

Esercitazione Statistica Bivariata

Aprile, 2005

Variabili

Variabile	Scala
genere	Dicotomica
Livello di felicità	Ordinale
Ore lettura	Rapporti (cardinale)
Genere lettura	Categoriale
Pratica sport	Dicotomica (dummy)
Ore TV	Rapporti (cardinale)
Lavora	Dicotomica (dummy)
Grado interesse politica	Ordinale

Esercitazione Statistica Bivariata

Aprile, 2005

"Strumenti"

Test	Tipologia	Var 1	Var 2
Chi quadro	connessione	categ.	Categ.
K Cohen	concordanza	categ.	categ.
Q Yule	concordanza	dummy	dummy
Correlazione tetrac.	concordanza	dummy	dummy
r_s Spearman	concordanza	ordinale	ordinale
Covarianza	concordanza	cardinale	cardinale
r_{xy} Pearson	concordanza	cardinale	cardinale
Test sulle mediane	*	dicotomica	ordinale
Test sulle medie	*	dicotomica	cardinale

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

C'è una differenza nei gusti letterari tra maschi e femmine?

→ Le variabili genere e genere letterario sono associate?

ident	genere	genere lettura
1	1	3
2	1	2
3	2	1
4	1	1
5	2	2
6	1	1
7	2	3
8	1	
9	2	4
10	1	1
11	1	3
12	2	2
13	1	2
14	1	
15	2	3
16	2	1
17	2	3
18	1	
19	2	2
20	1	2
	1=m	1=giallo
	2=f	2=classico
		3=fant.
		4=altro



	giallo	classico	fanta.	altro	totale
Maschi	3	3	2	0	8
Femmine	2	3	3	1	9
totale	5	6	5	1	17

Esercitazione Statistica Bivariata

Aprile, 2005

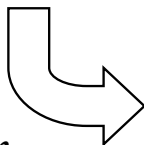
Alcuni esempi

Le variabili genere e genere letterario sono associate?

→ Variabili categoriali → test del chi quadro

Frequenze osservate

	giallo	classico	fanta.	altro	totale
Maschi	3	3	2	0	8
Femmine	2	3	3	1	9
totale	5	6	5	1	17



$$\hat{n}_{ij} = \frac{n_{i.} * n_{.j}}{N}$$

Frequenze attese

	giallo	classico	fanta.	altro	totale
Maschi	2,4	2,8	2,4	0,5	8
Femmine	2,6	3,2	2,6	0,5	9
totale	5	6	5	1	17

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

test del chi quadro

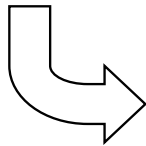
Frequenze osservate

	giallo	classico	fanta.	altro	totale
Maschi	3	3	2	0	8
Femmine	2	3	3	1	9
totale	5	6	5	1	17

Frequenze attese

	giallo	classico	fanta.	altro	totale
Maschi	2,4	2,8	2,4	0,5	8
Femmine	2,6	3,2	2,6	0,5	9
totale	5	6	5	1	17

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \hat{n}_{ij} \right)^2}{\hat{n}_{ij}}$$



$$\chi_{cal}^2 = \frac{(3-2,4)^2}{2,4} + \frac{(3-2,8)^2}{2,8} + \frac{2-2,4^2}{2,4} \dots = 1,35$$

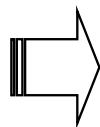
Esercitazione Statistica Bivariata

Aprile, 2005

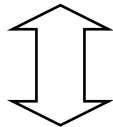
Alcuni esempi

→ Variabili categoriali → test del chi quadro

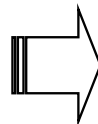
$$\alpha = .05$$
$$g.d.l. = 3$$



$$X_{crit}^2 = 7.815$$



$$X_{calc}^2 = 1.35$$



Accetto l'ipotesi nulla.
Le due variabili sono
indipendenti

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

C'è una relazione tra praticare sport e lavorare?

→ Le variabili sport/lavora sono concordanti o discordanti?

ident	sport	lavora
1	0	0
2	0	1
3	1	0
4	0	1
5	0	1
6	1	0
7	1	0
8	0	1
9	0	1
10	1	0
11	0	0
12	0	1
13	1	0
14	1	0
15	0	1
16	1	0
17	1	0
18	0	0
19	1	0
20	1	1
	0=no	0=no
	1=si	1=si



	lavoro si	lavoro no	totale
sport si	1	9	10
sport no	7	3	10
totale	8	12	20

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

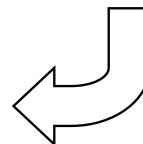
Le variabili lavora e sport sono concordanti o discordanti?

→ Variabili *dummy* → Q Yule

	lavoro si	lavoro no	totale
sport si	1	9	10
sport no	7	3	10
totale	8	12	20

$$Q = \frac{n_{11} n_{22} - n_{12} n_{21}}{n_{11} n_{22} + n_{12} n_{21}}$$

$$Q = \frac{1 \cdot 3 - 7 \cdot 9}{1 \cdot 3 + 7 \cdot 9} = \frac{3 - 63}{3 + 63} = \frac{-60}{+66} = -0,91$$



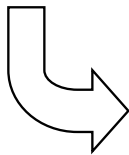
Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Q Yule

$$Q = \frac{1 \cdot 3 - 7 \cdot 9}{1 \cdot 3 + 7 \cdot 9} = \frac{3 - 63}{3 + 63} = \frac{-60}{+66} = -0,91$$



- 0,91 è un valore prossimo a -1. Quindi possiamo affermare che tra le due variabili c'è una forte associazione negativa

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

C'è una relazione tra il grado di felicità e l'interesse per la politica?

➔ Le variabili felicità e interesse per la politica sono correlate?

ident	livello di felicità	Grado di interesse per la politica
1	3	3
2	4	2
3	1	0
4	2	1
5	4	3
6	5	2
7	3	1
8	2	0
9	1	1
10	5	2
11	6	3
12	3	1
13	2	2
14	1	2
15	3	1
16	4	3
17	3	1
18	1	0
19	5	2
20	6	3



Livello Felicità	Interesse politica				totale
	0	1	2	3	
1	2	1	1	0	4
2	1	1	1	0	3
3	4	0	0	1	5
4	0	0	1	2	3
5	0	0	3	0	3
6	0	0	0	2	2
totale	7	2	6	5	20

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Le variabili felicità e interesse per la politica sono correlate?

→ Entrambe le variabili sono ordinali → R_s Spearman

$$r_s = 1 - \frac{6 \sum_{i=1}^N (x_i - y_i)^2}{N(N^2 - 1)}$$

ident	livello di felicità	Grado di interesse per la politica	rango 1	rango 2	D	D ²
1	3	3	8,5	18	-9,5	90,25
2	4	2	14	12,5	1,5	2,25
3	1	0	2,5	2	0,5	0,25
4	2	1	8,5	6,5	2	4
5	4	3	14	18	-4	16
6	5	2	17	12,5	4,5	20,25
7	3	1	8,5	6,5	2	4
8	2	0	8,5	2	6,5	42,25
9	1	1	2,5	6,5	-4	16
10	5	2	17	12,5	4,5	20,25
11	6	3	19,5	18	1,5	2,25
12	3	1	8,5	6,5	2	4
13	2	2	8,5	12,5	-4	16
14	1	2	2,5	12,5	-10	100
15	3	1	8,5	6,5	2	4
16	4	3	14	18	-4	16
17	3	1	8,5	6,5	2	4
18	1	0	2,5	2	0,5	0,25
19	5	2	17	12,5	4,5	20,25
20	6	3	19,5	18	1,5	2,25
					somma	384,5

Esercitazione Statistica Bivariata

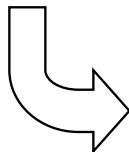
Aprile, 2005

Alcuni esempi

Le variabili felicità e interesse per la politica sono correlate?

→ Entrambe le variabili sono ordinali → R_s Spearman

$$r_s = 1 - \frac{6 \cdot 384,5}{20(400 - 1)} = 1 - \frac{2307}{7980} = 0,71$$



0,71 è un valore prossimo a 1. Quindi possiamo affermare che tra le due variabili c'è una forte associazione positiva

Esercitazione Statistica Bivariata

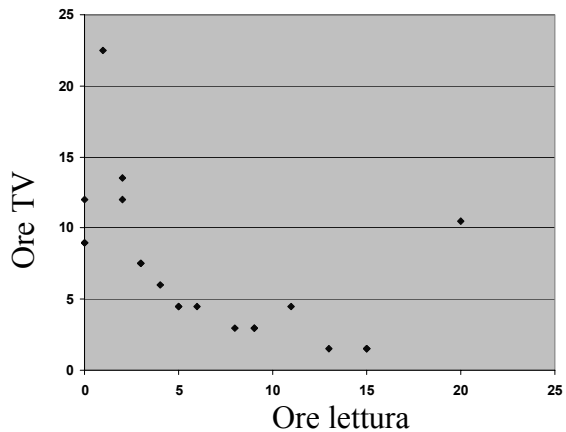
Aprile, 2005

Alcuni esempi

C'è una relazione tra ore passate davanti alla Tv ed ore di lettura?

→ Le variabili ore Tv ed ore lettura sono correlate?

ident	ore lettura settimana	ore TV
1	15	1,5
2	20	10,5
3	5	4,5
4	4	6,0
5	8	3,0
6	3	7,5
7	9	3,0
8	0	12,0
9	2	13,5
10	13	1,5
11	9	3,0
12	15	1,5
13	3	7,5
14	0	9,0
15	1	22,5
16	5	4,5
17	2	12,0
18	0	9,0
19	11	4,5
20	6	4,5



Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Le variabili ore Tv ed ore lettura sono cardinali → r pearson

$$S_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_x S_y}}$$



ident	ore lettura settimana	ore TV	scarto dalla media		scarto dalla media al quadrato	
1	15	1,5	8,45	-5,55	71,40	30,80
2	20	10,5	13,45	3,45	180,90	11,90
3	5	4,5	-1,55	-2,55	2,40	6,50
4	4	6,0	-2,55	-1,05	6,50	1,10
5	8	3,0	1,45	-4,05	2,10	16,40
6	3	7,5	-3,55	0,45	12,60	0,20
7	9	3,0	2,45	-4,05	6,00	16,40
8	0	12,0	-6,55	4,95	42,90	24,50
9	2	13,5	-4,55	6,45	20,70	41,60
10	13	1,5	6,45	-5,55	41,60	30,80
11	9	3,0	2,45	-4,05	6,00	16,40
12	15	1,5	8,45	-5,55	71,40	30,80
13	3	7,5	-3,55	0,45	12,60	0,20
14	0	9,0	-6,55	1,95	42,90	3,80
15	1	22,5	-5,55	15,45	30,80	238,70
16	5	4,5	-1,55	-2,55	2,40	6,50
17	2	12,0	-4,55	4,95	20,70	24,50
18	0	9,0	-6,55	1,95	42,90	3,80
19	11	4,5	4,45	-2,55	19,80	6,50
20	6	4,5	-0,55	-2,55	0,30	6,50
media	6,55	7,05		somma=	636,95	517,95
				varianza	31,85	25,90
				dev.st	5,64	5,09


Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Le variabili ore Tv ed ore lettura sono cardinali → r pearson

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{|s_x s_y|}$$


ident	ore lettura settimana	ore TV	scarto dalla media		Scarto x* scarto Y
1	15	1,5	8,45	-5,55	-46,90
2	20	10,5	13,45	3,45	46,40
3	5	4,5	-1,55	-2,55	3,95
4	4	6,0	-2,55	-1,05	2,68
5	8	3,0	1,45	-4,05	-5,87
6	3	7,5	-3,55	0,45	-1,60
7	9	3,0	2,45	-4,05	-9,92
8	0	12,0	-6,55	4,95	-32,42
9	2	13,5	-4,55	6,45	-29,35
10	13	1,5	6,45	-5,55	-35,80
11	9	3,0	2,45	-4,05	-9,92
12	15	1,5	8,45	-5,55	-46,90
13	3	7,5	-3,55	0,45	-1,60
14	0	9,0	-6,55	1,95	-12,77
15	1	22,5	-5,55	15,45	-85,75
16	5	4,5	-1,55	-2,55	3,95
17	2	12,0	-4,55	4,95	-22,52
18	0	9,0	-6,55	1,95	-12,77
19	11	4,5	4,45	-2,55	-11,35
20	6	4,5	-0,55	-2,55	1,40
media	6,55	7,05		somma	-307,05
				covarianza	-15,35
				Sx*Sy	28,71
				r	-0,53

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Controllo della significatività

$$t = \frac{r}{1-r^2} \sqrt{(N-2)}$$

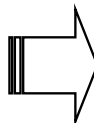


$$t = \frac{-0,535}{1-0,535^2} \sqrt{(20-2)} = \frac{-0,535}{1-0,286} \sqrt{(18)}$$

$$t = \frac{-0,535}{0,714} \sqrt{(18)} = -0,749 \cdot 4,24 = -3,18$$

$$t_{crit} (18ddl - \alpha = .05) = \pm 2,1$$

$$t_{calc} = -3,18$$



t calcolato è inferiore al limite Sx quindi cade nell'area di rifiuto.

→ Rifiuto l'ipotesi nulla.
r è diverso da zero

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Le variabili ore Tv ed ore lettura sono cardinali → r pearson

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{|s_x s_y|}$$



ident	ore lettura settimana	ore TV	scarto dalla media		scarto dalla media al quadrato	
1	15	1,5	9,157895	-5,36842	83,87	28,82
3	5	4,5	-0,84211	-2,36842	0,71	5,61
4	4	6,0	-1,84211	-0,86842	3,39	0,75
5	8	3,0	2,157895	-3,86842	4,66	14,96
6	3	7,5	-2,84211	0,631579	8,08	0,40
7	9	3,0	3,157895	-3,86842	9,97	14,96
8	0	12,0	-5,84211	5,131579	34,13	26,33
9	2	13,5	-3,84211	6,631579	14,76	43,98
10	13	1,5	7,157895	-5,36842	51,24	28,82
11	9	3,0	3,157895	-3,86842	9,97	14,96
12	15	1,5	9,157895	-5,36842	83,87	28,82
13	3	7,5	-2,84211	0,631579	8,08	0,40
14	0	9,0	-5,84211	2,131579	34,13	4,54
15	1	22,5	-4,84211	15,63158	23,45	244,35
16	5	4,5	-0,84211	-2,36842	0,71	5,61
17	2	12,0	-3,84211	5,131579	14,76	26,33
18	0	9,0	-5,84211	2,131579	34,13	4,54
19	11	4,5	5,157895	-2,36842	26,60	5,61
20	6	4,5	0,157895	-2,36842	0,02	5,61
media	5,842105	6,86842105			somma= 446,53	505,42
					varianza 23,50	26,60
					dev.st 4,85	5,16

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

Le variabili ore Tv ed ore lettura sono cardinali → r pearson

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{|s_x s_y|}$$



ident	ore lettura settimana	ore TV	scarto dalla media		Scarto x* scarto Y
1	15	1,5	9,157895	-5,368421	-49,16
3	5	4,5	-0,842105	-2,368421	1,99
4	4	6,0	-1,842105	-0,868421	1,60
5	8	3,0	2,157895	-3,868421	-8,35
6	3	7,5	-2,842105	0,631579	-1,80
7	9	3,0	3,157895	-3,868421	-12,22
8	0	12,0	-5,842105	5,131579	-29,98
9	2	13,5	-3,842105	6,631579	-25,48
10	13	1,5	7,157895	-5,368421	-38,43
11	9	3,0	3,157895	-3,868421	-12,22
12	15	1,5	9,157895	-5,368421	-49,16
13	3	7,5	-2,842105	0,631579	-1,80
14	0	9,0	-5,842105	2,131579	-12,45
15	1	22,5	-4,842105	15,63158	-75,69
16	5	4,5	-0,842105	-2,368421	1,99
17	2	12,0	-3,842105	5,131579	-19,72
18	0	9,0	-5,842105	2,131579	-12,45
19	11	4,5	5,157895	-2,368421	-12,22
20	6	4,5	0,157895	-2,368421	-0,37
media	5,842105	6,86842105			somma -355,89
					covarianza -17,79
					Sx*Sy 25,03
					r -0,71

Esercitazione Statistica Bivariata

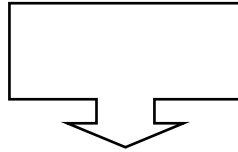
Aprile, 2005

Alcuni esempi

Outliers

20 casi $\Rightarrow r_{x,y} = -0,53$

19 casi $\Rightarrow r_{x,y} = -0,71$



Eliminando il soggetto "outlier"
il valore del coefficiente di
correlazione varia sensibilmente

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

C'è una differenza nella quantità di ore dedicate alla lettura tra maschi e femmine ?

→ La media di ore di letture dei maschi è diversa da quella delle femmine?

ident	genere	ore lettura settimana
1	1	15
2	1	20
3	2	5
4	1	4
5	2	8
6	1	3
7	2	9
8	1	0
9	2	2
10	1	13
11	1	9
12	2	15
13	1	3
14	1	0
15	2	1
16	2	5
17	2	2
18	1	0
19	2	11
20	1	6



	ore lettura settimana
	media
Maschi	6,6
Femmine	6,4

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

La media di ore di letture dei maschi è diversa da quella delle femmine?

→ Test della media

Varianze delle popolazioni		CAMPIONI	
		Grandi	Piccoli
Uguali	Note	A	B
	Ignote	C	D
Diverse	Note	E	F
	Ignote	G	H

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

La media di ore di letture dei maschi è diversa da quella delle femmine?

→ Test della media

$$H_0: (\mu_m = \mu_f) \quad \Rightarrow \quad t = \frac{(\bar{x}_m - \bar{x}_f)}{\sqrt{\frac{s_m^2}{n_m - 1} + \frac{s_f^2}{n_f - 1}}}$$

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

La media di ore di letture dei maschi è diversa da quella delle femmine?

→ Test della media

	ore lettura settimana maschi	ore lettura settimana femmine	scarto dalla media M	scarto dalla media F	M ²	F ²
	15	5	8,4	-1,4	70,0	2,1
	20	8	13,4	1,6	178,6	2,4
	4	9	-2,6	2,6	7,0	6,5
	3	2	-3,6	-4,4	13,2	19,8
	0	15	-6,6	8,6	44,0	73,2
	13	1	6,4	-5,4	40,5	29,6
	9	5	2,4	-1,4	5,6	2,1
	3	2	-3,6	-4,4	13,2	19,8
	0	11	-6,6	4,6	44,0	20,8
	0		-6,6		44,0	
	6		-0,6		0,4	
media	6,6	6,4		somma	460,5	176,2
				varianza	41,9	16,0
				dev.st	6,5	4,0

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

La media di ore di letture dei maschi è diversa da quella delle femmine?

→ Test della media

	m	f		m	f
media	6,6	6,4	somma	460,5	176,2
			varianza	41,9	16,0
			dev.st	6,5	4,0

$$t = \frac{(\bar{x}_m - \bar{x}_f)}{\sqrt{\frac{S_m^2}{n_m - 1} + \frac{S_f^2}{n_f - 1}}}$$



$$t = \frac{(6,6 - 6,4)}{\sqrt{\frac{41,9}{10} + \frac{16}{8}}} = \frac{0,2}{\sqrt{6,19}} = \frac{0,2}{2,49} = 0,08$$

Esercitazione Statistica Bivariata

Aprile, 2005

Alcuni esempi

La media di ore di letture dei maschi è diversa da quella delle femmine?

→ Test della media

