

# Settimana 3

G. M. Marchetti

2018

## Prima parte

Relazioni tra variabili e regressione lineare

## Una legge fisica approssimata



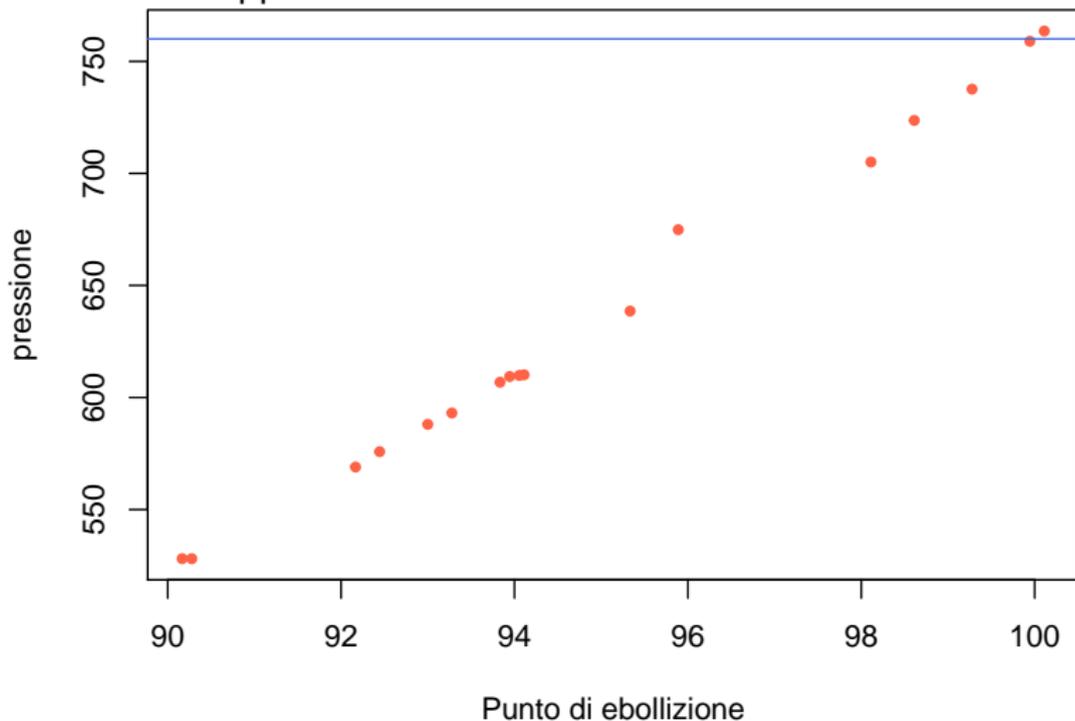
Il fisico scozzese Forbes

## L'esperimento di Forbes

- Sulla cima delle montagne l'acqua bolle a temperature più basse di 100 gradi.
- Forbes cercò di stimare l'altitudine sul livello del mare a partire dalla misura del punto di ebollizione dell'acqua.
- L'altitudine  $\rightarrow$  pressione  $\rightarrow$  temperatura di ebollizione. (Nota: il barometro era uno strumento delicato e pesante)
- In diverse zone delle Alpi misurò  $Y$ : la pressione atmosferica (con un barometro, in mm di Hg) e  $X$ : il punto di ebollizione dell'acqua (in gradi Celsius)

## Scatter del punto di ebollizione e della temperatura

- I dati sono rappresentati su uno **scatter**



## Andamento lineare

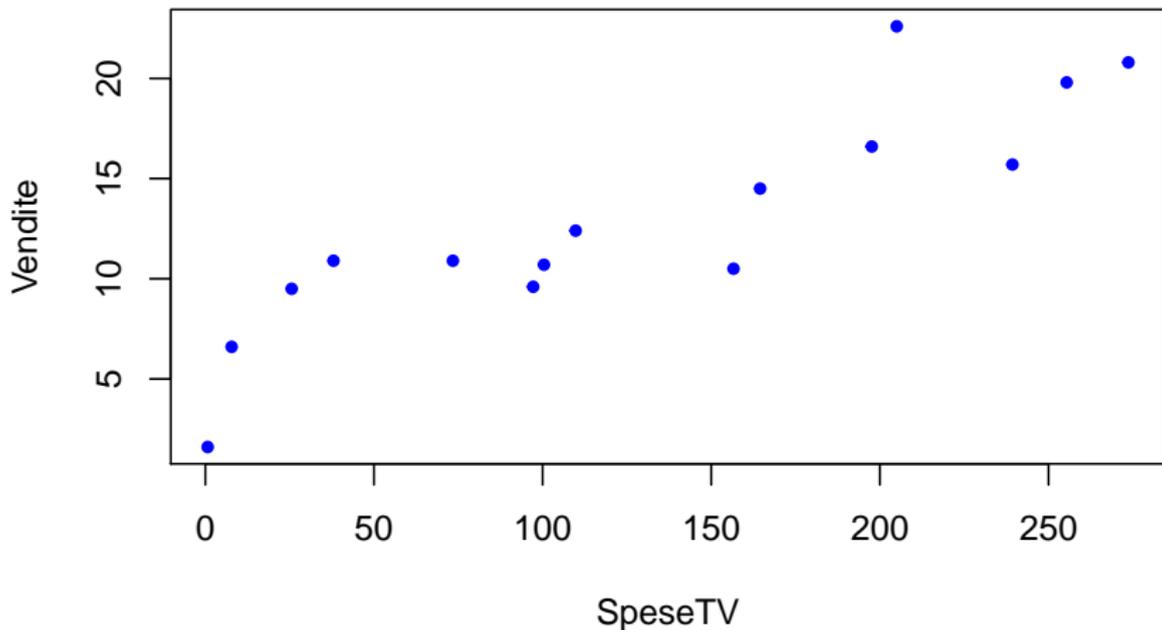
- A giudicare dallo scatter sembra i dati seguano una **legge lineare**

$$Pressione \approx \alpha + \beta \cdot (Punto\ di\ ebollizione)$$

- I punti non sono perfettamente allineati perché ci sono imprecisioni dovute ad **errori di misura**
- La pressione e la temperatura sono su due piani diversi
- La pressione è la variabile **dipendente**
- La temperatura è il **la variabile esplicativa**

## Una legge economica

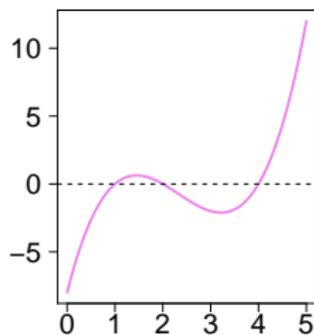
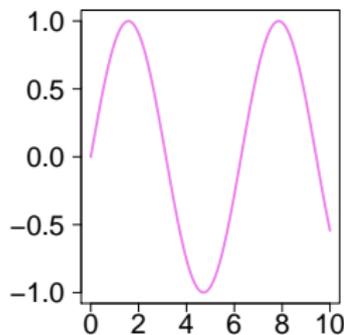
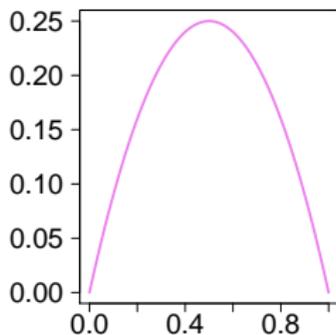
- Vendite e spese di pubblicità



$$Vendite \approx \alpha + \beta \cdot SpeseTV$$

## Relazioni matematiche

- Il concetto base è quello di **funzione**  $y = f(x)$ .
- Permettono di descrivere delle leggi semplici come una **retta**, una **parabola**, una **sinusoide**, un **polinomio**...



## La retta

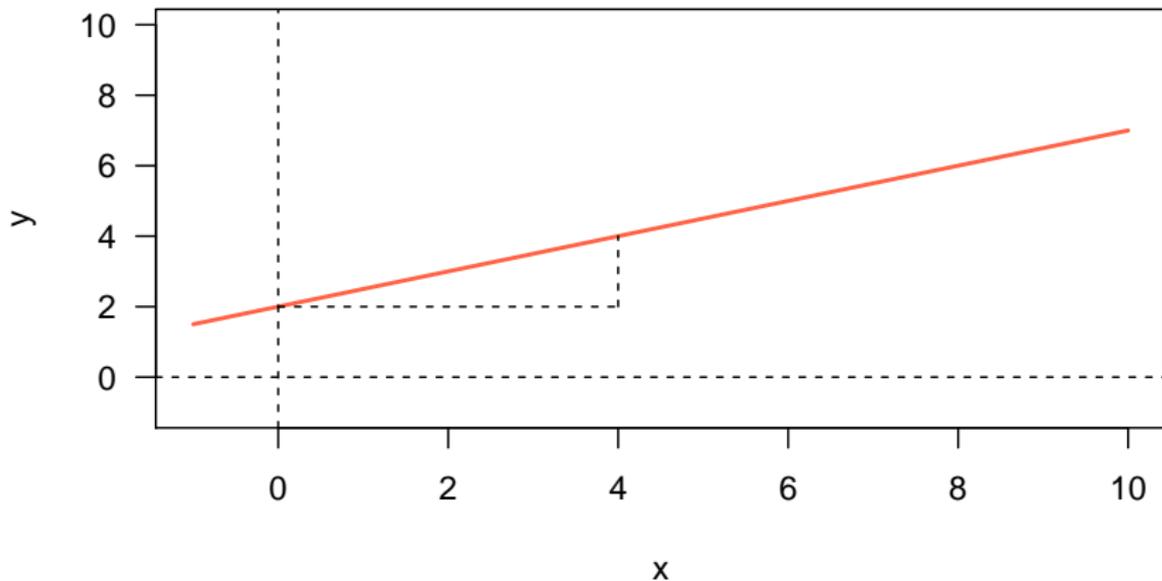
- La relazione più semplice è quella **lineare**

$$y = a + bx$$

dove  $a$  è l'**intercetta** e  $b$  è la **pendenza**.

## Leggere il grafico

- Trovare sulla figura

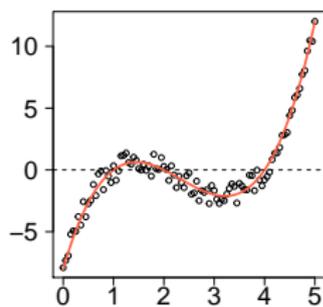
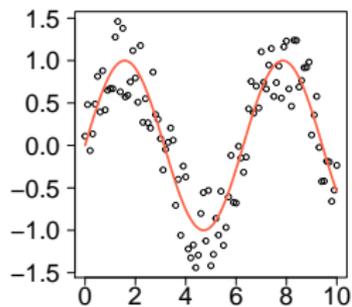
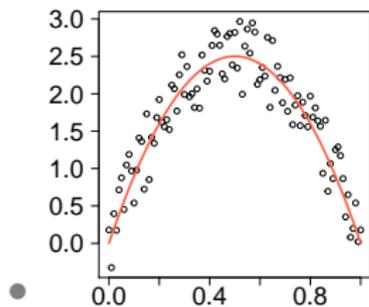


$$y = \text{intercetta} + \text{pendenza} \cdot x = 2 + \frac{1}{2}x$$

## Relazioni statistiche

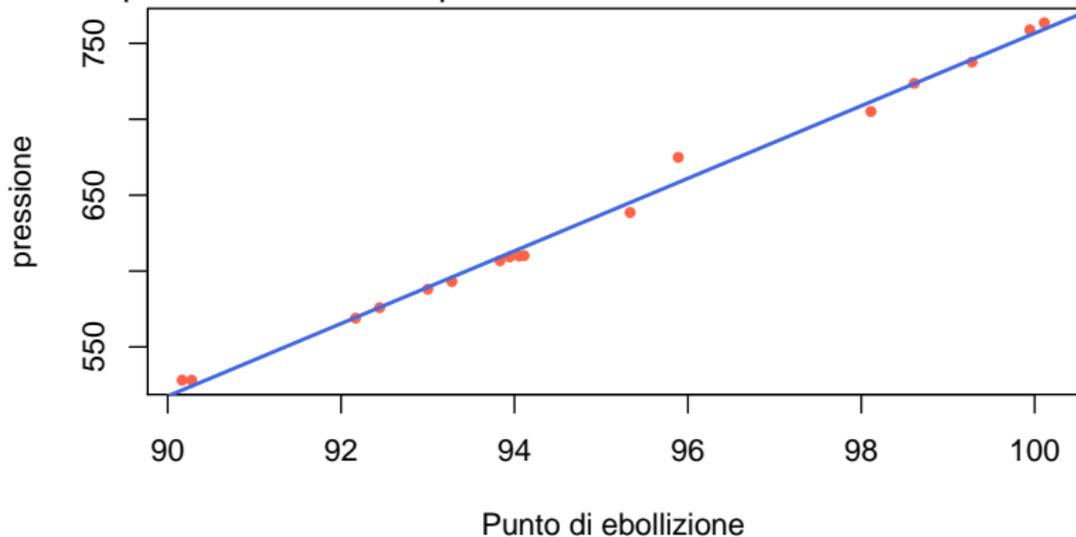
- Segnale più **rumore**

$$Y = f(x) + \text{residuo}$$



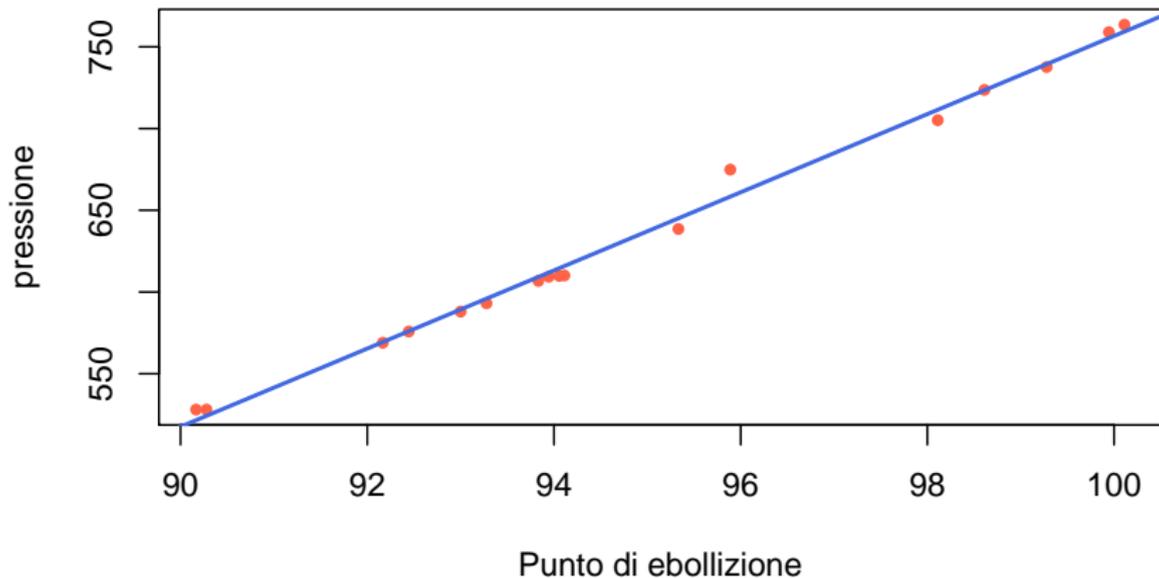
## Stima dei minimi quadrati

- Come si ottengono i coefficienti della legge lineare?
- Si adatta la **retta dei minimi quadrati**, la retta che passa FRA i punti ed è ad essi più vicina



- Si dice anche **retta di regressione** tra  $Y$  e  $X$

## Risultati e interpretazione

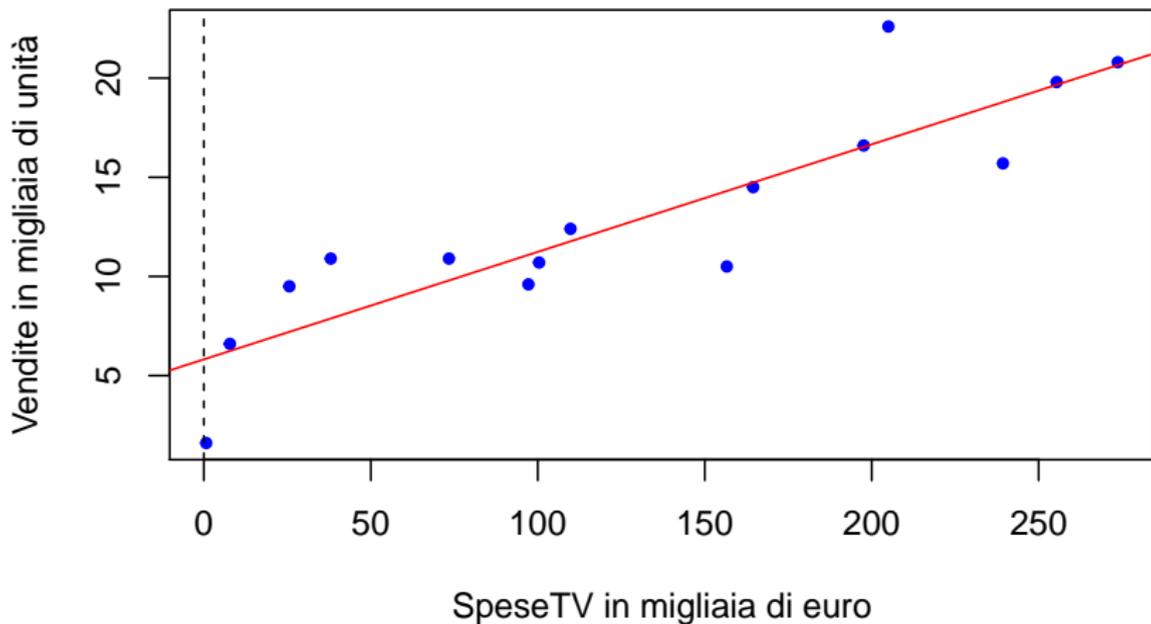


- Si stimano i due coefficienti  $a$  e  $b$  della retta dai dati

$$pressione = -1634.0 + 23.9 \cdot temperatura$$

- Per ogni grado di aumento del punto di ebollizione la pressione aumenta di 23.9 millimetri.

## Retta dei minimi quadrati tra vendite e spese



$$Vendite \approx 5.82 + 0.054 \cdot SpeseTV$$

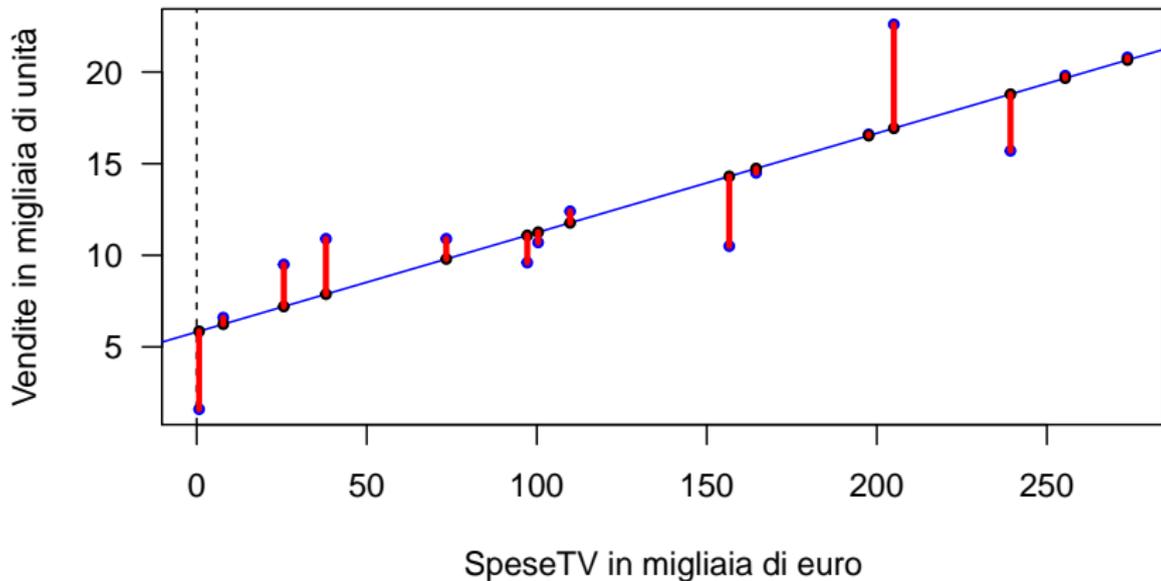
Per ogni 1000 euro in più le vendite aumentano di 54 unità

## Minimi quadrati

- Idea di Gauss



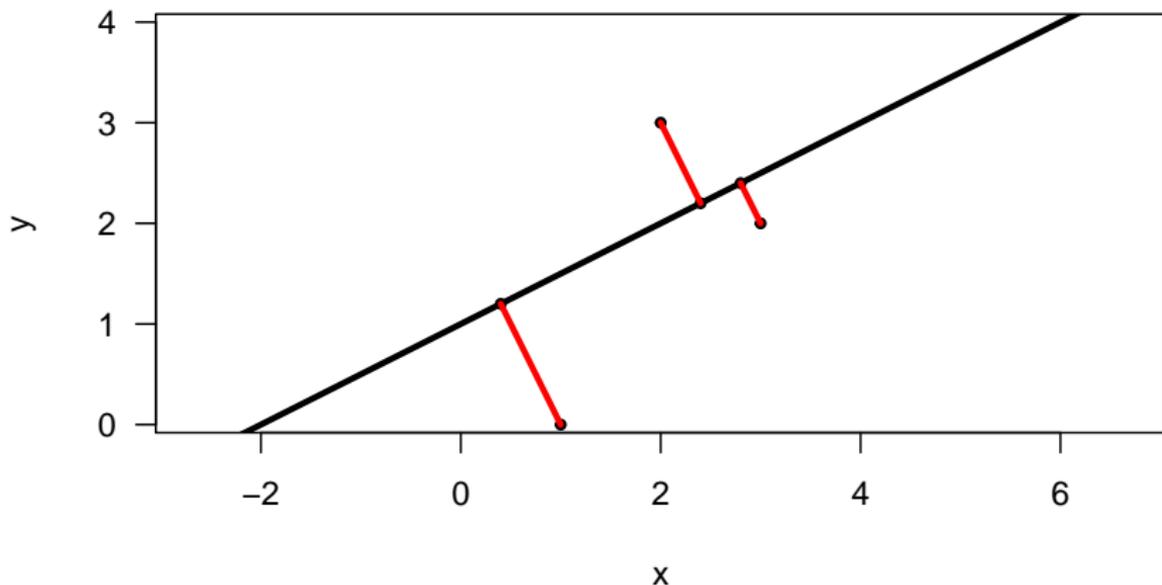
- La retta si trova minimizzando la somma dei quadrati delle distanze **verticali** tra i punti e la retta



- Vendite = variabile dipendente
- Spese TV = predittore

## Minimi quadrati

- **Non sono** le *distanze euclidee* tra i punti e la retta



- Sono scarti per la **variabile dipendente** tra osservato e previsto

## Minimi quadrati

La retta dei minimi quadrati ha **due proprietà**

- 1 Passa sicuramente per il punto  $(\bar{x}, \bar{y})$  (media di  $X$ , media di  $Y$ ).
- 2 ha pendenza

$$b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

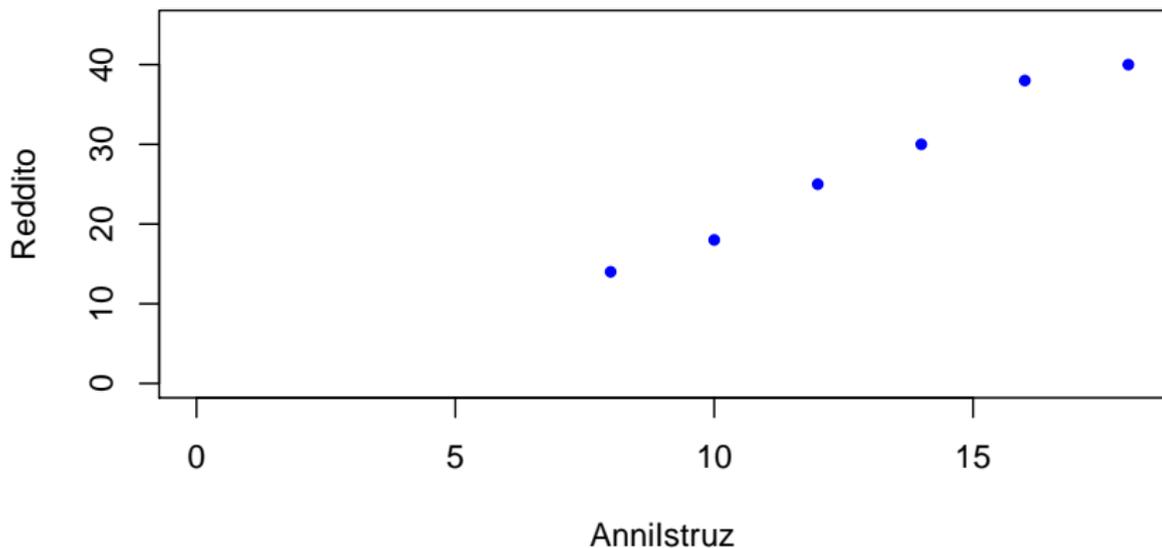
- Equazione

$$y \approx \bar{y} + b(x - \bar{x})$$

## Esercizio

- Relazione tra reddito e anni di istruzione

Unità	1	2	3	4	5	6
Anni	8	10	12	14	16	18
Reddito	14	18	25	30	38	40



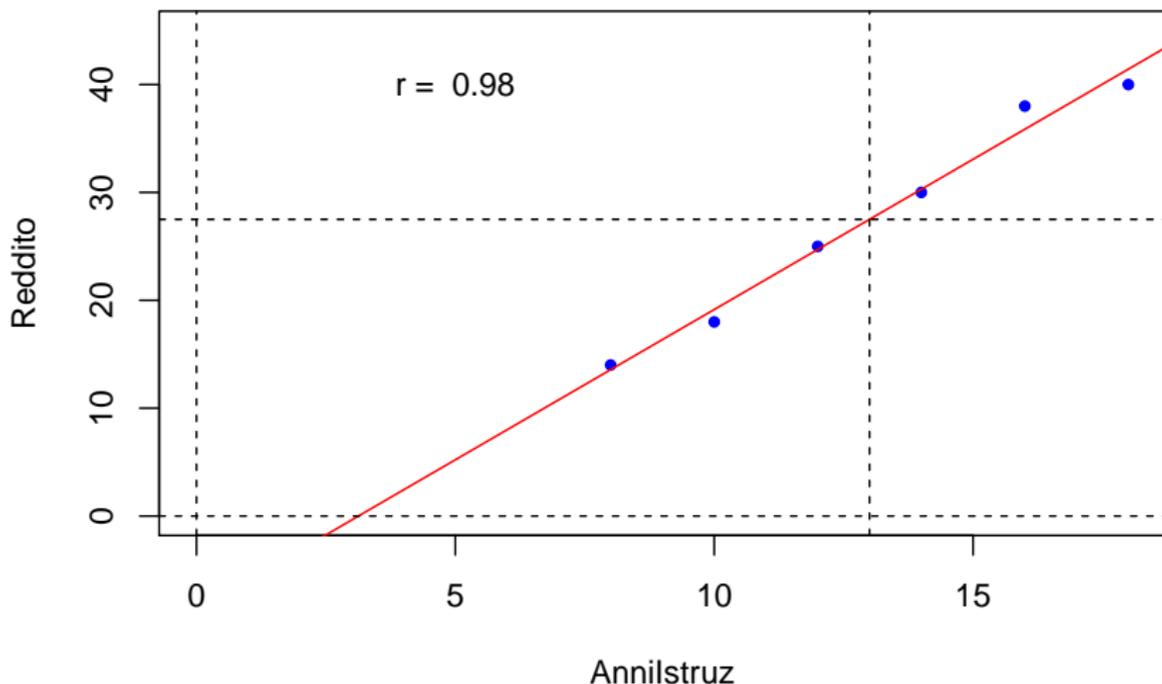
## Esercizio

- Relazione tra reddito e anni di istruzione

Unità	1	2	3	4	5	6	Media
Anni	8	10	12	14	16	18	13.0
Reddito	14	18	25	30	38	40	27.5
Scarti x	-5	-3	-1	1	3	5	0
Scarti y	-13.5	-9.5	-2.5	2.5	10.5	12.5	0
Prodotto	67.5	28.5	2.5	2.5	31.5	62.5	195/5*
Scarti x <sup>2</sup>	25	9	1	1	9	25	70/5*

- Pendenza:  $b = 195/70 = 2.79$
- Retta dei MQ:  $\hat{y} = 27.5 + 2.786(x - 13)$

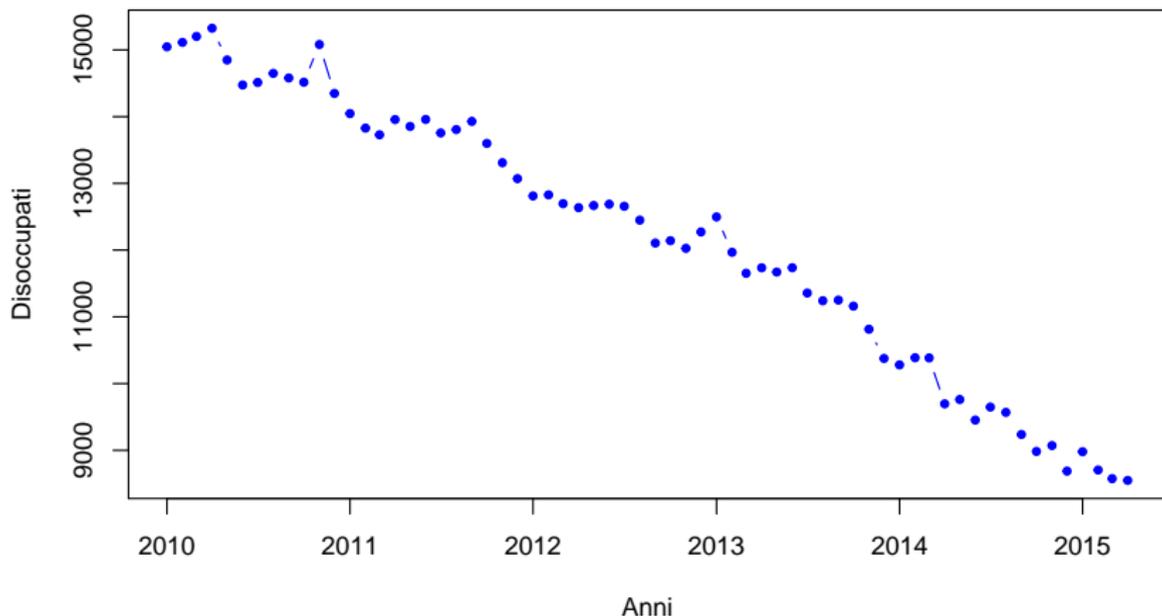
## Grafico e interpretazione



Per ogni anno di istruzione in più il reddito atteso aumenta di 2790 Euro all'anno.

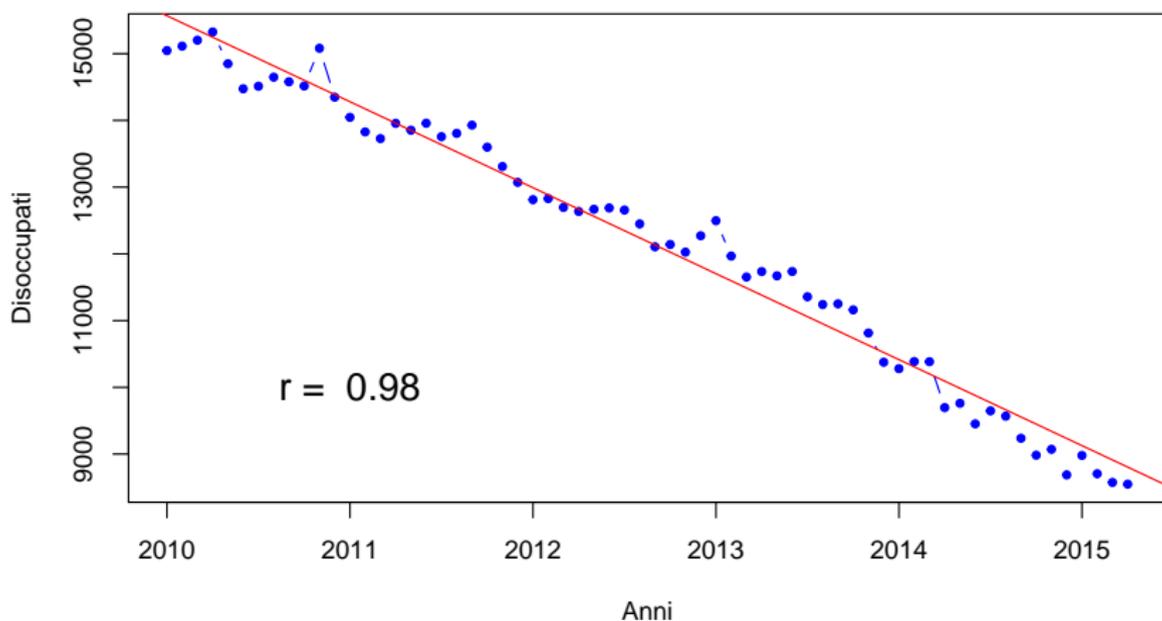
## Serie storiche

- Una serie storica (o temporale) è una serie di osservazioni su una unità statistica in tempi successivi.
- Esempio n. di disoccupati mensili (in migliaia), 2010-2015 (USA)



## Andamento tendenziale

- Usiamo la retta dei minimi quadrati



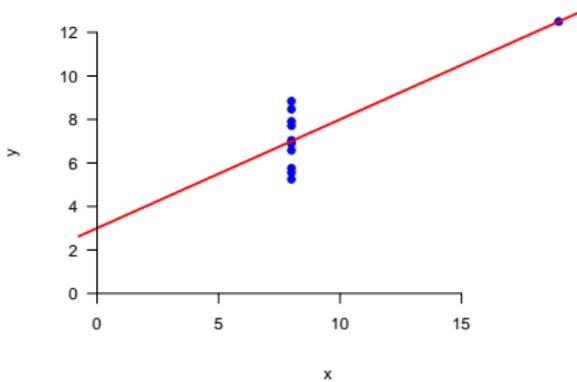
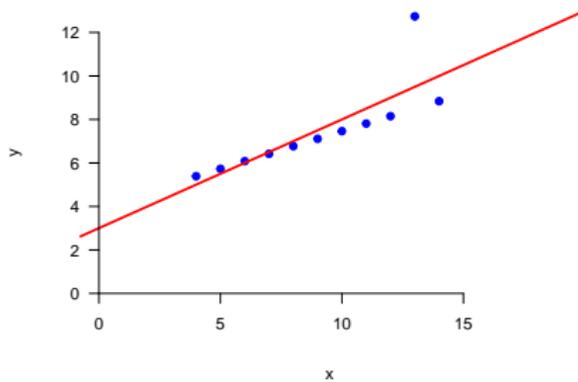
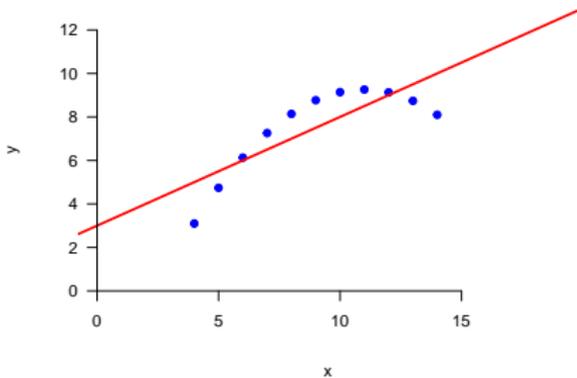
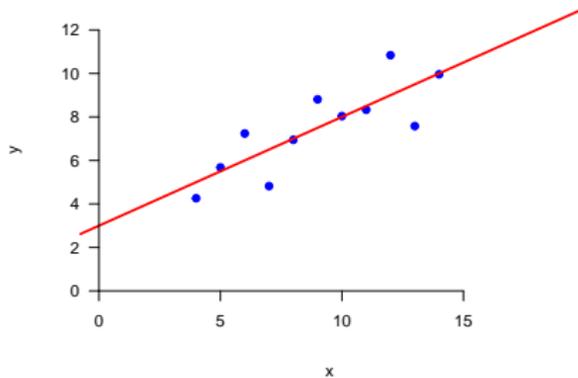
- Retta MQ:  $Disoccupazione \approx 2605180 - 1288 \cdot Anno$

## Relazioni tra correlazione e retta dei MQ

Disegnare i punti e visualizzare la retta dei minimi quadrati

- **Applet**
- $b = 0$  se e solo se  $r = 0$
- La retta è fallace se
  - ci sono outlier
  - la relazione è fortemente non lineare

## Quattro esempi da ricordare



## Dati di Anscombe

- Stessa retta di regressione  $y = 3 + 0.5x$
- Stesso coefficiente di correlazione  $r = 0.82$

	x1	y1	x2	y2	x3	y3	x4	y4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.10	4	5.39	19	12.50
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89