
Corso di Laurea in Economia Aziendale

Docente: Marta Nai Ruscone

Statistica

a.a. 2015/2016

Connessione e indipendenza stocastica

1) La connessione e l'indipendenza stocastica

Dipendenza fra due caratteri

La Connessione

■ Indipendenza stocastica

C'è indipendenza stocastica se le frequenze % di riga (e % di colonna) non variano al variare delle modalità di X (e delle modalità di Y).

| X | Y | y1 | y2 | y3 | |
|----|---|----|----|----|----|
| x1 | | 3 | 6 | 18 | 27 |
| x2 | | 2 | 4 | 12 | 18 |
| x3 | | 1 | 2 | 6 | 9 |
| | | 6 | 12 | 36 | 54 |

| X | Y | y1 | y2 | y3 | |
|----|---|--------|--------|--------|---------|
| x1 | | 11,11% | 22,22% | 66,67% | 100,00% |
| x2 | | 11,11% | 22,22% | 66,67% | 100,00% |
| x3 | | 11,11% | 22,22% | 66,67% | 100,00% |

→ Frequenze % di riga

Frequenze % di colonna ←

| X | Y | y1 | y2 | y3 |
|----|---|---------|---------|---------|
| x1 | | 50,00% | 50,00% | 50,00% |
| x2 | | 33,33% | 33,33% | 33,33% |
| x3 | | 16,67% | 16,67% | 16,67% |
| | | 100,00% | 100,00% | 100,00% |

Esempio Calcolo Indice di Connessione Età e Voto alle ultime elezioni

| Età | Voto ultime elezioni | |
|-------|----------------------|---------|
| | No | Sì |
| 21-30 | 22,22% | 38,10% |
| 31-40 | 33,33% | 28,57% |
| 41-50 | 22,22% | 4,76% |
| 51-60 | 22,22% | 28,57% |
| | 100,00% | 100,00% |

Non c'è indipendenza stocastica perché le frequenze % di riga (e % di colonna) variano al variare delle modalità di X (e delle modalità di Y).

| Età | Voto ultime elezioni | | Totale complessivo |
|-------|----------------------|--------|--------------------|
| | No | Sì | |
| 21-30 | 20,00% | 80,00% | 100,00% |
| 31-40 | 33,33% | 66,67% | 100,00% |
| 41-50 | 66,67% | 33,33% | 100,00% |
| 51-60 | 25,00% | 75,00% | 100,00% |

Correlazione lineare

1) Grafico a dispersione

2) La correlazione lineare

-la covarianza

funzione di excel: “covarianza()”

strumento di analisi: *covarianza*

-il coefficiente di correlazione lineare

funzione di excel “correlazione()”

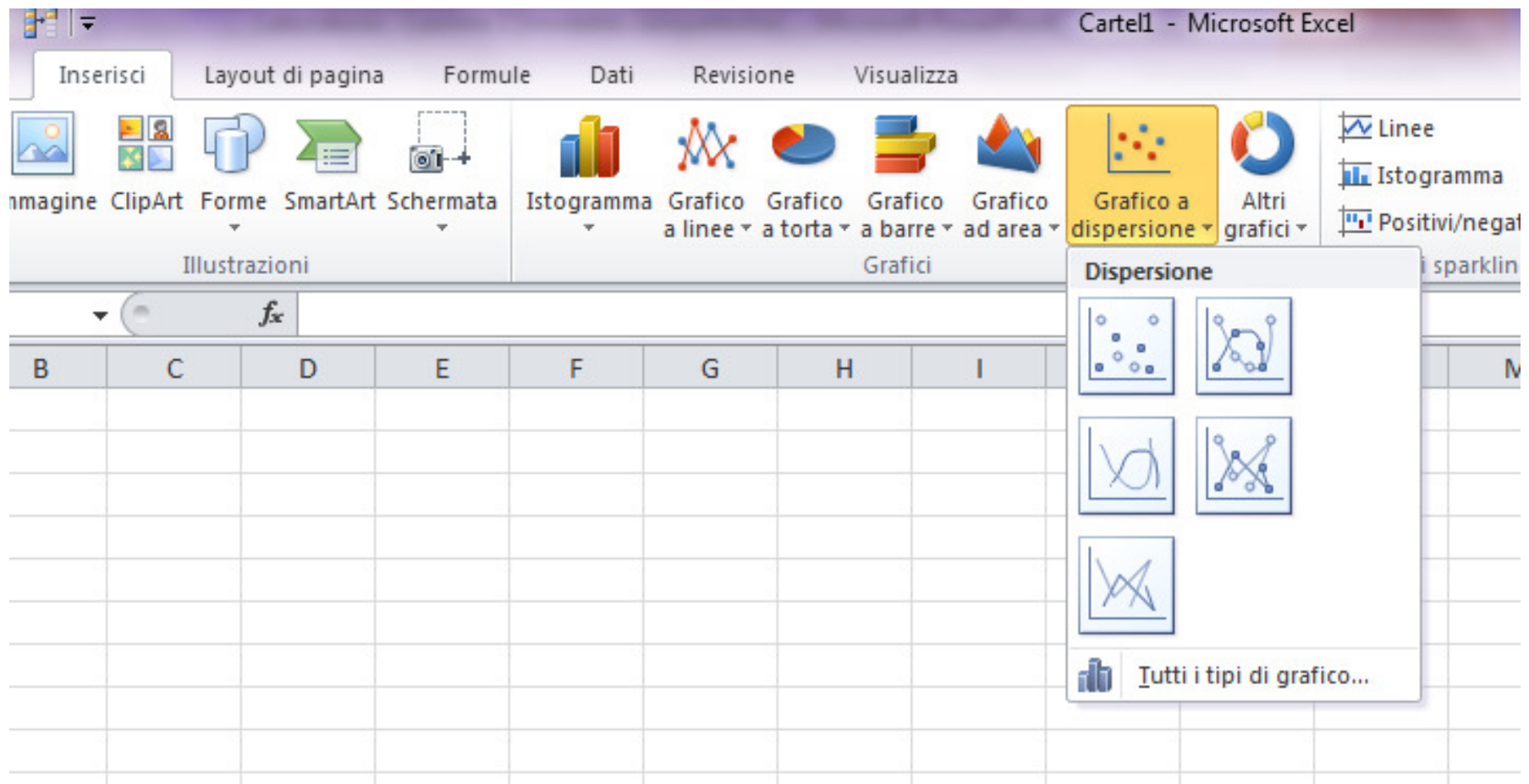
strumento di analisi: *correlazione*

Grafico a dispersione

Funzione:

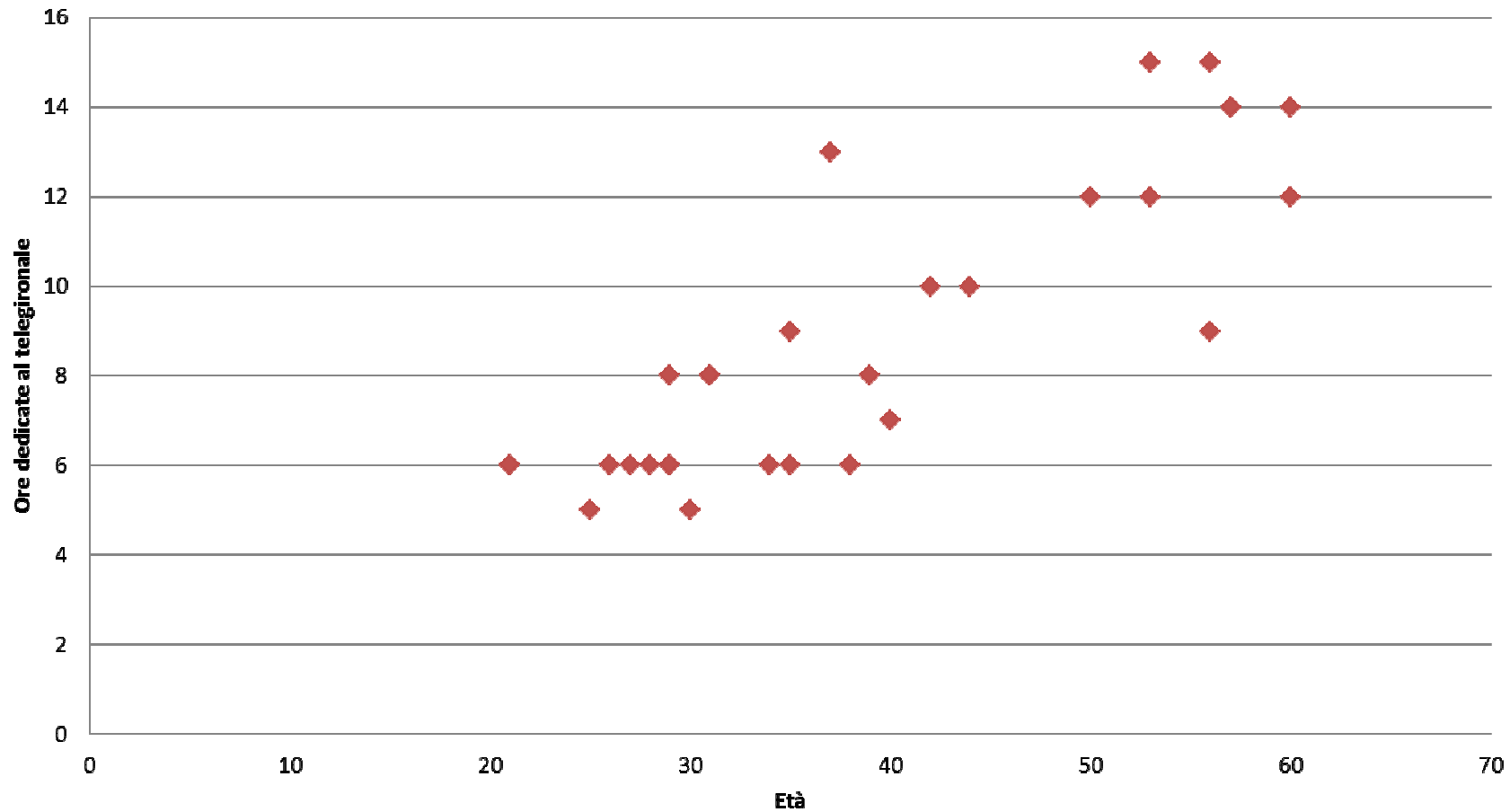
Scheda «inserisci» > grafici > grafico a dispersione

Costituisce un primo passo per valutare la relazione fra variabili quantitative






Esempio: Grafico a dispersione

v.s. Età e v.s. Ore dedicate ai telegiornali



Correlazione lineare fra X e Y

■ Covarianza
$$Cov(X, Y) = \frac{\sum_{i=1}^k (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

- $Cov(X, Y) > 0$  **Tendenza alla linearità positiva**
- $Cov(X, Y) < 0$  **Tendenza alla linearità negativa**
- $Cov(X, Y) = 0$  **Incorrelazione**

Funzione di excel: “Covarianza()”

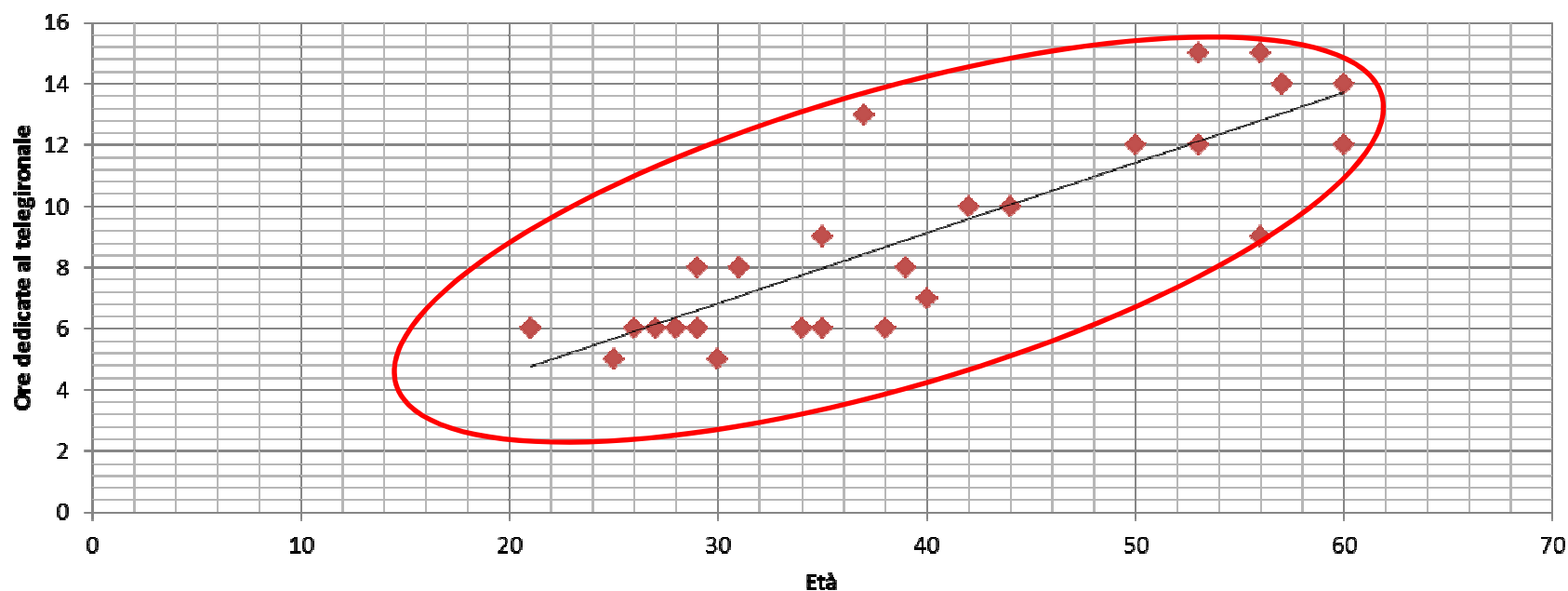
Strumento: Covarianza

Esempio: Covarianza positiva

v.s. Età e v.s. Ore dedicate ai telegiornali

- Calcolo covarianza = COVARIANZA(A2:A31;B2:B31)
= 34,37 > 0

Tendenza alla linearità positiva: retta interpolante crescente



■ Scheda «dati» > Analisi dati > Covarianza

| | <i>Età</i> | <i>Ore settimanali dedicate al telegiornale</i> |
|---|-------------|---|
| <i>Età</i> | 150,1566667 | |
| <i>Ore settimanali dedicate al telegiornale</i> | 34,37333333 | 10,72888889 |

Covarianza

Var.pop

Covarianza

Input

Intervallo di input:

Dati raggruppati per: Colonne Righe

Etichette nella prima riga

Opzioni di output

Intervallo di output:

Nuovo foglio di lavoro:

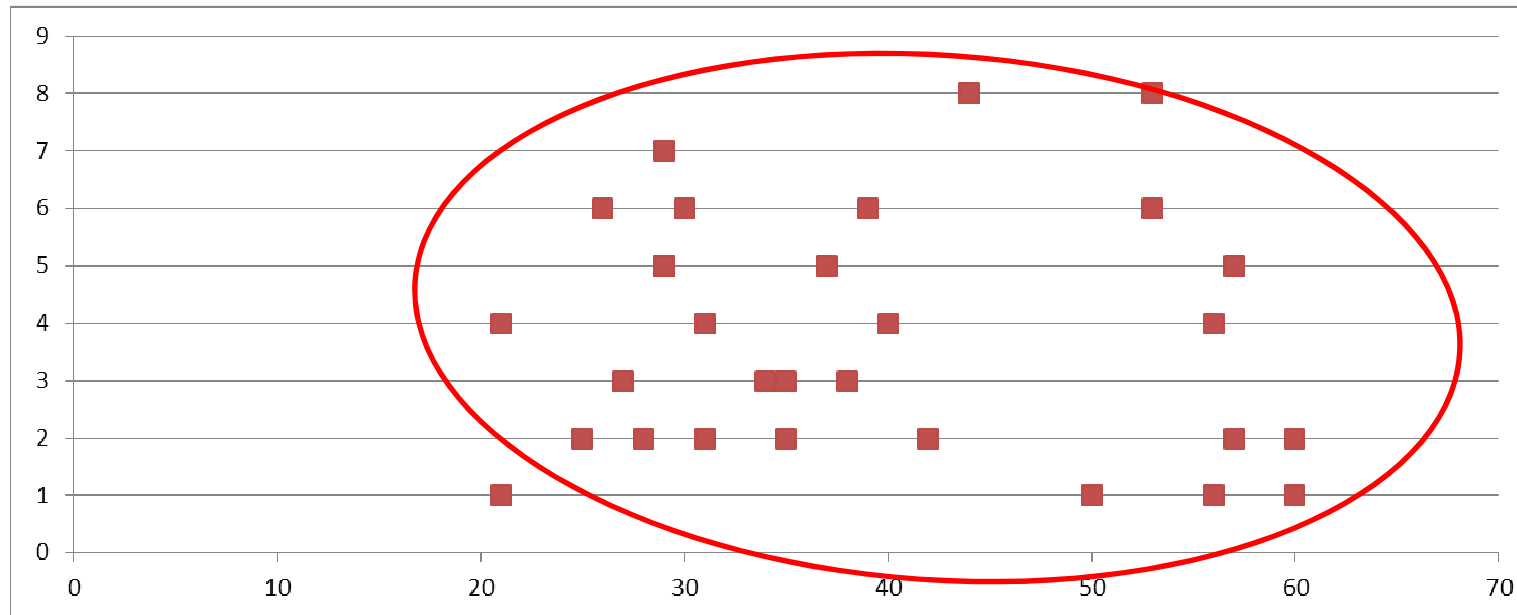
Nuova cartella di lavoro

OK
Annulla
?

Esempio Covarianza negativa *v.s. Età e Ore dedicate a News in internet*

- Calcolo covarianza = $\text{COVARIANZA}(A2:A31;B2:B31)$
= -1,81 < 0

Tendenza alla linearità negativa: retta interpolante decrescente, ma probabilmente bassissima correlazione (o incorrelazione)

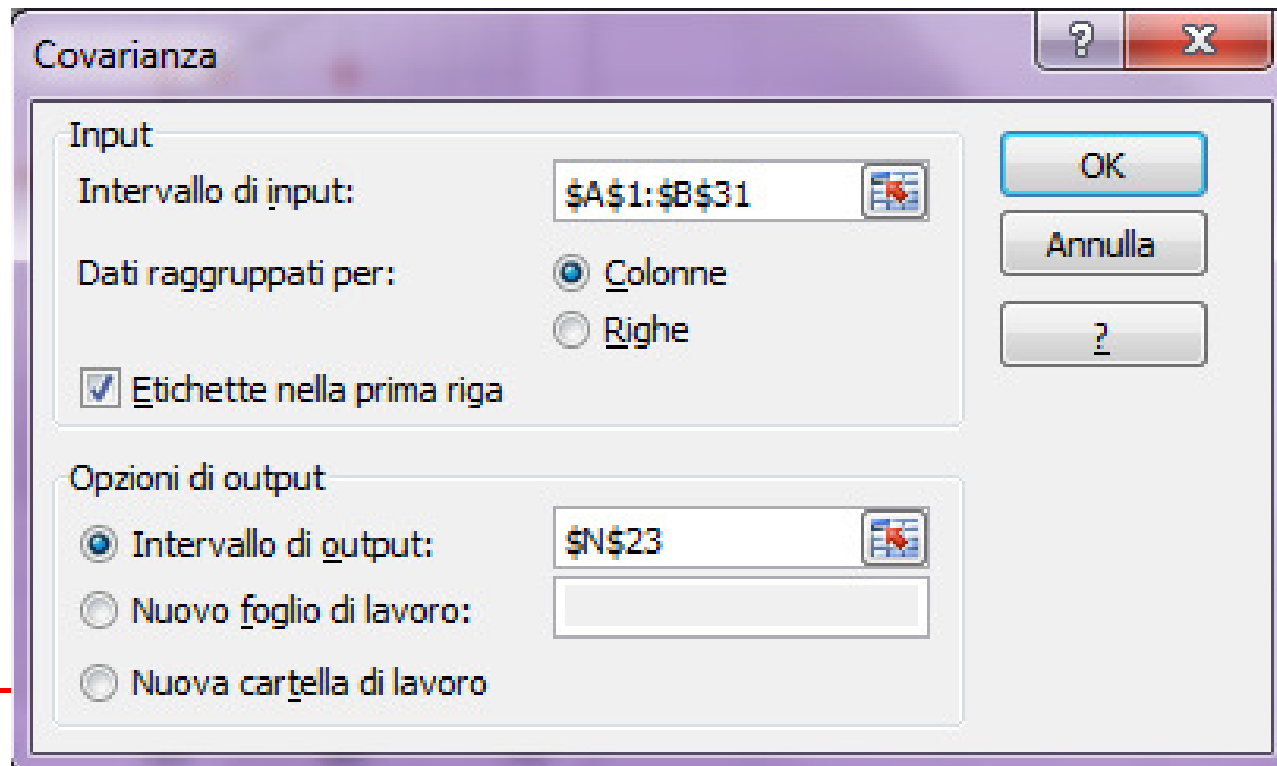


Strumenti > Analisi dati > Covarianza

| | <i>Età</i> | <i>Ore settimanali dedicate a News in Internet</i> |
|--|------------|--|
| <i>Età</i> | 150,1567 | |
| <i>Ore settimanali dedicate a News in Internet</i> | -1,81 | 4,24555556 |

Covarianza

Var.pop



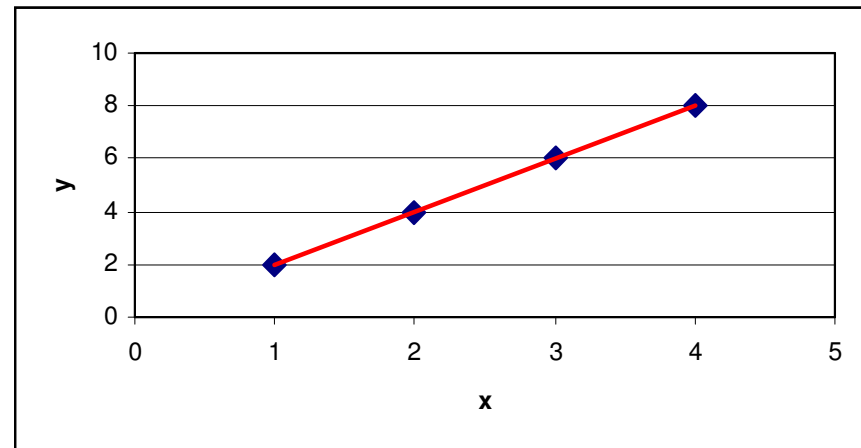
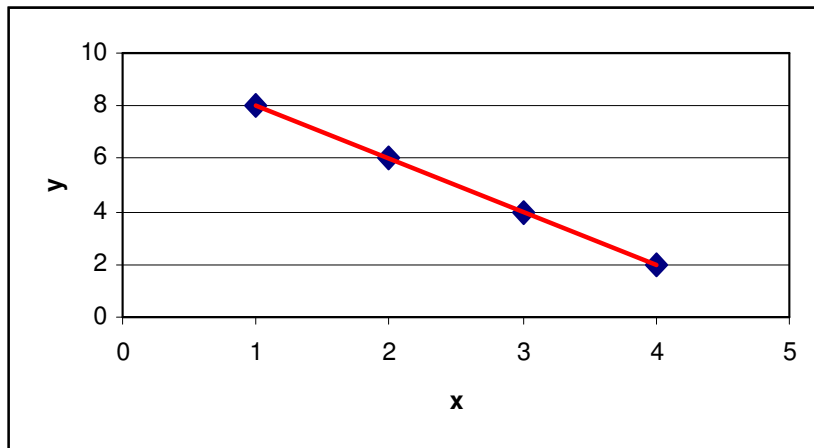
Coefficiente di correlazione lineare

- Coefficiente di correlazione lineare $\rho = \frac{Cov(X, Y)}{S_X S_Y}$

$$-1 \leq \rho \leq 1$$

Perfetta correlazione lineare
negativa

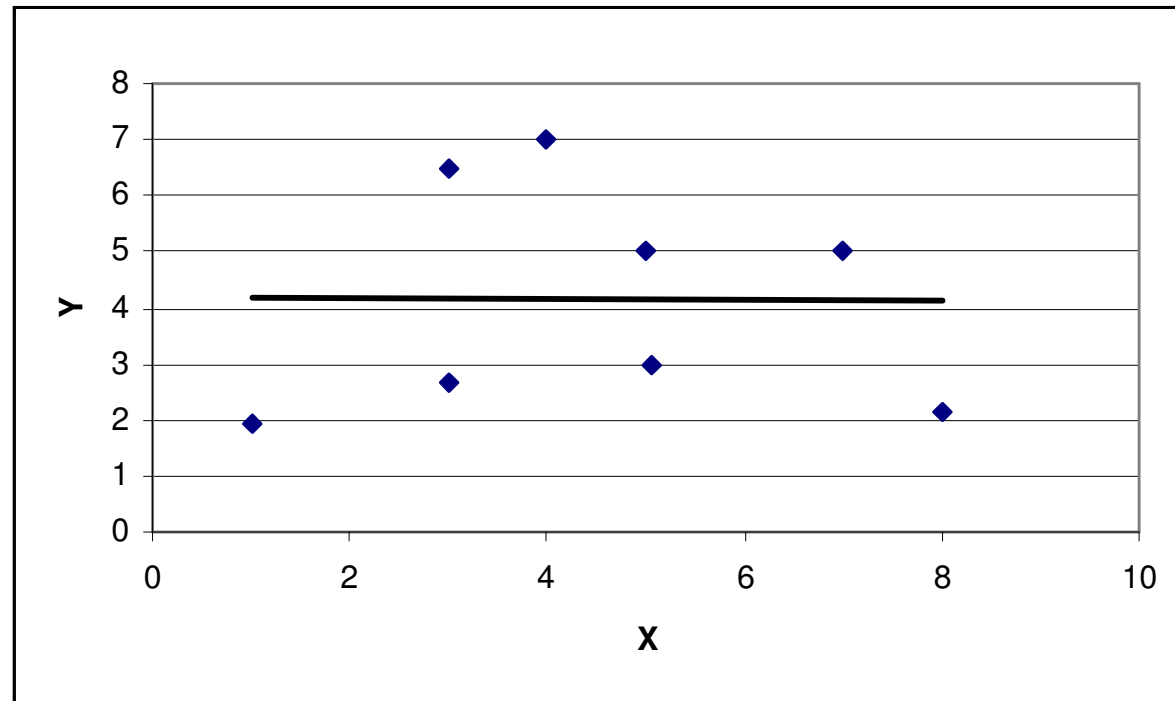
Perfetta correlazione lineare
positiva



$$\rho = 0$$



X e Y sono incorrelate



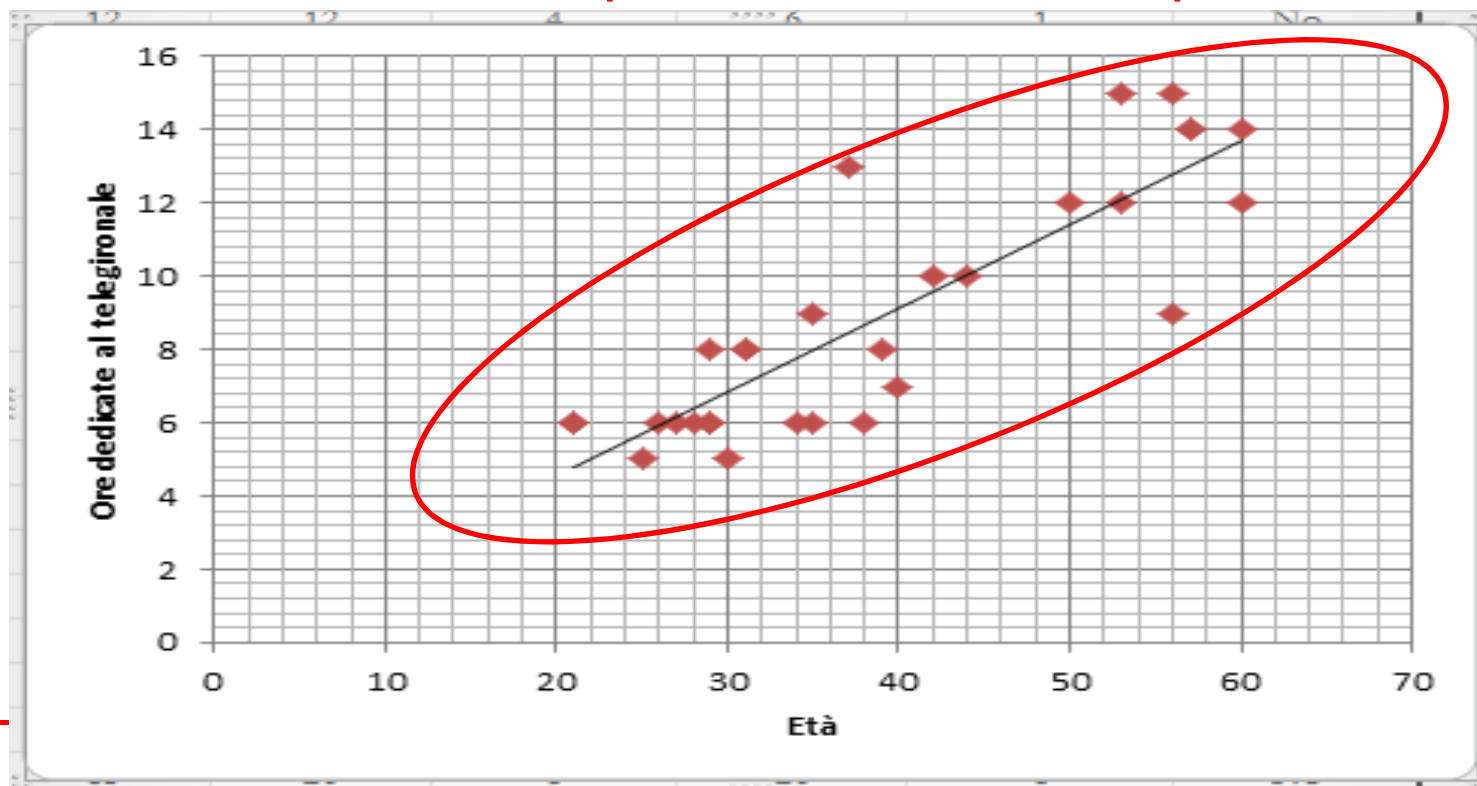
Funzioni di excel: "Correlazione()"
Strumento: Correlazione

Esempio Correlazione positiva

v.s. Età e v.s. Ore dedicate ai telegiornali

- Calcolo correlazione = $\text{CORRELAZIONE}(A2:A31;B2:B31)$
= 0,86 > 0

Tendenza alla linearità positiva: retta interpolante crescente



Scheda «dati» > Analisi dati > Correlazione

| | <i>Età</i> | <i>Ore settimanali dedicate al telegiornale</i> |
|---|------------|---|
| <i>Età</i> | 1 | |
| <i>Ore settimanali dedicate al telegiornale</i> | 0,856391 | 1 |

Coefficiente di correlazione lineare

Correlazione

Input

Intervallo di input:

Dati raggruppati per: Colonne Righe

Etichette nella prima riga

Opzioni di output

Intervallo di output:

Nuovo foglio di lavoro:

Nuova cartella di lavoro

OK

Annulla

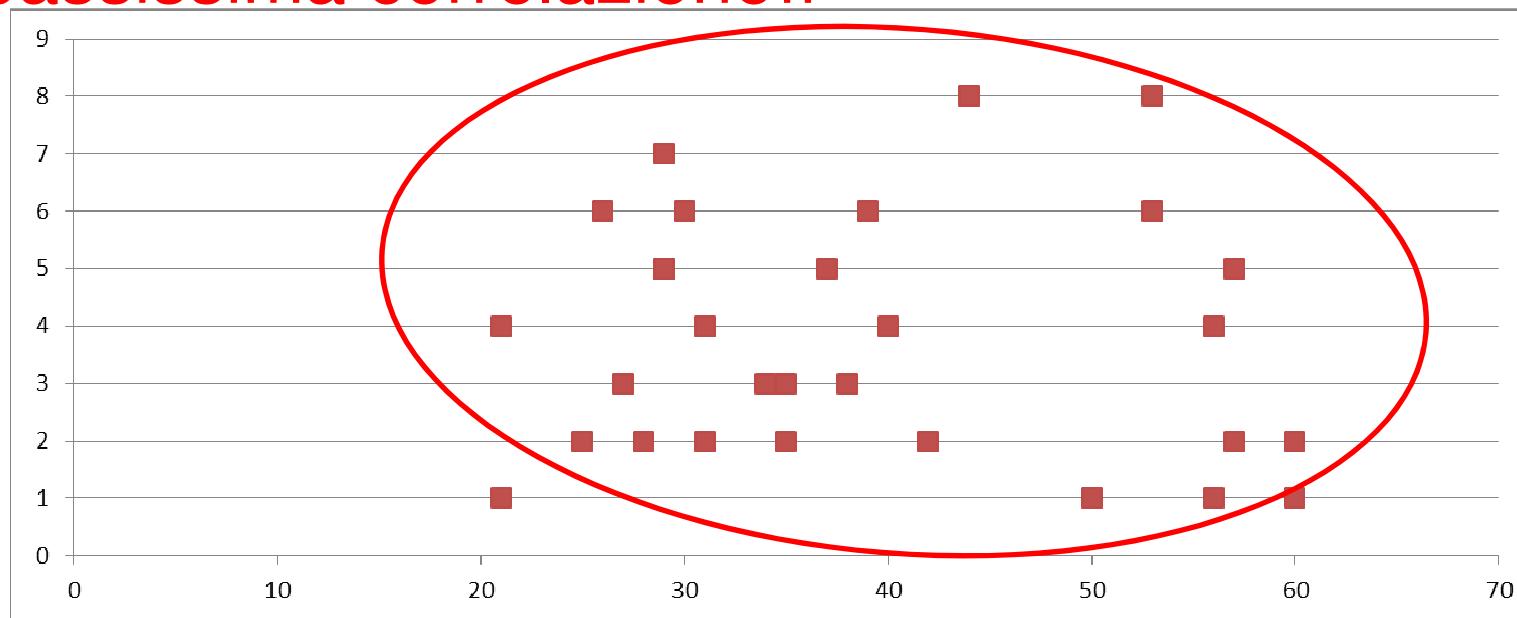
?

Esempio Correlazione negativa

v.s. Età e Ore dedicate a News in internet

- Calcolo correlazione = $\text{CORRELAZIONE}(A2:A31;B2:B31)$
= $-0,071 < 0$

Tendenza alla linearità negativa: retta interpolante decrescente
ma bassissima correlazione!!



Scheda «dati» > Analisi dati > Correlazione

| | <i>Età</i> | <i>Ore settimanali dedicate a News in Internet</i> | |
|--|-------------|--|---|
| Età | | 1 | |
| Ore settimanali dedicate a News in Interne | -0,07168677 | | 1 |

Coefficiente di correlazione lineare

Correlazione

Input

Intervallo di input:

Dati raggruppati per: Colonne Righe

Etichette nella prima riga

Opzioni di output

Intervallo di output:

Nuovo foglio di lavoro:

Nuova cartella di lavoro

OK

Annulla

?