STATISTICA A – K (63 ore)

Marco Riani

mriani@unipr.it

http://www.riani.it

Esercizio

- Calcolare la media geometrica, la media quadratica e la media troncata con $\alpha=0.4$ della distribuzione di un campione di 15 famiglie riportata nella tabella che segue

n. componenti	frequenza
2	2
3	4
4	4
5	2
6	2
7	1
Totale	15

Soluzione

- Media geometrica = 3.81
- Media aritmetica = 4.07
- Media quadratica = 4.31

- Media troncata con α=0.4 implica che devo togliere il 20% dei valori (i tre) più piccoli ed il 20% dei valori (i tre più grandi)

Soluzione: calcolo della media troncata

Distribuzione originale

n. componenti	frequenza
2	2
3	4
4	4
5	2
6	2
7	1
Totale	15

Distribuzione troncata con α=0.4

n. componenti	frequenza
3	3
4	4
5	2
Totale	9

Media troncata con $\alpha=0.4 = M_{[0.4]}=3.89$

Esercizio

ESERCIZIO 12

Nella seguente tabella è riportata la distribuzione dei dipendenti di una grande azienda in base alla retribuzione lorda mensile:

retribuzioni	Numero di dipendenti
1000 – 1200	30
1200 – 1500	130
1500 – 2000	150
2000 – 2500	50
2500 – 3500	30
3500 - 5000	20

i) Si calcoli lo scostamento quadratico medio e il MAD delle retribuzioni e si commenti il significato dei risultati ottenuti.

ii) Si dica, motivando la risposta, quale trasformazione subirebbero la media, la mediana, lo scostamento quadratico medio e il MAD delle retribuzioni, calcolati ai punti precedenti, se:

a) tutte le retribuzioni fossero aumentate di 50 euro,

b) tutte le retribuzioni fossero incrementate del 7%

c) tutte le retribuzioni fossero incrementate del 2% e, dopo questo aumento, aumentate di 100 euro .

iii) Si rappresenti graficamente la suddetta distribuzione e si dica quali informazioni si possono ricavare.

iv) Si calcoli l'indice di asimmetria di Fisher e lo si commenti

Soluzione calcolo di M e σ

classe di retribuzione		ni	f_i	Fi	xi	xi*fi	(xi-M)^2fi
1000	1200	30	0.073171	0.073171	1100	80.4878	41158.5366
1200	1500	130	0.317073	0.390244	1350	428.0488	79268.2927
1500	2000	150	0.365854	0.756098	1750	640.2439	3658.53659
2000	2500	50	0.121951	0.878049	2250	274.3902	19512.1951
2500	3500	30	0.073171	0.95122	3000	219.5122	96768.2927
3500	5000	20	0.04878	1	4250	207.3171	280975.61
		410	1			1850	521341.463

$$\overline{x} = \frac{\sum_{i=1}^{r} x_i n_i}{n} = \sum_{i=1}^{r} x_i f_i = M$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{r} (x_i - M)^2 n_i}{n}}$$

- M=1850 € σ=(521341.463)$^{0.5}$=722.04€

Soluzione: calcolo della mediana

classe di retribuzione		ni	f_i	Fi
1000	1200	30	0.073171	0.073171
1200	1500	130	0.317073	0.390244
1500	2000	150	0.365854	0.756098
2000	2500	50	0.121951	0.878049
2500	3500	30	0.073171	0.95122
3500	5000	20	0.04878	1
		410	1	

$$Me = \overline{x}_s + \frac{\overline{\overline{x}}_s - \overline{x}_s}{F(x_s) - F(x_{s-1})}\left[0{,}50 - F(x_{s-1})\right]$$

Me=1650 € (Ip. Variazione lineare all’interno delle classi)

Soluzione: calcolo del MAD

classe di retribuzione		ni	f_i	Fi	xi
1000	1200	30	0.073171	0.073171	1100
1200	1500	130	0.317073	0.390244	1350
1500	2000	150	0.365854	0.756098	1750
2000	2500	50	0.121951	0.878049	2250
2500	3500	30	0.073171	0.95122	3000
3500	5000	20	0.04878	1	4250
		410	1		

$$MAD = Me(|x_i - Me|)$$

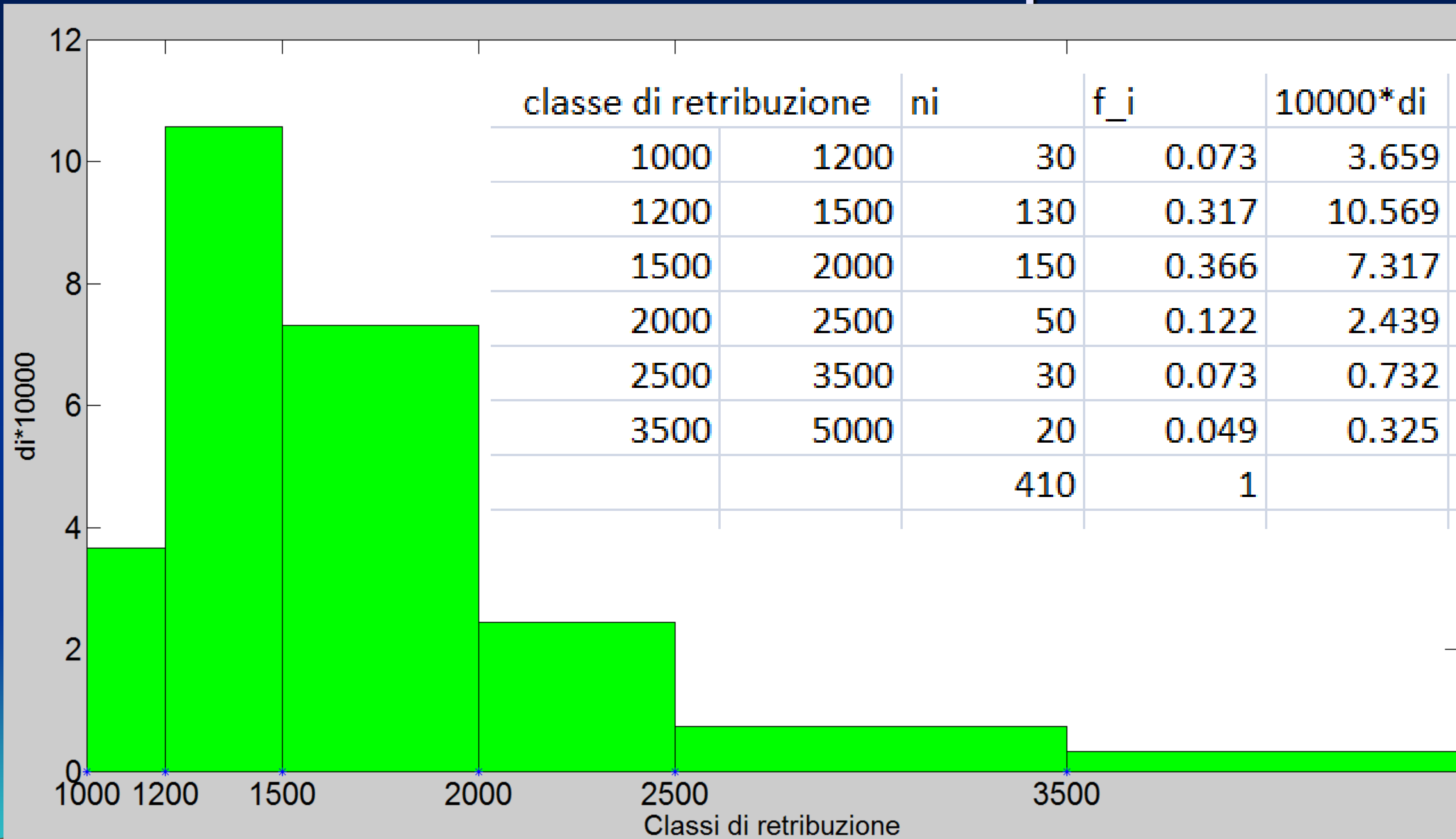
Me=1650 €

Scostamenti da Me in valore assoluto ordinati

|xi-Me|	f_i	F_i
100	0.365854	0.365854
300	0.317073	0.682927
550	0.073171	0.756098
600	0.121951	0.878049
1350	0.073171	0.95122
2600	0.04878	1

MAD=300 €

Soluzione: rappresentazione grafica tramite le densità di frequenza



classe di retribuzione		ni	f_i	10000*di
1000	1200	30	0.073	3.659
1200	1500	130	0.317	10.569
1500	2000	150	0.366	7.317
2000	2500	50	0.122	2.439
2500	3500	30	0.073	0.732
3500	5000	20	0.049	0.325
		410	1	

- Asimmetria positiva $M > Me > M_0$

Soluzione: indice di asimmetria di Fisher

$$\gamma(\mathbf{X}) = \frac{\sum_{i=1}^{r}(x_i - M(\mathbf{X}))^3 f_i}{[\sigma(\mathbf{X})]^3}$$

xi	xi*fi	(xi-M)^2fi	|xi-Me|	(xi-M)^3fi
1100	80.4878	41158.5366	550	-30868902.44
1350	428.0488	79268.2927	300	-39634146.34
1750	640.2439	3658.53659	100	-365853.6585
2250	274.3902	19512.1951	600	7804878.049
3000	219.5122	96768.2927	1350	111283536.6
4250	207.3171	280975.61	2600	674341463.4
	1850	521341.463		722560975.6

- $\gamma(X)=1.9195$ ➔ asimmetria positiva

Trasformazioni

ii) Si dica, motivando la risposta, quale trasformazione subirebbero la media, la mediana, lo scostamento quadratico medio e il MAD delle retribuzioni, calcolati ai punti precedenti, se:

a) tutte le retribuzioni fossero aumentate di 50 euro,
b) tutte le retribuzioni fossero incrementate del 7%
c) tutte le retribuzioni fossero incrementate del 2% e, dopo questo aumento, aumentate di 100 euro .

- M(a+bX)=a+bM(X) Me(a+bX)=a+bMe(X)

a) Se tutte le retribuzioni aumentano di 50 € (a=50; b=1) M e Me aumentano di 50 €

b) Se tutte le retribuzioni fossero incrementate del 7% (a=0; b=1.07) M e Me aumentano del 7%

c) b=1.02 a=100€

Me(100+1.02X)=100+1.02Me(X)

M(100+1.02X)=100+1.02M(X)

Trasformazioni

ii) Si dica, motivando la risposta, quale trasformazione subirebbero la media, la mediana, lo scostamento quadratico medio e il MAD delle retribuzioni, calcolati ai punti precedenti, se:

a) tutte le retribuzioni fossero aumentate di 50 euro,
b) tutte le retribuzioni fossero incrementate del 7%
c) tutte le retribuzioni fossero incrementate del 2% e, dopo questo aumento, aumentate di 100 euro .

- MAD(a+bX)=bMAD(X) σ(a+bX)=bσ(X)

a) Se tutte le retribuzioni aumentano di 50 € (a=50; b=1) MAD e σ rimangono invariati

b) Se tutte le retribuzioni fossero incrementate del 7% (a=0 b=1.07) MAD e σ aumentano del 7%

c) b=1.02 a=100€

MAD(100+1.02X)=1.02MAD(X)
σ(100+1.02X)=1.02σ(X)

Esercizio

- Si considerino i valori del reddito annuo di 16 individui come segue (dati in migliaia di Euro):
- 18; 18; 22; 23; 1; 5; 7; 13; 14; 15; 2400; 25; 27; 28; 29; 21.
- Si calcoli e si interpreti il valore della mediana e della media troncata con α=0.5.
- Si dica perché in questo caso la media troncata è preferibile rispetto alla media aritmetica.
- Si scriva e si rappresenti graficamente la distribuzione di massima concentrazione.

i	x_i
1	1
2	5
3	7
4	13
5	14
6	15
7	18
8	18
9	21
10	22
11	23
12	25
13	27
14	28
15	29
16	2400

- Me=?

Me= (18+21)/2 =19.5

i	x_i
5	14
6	15
7	18
8	18
9	21
10	22
11	23
12	25

- $M_{[0,5]}$?

Media troncata= 19.5

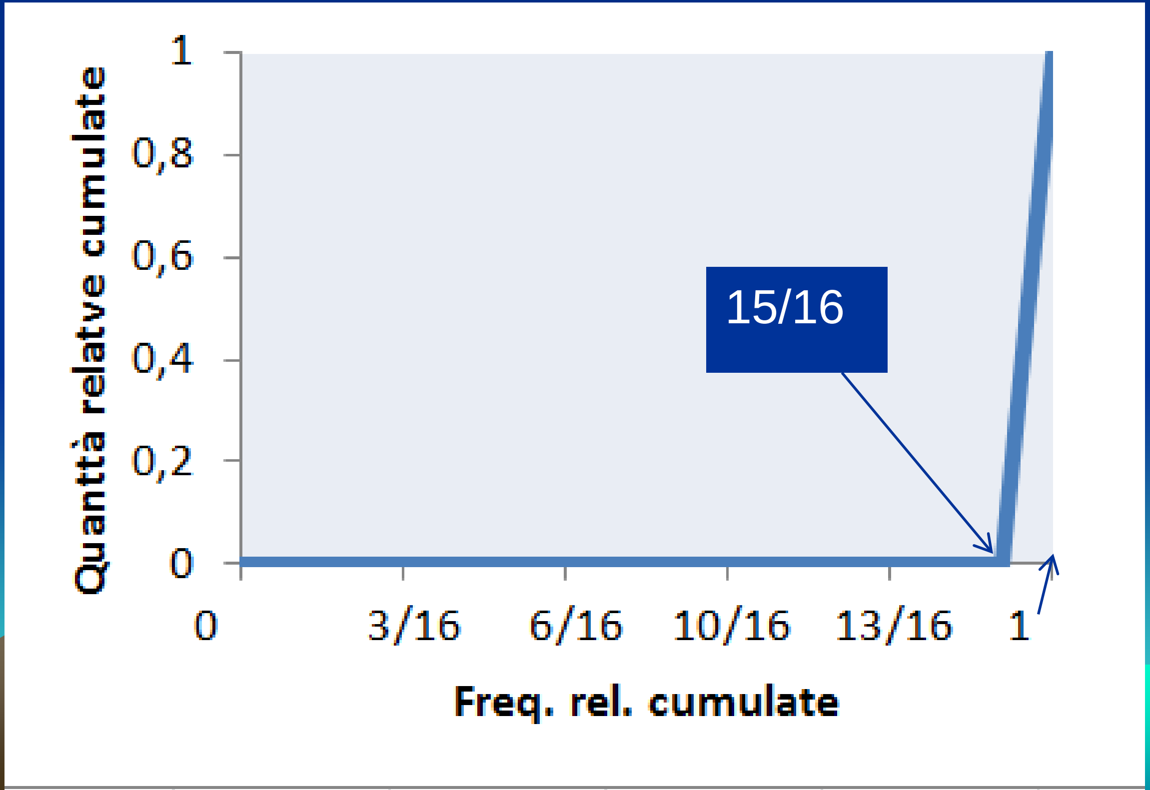
Commento

L'unità che presenta un reddito pari a 2400 è un chiaro valore anomale. Dato che la media aritmetica è influenzata da questo valore anomale è preferibile sintetizzare i dati tramite la mediana e/o la media troncata.

i	x_i
1	1
2	5
3	7
4	13
5	14
6	15
7	18
8	18
9	21
10	22
11	23
12	25
13	27
14	28
15	29
16	2400

Distribuzione di max concentrazione

xi	ni
0	15
2666	1
	16



Esercizio

- Si considerino i valori del reddito annuo di 14 individui come segue (dati in migliaia di Euro):
- 42; 13; 22; 23; 1; 5; 7; 13; 14; 15; 2400; 28; 29; 21.
- 1) Si calcoli l'indice di asimmetria di Gini commentando i risultati ottenuti
- 2) Si dica se in tale esempio è preferibile sintetizzare i dati con la media oppure con la mediana oppure con una media troncata

Indice di asimmetria di Gini

	$x_{(i)}$	$\lvert x_i\text{-Me}\rvert$
reddito 1	1	17
reddito 2	5	13
reddito 3	7	11
reddito 4	13	5
reddito 5	13	5
reddito 6	14	4
reddito 7	15	3
reddito 8	21	3
reddito 9	22	4
reddito 10	23	5
reddito 11	28	10
reddito 12	29	11
reddito 13	42	24
reddito 14	2400	2382
	2.633	2497

- M=188,07
- Me=(15+21)/2=18

$$S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me|}{n}.$$

S_{Me}=178,3571

$$AS_2 = \frac{M - Me}{S_{Me}}$$

AS_2= indice di asimm. di Gini=0,95

Commento

- As. Positiva molto elevata (95% del valore massimo possibile) causata dalla presenza di un chiaro valore anomalo (unità con reddito pari a 2400)

Esercizio

- I valori dell'ammontare dei debiti di 5 clienti di una banca (in migliaia di Euro) sono risultati i seguenti:

Cliente	Debiti
1	20
2	50
3	8
4	60
5	3

1. Si determini la differenza media assoluta dell’ammontare dei debiti e se ne commenti il significato.

2. Si scriva la distribuzione di massima concentrazione dei debiti.

3. Si determini l’indice di concentrazione dei debiti e lo si commenti.

4. Si illustri come cambia l’indice di concentrazione dei debiti se la banca decide di abbonare a ciascun cliente la somma di 1000 Euro.

5. Si illustri come cambia l’indice di concentrazione dei debiti se la banca decide di applicare a ciascun cliente una commissione pari al 2% dell’importo totale dei debiti.

Si determini la differenza media assoluta dell'ammontare dei debiti e se ne commenti il significato

$$\Delta = \frac{2\sum_{i=1}^{n} x_{(i)}[2i-(n+1)]}{n(n-1)}$$

Cliente	Debiti (x_i)
1	20
2	50
3	8
4	60
5	3

i	x_(i)	2*x_(i)*(2i-(n+1))
1	3	-24
2	8	-32
3	20	0
4	50	200
5	60	480
	141	624

Δ =differenza media assoluta =624/20=31,2

Distribuzione Max. concentrazione

Cliente	Debiti (x_i)
1	20
2	50
3	8
4	60
5	3

x_i	n_i
0	4
141	1
	5

Calcolo indice di concentrazione

$$R = \frac{n+1}{n-1} - \frac{2}{n-1}\sum_{i=1}^{n} q'_i.$$

		x_(i)	fi	q_i	q'_i
1					
2		x_(i)	fi	q_i	q'_i
3	debito 1	3	0,2	0,021276596	0,021276596
4	debito 2	8	0,2	0,056737589	0,078014184
5	debito 3	20	0,2	0,141843972	0,219858156
6	debito 4	50	0,2	0,354609929	0,574468085
7	debito 5	60	0,2	0,425531915	1
8		141			1,8936170

R=(6/4)-(2/4)*1,893617= 0,5531915

$$R = \frac{\Delta}{2M}$$

R=(31,2/ 2*28,2)=0,5531915

Si illustri come cambia l'indice di concentrazione dei debiti se la banca decide di abbonare a ciascun cliente la somma di 1000 Euro

$$R = \frac{\Delta}{2M}$$

	x_(i)
debito 1	3
debito 2	8
debito 3	20
debito 4	50
debito 5	60

- Abbuono debiti per 1000 Euro ➔
- Sottraggo 1 alla distribuzione degli x_i ➔ Δ non cambia M diminuisce ➔ il rapporto R aumenta

Si illustri come cambia l'indice di concentrazione dei debiti se la banca decide di applicare a ciascun cliente una commissione pari al 2% dell'importo totale dei debiti

$$R = \frac{\Delta}{2M}$$

	x_(i)
debito 1	3
debito 2	8
debito 3	20
debito 4	50
debito 5	60

- Commissione 2% → moltiplico ciascun x_i per 1.02 → R non cambia

Esercizio

Nella seguente tabella sono riportati, per alcune tipologie di gelato, il peso in gr. (Y) ed il contenuto energetico in kcal (X):

gelato	Peso	Contenuto energetico
Gelato nero Perugina	68	314
Maxibon	100	339
Coppa del nonno	70	179
Mottarello	52	172
Fortunello	80	237

I) Si determini la variabile Z = «contenuto energetico per 100 gr. di prodotto» per ciascun tipo di gelato.
II) Spiegando il significato dei simboli utilizzati, si scrivano le espressioni della media aritmetica e dello scostamento quadratico medio del la variabile Z .
III) Si calcoli il valore degli indici definiti al punto precedente e se ne illustri il significato.

Gelati	Peso (Y)	Contenuto energetico (X)	Z=(X/Y)·100
Gelato nero Perugina	68	314	461,76
Maxibon	100	339	339,0
Coppa del nonno	70	179	255,71
Mottarello	52	172	330,77
Fortunello	80	237	296,25
	370	1241	

$$M(Z) = \frac{\sum_{i=1}^{5} z_i \times y_i}{\sum_{i=1}^{5} y_i} = \frac{\sum_{i=1}^{5} x_i}{\sum_{i=1}^{5} y_i} \times 100$$

$$\sigma(Z) = \sqrt{\frac{\sum_{i=1}^{n} (z_i - M_Z)^2 \cdot y_i}{\sum_{i=1}^{n} y_i}}$$

M(Z) = (1241/370) 100 = 335,4

Scostamento quadratico medio:

$$\sigma(Z) = \sqrt{\frac{(461{,}76 - 335{,}4)^2 \cdot 68 + \ldots + (296{,}25 - 335{,}4)^2 \cdot 80}{370}} = 66{,}88$$

Esercizio

La seguente distribuzione di frequenze riporta l'ammontare (in euro) della spesa effettuata in una settimana dai titolari della carta fedeltà di un ipermercato.

Classi di spesa (€)	Frequenza
Sino a 50	381
50 – 100	482
100 – 200	264
200 – 500	58
Oltre 500	15

I) Si calcolino i quartili della distribuzione della spesa e se ne illustri l'interpretazione.
II) Si disegni il box plot della distribuzione della spesa.
III) Si illustrino tutte le informazioni traibili dal box plot disegnato al punto precedente.

$$x_z = \overline{x}_s + \frac{\overline{\overline{x}}_s - \overline{x}_s}{F(x_s) - F(x_{s-1})}\left[z - F(x_{s-1})\right].$$

i)

Classi di spesa (€)	n_i	f_i	F(x)
Sino a 50	381	0,3175	0,3175
50 – 100	482	0,4017	0,7192
100 – 200	264	0,2200	0,9392
200 – 500	58	0,0483	0,9875
Oltre 500	15	0,0125	1
	1200		

$x_{25\%}$ = 0 + (50/0,3175)·0,25 = 39,37

$x_{50\%}$ =50+(50/0,4017)·(0,5 – 0,3175)=72, 716

$x_{75\%}$ =100 + (100/0,22)·(0,75 – 0,7192) = 114

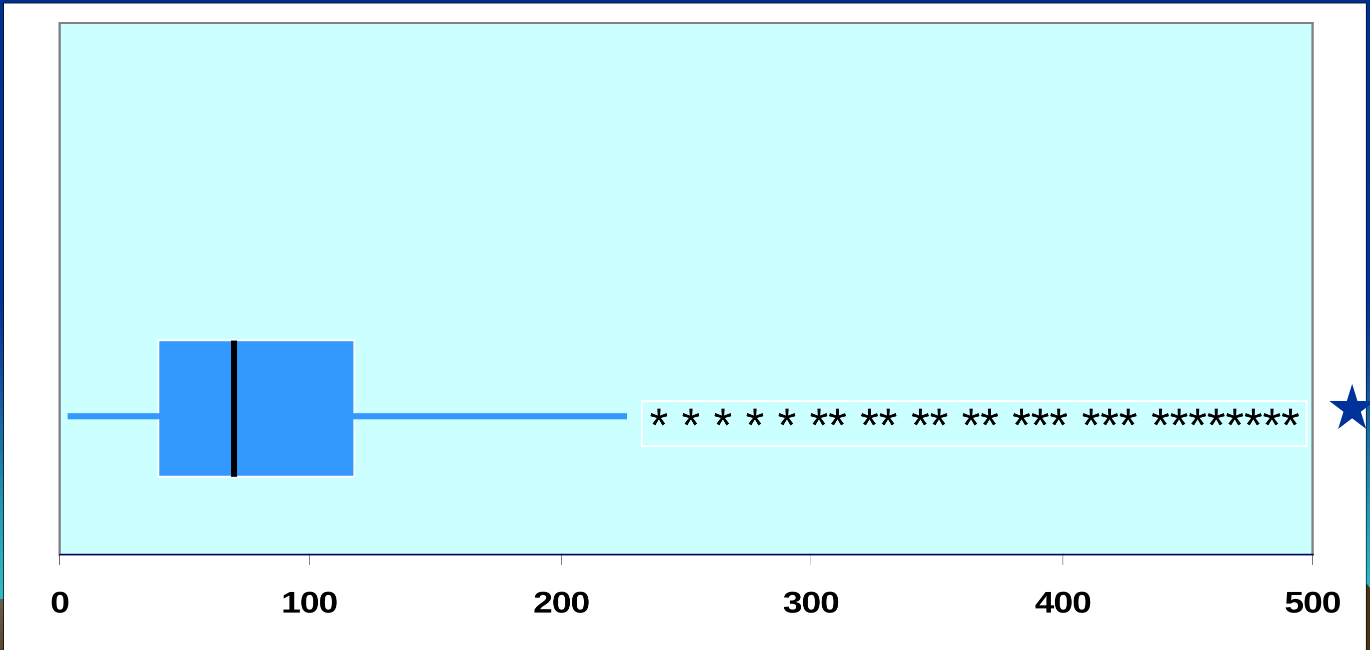
ii) DI = 114 – 39,37 = 74,63

t. inf intermedio: 39,37 – 1,5·74,63 = -72,575

t. sup intermedio: 114 + 1,5·74,63 = 225,94

Punto t. inf = 0

Punto t. sup. = 225,94



Classi di spesa (€)	n_i
Sino a 50	381
50 – 100	482
100 – 200	264
200 – 500	58
Oltre 500	15
	1200

Esercizio

Gli studenti immatricolati in una Facoltà di Economia nell'anno accademico 2006-2007 sono stati classificati in base al voto alla maturità (X) e al numero di esami sostenuti al primo anno (Y):

X \ Y	≤ 2	3 – 4	5 – 6	totale
60 – 70	40	80	30	150
71 – 80	60	150	90	300
81 – 90	20	110	220	350
91 - 100	0	60	120	180
totale	120	400	460	980

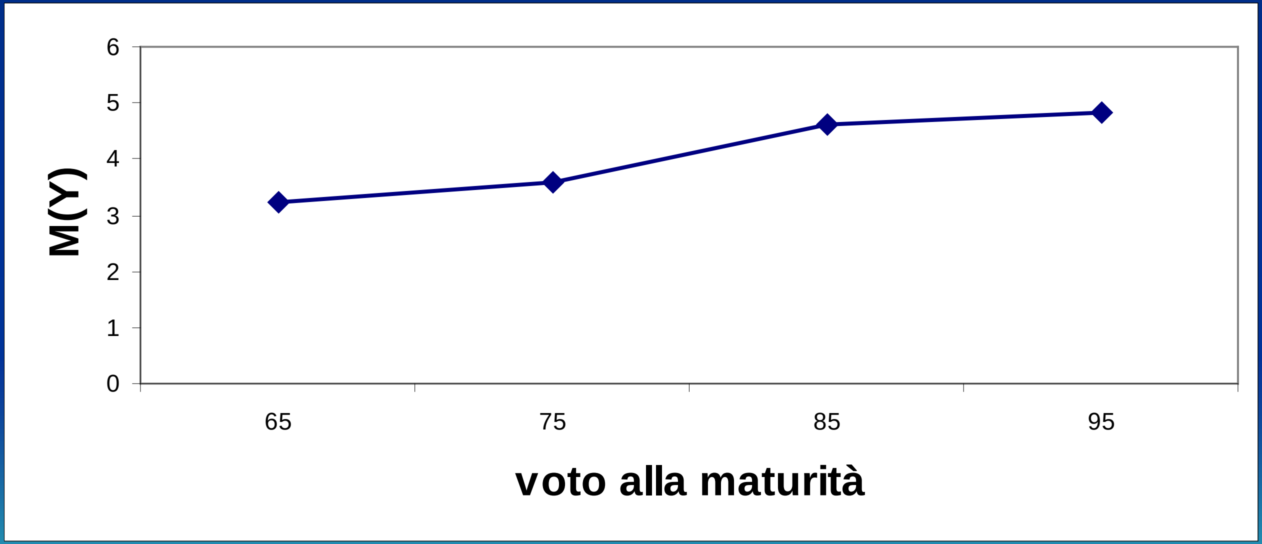
I) calcolino le medie parziali di Y e si commenti il significato di una di esse.
II) Si riportino le medie ottenute al punto precedente in un grafico opportuno.
III) Applicando un'opportuna proprietà della media aritmetica, si determini il numero medio di esami sostenuti da tutti gli studenti.

i) $M(Y)_1 = (1 \cdot 40 + 3{,}5 \cdot 80 + 5{,}5 \cdot 30)/150 = 3{,}23$

- $M(Y)_2 = 3{,}6$
- $M(Y)_3 = 4{,}6$
- $M(Y)_4 = 4{,}83$

X \ Y	≤ 2	3 – 4	5 – 6	totale
60 – 70	40	80	30	150
71 – 80	60	150	90	300
81 – 90	20	110	220	350
91 - 100	0	60	120	180
totale	120	400	460	980

ii)



iii) (proprietà associativa)

$M = (3{,}23 \cdot 150 + 3{,}6 \cdot 300 + 4{,}6 \cdot 350 + 4{,}83 \cdot 180)/980 = 4{,}12$

Esercizio

Una ditta che produce elettrodomestici per controllare la qualità del proprio output fa ispezionare in una settimana un campione di 1.000 lavatrici. Nella tabella seguente sono riportati i risultati dell'ispezione per tipo di difetto:

Tipo di difetto	n. lavatrici
Rifiniture	230
Chiusura	80
Verniciatura	610
Parti meccaniche	60
Altro	20
Totale	1.000

i) Si disegni il diagramma di Pareto per tipo di difetto.

Supponendo che il costo unitario di ogni intervento per eliminare i difetti di produzione sia il seguente:

verniciatura: 3,7 euro rifiniture: 12,30 euro chiusura: 9,60 euro

parti meccaniche: 35,40 euro

altro: 5,30 euro,

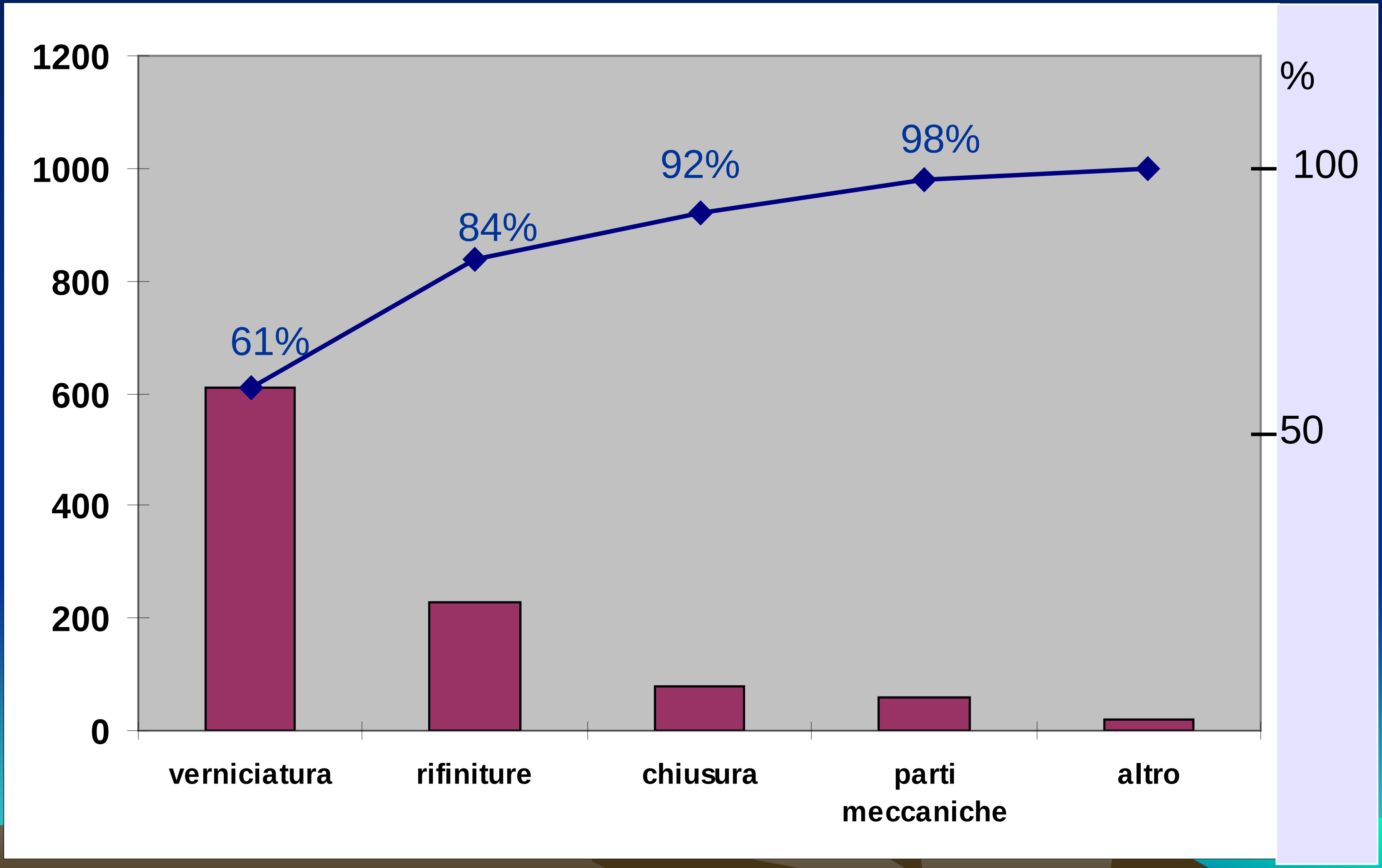
ii) si disegni il diagramma di Pareto per costo complessivo dei difetti.

iii) Si commentino i due grafici in termini comparati.

Ordinamento non crescente in base alla frequenza del tipo di difetto

Tipo di difetto	n. lavatrici	f_i	F(x)
Verniciatura	610	0,61	0,61
Rifiniture	230	0,23	0,84
Chiusura	80	0,08	0,92
Parti meccaniche	60	0,06	0,98
Altro	20	0,02	1
Totale	1.000	1,00	

Diagramma di Pareto per tipo di difetto



Ordinamento non crescente in base alla frequenza del costo totale del tipo di difetto

Tipo di difetto	costo totale	f_i	F(x)
Rifiniture	2829	0,35	0,35
Verniciatura	2257	0,28	0,63
Parti meccaniche	2124	0,26	0,89
Chiusura	768	0,10	0,99
Altro	106	0,01	1
Totale	8084	1,00	

230·12,3

Diagramma di Pareto per costo complessivo dei difetti

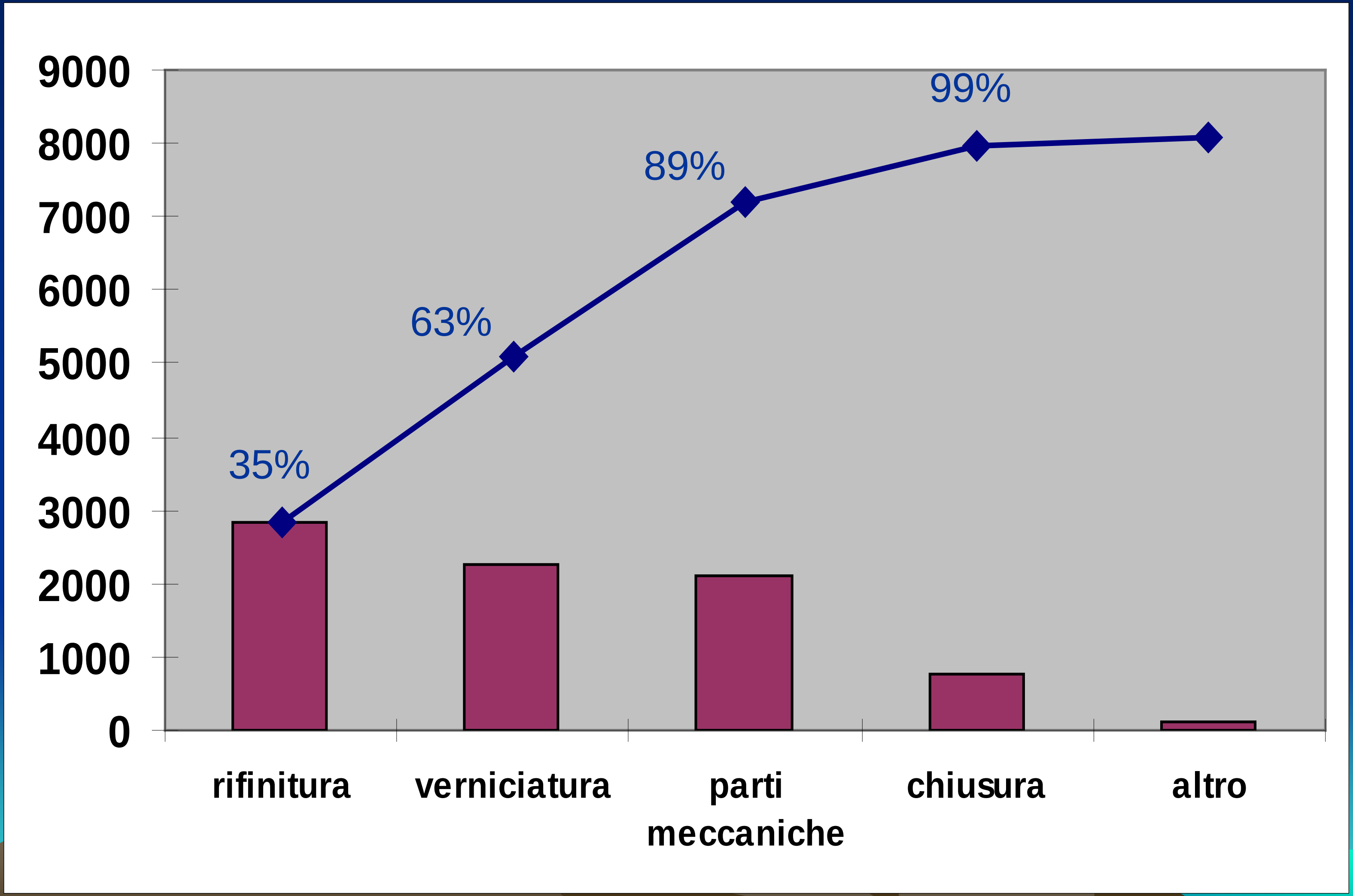
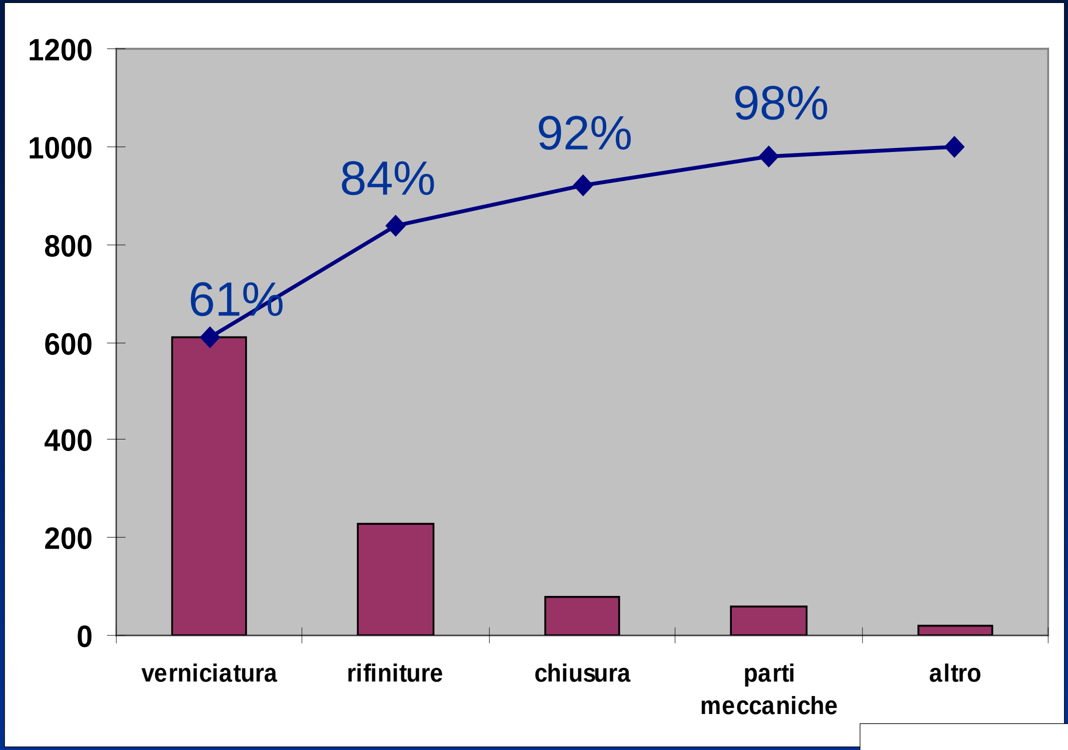


Diagramma di Pareto per tipo di difetto



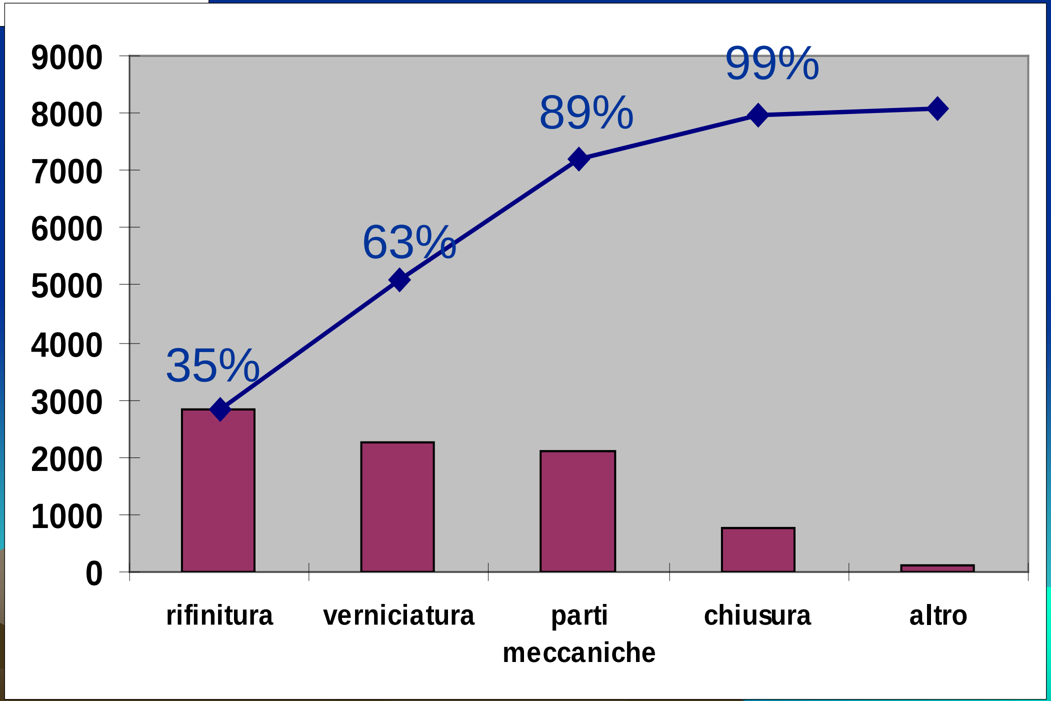


Diagramma di Pareto per costo complessivo dei difetti

Esercizio

- Data la seguente serie storica di NI dei matrimoni civili dal 2001 al 2006 con base 2001=100
  - si calcoli la serie dei NI con base 2003=100
  - Sapendo che i matrimoni civili nel 2005 sono stati pari a 4552 si ricostruisca la serie storica originaria

2001	100
2002	100.5
2003	95.8
2004	101.6
2005	105.5
2006	103.5

Soluzione

	Input	Input	Obiettivo	Obiettivo
2001	1	X_{01}/X_{01}	X_{01}/X_{03}	1.044
2002	1.005	X_{02}/X_{01}	X_{02}/X_{03}	1.049
2003	0.958	X_{03}/X_{01}	X_{03}/X_{03}	1
2004	1.016	X_{04}/X_{01}	X_{04}/X_{03}	1.061
2005	1.055	X_{05}/X_{01}	X_{05}/X_{03}	1.101
2006	1.035	X_{06}/X_{01}	X_{06}/X_{03}	1.080

- Si dividono tutti i valori per 0.958= X_{03}/X_{01}
- Ad esempio: $X_{02}/X_{03}=(X_{02}/X_{01})/X_{03}/X_{01}$

Soluzione

- Ricostruzione (approssimata) serie originaria

	Input	Input	Obiettivo	Obiettivo
2001	1	X_{01}/X_{01}	X_{01}	4315≈
2002	1.005	X_{02}/X_{01}	X_{02}	4336≈
2003	0.958	X_{03}/X_{01}	X_{03}	4133≈
2004	1.016	X_{04}/X_{01}	X_{04}	4384≈
2005	1.055	X_{05}/X_{01}	X_{05}	4552
2006	1.035	X_{06}/X_{01}	X_{06}	4466≈

X_{05}=input =4552

Ad esempio: $X_{01}=[(X_{05}/X_{01})(1/X_{05})]^{-1}$
$=[1.055(1/4552)]^{-1}\approx 4315$

Esercizio

La seguente tabella riporta: la serie storica del reddito netto (in euro) dichiarato da un lavoratore dipendente; la serie storica delle variazioni percentuali rispetto all'anno precedente dell'indice dei prezzi al consumo per le famiglie di operai e di impiegati (FOI), determinata dall'ISTAT.

Anno	Reddito netto	Variazioni % rispetto all'anno precedente dei n.i. FOI
2004	45.300	–
2005	47.400	+1,7%
2006	50.600	+2,0%
2007	51.100	+1,7%
2008	53.700	+3,2%

1. Si calcoli la serie storica dei numeri indici dei prezzi al consumo FOI con base 2008=100.
2. Si determini la serie storica del reddito netto a prezzi costanti del 2008.
3. Si calcolino il tasso medio annuo di variazione del reddito netto a prezzi correnti ed il tasso medio annuo di variazione del reddito netto a prezzi costanti e si commentino i risultati ottenuti.
4. Si determini quale avrebbe dovuto essere la retribuzione del 2004 che ha lo stesso potere di acquisto della retribuzione di 53700 Euro percepita nel 2008.

Esercizio

- Reddito netto di un lav. e var % a base mobile FOI

Anno	Reddito netto	Var % a base mobile FOI
2004	45300	-
2005	47400	1.7%
2006	50600	2.0%
2007	51100	1.7%
2008	53700	3.2%

- Calcolare (v. slide seguente)

Calcolare

- Serie storica FOI con base 2008=100
- Serie storica del reddito netto a prezzi costanti del 2008
- Tax medio annuo di variazione del reddito netto a prezzi correnti ed il Tax medio annuo di variazione del reddito netto a prezzi costanti. Commentare i risultati
- Quale avrebbe dovuto essere la retribuzione del 2004 che ha lo stesso potere di acquisto della retribuzione di 53700 € percepita nel 2008

Soluzione: NI FOI con base 2008=100 e reddito a prezzi costanti

Anno	Reddito netto	NI a base mobile FOI	NI a base mobile FOI	NI FOI base 2008=100
2004	45300	-	-	X_{04}/X_{08}
2005	47400	101.7	X_{05}/X_{04}	X_{05}/X_{08}
2006	50600	102.0	X_{06}/X_{05}	X_{06}/X_{08}
2007	51100	101.7	X_{07}/X_{06}	X_{07}/X_{08}
2008	53700	103.2	X_{08}/X_{07}	1

Es. $x_{06}/x_{08}=0.9528=(x_{07}/x_{06}\ x_{08}/x_{07})^{-1}=(1.017\ 1.032)^{-1}$

$x_{05}/x_{08}=0.9341=(x_{06}/x_{05}\ x_{07}/x_{06}\ x_{08}/x_{07})^{-1}=(1.02\times1.017\times1.032)^{-1}$

Soluzione: NI FOI con base 2008=100 e reddito a prezzi costanti

Anno	Reddito netto	NI a base mobile FOI	NI FOI base 2008=100	Reddito a prezzi costanti 2008
2004	45300	-	91.85	49320
2005	47400	101.7	93.41	50744
2006	50600	102.0	95.28	53107
2007	51100	101.7	96.90	52735
2008	53700	103.2	100	53700

Reddito a prezzi costanti: es. 49320=45300/0.9185

Calcolo dei tassi medi annui di variazione

Anno	Reddito netto	NI % a base mobile FOI	NI FOI base 2008=100	Reddito a prezzi costanti 2008
2004	45300	-	91.85	49320
2005	47400	101.7	93.41	50744
2006	50600	102.0	95.28	53107
2007	51100	101.7	96.90	52735
2008	53700	103.2	100	53700

- Reddito a prezzi correnti $(53700/45300)^{0.25}-1=0.0434 \rightarrow 4.34\%$
- Reddito a prezzi costanti $(53700/49320)^{0.25}-1=0.0215 \rightarrow 2.15\%$

Calcolo del tasso di inflazione nel periodo

Anno	Reddito netto	NI % a base mobile FOI	NI FOI base 2008=100	Reddito a prezzi costanti 2008
2004	45300	-	91.85	49320
2005	47400	101.7	93.41	50744
2006	50600	102.0	95.28	53107
2007	51100	101.7	96.90	52735
2008	53700	103.2	100	53700

- Inflazione del periodo: 100/91.85=1.088732 ➔ 8.87%
- Retribuzione del 2004 che ha lo stesso potere di acquisto della retribuzione di 53700 € percepita nel 2008= Reddito 2004*1.088732➔ Reddito 2004 a prezzi 2008=49320 Euro

Esercizio

- Distribuzione dei nati maschi in Italia secondo il peso in un determinato anno

- Si calcoli la moda,
- Si disegni il boxplot,
- Si calcoli l'indice di asimmetria di Bowley (v. p. 157) e l'indice di curtosi

Commentare i risultati

classi di peso		ni
0	500	39
500	1000	548
1000	1500	1149
1500	2000	2020
2000	2500	7691
2500	2750	12772
2750	3000	26922
3000	3250	53178
3250	3500	58093
3500	3750	50640
3750	4000	30051
4000	4500	20257
4500	5000	2369
5000	5500	192
5500	6000	27
>6000		13
		265961

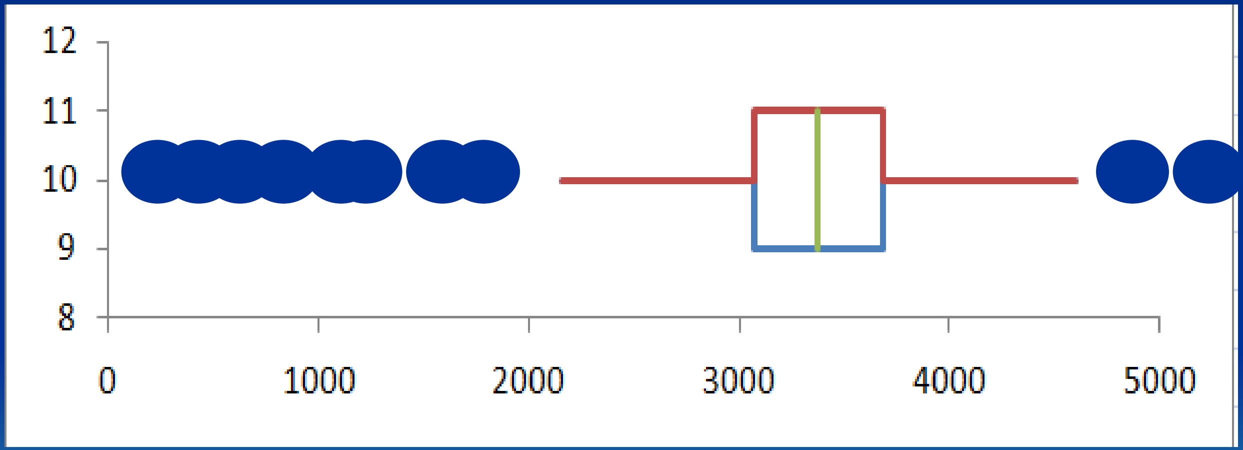
Soluzione

	A	B	C	D	E	F
1	classi di peso		ni	f_i	10000*d_i	Fi
2	0	500	39	0.000147	0.00293276	0.000147
3	500	1000	548	0.00206	0.04120905	0.002207
4	1000	1500	1149	0.00432	0.08640365	0.006527
5	1500	2000	2020	0.007595	0.15190197	0.014122
6	2000	2500	7691	0.028918	0.57835547	0.04304
7	2500	2750	12772	0.048022	1.92088314	0.091062
8	2750	3000	26922	0.101225	4.04901471	0.192288
9	3000	3250	53178	0.199947	7.99786435	0.392234
10	3250	3500	58093	0.218427	8.73707047	0.610661
11	3500	3750	50640	0.190404	7.61615425	0.801065
12	3750	4000	30051	0.11299	4.51961002	0.914055
13	4000	4500	20257	0.076165	1.52330605	0.99022
14	4500	5000	2369	0.008907	0.17814642	0.999128
15	5000	5500	192	0.000722	0.01443821	0.99985
16	5500	6000	27	0.000102	0.00203037	0.999951
17		>6000	13	4.89E-05		1
18			265961	1		

Mo = 3375 gr.; Me=3373.343 gr

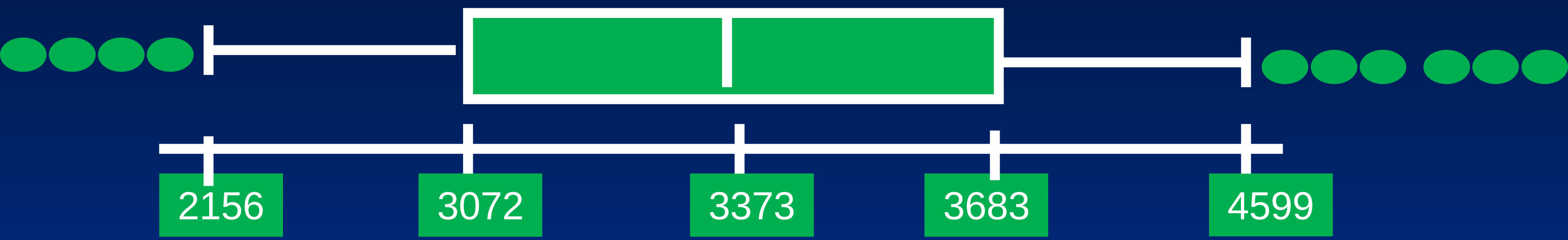
Boxplot

- PT_{inf} = 3072.16-1.5(3682.952-3072.16)≈2156
- PT_{sup} =3682.952+1.5(3682.952-3072.16)≈4599
- Tutti i bambini con peso inferiore a 2156 gr o superiore a 4.6 Kg circa sono dichiarati “outliers”



- La distribuzione dei valori è quasi simmetrica

Boxplot



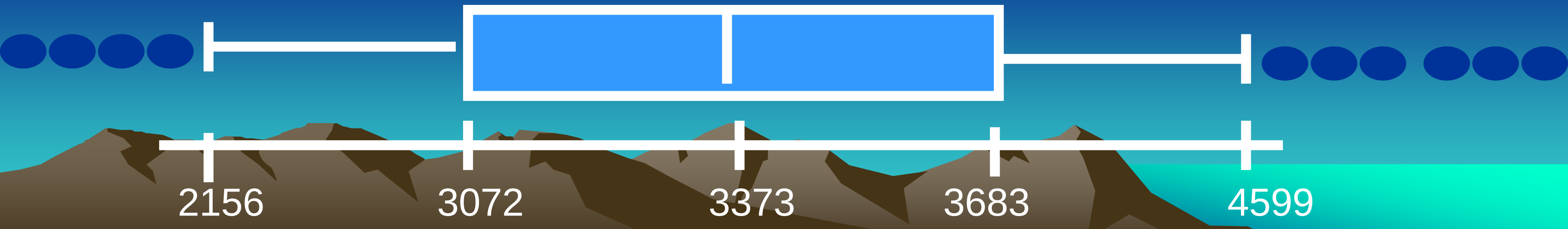
classi di peso		ni
0	500	39
500	1000	548
1000	1500	1149
1500	2000	2020
2000	2500	7691
2500	2750	12772
2750	3000	26922

2750	3000	26922
3000	3250	53178
3250	3500	58093
3500	3750	50640
3750	4000	30051
4000	4500	20257
4500	5000	2369
5000	5500	192
5500	6000	27
	>6000	13
		265961

In questo esempio la percentuale di valori anomali è molto maggiore di 0.7%

Boxplot

- Osservazione: in presenza di una distribuzione normale i valori fuori dai punti di troncamento del boxplot sono circa il 0.7%
- In questo esempio la % di valori anomali molto maggiore di 0.7%. La distribuzione sembra avere quindi code leggermente più lunghe di quelle della normale (distribuzione ipernormale)



Calcolo indice di asimmetria di Bowley

$$AS_r = \frac{(x_{0,75} - Me) - (Me - x_{0,25})}{x_{0,75} - x_{0,25}}$$

- (3682.952-2*3373.343+3072.16)/(3682.952-3072.16)=0.014

- Anche tramite questo indice si segnala la quasi simmetria della distribuzione

Calcolo indice di curtosi (p. 165)

$$Ku = \frac{\sum_{i=1}^{r}(x_i - M)^4 n_i/n}{\left[\sum_{i=1}^{r}(x_i - M)^2 n_i/n\right]^2} = \frac{\sum_{i=1}^{r}(x_i - M)^4 n_i/n}{\sigma^4}$$

xi	xi*fi	(xi-M)^2fi	(xi-M)^4fi
250	0.03666	1421.20499	13774213026
750	1.545339	14070.2939	96082376907
1250	5.400228	19292.0077	86149503690
1750	13.29142	19765.2438	51436444365
2250	65.06499	35834.4006	44405364059
2625	126.058	26168.1059	14259477878
2875	291.0229	24124.5706	5749496574
3125	624.8332	11343.463	643542559.8
3375	737.1903	30.4871515	4255.277148
3625	690.214	13051.5545	894640889.8
3875	437.8372	29598.2264	7753368123
4250	323.7025	59899.401	47107253931
4750	42.30977	17131.0382	32947332845
5250	3.790029	2570.04992	9149552328
5750	0.583732	578.339752	3294733428
6250	0.305496	407.345638	3394694885
	3363.186	275285.733	4.17042E+11

- Ponendo il valore superiore dell'ultima classe pari a 6500 si ottiene M=3363.186
- Ku=4.17042E+11/ 275285.7^2=5.47
- Distribuzione ipernormale

Esercizio

- Dimostrare che se i dati provengono da una distribuzione normale i valori esterni ai punti di troncamento del boxplot sono circa il 0.7%

Soluzione

- Senza perdita di generalità possiamo fare riferimento alla normale standardizzata
- $x_{0.75}$=0.6745 (=inv.norm.st(0.75))
- $x_{0.25}$=-0.6745 (=inv.norm.st(0.25))
- $x_{0.75+}1.5(x_{0.75}-x_{0.25})=0.6745+3*0.6745=2.698$
- 1-F(2.698)=1-0.9965=0.0035
- =DISTRIB.NORM.ST(2.698)=0.9965
- 0.35% punti fuori dal punto di troncamento superiore ➔ 0.7% punti fuori dai due punti di troncamento

Esercizio

- Un'organizzazione indice una lotteria con i seguenti premi: un premio da 500€, 5 premi da 100€ e 50 premi da 50€. Si stabiliscono di vendere 5000 biglietti per la lotteria.
  - Calcolare il prezzo equo del biglietto (ossia il prezzo che garantisce allo scommettitore un guadagno atteso pari a 0).
  - Calcolare il prezzo del biglietto in modo tale che esso sia tre volte più grande del prezzo equo

Soluzione: Calcolo del prezzo equo del biglietto

- Un’organizzazione indice una lotteria con i seguenti premi: un premio da 500€, 5 premi da 100€ e 50 premi da 50€. Si stabiliscono di vendere 5000 biglietti per la lotteria.

E(X)=500(1/5000)+100(5/5000)+50(50/5000)=0.7
Il costo equo del biglietto sarebbe di 70 centesimi

Per garantire il profitto della lotteria si può decidere di fare pagare il biglietto 3 volte il prezzo equo =2.10 €

Esercizio

Un’azienda assembla televisori in tre differenti stabilimenti nelle percentuali del 50%, 10% e 40%. E’ noto che ogni stabilimento produce unità difettose nelle percentuali 1%, 5% e 2%.

Dall’intera produzione si estrae un televisore osservandolo difettoso.

Quali sono le probabilità che esso provenga dal primo, secondo o terzo stabilimento?

Soluzione

- A_i = il televisore è stato assemblato nello stabilimento i (i = 1, 2, 3)
- $P(A_1) = 0,5$ $P(A_2) = 0,1$ $P(A_3) = 0,4$

- B = il televisore è difettoso

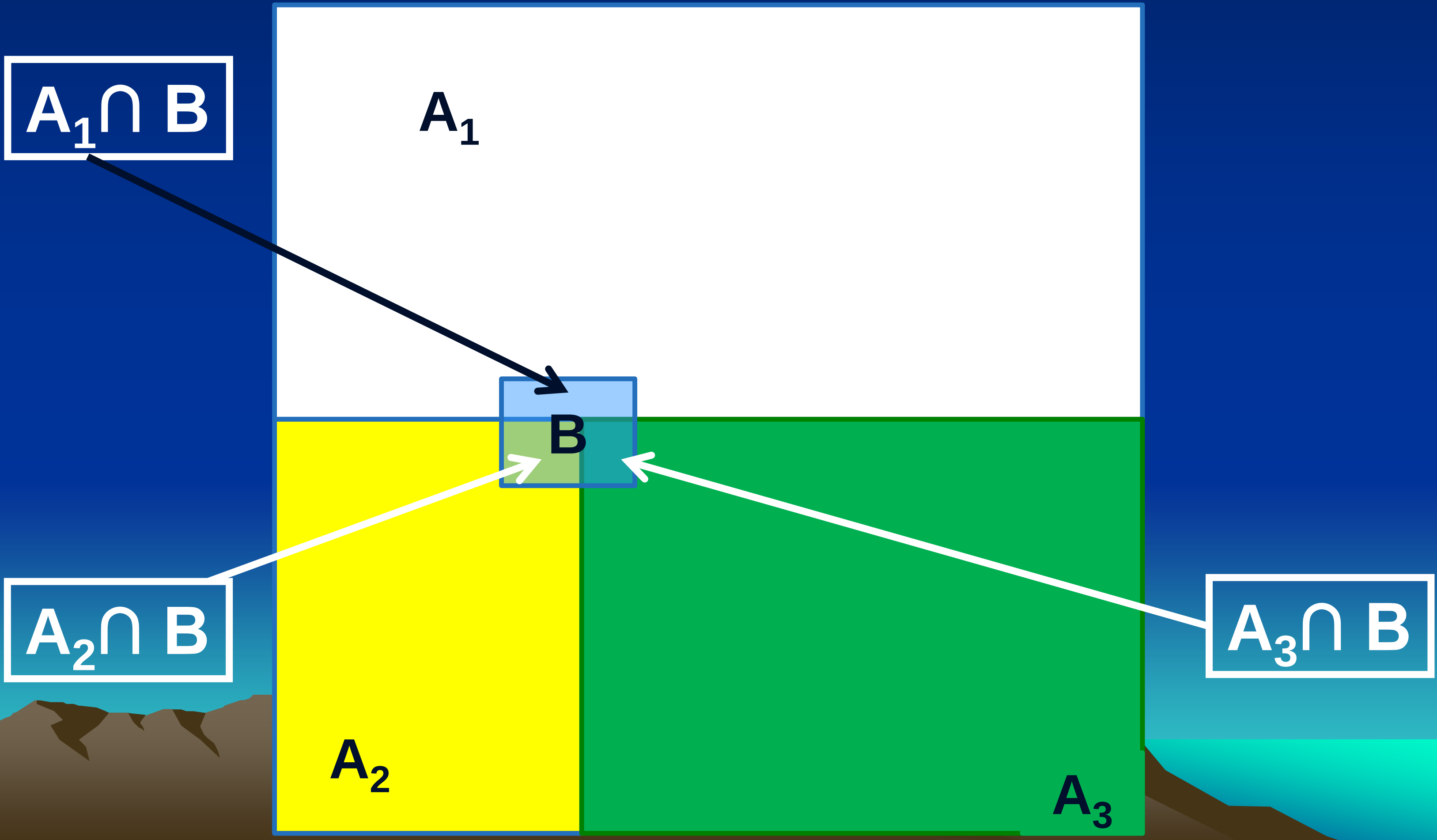
$P(B|A_1) = 0,01$ $P(B|A_2) = 0,05$ $P(B|A_3) = 0,02$

- $P(A_i \mid B) = ?$
- $P(A_i \mid B) = P(B|A_i)\, P(A_i) / P(B)$

- $B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$

$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$



OBIETTIVO FINALE: $P(A_i \mid B) = P(B|A_i)\, P(A_i) / P(B)$

$P(A_1) = 0{,}5 \quad P(A_2) = 0{,}1 \quad P(A_3) = 0{,}4$

$P(B|A_1) = 0{,}01 \quad P(B|A_2) = 0{,}05 \quad P(B|A_3) = 0{,}02$

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$

$P(B \cap A_1) = P(B|A_1) \cdot P(A_1) = 0{,}01 \cdot 0{,}5 = 0{,}005$

$P(B \cap A_2) = P(B|A_2) \cdot P(A_2) = 0{,}05 \cdot 0{,}1 = 0{,}005$

$P(B \cap A_3) = P(B|A_3) \cdot P(A_3) = 0{,}02 \cdot 0{,}4 = 0{,}008$

$P(B) = 0{,}005 + 0{,}005 + 0{,}008 = 0{,}018$

$P(A_i \mid B)$ = televisore assemblato dallo stabilimento i dato che è difettoso

- $P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0{,}005}{0{,}018} = 0,278$
- $P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{0{,}005}{0{,}018} = 0,278$
- $P(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = \frac{0{,}008}{0{,}018} = 0,444$

Esercizio

In un'indagine condotta su un campione di 800 individui (maschi e femmine) residenti in un'area metropolitana è stato chiesto, fra l'altro, se amano fare shopping. 260 dei 380 uomini intervistati e 360 delle 420 donne hanno risposto affermativamente.

1. Sapendo che il soggetto scelto è una donna, si calcoli la probabilità che non ami fare shopping.

2. Sapendo che il soggetto scelto ama fare shopping, si calcoli la probabilità che si tratti di un uomo.

3. Si dica se gli eventi "il soggetto ama fare shopping" e "il soggetto è un uomo" sono stocasticamente indipendenti.

Soluzione

	D	U	tot
S	**360**	**260**	?
NS	?	?	?
Tot	**420**	**380**	**800**

S = il soggetto ama fare shopping

NS = il soggetto non ama fare shopping

U= Uomo

D = Donna

Soluzione

	D	U	tot
S	**360**	**260**	620
NS	60	120	180
tot	**420**	**380**	**800**

- P(NS|D)? P(U|S)? P(U|S)=P(U)?

P(NS|D) = 60/420 = 0,143

P(U|S) = 260/620 = 0,419

I due eventi S e U non sono stocasticamente indipendenti poiché

P(U|S) = 0,419 ≠ P(U) = 380/800 = 0,475

Esercizio

- Una compagnia che produce sementi sviluppa una nuova carota. Dopo aver misurato 5000 di queste carote è stato possibile affermare che la lunghezza delle carote ~N(11.5 cm, 1.15^2)
- L'azienda può dichiarare nel suo catalogo
  - che almeno l'80% di carote presenta un'altezza tra i 10 cm ed i 13 cm?
  - Che almeno il 90% delle carote siano di 10 cm e oltre?

Esercizio: continua

- Qual è la lunghezza maggiore o uguale al 96% delle altre lunghezze?
- Se vengono raccolte due nuove carote, qual è la probabilità che entrambe siano più lunghe di 13.5 cm?

Soluzione

- L’azienda può dichiarare nel suo catalogo
  – che almeno l’80% di carote presenta un’altezza tra i 10 cm ed i 13 cm?

$$X \sim N(11.5 \text{ cm}, 1.15^2)$$

$$Pr(10<X<13)=Pr((10-11.5)/1.15<Z<(13-11.5)/1.15$$

$$=P(-1.3<Z<1.3)=0.90320-0.09689\approx 0.8063$$

- L’affermazione è corretta

Soluzione

- L'azienda può dichiarare nel suo catalogo
  – Che almeno il 90% delle carote siano di 10 cm e oltre?

$$X \sim N(11.5\text{ cm}, 1.15^2)$$

$$Pr(X>10)=1-Pr(Z<(10-11.5)/1.15)$$

$$=1-P(Z<-1.3)=1-0.09689 \approx 0.90311$$

L'affermazione è corretta

Soluzione

- Qual è la lunghezza maggiore o uguale al 96% delle altre lunghezze?

$$X \sim N(11.5\text{ cm}, 1.15^2)$$

$$F((x-11.5)/1.15)=0.96$$

$$(x-11.5)/1.15=1.75$$

$$x \approx 13.5\text{ cm}$$

La lunghezza di 13.5 cm (circa) è maggiore o uguale al 96% delle altre lunghezze

Soluzione

Se vengono raccolte due nuove carote, qual è la probabilità che entrambe siano più lunghe di 13.5 cm?

Osservazione: le due nuove carote raccolte possono essere considerate due eventi indipendenti

Indicando con X_i=la lunghezza della carota i (i=1,2),

dato che $P(X>13.5) \approx 0.04$

$$P(X_1>13.5 \cap X_2>13.5) = P(X_1>13.5) \cap P(X_2>13.5)=0.04^2=0.0016$$

Esercizio

- La segretaria di uno studio dentistico sa per esperienza che il 10% dei pazienti arriva in ritardo rispetto all'appuntamento stabilito. Si scrivano le espressioni e si calcolino:

A. la probabilità che su sette pazienti scelti a caso due arrivino in ritardo.

B. la prob. che almeno due pazienti su 7 arrivino in ritardo

Soluzione

- U~Bernoulliano con π=1/10
- X = numero di pazienti che arrivano in ritardo. Distribuzione Binomiale (n=7)

A. $$Pr(X=2)=\binom{7}{2}0.1^2 0.9^5=0.124$$

B. $$Pr(X\geq 2)=\sum_{s=2}^{7}\binom{7}{s}0.1^s 0.9^{7-s}=0.1497$$

$$Pr(X\geq 2)=1-Pr(X=1)-Pr(X=0)$$

$$1-\sum_{s=0}^{1}\binom{7}{s}0.1^s 0.9^{7-s}=0.1497$$

Calcoli in Excel

- Visualizzazione output
- Visualizzazione formule

	A	B
1	X	prob
2	0	0.4782969
3	1	0.3720087
4	2	0.1240029
5	3	0.0229635
6	4	0.0025515
7	5	0.0001701
8	6	0.0000063
9	7	0.0000001
10		1
11		
12	P(X>=2)	0.1496944
13		0.1496944

	A	B
1	X	prob
2	0	=COMBINAZIONE(7;A2)*0.1^A2*0.9^(7-A2)
3	1	=COMBINAZIONE(7;A3)*0.1^A3*0.9^(7-A3)
4	2	=COMBINAZIONE(7;A4)*0.1^A4*0.9^(7-A4)
5	3	=COMBINAZIONE(7;A5)*0.1^A5*0.9^(7-A5)
6	4	=COMBINAZIONE(7;A6)*0.1^A6*0.9^(7-A6)
7	5	=COMBINAZIONE(7;A7)*0.1^A7*0.9^(7-A7)
8	6	=COMBINAZIONE(7;A8)*0.1^A8*0.9^(7-A8)
9	7	=COMBINAZIONE(7;A9)*0.1^A9*0.9^(7-A9)
10		=SOMMA(B2:B9)
11		
12	P(X>=2)	=SOMMA(B4:B9)
13		=1-DISTRIB.BINOM(1;7;0.1;1)

Esercizio

Il tasso di disoccupazione in una città è pari al 7.6%. Un campione casuale di 200 unità viene estratto dalle forze di lavoro. Si scriva l'espressione della probabilità che nel campione vi siano:

A) meno di 15 disoccupati

B) almeno 18 disoccupati

Calcolare con Excel o con la calcolatrice queste probabilità

Si calcolino le precedenti probabilità utilizzando l'approssimazione normale

Soluzione

- U~Bernoulliano con π=0.076
- X = numero di disoccupati in un campione di 200 persone Distribuzione Binomiale (n=200)

$$Pr(X \leq 14) = \sum_{s=0}^{14} \binom{200}{s} 0.076^s (1-0.076)^{200-s} =$$

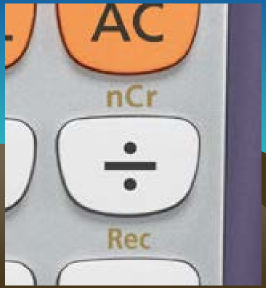
$$Pr(X > 17) = \sum_{s=18}^{200} \binom{200}{s} 0.076^s (1-0.076)^{200-s} =$$

- Utilizzando le funzioni di Excel =DISTRIB.BINOM(14;200;0.076;1)
- =1-DISTRIB.BINOM(17;200;0.076;1) o la calcolatrice si ottiene Pr(X<=14)=0.4405 e Pr(X>17)=0.2620

Utilizzo della calcolatrice

- ∑ = sommatoria

- Indice della sommatoria (X)



- nCr= combinazioni di n elementi di classe r

Soluzione

- Quando n è grande

$$X = \text{Binomiale}(n,\pi) \sim N(n\pi,\ n\pi(1-\pi))$$

$$X \sim N(200*0.076,\ 200*0.076*(1-0.076))$$

$$X \sim N(15.2,\ 3.7476^2)$$

Obiettivo: prob. di estrarre meno di 15 disoccupati

$$\Pr(X<=14) = \Pr(Z<(14-15.2)/3.7476)$$

$$= \Pr(Z<-0.3202) = 0{,}37448$$

Osservazione: la prob. esatta era 0.44

Soluzione

$X \sim N(15.2, 3.7476^2)$

Obiettivo: prob. di estrarre meno di 15 disoccupati

14.5 è il valore centrale nella classe 14-15

$Pr(X<14.5)= Pr(Z<(14.5-15.2)/3.7476)$

$= Pr(Z<-0.18679)$=0,42465 (dalle tavole)

$= Pr(Z<-0.18679)= 0.4259$ (dalla funzione di Excel DISTRIB.NORM.ST)

Osservazione: la prob. esatta era 0.44

Soluzione

$X \sim N(15.2, 3.7476^2)$

Obiettivo: prob. di estrarre almeno 18 disoccupati

$Pr(X>18) = 1 - Pr(Z<(18-15.2)/3.7476)$

$= 1 - Pr(Z<0.747) = 1-0.77337 = 0.22663$ (dalle tavole)

Approssimazione più accurata

$Pr(X>17.5) = 1 - Pr(Z<(17.5-15.2)/3.7476)$

$= 1-Pr(Z<0.6137) = 1-0.72907 \approx 0.27$

Osservazione: la prob. esatta era 0.26

Approfondimenti sull'approssimazione normale della Binomiale

- Se X~B con media nπ e varianza nπ(1-π) se n è sufficientemente grande

$$Pr(a \leq X \leq b) = F\left(\frac{b - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - F\left(\frac{a - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

L'approssimazione migliora se si considera b+0.5 e a-0.5 (correzione per continuità)

$$Pr(a \leq X \leq b) = F\left(\frac{b + 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - F\left(\frac{a - 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

Nell’esempio precedente

$$Pr(a \leq X \leq b) = F\left(\frac{b+0.5-n\pi}{\sqrt{n\pi(1-\pi)}}\right) - F\left(\frac{a-0.5-n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

– La condizione: meno di 15 disoccupati → (Pr $X \leq 14$) → $b=14$ → $F(14.5)$

– La condizione almeno 18 disoccupati → $Pr(X \geq 18)$ → $a=18$ → $F(17.5)$

Esercizio

- Un universo statistico X è distribuito secondo la seguente funzione di densità:
- $f(x)=\exp(-x)=e^{-x}$ per $x \geq 0$
- Verificare che f(x) sia effettivamente una densità
- Scrivere la distribuzione della v.c. quindicesimo elemento del campione
- Calcolare la probabilità che $X<2$.
- Calcolare media e varianza dell'universo.
- Calcolare la funzione di ripartizione.
- Rappresentare graficamente F(X) e f(x)
- Calcolare la $P(X>x)$

Verificare che f(x) sia effettivamente una densità

- $f(x)=\exp(-x)=e^{-x}$ per $x \geq 0$
- $f(x) >= 0$ sempre

$$\int_0^{+\infty} f(x)dx = \int_0^{+\infty} e^{-x}dx = -e^{-x}\Big|_0^{+\infty} = 1$$

Scrivere la distribuzione della v.c. quindicesimo elemento del campione

- X_1 X_2, …, X_n = campione casuale = v.c. IID con la stessa distribuzione del fenomeno nell’universo
- $f(x_{15})= f(x)= \exp(-x)=e^{-x}$ per $x \geq 0$

Calcolare la probabilità che X<2.

$$P(X < 2) = \int_0^2 f(x)dx = \int_0^2 e^{-x}dx$$

$$P(X < 2) = -e^{-x}\Big|_0^2 = 1 - e^{-2}$$

Calcolare media e varianza dell'universo.

$$E(X) = \int_0^{+\infty} x f(x) dx = \int_0^{+\infty} x e^{-x} dx$$

- Integrando per parti si ottiene

$$\int_a^b f(x) g'(x) dx = f(x) g(x)|_a^b - \int_a^b f'(x) g(x) dx$$

$$E(X) = \int_0^{+\infty} x f(x) dx = -x e^{-x}|_0^{+\infty} + \int_0^{+\infty} e^{-x} dx$$

$$= 0 - e^{-x}|_0^{+\infty} = 1$$

Calcolare media e varianza dell'universo.

$$E(X^2) = \int_0^{+\infty} x^2 f(x) dx = \int_0^{+\infty} x^2 e^{-x} dx$$

- Integrando per parti si ottiene

$$E(X^2) = \int_0^{+\infty} x^2 e^{-x} dx = -x^2 e^{-x} dx \Big|_0^{+\infty} + \int_0^{+\infty} 2x e^{-x} dx$$

$$= 0 + 2E(X) = 2$$

$$var(X) = E(X^2) - [E(X)]^2 = 1$$

Calcolare la funzione di ripartizione $F(x) = P(X<x)$

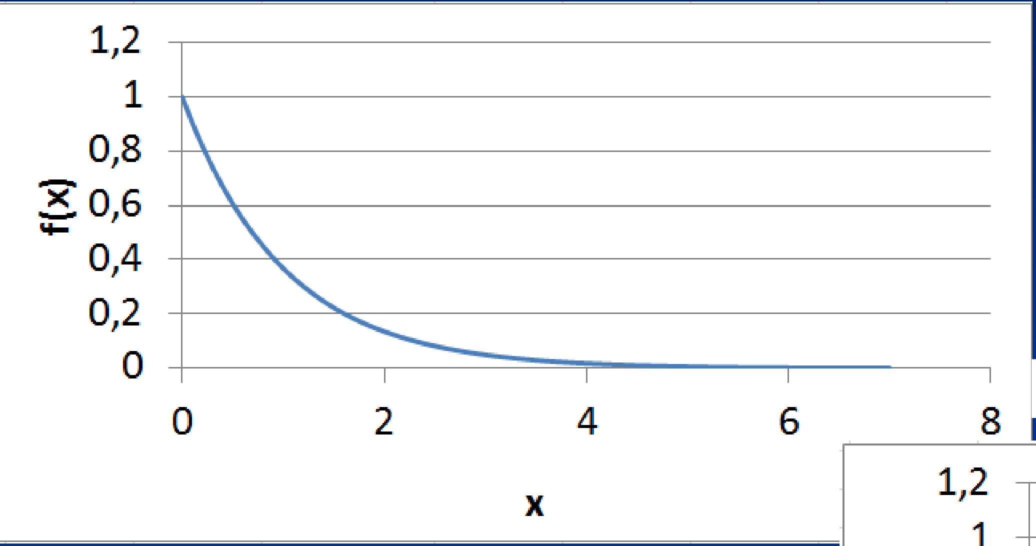
$$F(x) = \int_0^x f(t)dt = \int_0^x e^{-t}dx$$

$$F(x) = -e^{-t}\Big|_0^x = 1 - e^{-x}$$

- Quindi $P(X>x)=1-F(X)=e^{-x}$

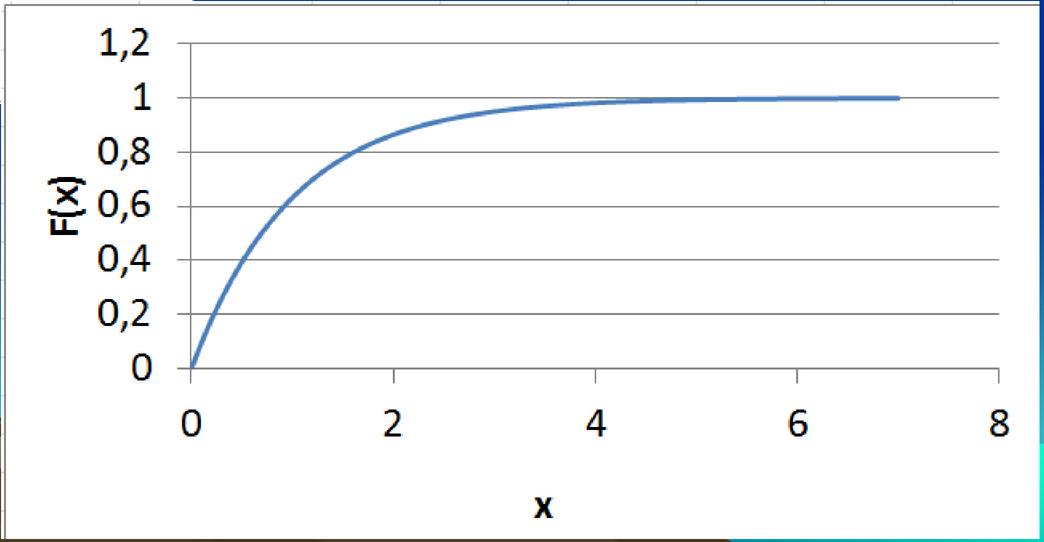
Rappresentare graficamente F(X) e f(x)

Funzione di densità



$$f(x) = e^{-x}$$

Funzione di ripartizione



$$F(x) = 1 - e^{-x}$$

Esercizio

Un rappresentante di strumenti musicali valuta le seguenti probabilità per il numero di strumenti che riuscirà a vendere nella settimana seguente

Numero di strumenti musicali venduti	Probabilità
0	0.10
1	0.2
2	0.6
3	0.1

1) Calcolare il numero atteso di strumenti musicali che riuscirà a vendere nella settimana seguente
2) Calcolare lo scarto quadratico medio del numero di strumenti musicali che riuscirà a vendere nella settimana seguente.
3) Il rappresentante riceve un compenso fisso di 200 Euro più un ammontare di 50 Euro per ogni strumento musicale venduto. Trovare la media e lo scarto quadratico medio del suo compenso complessivo per la settimana seguente.
4) Qual è la probabilità che il venditore, nella settimana seguente, venda un numero di strumenti musicali minore o uguale a 2?
5) Qual è la probabilità che il venditore, nella settimana seguente, abbia un compenso superiore a 300 Euro?

Soluzione

- Calcoli per ottenere E(X) e σ(X)

xi	pi	xi*pi	(xi-E(X))^2*pi
0	0,1	0	0,28900
1	0,2	0,2	0,09800
2	0,6	1,2	0,05400
3	0,1	0,3	0,16900
		1,7	0,61000

$E(X)=1{,}7$ $\sigma(X)=0{,}78$

Y= Compenso complessivo E(Y)? e σ(Y)=

- $Y=200+50X$
- $E(Y) = 200+50E(X)=285$
- $\sigma(Y) = 50\ \sigma(X) =39{,}05$

Probabilità che il venditore, nella settimana seguente, venda un numero di strumenti musicali minore o uguale a 2?

xi	pi
0	0,1
1	0,2
2	0,6
3	0,1

$$P(X \leq 2) = 0.9$$

Probabilità che il venditore, nella settimana seguente, abbia un compenso Y superiore a 300 Euro (Y=200+50X)

$P(Y > 300) = P(200+50X > 300) = P(X > 2)$

xi	pi
0	0,1
1	0,2
2	0,6
3	0,1

$P(Y > 300) = P(X > 2) = 0.1$

Esercizio

- Nei ragazzi (maschi) di 17 anni, la media deviazione standard della pressione diastolica seguono una v.a. gaussiana con media 63.7 e s.q.m. 11.4. Ci sono diversi approcci per diagnosticare una «pressione elevata».
- 1. Un approccio diagnostica «pressione elevata» se la pressione diastolica supera il quantile di ordine 0.9 della distribuzione citata sopra. Dire a quanto corrisponde questa soglia.

Continua

- 2. Un altro approccio diagnostica «pressione elevata» se la pressione diastolica supera i 90 mm Hg. Che percentuale di ragazzi di 17 anni avrà una pressione elevata con questo
- approccio?
- 3. Qual e la probabilità di trovare, tra 5 ragazzi, più di uno con la pressione elevata se si segue l'approccio del punto 2.?

Soluzione

- Calcolo del quantile 0.90 ($x_{0.90}$) in una distribuzione $X \sim N(63.7, 11.4^2)$

$$\frac{X-63.7}{11.4} \sim N(0,1)$$

$$\frac{x_{0.90}-63.7}{11.4} = 1.28$$

- $x_{0.90} = 63.7 + 11.4 \cdot 1.28 = 78.3$

Continua

- Un altro approccio diagnostica «pressione elevata» se la pressione diastolica supera i 90 mm Hg. Che percentuale di ragazzi di 17 anni avrà una pressione elevata con questo approccio?

In una X~N(63.7, 11.4^2) occorre calcolare P(X>90)

$$1 - F\left(\frac{90 - 63.7}{11.4}\right) = 0.0107$$

3. Qual e la probabilità di trovare, tra 5 ragazzi, più di uno con la pressione elevata se si segue l'approccio del punto 2.?

- Il numero di ragazzi con “pressione alta” X è una variabile aleatoria binomiale con parametri n = 5 e π = 0.0107 (se si segue il criterio del punto 2) P(X>90)=P(successo)=0.0107

P(più di uno) = 1- P(nessuno) – P(uno)=

$$1-\binom{5}{0}0{,}0107^{0}\cdot(1-0.0107)^{5}-\binom{5}{1}0{,}0107^{1}\cdot(1-0.0107)^{4}=0.001121$$

Esercizio

- Il produttore di una marca di sigarette desidera controllare il quantitativo medio di catrame in esse contenuto. A tale scopo osserva un campione di 30 sigarette e trova una media campionaria=10.92 mg e s_{cor}=0.51. Calcolare l'intervallo di confidenza di μ al 99%.

Soluzione

- Ip. Distribuzione normale nell'universo
- t(0.005) con 29 gradi di libertà=2.756

$$P\left\{ \overline{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \leq \mu \leq \overline{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1 - \alpha$$

- Estremo inferiore = $10.92-2.756*0.51/30^{0.5}$
- Estremo superiore = $10.92+2.756*0.51/30^{0.5}$
- $Pr(10.66 < \mu < 11.18)=0.99$

Esercizio

- Nel processo di controllo del peso delle confezioni di spaghetti l’azienda esamina un campione di 830 confezioni e trova che 19 di esse hanno un peso fuori norma. Si determini l’intervallo di confidenza del 95% della proporzione di pezzi fuori norma

Soluzione

- p=0.023 n=830

$$P\left\{p - z(\alpha/2)\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z(\alpha/2)\sqrt{\frac{p(1-p)}{n}}\right\} = 1-\alpha$$

- Estremo inf= $0.023-1.96(0.023*0.977/830)^{0.5}$
- Estremo sup= $0.023+1.96(0.023*0.977/830)^{0.5}$
- Pr(0,013<π< 0,033)=0.95
- C'è una prob. del 95% che la proporzione di confezioni difettose sia compresa nell'intervallo (0.013 0.033)

Esercizio

- Un ricercatore desidera stimare la media di un popolazione che presenta una deviazione standard σ con un campione di numerosità n in modo tale che sia uguale a 0.95 la probabilità che la media del campione non differisca dalla media della popolazione per più di 5% della deviazione standard.
- Si determini n

Soluzione

Se l'intervallo di confidenza è al 95%

$$P\left\{\bar{X} - 1{,}96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right\} = 0{,}95$$

$$P\left\{|\bar{X} - \mu| \le 1{,}96\frac{\sigma}{\sqrt{n}}\right\} = 0{,}95$$

Se vogliamo che l'errore di stima della media non superi 0.05 σ

$$1{,}96\frac{\sigma}{\sqrt{n}} \le 0.05\sigma$$

$$n \ge \left(\frac{1{,}96}{0.05}\right)^2 \approx 1537$$

Esercizio

- Obiettivo: verificare in che modo la vendita di prodotti di pulizia per la casa è influenzata dallo spazio sugli scaffali destinati ai prodotti
- X = spazio sugli scaffali (in metri)
- Y = ammontare di vendite settimanali (in migliaia di Euro) in 6 supermercati

Supermercato	X	Y
1	1.55	3.2
2	3.05	6.0
3	3.35	6.2
4	1.80	4.2
5	2.00	3.9
6	2.55	4.3

Quesiti

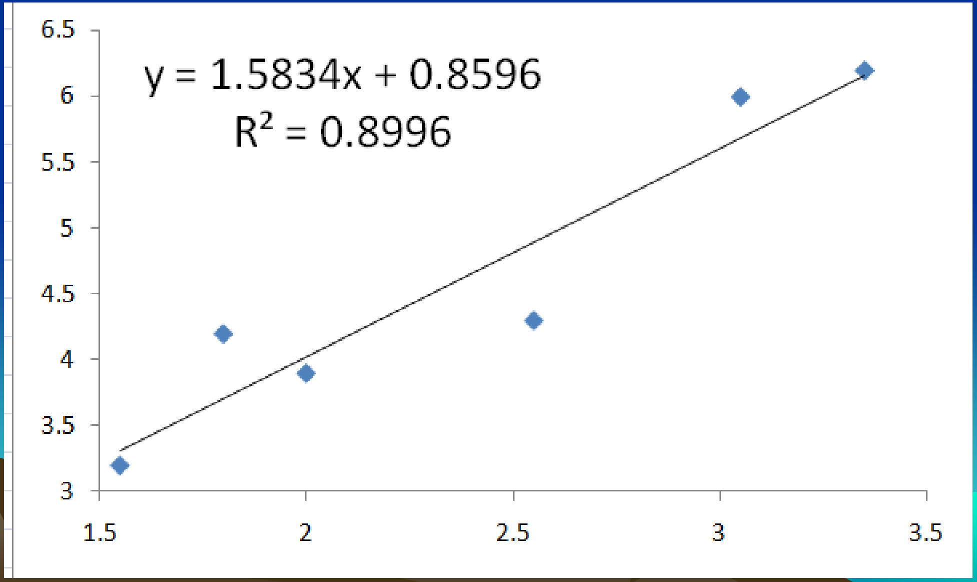
- Scegliendo opportunamente la variabile indipendente e la variabile dipendente si calcolino i parametri della retta di regressione e si illustri il significato dei parametri così calcolati
- Si determini la bontà di adattamento e si commenti il risultato ottenuto
- Si determini l'ammontare settimanale delle vendite di prodotti per la pulizia della casa per un supermercato che ha destinato a tali prodotti 4.85 metri chiarendo l'attendibilità del risultato

Quesiti

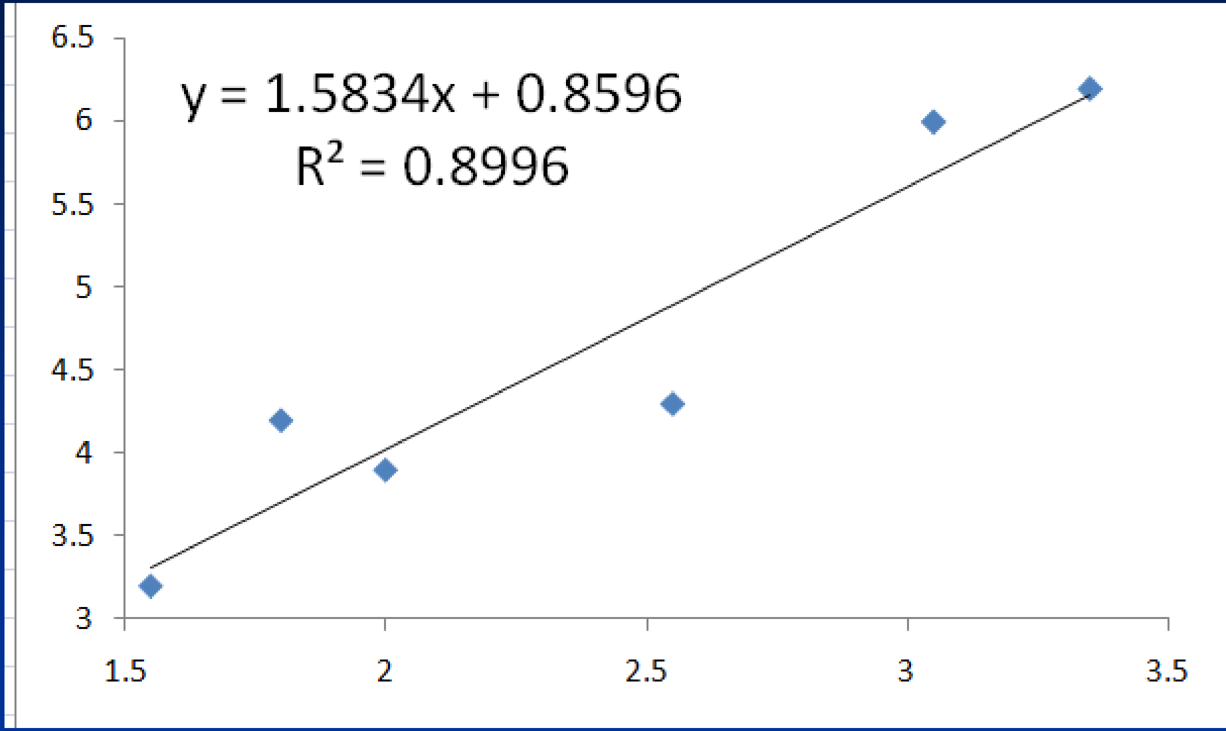
- Si costruisca un intervallo al 95% ed al 99% per α e β
- Si testino le due ipotesi nulle H_0: $\beta=0$ e $\beta=1$
- Si testi l'ipotesi nulla H_0: $\alpha=0$ e si fornisca il relativo p-value
- Si calcoli la tabella di analisi della varianza

Soluzione

- X= spazio sugli scaffali (in metri), (esplicativa)
- Y= ammontare di vendite settimanali (in migliaia di Euro) in 6 supermercati (dipendente)



Calcolo del valore previsto



- $x=4.85$
  vendite previste=$1.5834*4.85+0.8596\approx 8.54$
- Attendibilità scarsa poiché $x=4.85$ è molto al di fuori del range dei valori utilizzati per l'adattamento

Costruzione di un intervallo di confidenza al 95% per il coeff. angolare

$$P\left\{\hat{\beta} - t(\alpha/2)SE(\hat{\beta}) \le \beta \le \hat{\beta} + t(\alpha/2)SE(\hat{\beta})\right\} = 1 - \alpha$$

- $t_4(0.025)=+2.78$ (=INV.T(0.05;4))
- (Oss: Pr.($T_4>2.78$)=0.025)

$$\hat{\beta} = 1.5834$$

$$s_{\hat{\beta}} = SE(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = 0.264$$

Intervallo al 95% ed al 99% per i coefficienti del modello

- Pr(-0.956 < α < 2.675)=0.95
- Pr(-2.150 < α < 3.870)=0.99

- Pr(0.849< β <2.318)=0.95
- Pr(0.366 < β < 2.801)=0.99

- Si noti che mentre l’intervallo di confidenza di β non contiene il valore 0 mentre quello di α lo contiene

Dato che

$$\frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim T_{n-2}$$

Dato che

$$\frac{\hat{\alpha}-\alpha}{SE(\hat{\alpha})} \sim T_{n-2}$$

Sotto H_0: $\beta = 0$

$$\frac{\hat{\beta}}{SE(\hat{\beta})} \sim T_{n-2}$$

Sotto H_0: $\alpha = 0$

$$\frac{\hat{\alpha}}{SE(\hat{\alpha})} \sim T_{n-2}$$

Test sulle ipotesi nulle $H_0: \alpha=0$ $H_0: \beta=0$

	Coefficienti	Errore standard	Stat t	p-value
Intercetta	0.860	0.654	1.315	0.259
Spazio	1.583	0.264	5.987	0.004

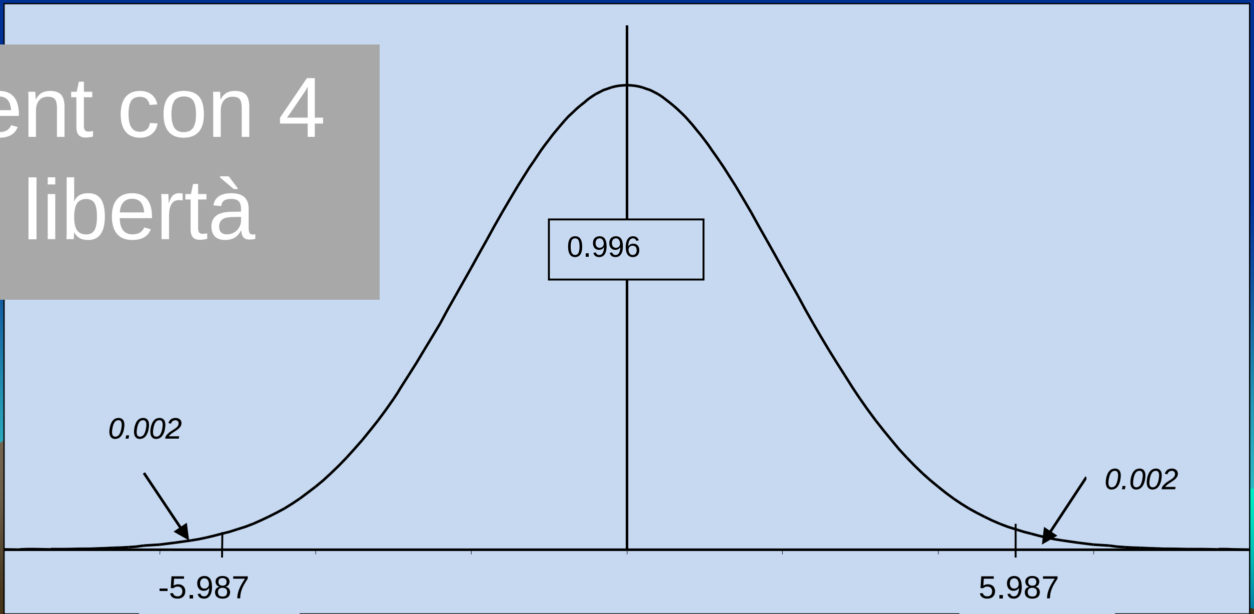
- A livello di significatività dell'1%
  - posso rifiutare l'ipotesi nulla di assenza di relazione tra spazio e vendite
  - non posso rifiutare l'ipotesi nulla che l'intercetta nell'universo sia uguale a 0

Test sulle ipotesi nulle H_0:α=0 H_0: β=0

	Coefficienti	*Errore standard*	*Stat t*	*p-value*
Intercetta	0.860	0.654	1.315	0.259
Spazio	1.583	0.264	5.987	0.004

Visualizzazione grafica del p-value per H_0: β=0

T di student con 4 gradi di libertà



Test sull'ipotesi nulla H_0: $\beta=1$

- Se è vera l'ipotesi sulla $\beta=1$ allora

$$\frac{\hat{\beta} - E(\hat{\beta})}{s(\hat{\beta})} \sim T(n-2)$$

$$\frac{1.583 - 1}{0.264} = 2.21$$

Il valore 2.21 si trova tra il valore critico con $\alpha=0.1$ (2.132) e quello con $\alpha=0.05$ (2.776)

(Dalla funzione di Excel DISTRIB.T(2.21;4)=0.0916)

Visualizzazione grafica del p-value per H_0: $\beta=1$

T di student con 4 gradi di libertà

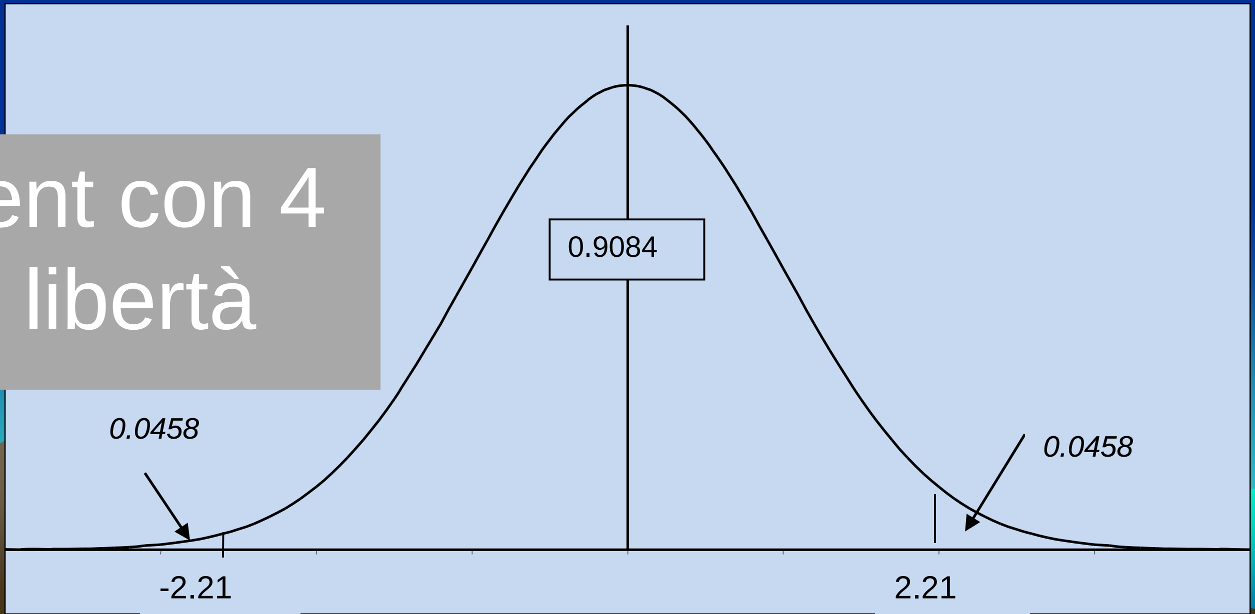


Tabella di analisi della varianza

	Gradi di libertà	*Somme dei quadrati*	*Medie dei quadrati*	*F*	*P-value*
Regressione	1	6.489	6.489	35.848	0.004
Residuo	4	0.724	0.181		
Totale	5	7.213333333			

- Si noti che $35.848^{0.5}$=5.987 (valore del test *t* per il coefficiente di regressione elevato al quadrato)

Esercizio

- 2 generici indicatori di redditività (MOL e ROS) in 7 aziende operanti nel settore della moda

MOL	ROS
2.8	6.2
2.3	4.5
4.1	9.1
1.3	7.2
0.8	0.2
-0.1	-2.5
2.1	4.8

- Si calcoli e si commenti la covarianza ed il coefficiente di correlazione lineare
- Si costruisca il diagramma di dispersione e si dica se le informazioni che esso fornisce sono in accordo con quelle del coeff. di corr. lineare
- Si dica perché il calcolo del coeff. di corr. lineare è preferibile in questo esempio al calcolo della retta di regressione

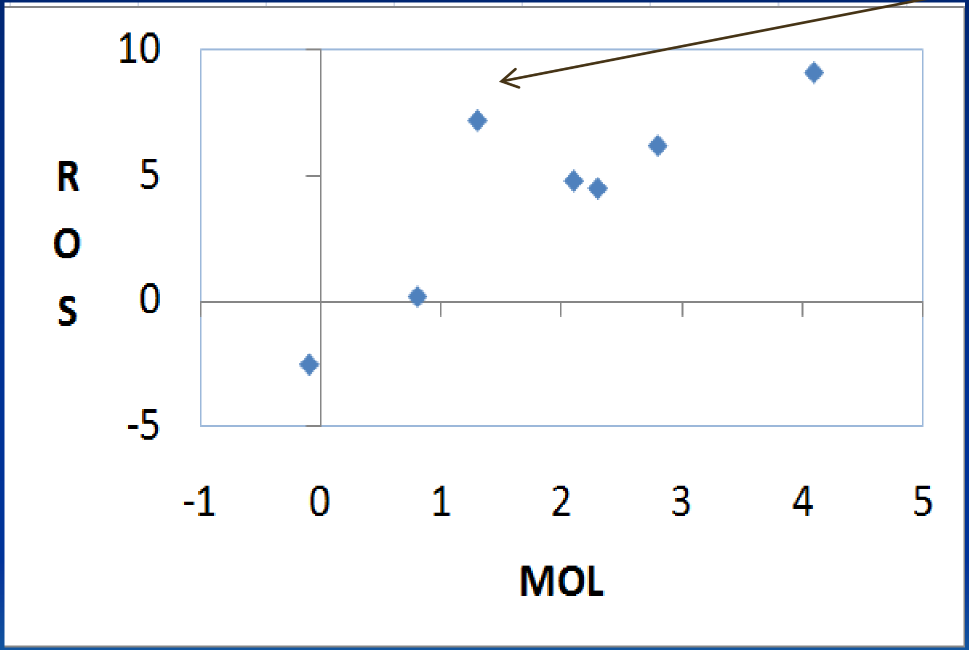
Esercizio continua

- Supponiamo che il direttore dell’azienda decida di effettuare una regressione in cui il MOL agisce da variabile esplicativa.
- Si costruisca un intervallo di confidenza del coefficiente angolare al 99%. Si testi al livello di significatività dell’1% l’ipotesi H_0:β=0
- Si costruisca la tabella di analisi della varianza per il modello finale

Soluzione

COV(MOL,ROS)=4.117 $r_{MOL,ROS}$=0.86

Diagramma di dispersione



C'è un chiaro outlier bivariato (il punto di coordinate 1.3 7.2). Se tale punto viene eliminato r diventa pari a 0.99

I due indicatori, essendo due generiche misure di redditività, sono sullo stesso piano logico. E' preferibile, quindi, un'analisi simmetrica che non individua una variabile esplicativa ed una variabile dipendente

Soluzione

- Intervallo di confidenza con tutte le osservazioni
- $Pr(-0.177 < \beta < 5.224)=0.99$
- Si noti che l'intervallo di confidenza di β contiene il valore 0

- Intervallo di confidenza escludendo l'outlier
- $Pr(2.085 < \beta < 3.535)=0.99$
- Si noti che l'intervallo di confidenza di β non contiene il valore 0

Test sull'ipotesi nulla H_0: $\beta=0$ (tutte le osservazioni)

	Coefficienti	*Errore standard*	*Stat t*	*p-value*
Intercetta	-0.581	1.533	-0.379	0.720
MOL	2.524	0.670	3.768	0.013

Al livello di significatività dell'1% non posso rifiutare l'ipotesi nulla che nell'universo MOL e ROS non presentino una relazione lineare

- p-value>0.01

Test sull’ipotesi nulla H_0: $\beta=0$ (escludendo l’outlier)

	Coefficienti	*Errore standard*	*Stat t*	*p-value*
Intercetta	-1.903	0.380	-5.004	0.007
MOL	2.810	0.157	17.843	0.0001

Al livello di significatività dell’1% posso sicuramente rifiutare l’ipotesi nulla che nell’universo MOL e ROS non presentino una relazione lineare

- p-value>0.0001

Tabella di analisi della varianza per il modello finale (senza l'outlier)

	gdl	*SQ*	*MQ*	*F*
Regressior	1	86.857	86.857	318.381
Residuo	4	1.091	0.273	
Totale	5	87.948		

- Si noti che $318.381^{0.5}=17.843$ (valore del test *t* per il coefficiente di regressione)