

Probabilità e Statistica

a.a. 2016/2017

C.d.L.: Ingegneria per l'Ambiente ed il Territorio, Ingegneria Meccanica e dei Materiali, Ingegneria Gestionale, Ingegneria Informatica

C.d.L.: Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria dell'automazione Industriale

Statistica descrittiva: campione bivariato coefficiente di correlazione campionaria

Marco Pietro Longhi

Esempio

Esempio 1. Determinare il coefficiente di correlazione (lineare) campionaria per il campione bivariato di 18 unità i cui dati sono riportati nella tabella seguente.

x_i	5	6	7	8	9	10	12	13	15
y_i	20,0	24,4	14,0	27,1	12,0	28,7	29,5	10,2	10,0

x_i	16	18	19	20	21	22	23	24	25
y_i	29,9	10,5	29,2	11,3	28,0	12,9	26,0	15,6	20,0

Fonte: fantasia (esempio didattico)

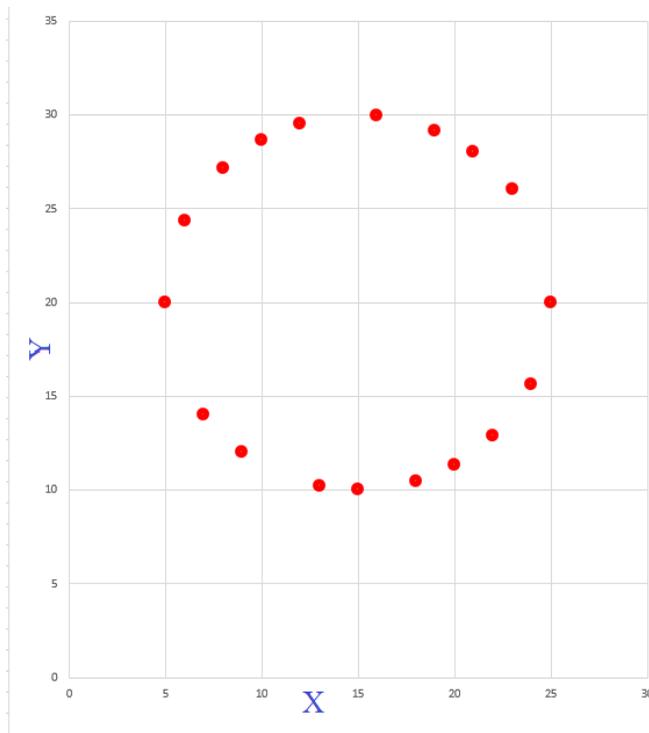
x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
5	20,0	-10,2	0,04	-0,39	103,3611	0,0014
6	24,4	-9,2	4,40	-40,30	84,0278	19,3322
7	14,0	-8,2	-5,96	48,69	66,6944	35,5462
8	27,1	-7,2	7,18	-51,45	51,3611	51,5433
9	12,0	-6,2	-7,96	49,10	38,0278	63,3944
10	28,7	-5,2	8,70	-44,94	26,6944	75,6586
12	29,5	-3,2	9,58	-30,33	10,0278	91,7253
13	10,2	-2,2	-9,76	21,15	4,6944	95,2580
15	10,0	-0,2	-9,96	1,66	0,0278	99,2426
16	29,9	0,8	9,99	8,32	0,6944	99,7565
18	10,5	2,8	-9,50	-26,92	8,0278	90,2776
19	29,2	3,8	9,20	35,28	14,6944	84,6969
20	11,3	4,8	-8,62	-41,67	23,3611	74,3443
21	28,0	5,8	8,04	46,89	34,0278	64,6085
22	12,9	6,8	-7,10	-48,54	46,6944	50,4595
23	26,0	7,8	6,04	47,30	61,3611	36,4567
24	15,6	8,8	-4,32	-38,17	78,0278	18,6707
25	20,0	9,8	0,04	0,37	96,6944	0,0014
			somma	-63,96	748,5000	1050,9741

$$\bar{x} = 15,2 \quad \bar{y} = 19,96 \quad S_x = \sqrt{\frac{748,5000}{(18-1)}} = 6,64 \quad s_y = \sqrt{\frac{1050,9741}{(18-1)}} = 7,86$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n-1) \cdot S_x \cdot S_y} = \frac{-63,96}{(18-1) \cdot 6,64 \cdot 7,86} = -0,0721$$

Dal valore del coefficiente di correlazione possiamo dedurre che fra i due caratteri, il carattere X e il carattere Y , non esiste alcuna relazione \dots LINEARE

Rappresentiamo i dati in un diagramma a dispersione



Dal grafico si intuisce che fra i dati é presente una dipendenza **non lineare**: le coppie di valori si distribuiscono su un arco di circonferenza.