

## Correlazione (quesiti da 1 a 2)

Le tabelle di contingenza con il corrispondente calcolo dell'indice chi\_quadro ci forniscono informazioni sulla relazione di dipendenza o indipendenza di due caratteri statistici.

**La Correlazione** fornisce informazioni sul grado di dipendenza lineare fra due variabili quantitative. E' un indice compreso tra -1 e 1, 0 significa che non c'è nessun tipo di relazione lineare, 1 e -1 massima relazione. (1 quando i dati sono allineati su di una retta con coefficiente angolare positivo, -1 quando i dati sono allineati su di una retta con coefficiente angolare negativo)

Il chi\_quadro è applicabile quindi sia a variabili qualitative che a variabili quantitative, invece la correlazione è applicabile solo a variabili quantitative.

Siano X e Y le due variabili quantitative. Per determinare la correlazione devo prima calcolare

1) la covarianza di X e Y (definita come la media del prodotto degli scarti di X e di Y)

2) la varianza di X (definita come la media degli scarti al quadrato di X)

ricordo che la radice quadrata della varianza è chiamata deviazione standard di X (dev.stdX)

3) La varianza di Y (definita come la media degli scarti al quadrato di Y)

ricordo che la radice quadrata della varianza è chiamata deviazione standard di Y (dev.stdY)

Covarianza(X,Y)

Correlazione(X,Y)=  
-----  
(dev.stdX)\*(dev.stdY)

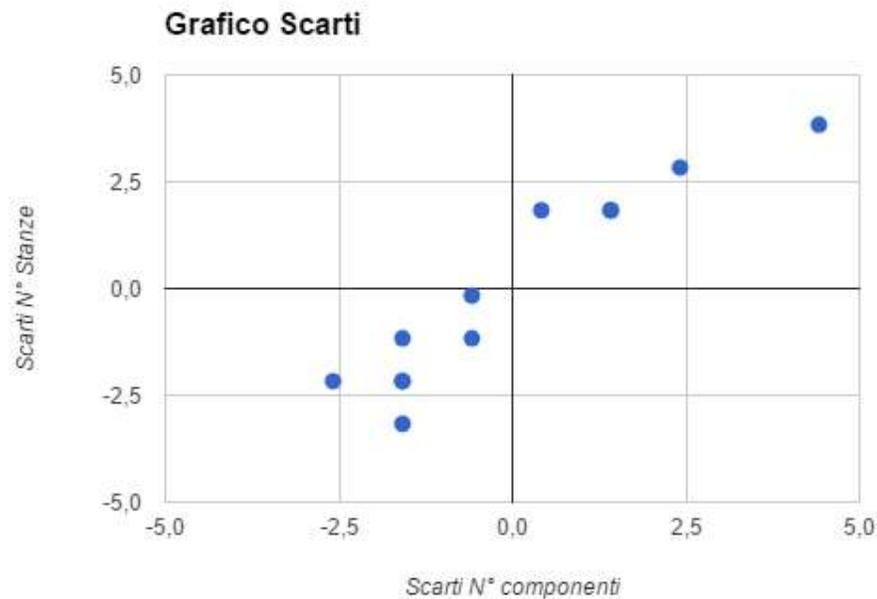
Qui sotto sono riportati i dati relativi ad un piccolo campione di famiglie. Sono registrati il numero di componenti della famiglia ed il numero di stanze in cui vive la famiglia stessa

	Numero	Numero	Scarti	Scarti	Prodotto scarti	Scarti^2	Scarti^2
	Componenti	Stanze	N° Componenti	N° Stanze		N° Componenti	N° Stanze
	2	3	-1,6	-2,2	3,4	2,5	4,7
	3	4	-0,6	-1,2	0,7	0,3	1,4
	5	7	1,4	1,8	2,6	2,0	3,4
	2	2	-1,6	-3,2	5,0	2,5	10,0
	1	3	-2,6	-2,2	5,6	6,7	4,7
	2	3	-1,6	-2,2	3,4	2,5	4,7
	6	8	2,4	2,8	6,8	5,8	8,0
	5	7	1,4	1,8	2,6	2,0	3,4
	4	7	0,4	1,8	0,8	0,2	3,4
	3	5	-0,6	-0,2	0,1	0,3	0,0
	2	4	-1,6	-1,2	1,8	2,5	1,4
	8	9	4,4	3,8	16,9	19,5	14,7
<b>Medie</b>	<b>3,6</b>	<b>5,2</b>	<b>0,0</b>	<b>0,0</b>	<b>4,2</b>	<b>3,9</b>	<b>5,0</b>

Covarianza (X,Y)=4,2  
Varianza(X)=3,9  
Varianza(Y)=5  
Correlazione(X,Y)=0,94

La Correlazione è quasi 1, significa che è presente un legame lineari quasi perfetto fra le due variabili.

Mostro qui sotto il



Il grafico qui presentato esprime una relazione lineare molto forte e positiva. I punti infatti si trovano tutti nel primo e nel terzo quadrante, qui il prodotto degli scarti è sempre positivo, quindi quando è eseguita la media (per determinare la covarianza e di conseguenza la correlazione), il calcolo è eseguito solo su quantità positive.

### Quesito 1

Utilizzando il foglio elettronico modifica la seconda colonna (Numero stanze). assegna un significato a tuo piacere alla seconda colonna; fornisci dati di fantasia relativi all'interpretazione che tu hai fornito. I dati che fornirai devono permettere la costruzione di un modello con correlazione praticamente nulla.

### Quesito 2

Utilizzando il foglio elettronico modifica la seconda colonna (Numero stanze). assegna un significato a tuo piacere alla seconda colonna; fornisci dati di fantasia relativi all'interpretazione che tu hai fornito. I dati che fornirai devono permettere la costruzione di un modello con correlazione negativa molto alta.