$

TAVOLA A $\rightarrow$ 0.2090 probabilità

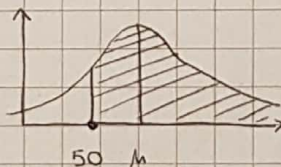
$L > 0.21$

• proporzione uso < 76 ?

$$z = 0.81 \quad \mu > 0.2090 \text{ prob.}$$

$$1 - 0.2090 = 0.791$$

$L > 0.79$

• proporzione uso > 50 ?

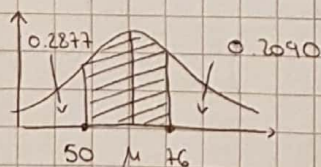
$$z = \frac{50 - 60.6}{18.9} = -0.56$$

TAVOLA A $\rightarrow$ 0.2877

$$1 - 0.2877 = 0.7123$$

$L > 0.71$

• proporzione intervallo 50-76?

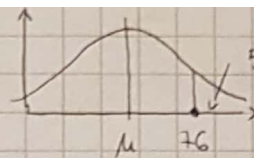


$$\Rightarrow 1 - (0.2877 + 0.2090) = 1 - 0.4967 = 0.5033$$

$$\Downarrow$$

$$0.50$$

- b) $\sigma = 18.9$ litri.
5% usa > 76 litri a settimana $\rightarrow$



5% $\rightarrow$ 0.05 probab.
 $\rightarrow$ TAVOLA A
 $z = 1.64$

$$\mu + z\sigma = 76$$

$$\mu + 1.64 \cdot 18.9 = 76$$

$$\mu + 31 = 76$$

$$\mu = 76 - 31 = \textcircled{45} \text{ litri} \rightarrow \text{in galloni } 11.88$$

c) Asimmetrica positiva

Es. 4.24 PAG. 105

$$\mu = 2500$$

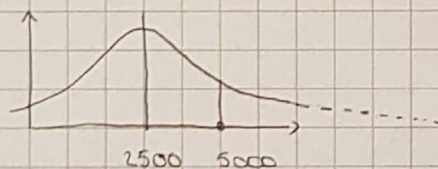
$$\sigma = 1500$$

$$x_i = 5000$$

a) z-score corrispondente a x_i ? $z = \frac{x_i - \mu}{\sigma} = \frac{5000 - 2500}{1500} = 1.67$

b) TAVOLA A $\rightarrow$ probab. corrispondente a $z = 1.67$
0.0475

c) Asimmetrica positiva (coda di dx maggiore)



2500 5000
è a 1.67 dev. st. dalla media

0 2500 2500

$\rightarrow$ sotto lo zero non posso andare

Es. 4.30 PAG. 106

$$\mu = 13.6$$

$$\sigma = 3.0$$

$$\bar{y} = ?$$

$$\sigma_{\bar{y}} = ?$$

a) $n = 9$

$$\bar{y} = 13.6$$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{3.0}{\sqrt{9}} = 1$$

b) $n = 36$

$$\bar{y} = 13.6$$

$$\sigma_{\bar{y}} = \frac{3.0}{\sqrt{36}} = 0.5$$

c) $n = 100$

$$\bar{y} = 13.6$$

$$\sigma_{\bar{y}} = \frac{3.0}{\sqrt{100}} = 0.3$$

L'errore standard diminuisce all'aumentare di $n \rightarrow$ ampiezza della distrib. campionaria diminuisce ($\bar{y}$ tende a unire meno tra tutti i possibili campioni).

DISTRIB. CAMPIONARIA DELLA PROPORZIONE CAMPIONARIA dell'uscita TESTA



distribuzione di probabilità che fornisce le probabilità per i possibili valori che la statistica può assumere.

↳ riportare le probabilità per tutte le possibili combinazioni

 $y = 1$ TESTA
 $y = 0$ CROCE
a) Una volta ($n=1$)

		prob.
T	(1)	$1/2$
C	(0)	$1/2$

$$P(0) = P(1) = 0.5$$

b) Due volte ($n=2$)

		prob.
TT	(1,1)	$1/4$
TC	(1,0)	$1/4$
CT	(0,1)	$1/4$
CC	(0,0)	$1/4$

Per un campione di ampiezza 2 le proporz. possono essere 1, 0.5, 0

$$P(1) = 0.25$$

$$P(0.5) = 0.50$$

$$P(0) = 0.25$$

c) Tre volte ($n=3$)

TTT	(1,1,1)	prob.
TTC	(1,1,0)	$1/8$
TCT	(1,0,1)	$3/8$
CTT	(0,1,1)	
TCC	(1,0,0)	
CTC	(0,1,0)	$3/8$
CCT	(0,0,1)	
CCC	(0,0,0)	

Per un campione di ampiezza 3 le proporzioni possono essere 1, $2/3$, $1/3$, 0

$$P(1) = 1/8 = 0.125$$

$$P(2/3) = 3/8 = 0.375$$

$$P(1/3) = 3/8 = 0.375$$

$$P(0) = 1/8 = 0.125$$

...

e) Al crescere dei lanci (di n) la forma della distribuzione campionaria diventa sempre più campanulata.