

**Corso di Laurea in Economia Aziendale**

**Docente: Marta Nai Ruscone**

# **Statistica**

a.a. 2015/2016

1

## ***Lezione 3***

### **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*

### **3) Il modello di regressione lineare**

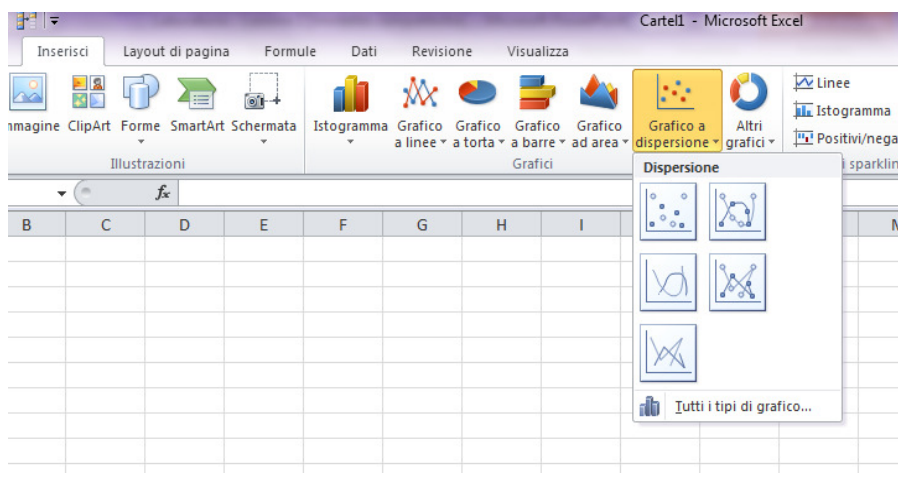
funzione di excel: "intercetta", "pendenza"  
strumento di analisi: *regressione*

## Grafico a dispersione

### Funzione:

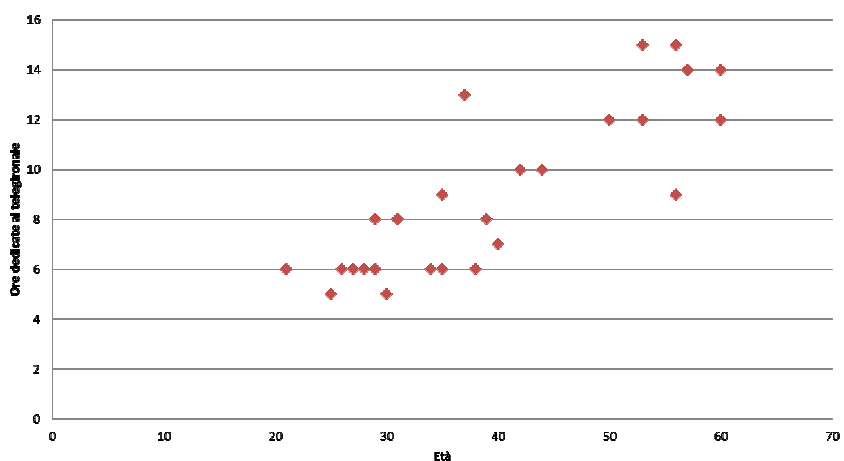
Scheda «inserisci» > grafici > grafico a dispersione

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*



## Customizzazione del grafico

1) [modalità compatibilità] - Microsoft Excel

Strumenti grafico

Progettazione Layout Formato

Stili grafici

	G	H	I	J	K	L	M	N	O
6		7	2	13	4	Si			
12		12	4	6	1	No			

Ore dedicate al telegrafone

5

## Correlazione lineare fra X e Y

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

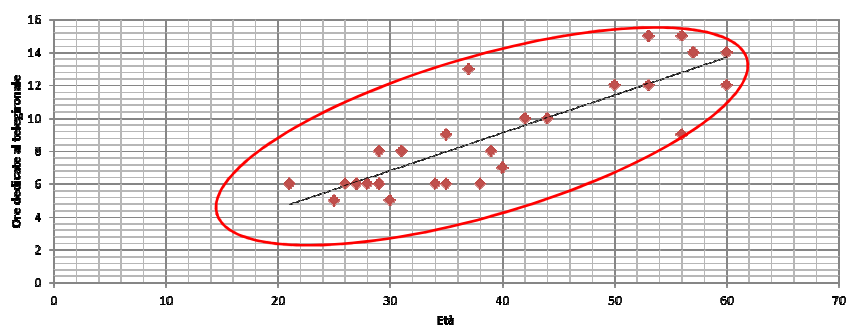
- $Cov(X, Y) > 0$   $\longrightarrow$  **Tendenza alla linearità positiva**
- $Cov(X, Y) < 0$   $\longrightarrow$  **Tendenza alla linearità negativa**
- $Cov(X, Y) = 0$   $\longrightarrow$  **Incorrelazione**

Funzione di excel: "Covarianza()"  
Strumento: Covarianza

## **Esempio: Covarianza positiva** v.s. Età e v.s. Ore dedicate ai telegiornali

- Calcolo covarianza =  $\text{COVARIANZA}(A2:A31;B2:B31)$   
= 34,37 > 0

Tendenza alla linearità positiva: retta interpolante crescente



7

## ■ Scheda «dati» > Analisi dati > Covarianza

	Età	Ore settimanali dedicate al telegiornale
Età	150,1566667	
Ore settimanali dedicate al telegiornale	34,37333333	10,72888889

Covarianza

Var.pop

**Covarianza** [?] [X]

**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

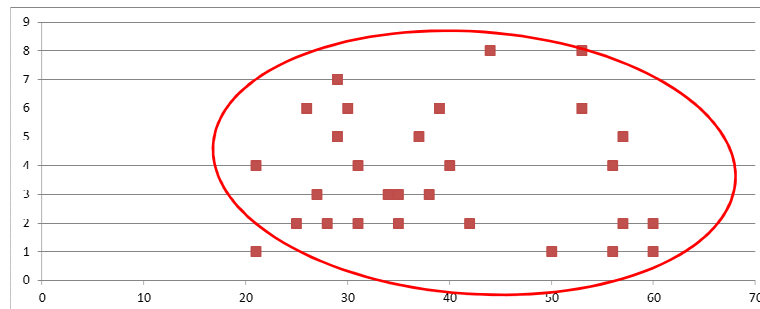
[OK] [Annulla] [?]

8

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

- Calcolo covarianza = COVARIANZA(A2:A31;B2:B31)  
= -1,81 < 0

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



9

## Strumenti > Analisi dati > Covarianza

	<i>Età</i>	<i>Ore settimanali dedicate a News in Internet</i>
Età	150,1567	
Ore settimanali dedicate a News in Internet	-1,81	4,245555556

Covarianza

Var.pop

10

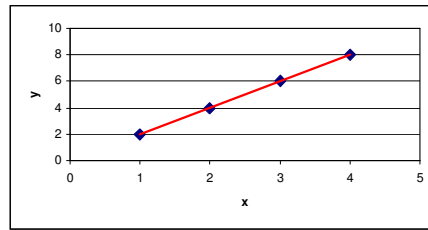
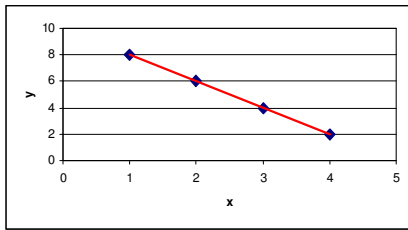
## Coefficiente di correlazione lineare

▪ Coefficiente di correlazione lineare  $\rho = \frac{Cov(X, Y)}{S_x S_y}$

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

Perfetta correlazione lineare  
negativa

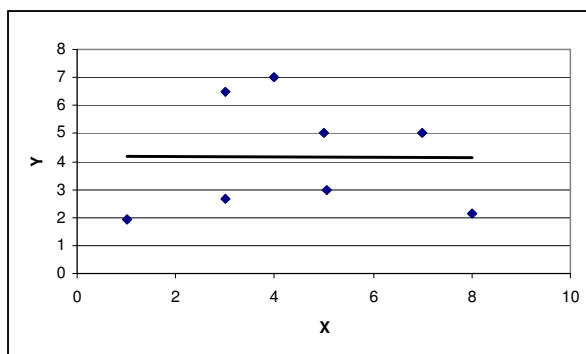
Perfetta correlazione lineare  
positiva



11

$$\rho = 0$$

X e Y sono incorrelate



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

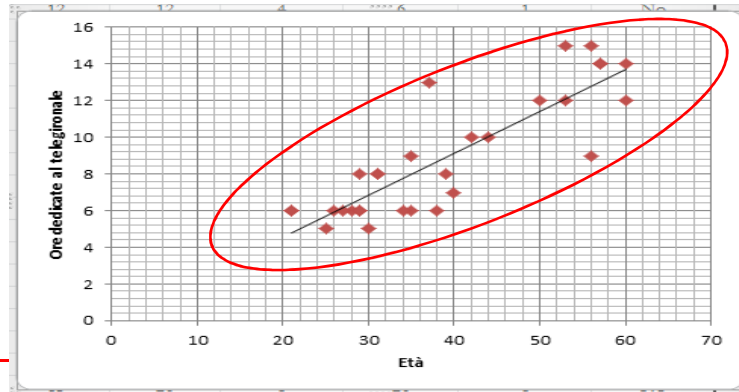
12

## 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



13

## Scheda «dati»>Analisi dati> Correlazione

Età	Ore settimanali dedicate al telegiornale
1	0,856391

Coefficiente di correlazione lineare

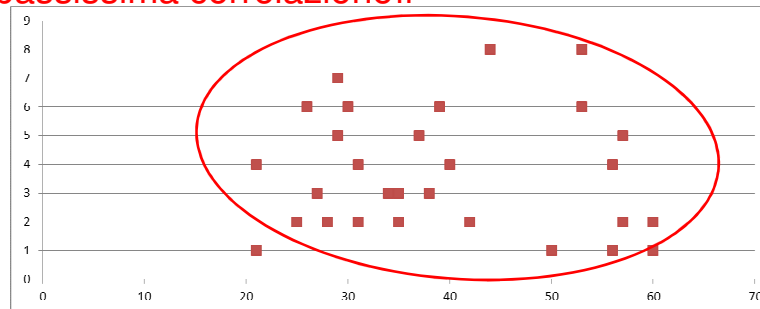
14

## Esempio Correlazione negativa

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

- Calcolo correlazione = `CORRELAZIONE(A2:A31;B2:B31)`  
= -0,071 < 0

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



15

## Scheda «dati»>Analisi dati> Correlazione

Età	Ore settimanali dedicate a News in Internet
1	1
Ore settimanali dedicate a News in Internet	-0,07168677

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 ?

16



## Il modello di regressione lineare

$$Y = a + bX + e$$

Variabile dipendente
Variabile indipendente

a= costante

b= pendenza

Criterio dei minimi quadrati ordinari

Stima dei parametri:

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

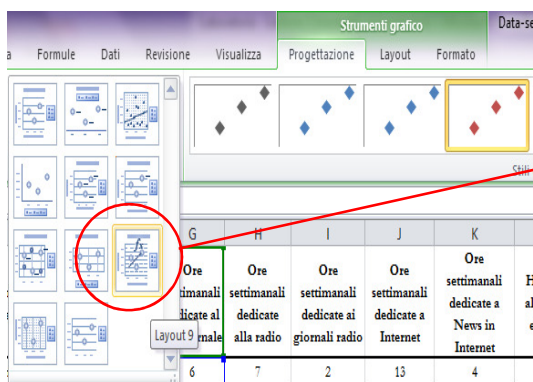
$$\hat{b} = \frac{Cov(x, y)}{s_x^2}$$

17

## La regressione lineare in excel

- **Il comando:** aggiungi linea di tendenza

Nel grafico selezionare la serie dei dati, premere il tasto destro del mouse e selezionare il comando "Aggiungi linea di tendenza"

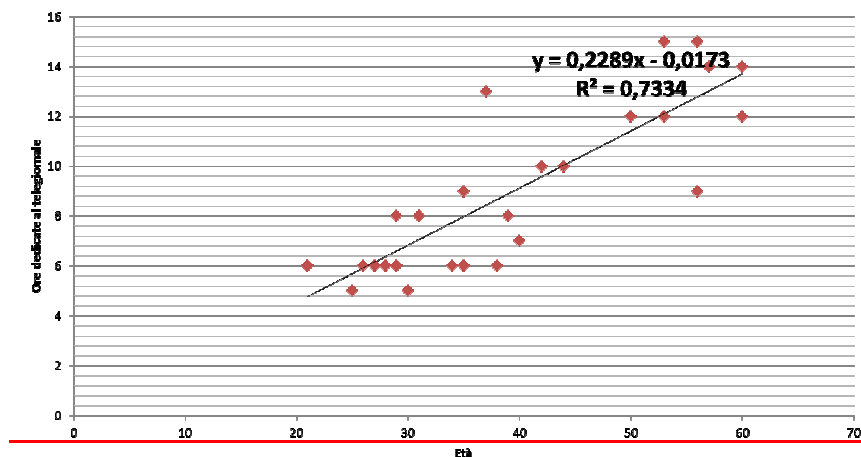


Nella scheda Progettazione – Layout Grafici si può scegliere di visualizzare sul grafico l'equazione della retta e il valore dell'indice di adattamento.

18

## Esempio "Aggiungi linea di tendenza"

*Età e Ore dedicate al telegiornale*



19

## La regressione lineare in excel

### ■ Le funzioni di excel:

a="intercetta(y\_nota;x\_nota)"

b="pendenza(y\_nota;x\_nota)"

$R^2$ ="RQ(y\_nota;x\_nota)"

per stimare i parametri della retta e calcolare l'indice di adattamento del modello ai dati.

Esempio: *Età e Ore dedicate al telegiornale*

a=INTERCETTA(B2:B31;A2:A31)=-0,0173

b=PENDENZA(B2:B31;A2:A31)=0,2289

$R^2$ =RQ(B2:B31;A2:A31)=0,7334

20

## La regressione lineare in excel

### ■ Lo strumento di analisi: Regressione

Strumenti>Analisi dati  
> regressione

Var. dipendente

Var. indipendente

Se sono selezionati i  
nomi delle variabili

Per calcolare gli errori

Per disegnare il grafico

## Output

### OUTPUT RIEPILOGO

<i>Statistica della regressione</i>	
R multiplo	0,856391
R al quadrato	0,733405
R al quadrato corre	0,723884
Errore standard	1,750592
Osservazioni	30

Indice di adattamento

<i>Coefficienti</i>	
Intercetta	-0,017300449
Età	0,228916465

Stima dei coefficienti

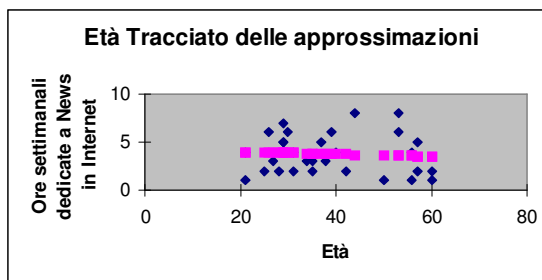
Modello stimato:  $y = -0,0173 + 0,2289x$

## Esercizi di riepilogo

- 1) Commentare l'output della regressione per valutare la relazione lineare esistente fra la variabile dipendente *Ore dedicate ai news in internet* e la variabile indipendente *Età*

### OUTPUT RIEPILOGO

<i>Statistica della regressione</i>	
R multiplo	0,071687
R al quadrato	0,005139
R al quadrato corretto	-0,030392
Errore standard	2,127307
Osservazioni	30



<i>Coefficienti</i>	
Intercetta	4,237981072
Età	-0,012054077