

Corso di

Statistica

La correlazione

Lezione sette

Cristina Davino

cdavino@unimc.it



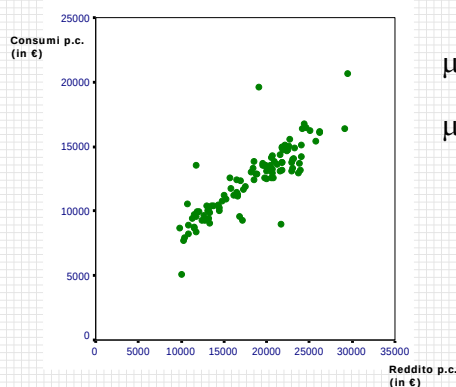
Lo studio della interdipendenza tra caratteri quantitativi: la correlazione

	Reddito p.c.	Consumi p.c.
AG	9864,3	8676,5
AL	20555,0	14150,9
AN	22310,9	14667,4
AO	24583,4	16475,0
AP	18179,3	13014,7
AQ	16165,1	11258,8
AR	19780,3	12601,6
AT	18540,8	12395,0
AV	13169,7	9451,2
BA	14770,7	10794,0
BG	20606,6	12963,1
BI	23963,6	13169,7
BL	20606,6	14305,9
BN	12653,2	9657,7
BO	29489,7	20658,3
BR	11775,2	13531,2
BS	21639,5	8986,4
BZ	24170,2	16371,7
:	:	:

La correlazione

	Reddito p.c.	Consumi p.c.
AG	9864,3	8676,5
AL	20555,0	14150,9
AN	22310,9	14667,4
AO	24583,4	16475,0
AP	18179,3	13014,7
AQ	16165,1	11258,8
AR	19780,3	12601,6
AT	18540,8	12395,0
AV	13169,7	9451,2
BA	14770,7	10794,0
BG	20606,6	12963,1
BI	23963,6	13169,7
BL	20606,6	14305,9
BN	12653,2	9657,7
BO	29489,7	20658,3
BR	11775,2	13531,2
BS	21639,5	8986,4
BZ	24170,2	16371,7
:	:	:

Il diagramma di dispersione



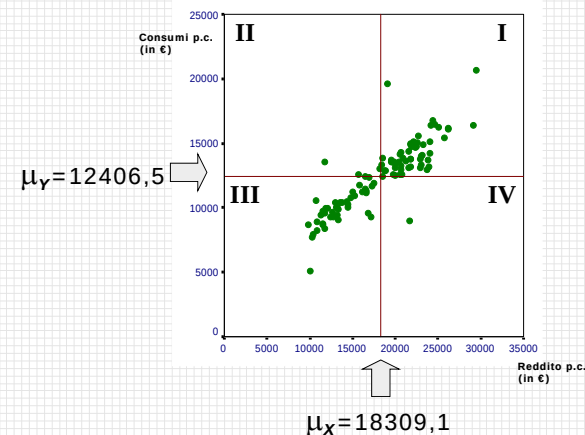
$$\mu_X = 18309,1$$

$$\mu_Y = 12406,5$$

La correlazione

	Reddito p.c.	Consumi p.c.
AG	9864,3	8676,5
AL	20555,0	14150,9
AN	22310,9	14667,4
AO	24583,4	16475,0
AP	18179,3	13014,7
AQ	16165,1	11258,8
AR	19780,3	12601,6
AT	18540,8	12395,0
AV	13169,7	9451,2
BA	14770,7	10794,0
BG	20606,6	12963,1
BI	23963,6	13169,7
BL	20606,6	14305,9
BN	12653,2	9657,7
BO	29489,7	20658,3
BR	11775,2	13531,2
BS	21639,5	8986,4
BZ	24170,2	16371,7
:	:	:

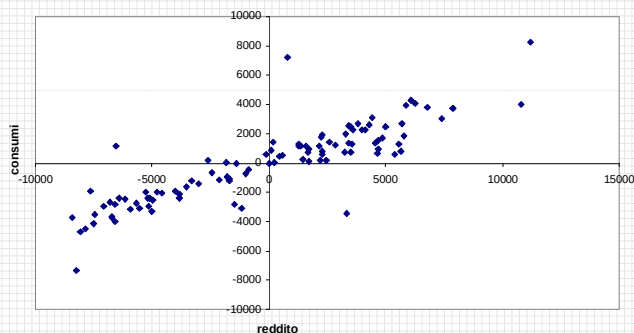
Il diagramma di dispersione



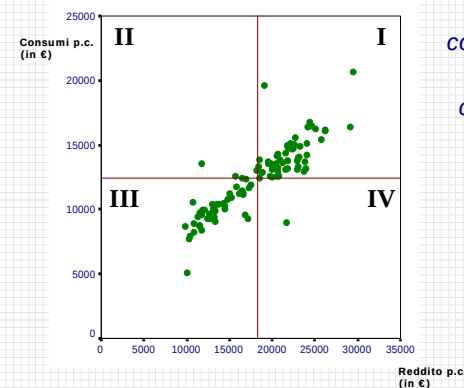
La correlazione

	Reddito p.c.	Consumi p.c.
AG	9864,3	8676,5
AL	20555,0	14150,9
AN	22310,9	14667,4
AO	24583,4	16475,0
AP	18179,3	13014,7
AQ	16165,1	11258,8
AR	19780,3	12601,6
AT	18540,8	12395,0
AV	13169,7	9451,2
BA	14770,7	10794,0
BG	20606,6	12963,1
BI	23963,6	13169,7
BL	20606,6	14305,9
BN	12653,2	9657,7
BO	29489,7	20658,3
BR	11775,2	13531,2
BS	21639,5	8986,4
BZ	24170,2	16371,7
:	:	:

Gli scostamenti dalle medie



La codevianza e la covarianza



$$codev(XY) = \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

$$cov(XY) = \sigma_{xy} = \frac{\sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)}{n}$$

Il segno

>0 ➡ Prevalenza di punti nei quadranti I e III
 Relazione diretta

<0 ➡ Prevalenza di punti nei quadranti II e IV
 Relazione inversa

$=0$ ➡ Nessuna prevalenza
 Mancanza di relazione lineare

L'intensità

$$Max[\sigma_{x,y}] = \sigma_x \cdot \sigma_y$$

La codevianza e la covarianza

	Reddito p.c.	Consumi p.c.
AG	9864,3	8676,5
AL	20555,0	14150,9
AN	22310,9	14667,4
AO	24583,4	16475,0
AP	18179,3	13014,7
AQ	16165,1	11258,8
AR	19780,3	12601,6
AT	18540,8	12395,0
AV	13169,7	9451,2
BA	14770,7	10794,0
BG	20606,6	12963,1
BI	23963,6	13169,7
BL	20606,6	14305,9
BN	12653,2	9657,7
BO	29489,7	20658,3
BR	11775,2	13531,2
BS	21639,5	8986,4
BZ	24170,2	16371,7
:	:	:

$$\mu_x = 18309,1 \quad \sigma_x = 4828,80$$

$$\mu_y = 12406,5 \quad \sigma_y = 2631,75$$

La codevianza e la covarianza

	Reddito p.c.	Consumi p.c.	X-μ _x	Y-μ _y	(X-μ _x)(Y-μ _y)
AG	9864,3	8676,5	-8444,8	-3730,0	31499328,5
AL	20555,0	14150,9	2245,8	1744,4	3917676,4
AN	22310,9	14667,4	4001,8	2260,9	9047576,4
AO	24583,4	16475,0	6274,2	4068,5	25526472,4
AP	18179,3	13014,7	-129,9	608,2	-78988,0
AQ	16165,1	11258,8	-2144,0	-1147,7	2460807,0
AR	19780,3	12601,6	1471,2	195,1	286950,8
AT	18540,8	12395,0	231,7	-11,5	-2670,5
AV	13169,7	9451,2	-5139,5	-2955,3	15188957,8
BA	14770,7	10794,0	-3538,5	-1612,5	5705967,9
BG	20606,6	12963,1	2297,5	556,6	1278713,3
BI	23963,6	13169,7	5654,5	763,2	4315204,9
BL	20606,6	14305,9	2297,5	1899,4	4363747,5
BN	12653,2	9657,7	-5656,0	-2748,8	15546863,7
BO	29489,7	20658,3	11180,5	8251,8	92259384,2
BR	11775,2	13531,2	-6533,9	1124,7	-7348526,6
BS	21639,5	8986,4	3330,4	-3420,1	-11390429,8
BZ	24170,2	16371,7	5861,0	3965,2	23240053,2
:	:	:	:	:	:

$$\sigma_{xy} = \frac{\sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)}{n} = 11.068.848,4$$

$$\mu_x = 18309,1 \quad \sigma_x = 4828,80$$

$$\mu_y = 12406,5 \quad \sigma_y = 2631,75$$

La correlazione

	Reddito p.c.	Consumi p.c.	X-μ _x	Y-μ _y	(X-μ _x)(Y-μ _y)
AG	9864,3	8676,5	-8444,8	-3730,0	31499328,5
AL	20555,0	14150,9	2245,8	1744,4	3917676,4
AN	22310,9	14667,4	4001,8	2260,9	9047576,4
AO	24583,4	16475,0	6274,2	4068,5	25526472,4
AP	18179,3	13014,7	-129,9	608,2	-78988,0
AQ	16165,1	11258,8	-2144,0	-1147,7	2460807,0
AR	19780,3	12601,6	1471,2	195,1	286950,8
AT	18540,8	12395,0	231,7	-11,5	-2670,5
AV	13169,7	9451,2	-5139,5	-2955,3	15188957,8
BA	14770,7	10794,0	-3538,5	-1612,5	5705967,9
BG	20606,6	12963,1	2297,5	556,6	1278713,3
BI	23963,6	13169,7	5654,5	763,2	4315204,9
BL	20606,6	14305,9	2297,5	1899,4	4363747,5
BN	12653,2	9657,7	-5656,0	-2748,8	15546863,7
BO	29489,7	20658,3	11180,5	8251,8	92259384,2
BR	11775,2	13531,2	-6533,9	1124,7	-7348526,6
BS	21639,5	8986,4	3330,4	-3420,1	-11390429,8
BZ	24170,2	16371,7	5861,0	3965,2	23240053,2
:	:	:	:	:	:

$$\mu_X = 18309,1 \quad \sigma_X = 4828,80$$

$$\mu_Y = 12406,5 \quad \sigma_Y = 2631,75$$

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X) \cdot (y_i - \mu_Y)}{n} = 11.068.848,4$$

$$\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = r_{XY} \text{ Coefficiente di correlazione}$$

$$-1 \leq \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \leq +1$$

La correlazione

	Reddito p.c.	Consumi p.c.	X-μ _x	Y-μ _y	(X-μ _x)(Y-μ _y)
AG	9864,3	8676,5	8444,8	3730,0	31499328,5
AL	20555,0	14150,9	2245,8	1744,4	3917676,4
AN	22310,9	14667,4	4001,8	2260,9	9047576,4
AO	24583,4	16475,0	6274,2	4068,5	25526472,4
AP	18179,3	13014,7	-129,9	608,2	-78988,0
AQ	16165,1	11258,8	-2144,0	-1147,7	2460807,0
AR	19780,3	12601,6	1471,2	195,1	286950,8
AT	18540,8	12395,0	231,7	-11,5	-2670,5
AV	13169,7	9451,2	-5139,5	-2955,3	15188957,8
BA	14770,7	10794,0	-3538,5	-1612,5	5705967,9
BG	20606,6	12963,1	2297,5	556,6	1278713,3
BI	23963,6	13169,7	5654,5	763,2	4315204,9
BL	20606,6	14305,9	2297,5	1899,4	4363747,5
BN	12653,2	9657,7	-5656,0	-2748,8	15546863,7
BO	29489,7	20658,3	11180,5	8251,8	92259384,2
BR	11775,2	13531,2	-6533,9	1124,7	-7348526,6
BS	21639,5	8986,4	3330,4	-3420,1	-11390429,8
BZ	24170,2	16371,7	5861,0	3965,2	23240053,2
:	:	:	:	:	:

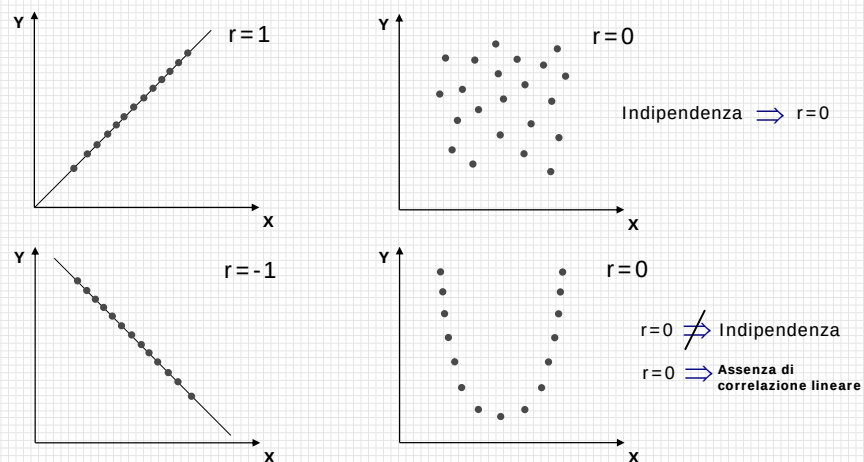
$$\mu_X = 18309,1 \quad \sigma_X = 4828,80$$

$$\mu_Y = 12406,5 \quad \sigma_Y = 2631,75$$

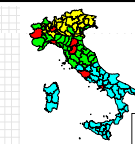
$$\sigma_{XY} = \frac{\sum (x_i - \mu_X) \cdot (y_i - \mu_Y)}{n} = 11.068.848,4$$

$$r_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{11.068.848,4}{4.828,80 \times 2.631,75} = 0,87$$

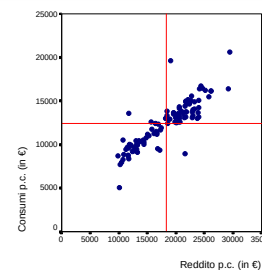
L'interpretazione di r



La correlazione

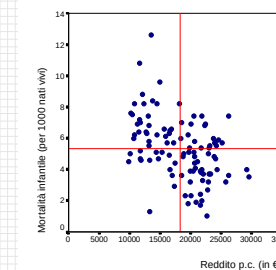


X=Reddito ; Y=Consumi



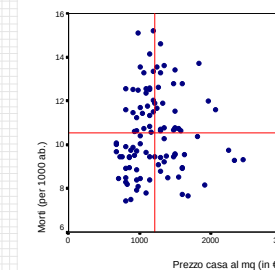
$$r=0,870$$

X=Reddito ; Y=Mortalità inf.



$$r=-0,438$$

X=Prezzo casa ; Y=Mortalità

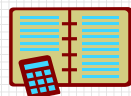


$$r=0,057$$

Dove e come studiare



- S. Borra, A. Di Ciaccio (2008) – *Statistica – Metodologie per le scienze economiche e sociali* – McGraw-Hill. Paragrafo 6.9
- D. Piccolo (2004) – *Statistica per le decisioni* – Il Mulino. Paragrafo 7.5



File “esercizi statistiche bivariate.pdf” { Esercizio n. 2

T3: Test di autovalutazione

(domande 1, 2, 3, 5, 7, 8, 11, 15, 19, 20, 21)

Riepilogo



- ✓ **Analisi della interdipendenza tra due caratteri quantitativi**
 - ✓ Codevarianza
 - ✓ Covarianza
 - ✓ Coefficiente di correlazione