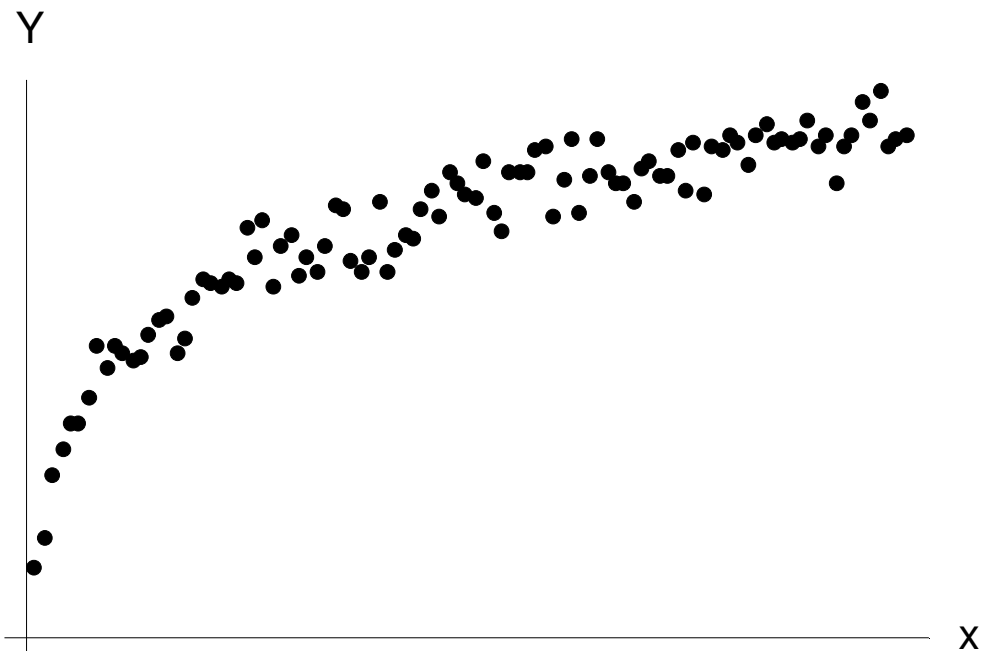


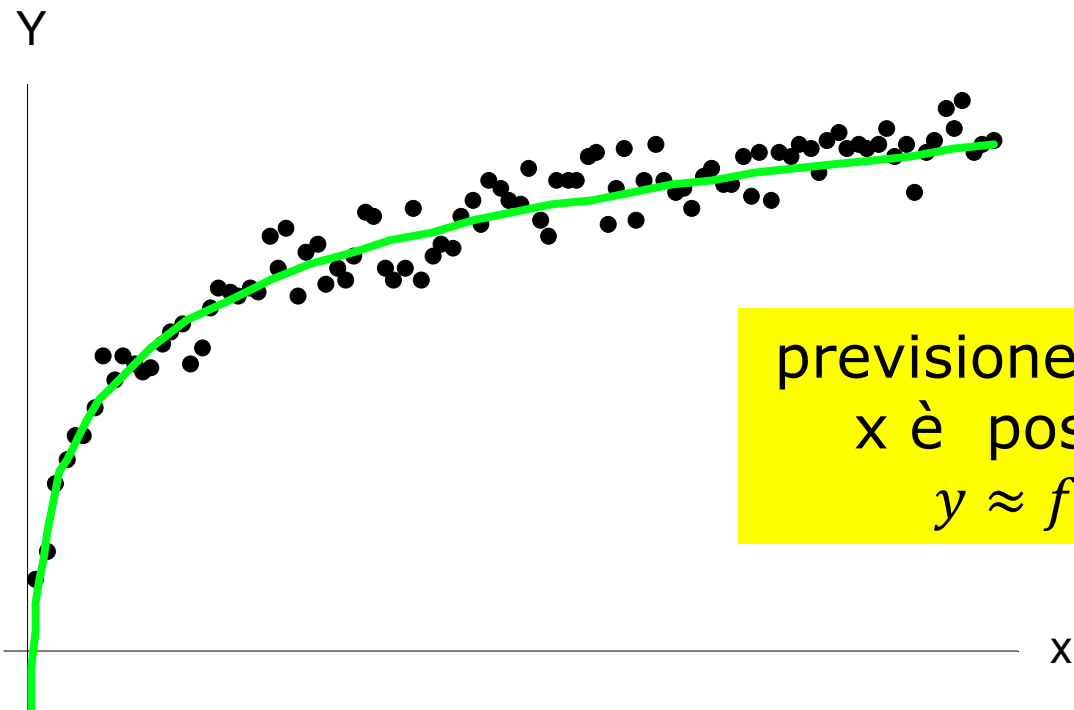
STATISTICA

Regressione-1

Obiettivo generale



Obiettivo generale

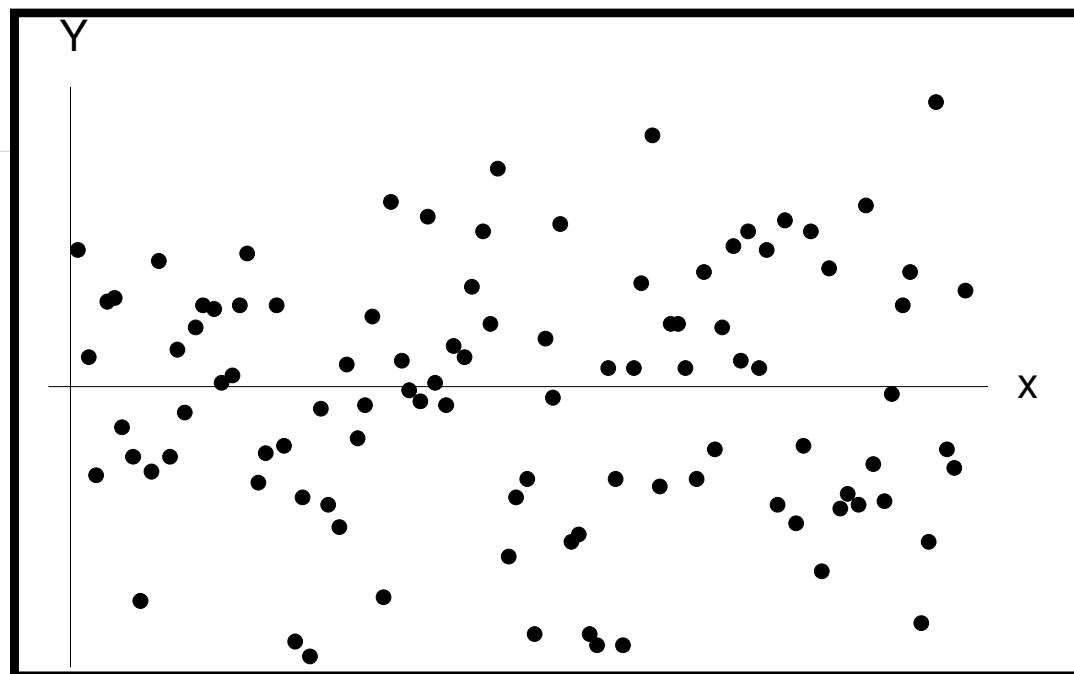


previsione di y da
x è possibile
 $y \approx f(x)$

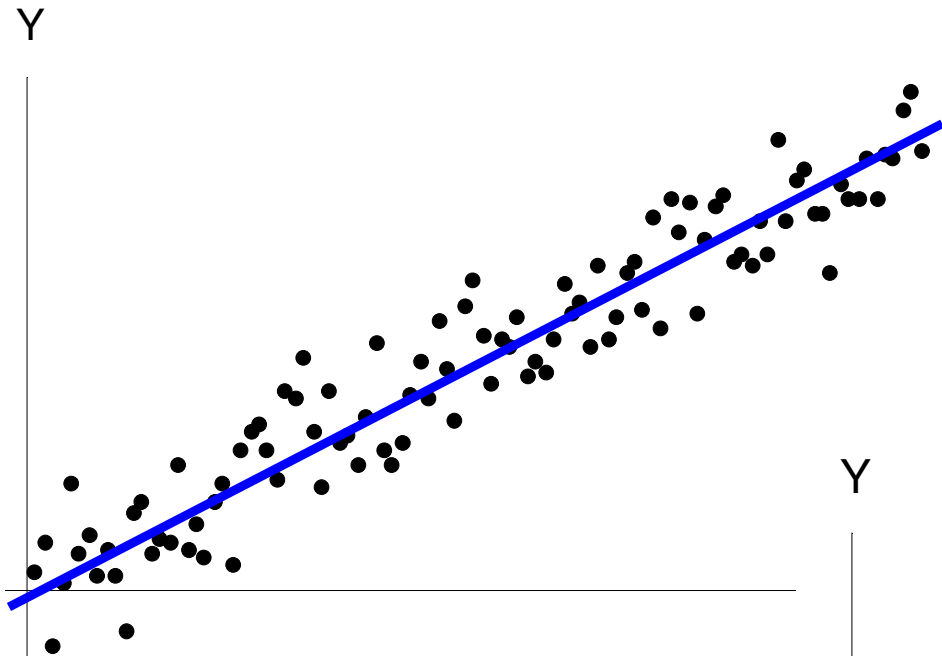
Obiettivo generale



previsione di y da x
non è possibile

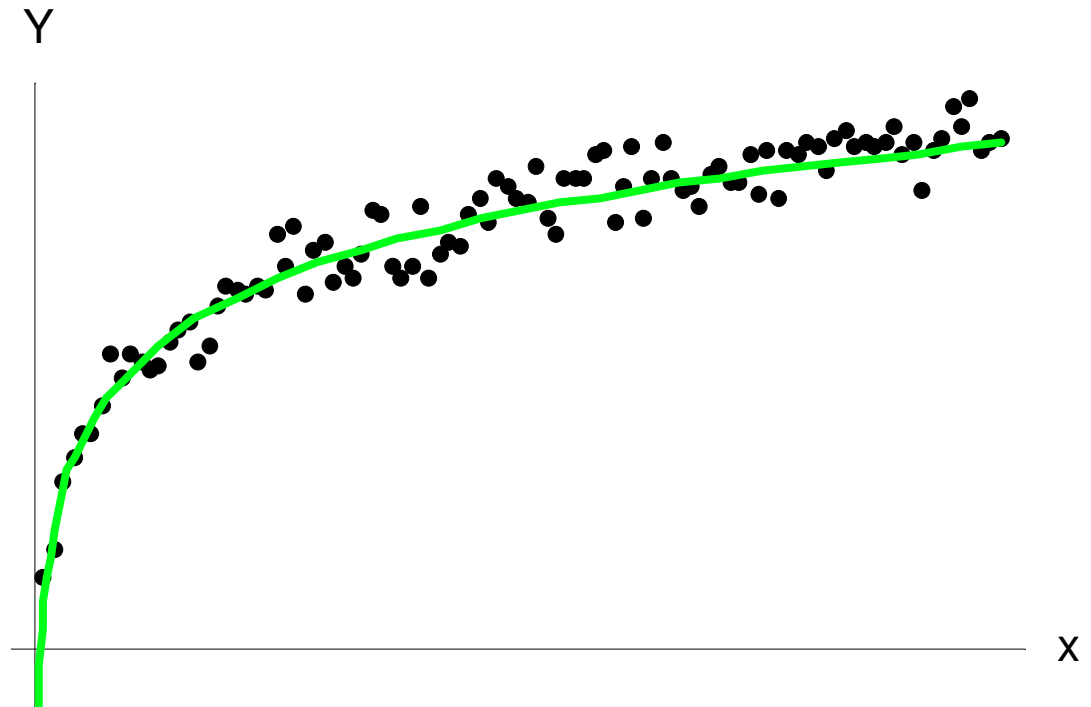


Obiettivo generale

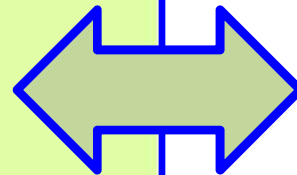
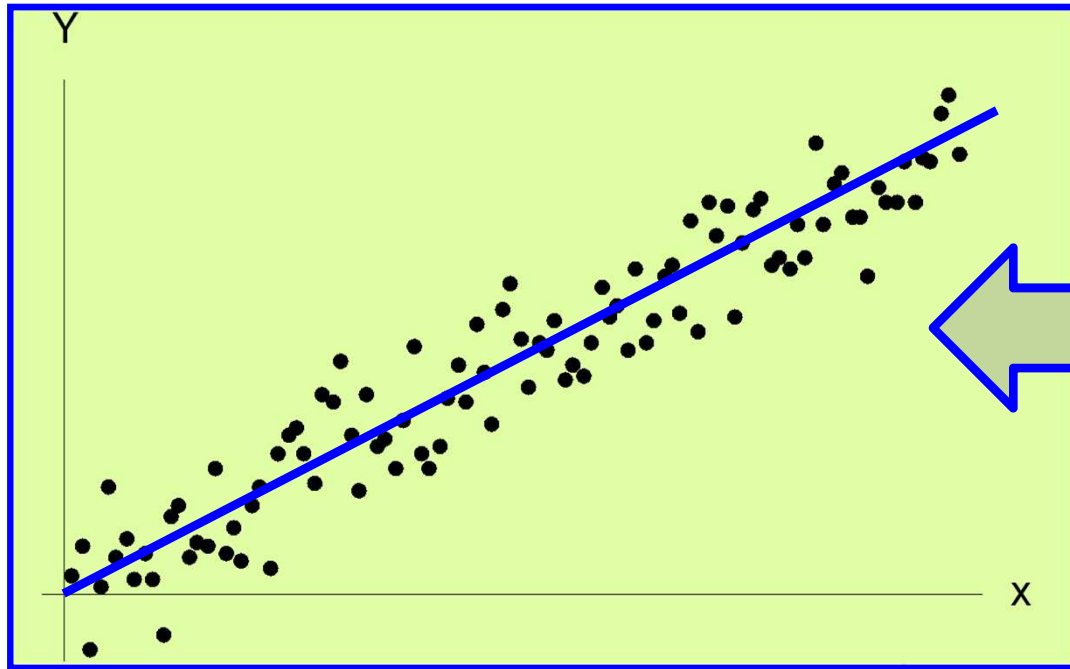


Un **modello** per la previsione di una variabile Y che dipende da un'altra variabile, X:

$$f(x) \quad \& \quad y \approx f(x)$$



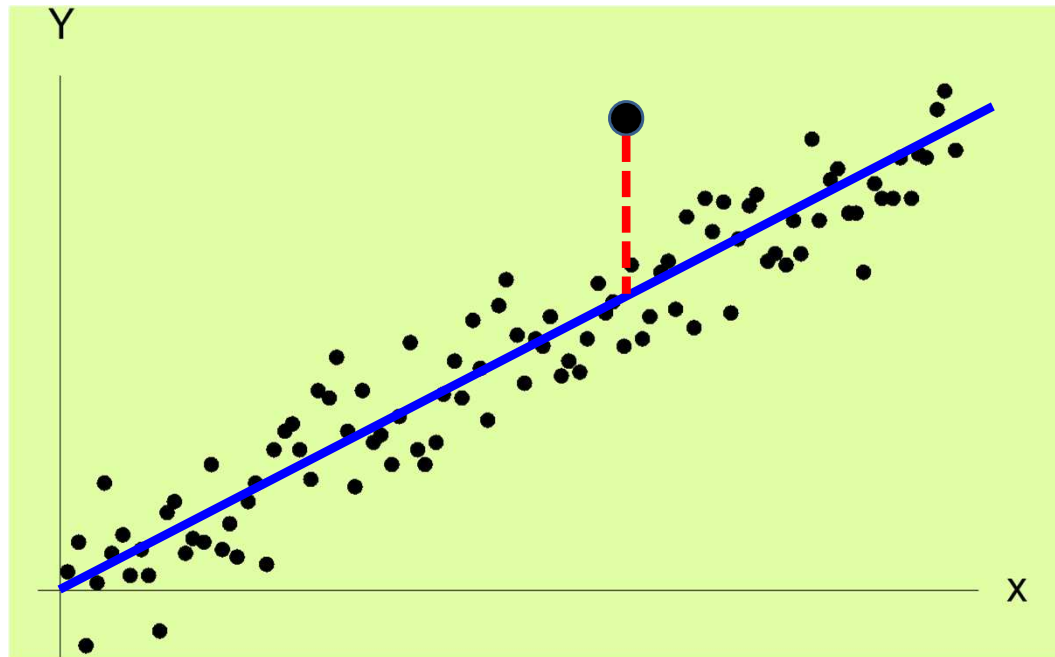
Obiettivo specifico



Il modello della
**regressione lineare
semplice:**



Obiettivo specifico



Il modello della
**regressione lineare
semplice:**

$$f(x) = a + bx$$
$$\approx \Leftrightarrow \varepsilon_i \sim N(0, \sigma^2)$$

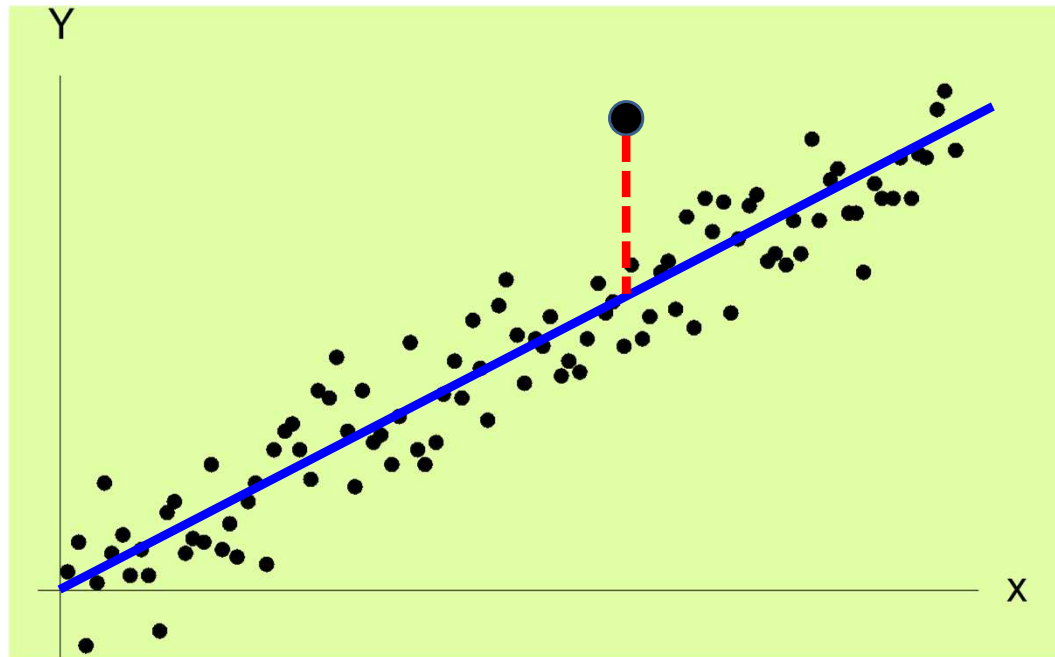
ε_i indipendenti

$$Y_i = a + bx_i + \varepsilon_i$$

x_i punti del disegno: *come se avessimo scelto di misurare Y in quei valori di X*

Y_i variabile casuale

Obiettivo specifico



Il modello della
**regressione lineare
semplice:**

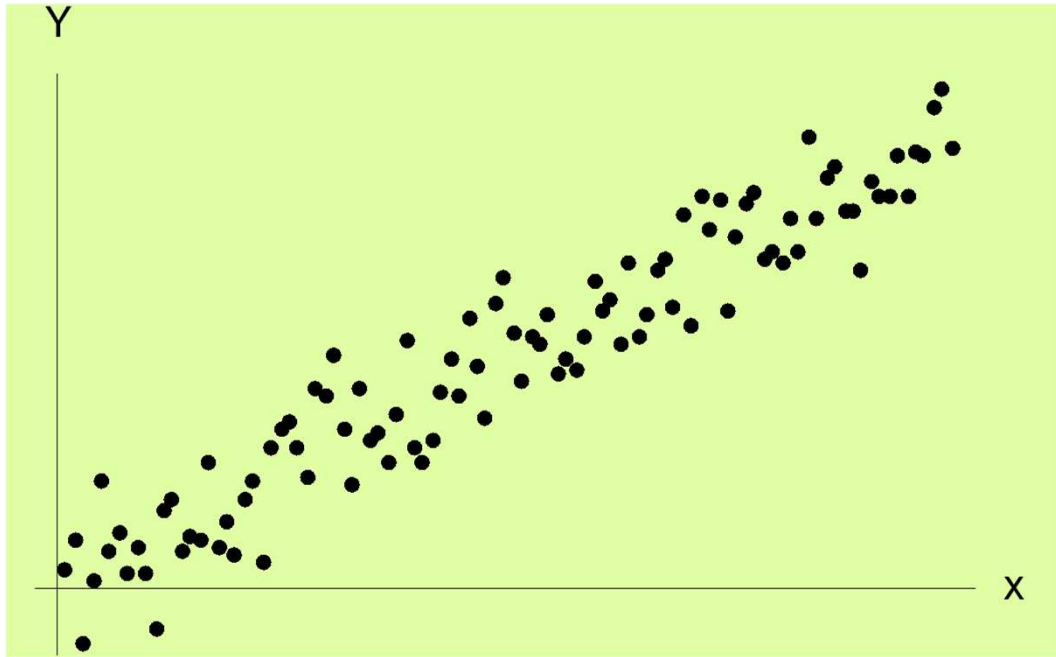
$$f(x) = a + bx$$
$$\approx \Leftrightarrow \varepsilon_i \sim N(0, \sigma^2)$$

ε_i indipendenti

$$Y_i = a + bx_i + \varepsilon_i$$

Supponiamo l'esistenza di una dipendenza lineare, **nella popolazione** da cui è estratto il campione, della variabile Y dalla variabile X , a meno di un errore casuale. Dal campione vogliamo stimare l'equazione della retta e la varianza dell'errore.

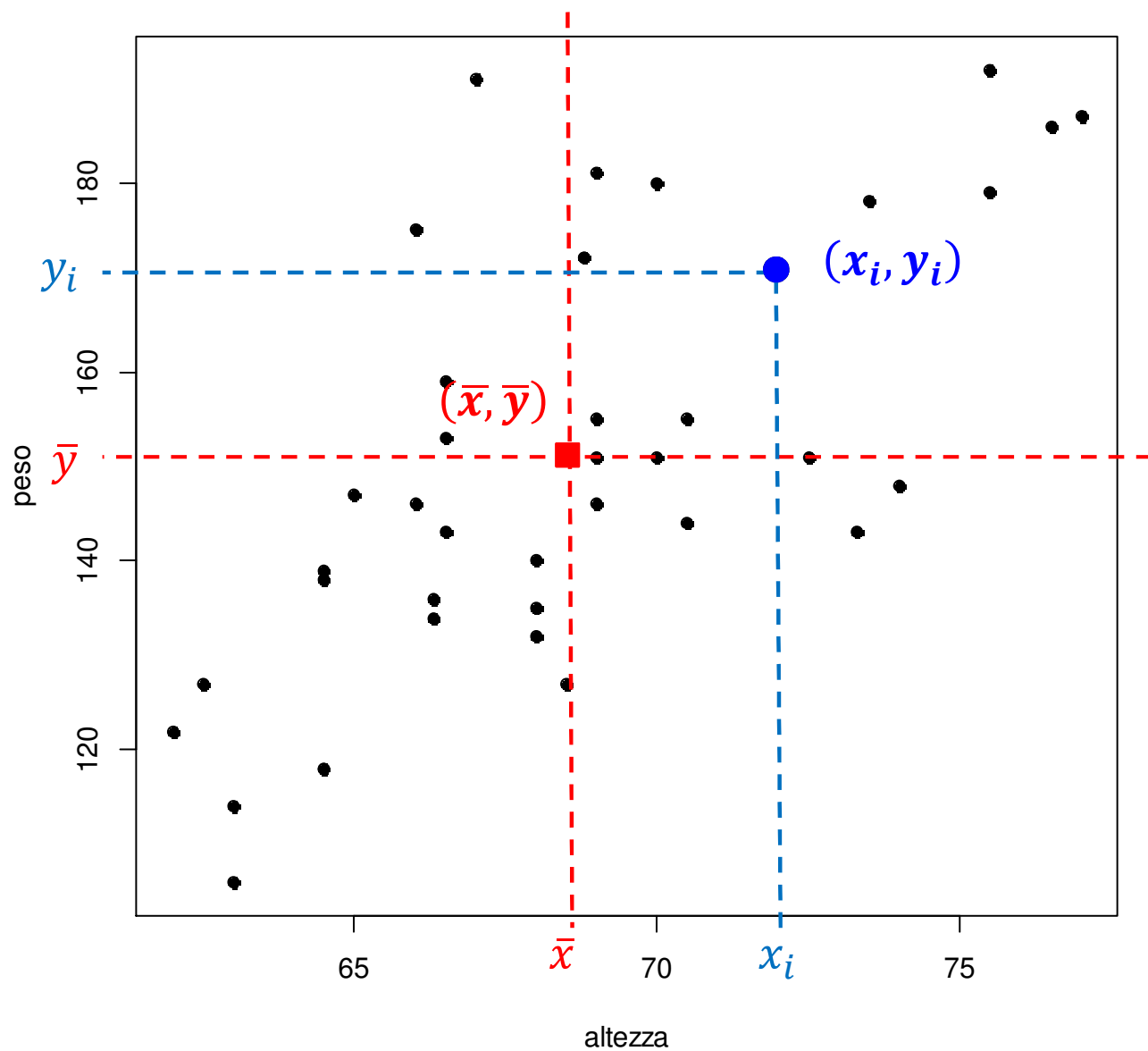
Regressione lineare



1. GRAFICO di dispersione

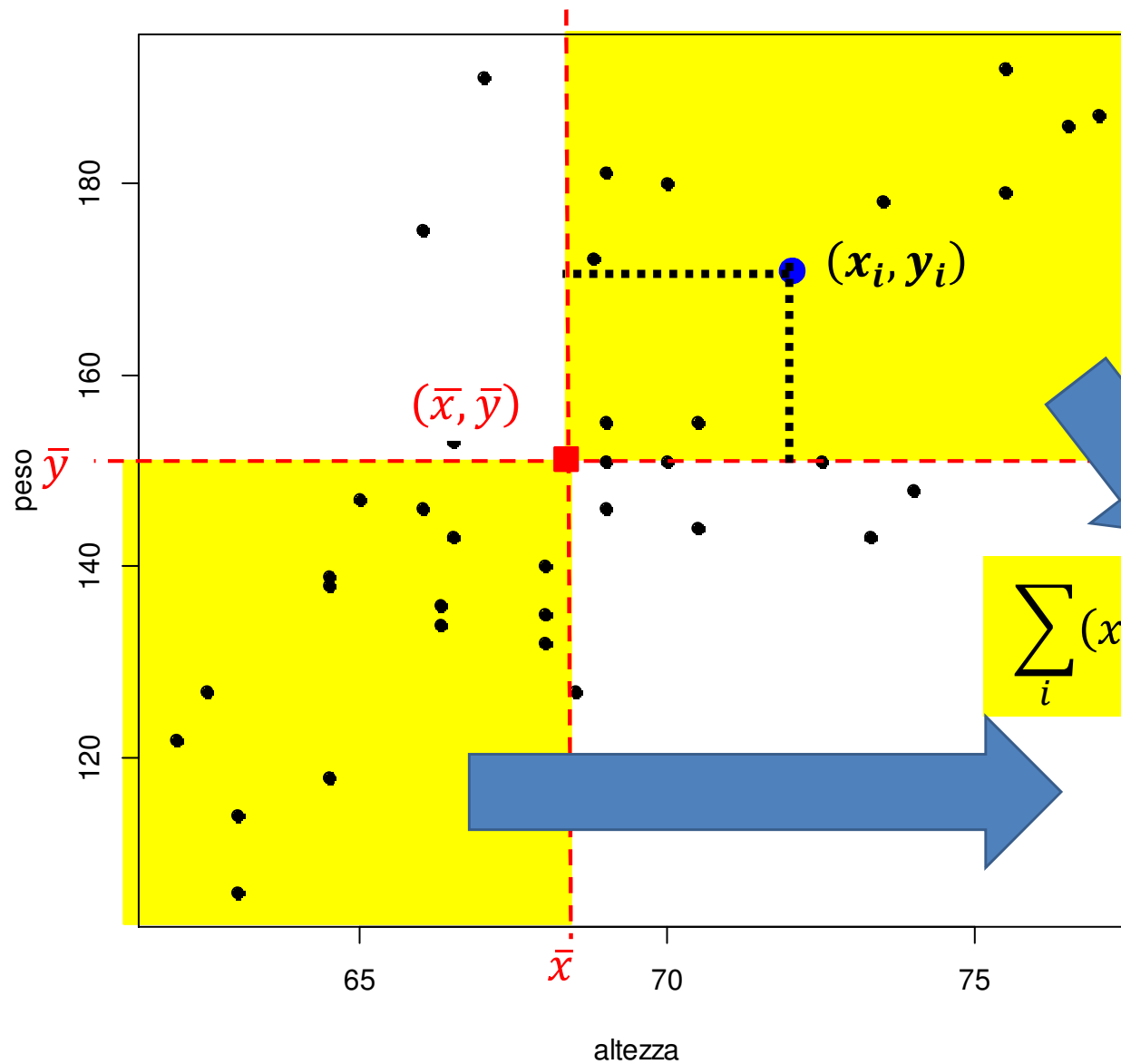
2. OPPORTUNI INDICI STATISTICI

La covarianza



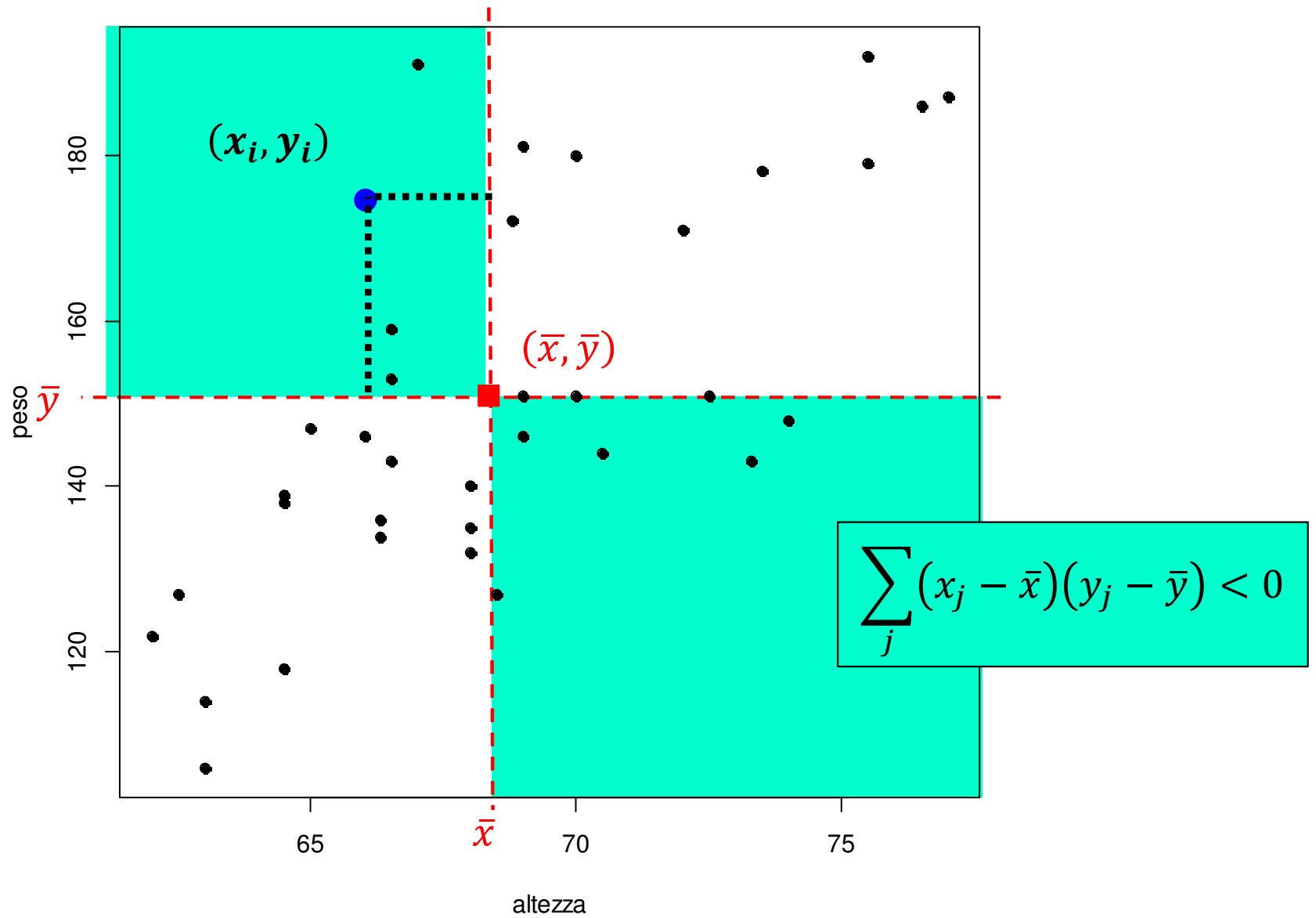
**grafico
di
dispersione
attorno
al
baricentro
(centroide)**

La covarianza

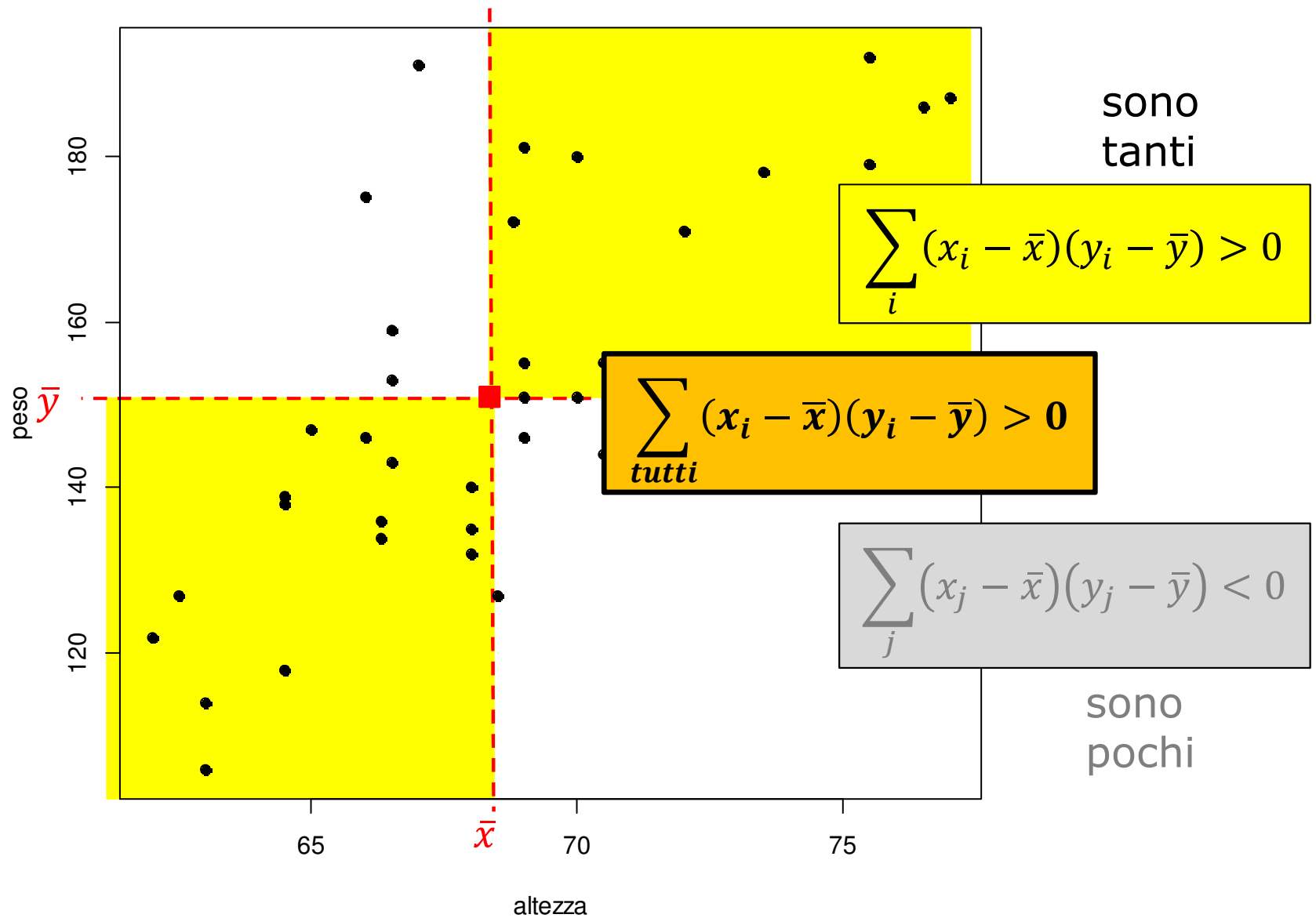


$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) > 0$$

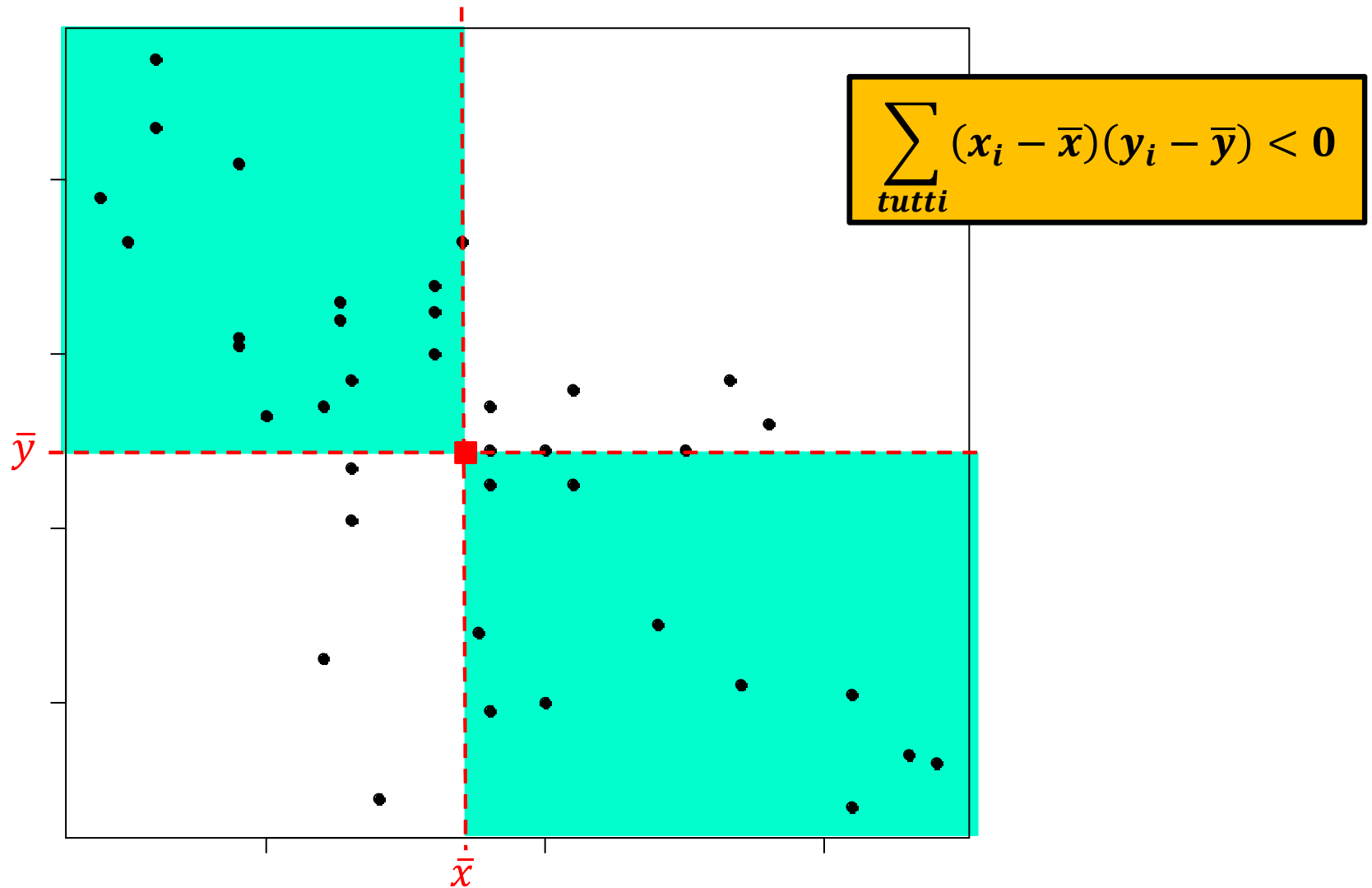
La covarianza



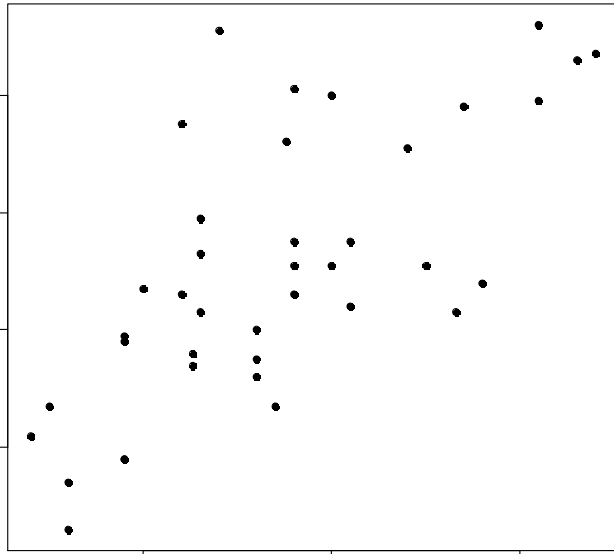
La covarianza



La covarianza

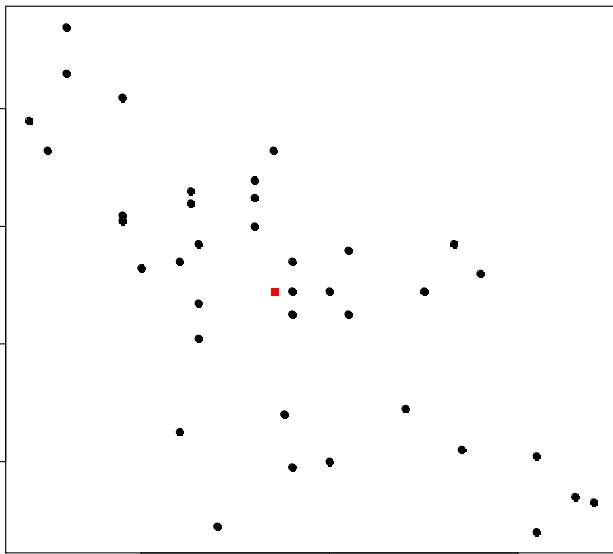


La covarianza



$$\sigma_{xy} = \text{cov}(x, y) = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$\text{cov}(x, y) > 0$ \longleftrightarrow quando x tende a crescere y fa lo stesso

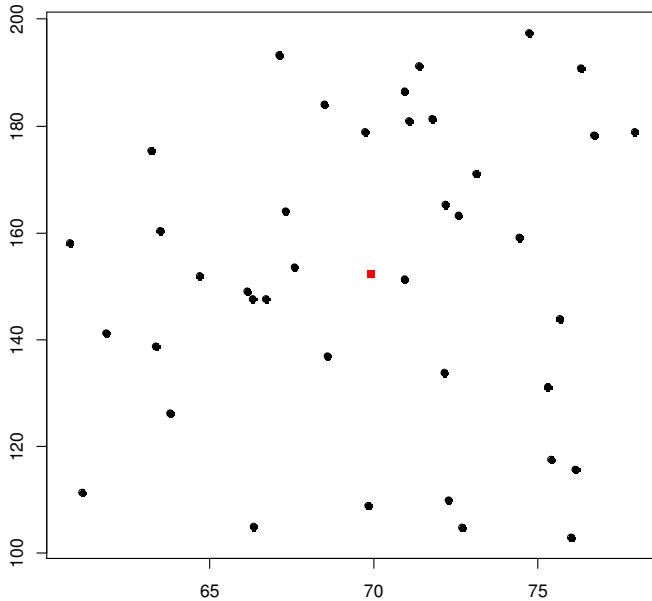


$\text{cov}(x, y) < 0$ \longleftrightarrow quando x tende a crescere y tende a decrescere

$$\sigma_{xy} = \text{cov}(x, y) = \left[\frac{1}{n} \sum_i x_i y_i \right] - \bar{x} \times \bar{y}$$

La covarianza

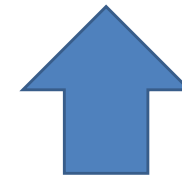
$$\sigma_{xy} = cov(x, y) = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$



$$cov(x, y) = 0$$



$$\left[\frac{1}{n} \sum_i x_i y_i \right] = \bar{x} \times \bar{y}$$



$$\sigma_{xy} = cov(x, y) = \left[\frac{1}{n} \sum_i x_i y_i \right] - \bar{x} \times \bar{y}$$

La correlazione lineare

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$\sigma_{xy} = \text{cov}(x, y) = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$-1 \leq \rho_{xy} \leq 1$$

$$\rho_{xy} > 0$$

$$\rho_{xy} < 0$$

$$\rho_{xy} = 0$$

$$\text{cov}(x, y) > 0 \iff$$

quando x tende a crescere
 y fa lo stesso

$$\text{cov}(x, y) < 0 \iff$$

quando x tende a crescere
 y tende a decrescere

$$\text{cov}(x, y) = 0$$

**coeff. di
correlazione
lineare**

$$\sigma_{xy} = \text{cov}(x, y) = \left[\frac{1}{n} \sum_i x_i y_i \right] - \bar{x} \times \bar{y}$$

$$-\sigma_x \sigma_y \leq \sigma_{xy} \leq \sigma_x \sigma_y$$

La correlazione I

$$\sigma^2_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$\sigma_{xy} = \text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$-1 \leq \rho_{xy} \leq 1$$

$$\rho_{xy} > 0$$

$$\rho_{xy} < 0$$

$$\rho_{xy} = 0$$

$$\text{cov}(x, y) > 0$$

$$\text{cov}(x, y) < 0$$

$$\text{cov}(x, y) = 0$$

anche
 $n - 1 \dots$

quando x tende a crescere
 y fa lo stesso

quando x tende a crescere
 y tende a decrescere

coeff. di
correlazione
lineare

$$\sigma_{xy} = \text{cov}(x, y) = \left[\frac{1}{n} \sum_{i=1}^n x_i y_i \right] - \bar{x} \times \bar{y}$$

$$-\sigma_x \sigma_y \leq \sigma_{xy} \leq \sigma_x \sigma_y$$

La correlazione lineare

A spanne...		
Valore di $ \rho $	Interpretazione fine	Meno fine
.90 - 1.00	Molto alta	forte (talvolta 0.5->0.6)
.70 - .90	Alta	
.50 - .70	Moderata	
.30 - .50	Bassa	moderata
.00 - .30	Poca o non correlazione	> 0.10 debole

Hinkle, Wiersma, & Jurs (2003). Applied Statistics for the Behavioral Sciences (5th ed.).



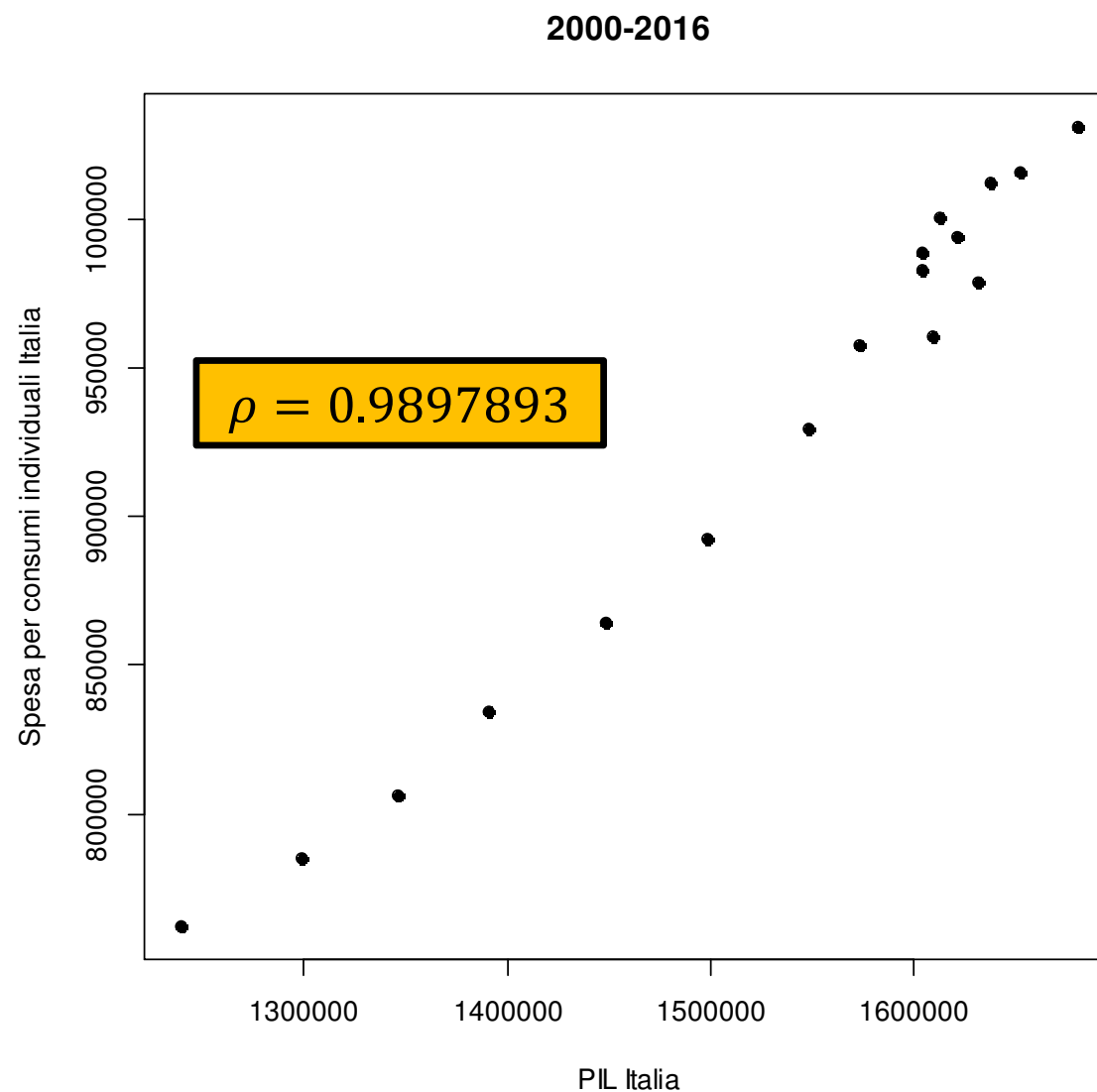
Esempio 1

Istat (2017)

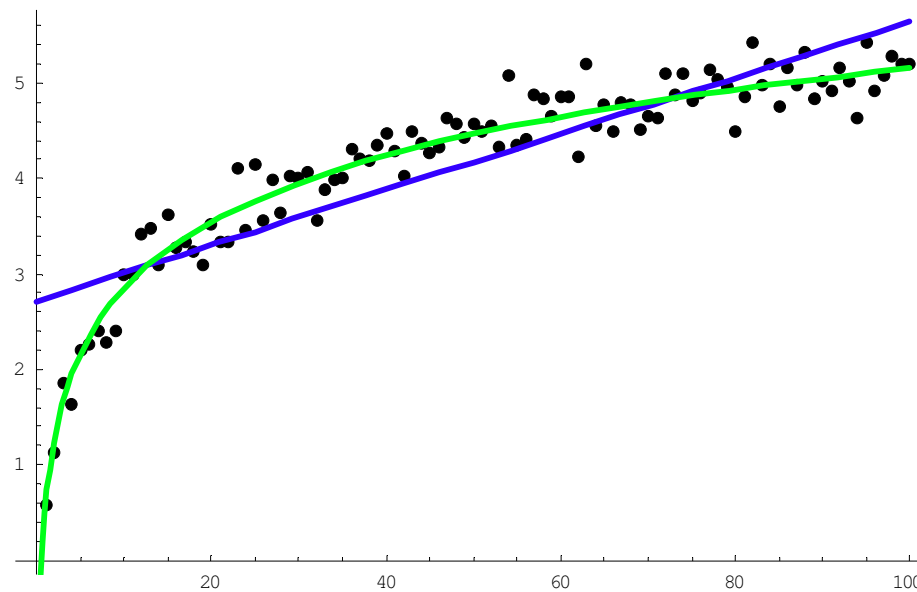
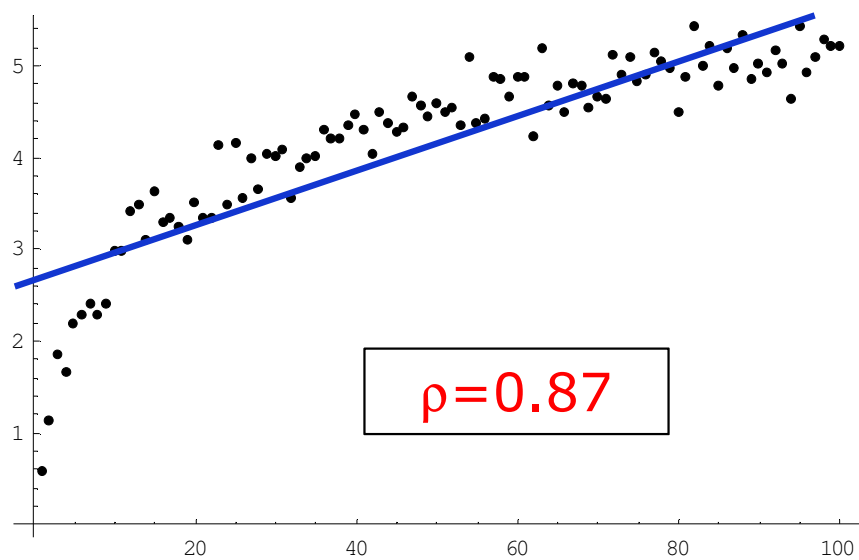
Conti e aggregati
economici nazionali
annuali

Spese per consumi
finali delle famiglie

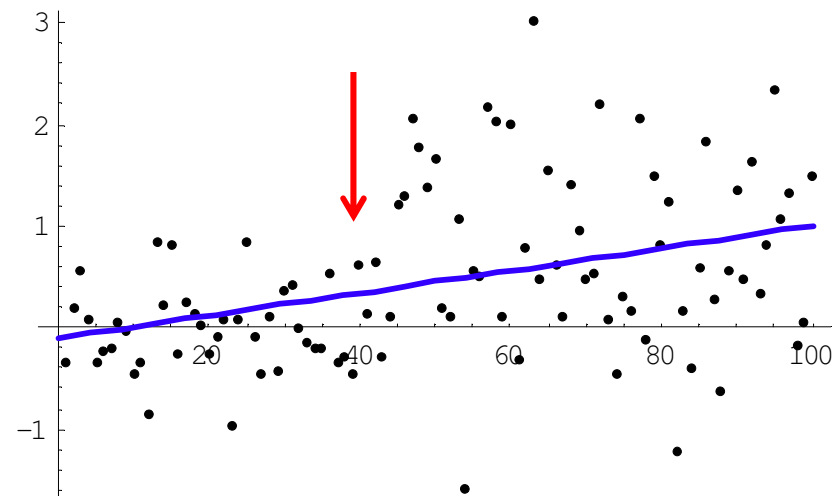
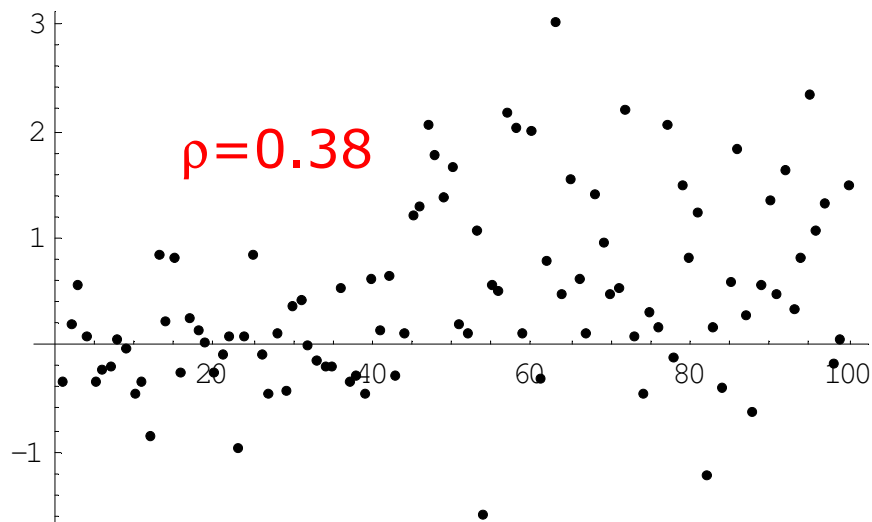
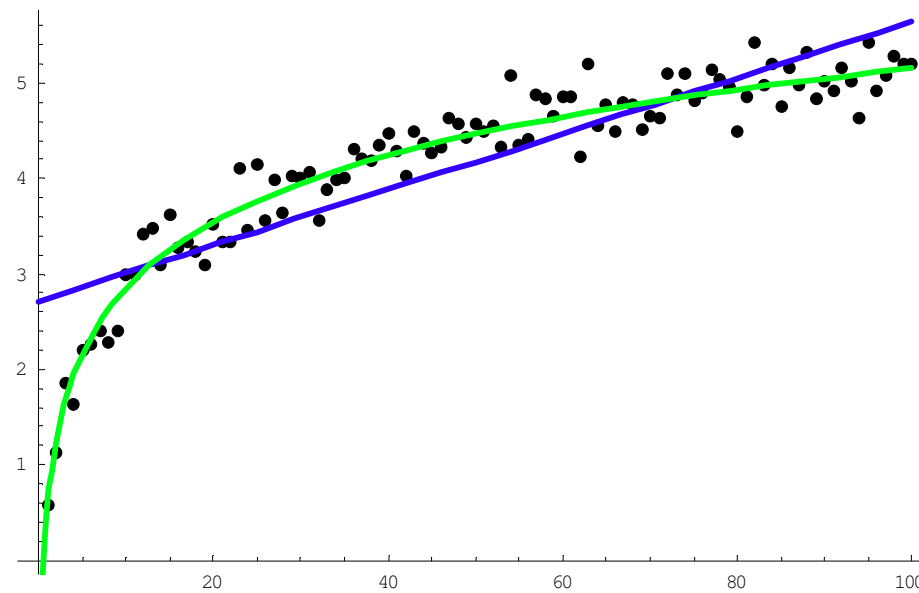
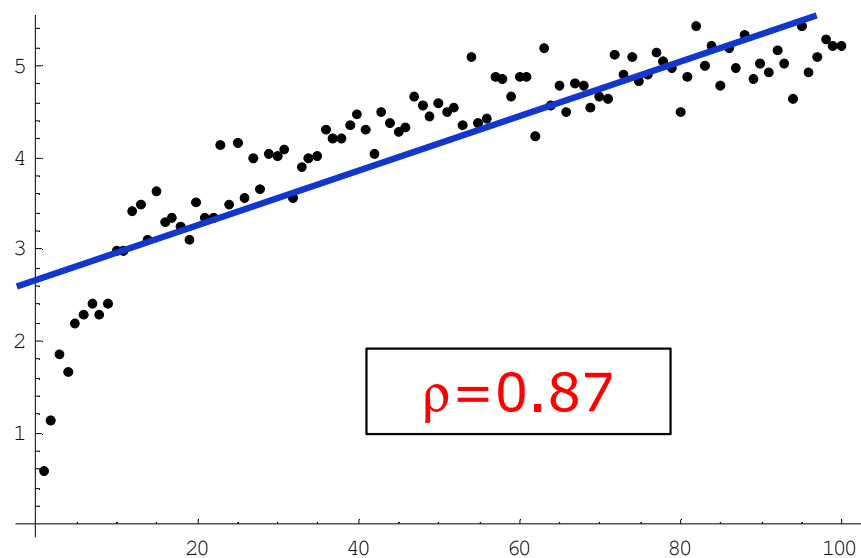
<http://dati.istat.it/>



Interpretare l'associazione



Interpretare l'associazione



Interpretare l'associazione



Il tempo giusto per tutti L'editoriale del n.160 di Mind di Marco Cattaneo

Il tempo giusto per tutti



di MARCO CATTANEO

Gufo, decisamente gufo. Secondo i parametri che assegnano un profilo ai nostri ritmi biologici, sono un gufo. Non nel senso di menagramo, ma di inguaribile tiratardi che non andrebbe mai a dormire la sera e non si alzerebbe mai dal letto al mattino.

Nel test che trovate a pagina 42, d'altra parte, ho totalizzato 22 punti. Considerato che si è classificati come gufi sotto i 53 punti, il mio «cronotipo» non è nemmeno in discussione. Ci vogliono almeno 64 punti invece per essere un'allodola, vale a dire uno che ama alzarsi alle prime luci dell'alba e che non ha perso l'abitudine, e il gusto, di coricarsi dopo Carosello. (Che, a beneficio dei nati dopo il 1977, era un programma televisivo di lunghe clip pubblicitarie con siparietti comici e cartoni animati molto amato dai bambini).

Certo, quel test è poco più che un gioco, ma è evidente che ognuno di noi ha i suoi tempi. A cominciare proprio dall'orologio interno che regola il ciclo sonno/veglia, il più conosciuto dei ritmi circadiani. È talmente ben codificato nel nostro «libretto delle istruzioni» che si conserva anche in condizioni di isolamento: in assenza di luce naturale e di orologi, il nostro corpo segue automaticamente un ciclo di circa 25 ore.

Ma non è l'unico, anzi. Diverso da persona a persona è l'orologio che scandisce i tempi della fame, per esempio. E

diversi sono orologi molto meno conosciuti e che, soprattutto, non percepiamo, perché non sono sintomatici. Così, mentre la melatonina regola il ritmo giorno/notte, anche altri ormoni seguono un andamento periodico, che governa i nostri comportamenti. «È probabile - racconta Henrik Oster a pagina 26 - che ciascuna cellula del nostro organismo contenga addirittura un proprio orologio che, in base alla funzione dell'organo in cui si trova, attiva o disattiva durante il giorno e in modo coordinato i programmi biologici adeguati».

Ma c'è di più. Da una parte i nostri ritmi biologici non rimangono inalterati per tutta la vita. In genere siamo allodole da bambini e diventiamo gufi da adolescenti, per poi ritornare gradualmente allodole. E con l'età «si riducono le differenze tra il giorno e la notte nei numerosi processi fisiologici guidati dai ritmi circadiani». Dall'altra, ci sono fattori esterni che influenzano sensibilmente i nostri orologi biologici. La luce e l'assunzione di cibo sono i due principali.

A questo proposito, gli studi su chi lavora di frequente facendo turni di notte

hanno evidenziato che lo sfasamento dei ritmi circadiani può portare a una maggiore frequenza di malattie anche gravi, come diabete di tipo 2 e disturbi cardiovascolari. Se poi si tratti soltanto di correlazioni oppure di fattori causali, è ancora da stabilire.

È invece associato che il lungo inverno artico porti a una maggiore frequenza di depressione nelle popolazioni che vivono ad alte latitudini, a causa di quello che è chiamato «disturbo affettivo stagionale». Questo genere di disturbi, cui contribuisce il malfunzionamento del nostro orologio interno, può essere trattato con la terapia della luce, come racconta Irene Campagna a pagina 36 in un'intervista a Cristina Anna Colombo, direttore del Centro disturbi dell'umore dell'IRCCS Ospedale San Raffaele Turro, a Milano. Diffusa soprattutto nei paesi scandinavi, è arrivata però anche da noi, e viene usata per la depressione grave ricorrente.

Anche se non soffriamo di disturbi dell'umore, però, assecondare il ritmo dei nostri orologi biologici è importante per il benessere di tutti. Sia che siate gufi, sia che siate allodole.

Interpretare l'associazione

A questo proposito, gli studi su chi lavora di frequente facendo turni di notte hanno evidenziato che lo sfasamento dei ritmi circadiani può portare a una maggiore frequenza di malattie anche gravi, come diabete di tipo 2 e disturbi cardiocircolatori. **Se poi si tratti soltanto di correlazioni oppure di fattori causali, è ancora da stabilire.**

...o diversi da persona a persona e cambiano
...do la nostra salute fisica e mentale

...sto per tutti



di MARCO CATTANEO

...assegnano
...n nel senso
...andrebbe mai
...il mattino.

...sono-
...si,
...mo
...guo-
...gover-
...e proba-
...a pagina
...il nostro or-
...tura un pro-
...alla funzione
...a, attiva o disat-
...in modo coordi-
...gici adeguati.
...a parte i nostri rit-
...nangono inalterati
...nere siamo allodo-
...entiamo gufi da ado-
...tornare gradualmente
...età «si riducono le diffe-
...giorno e la notte nei nune-
...essi fisiologici guidati dai rit-
...circadiani». Dall'altra, ci sono fattori
...esterni che influenzano sensibilmente
...i nostri orologi biologici. La luce e l'as-
...sunzione di cibo sono i due principali.

hanno evidenziato che lo sfasamento dei ritmi circadiani può portare a una maggiore frequenza di malattie anche gravi, come diabete di tipo 2 e disturbi cardiocircolatori. Se poi si tratti soltanto di correlazioni oppure di fattori causali, è ancora da stabilire.

È invece associato che il lungo inverno artico porti a una maggiore frequenza di depressione nelle popolazioni che vivono ad alte latitudini, a causa di quello che è chiamato «disturbo affettivo stagionale». Questo genere di disturbi, cui contribuisce il malfunzionamento del nostro orologio interno, può essere trattato con la terapia della luce, come racconta Irene Campagna a pagina 36 in un'intervista a Cristina Anna Colombo, direttore del Centro disturbi dell'umore dell'IRCCS Ospedale San Raffaele Turro, a Milano. Diffusa soprattutto nei paesi scandinavi, è arrivata però anche da noi, e viene usata per la depressione grave ricorrente.

Anche se non soffriamo di disturbi dell'umore, però, assecondare il ritmo dei nostri orologi biologici è importante per il benessere di tutti. Sia che siate gufi, sia che siate allodole.

orologi, il nostro corpo segue automaticamente un ciclo di circa 25 ore.
Ma non è l'unico, anzi. Diverso da persona a persona è l'orologio che scandisce i tempi della fame, per esempio. E

A questo proposito, gli studi su chi lavora di frequente facendo turni di notte

80112
MILANO ANONIMA S.p.A. - VIA S. PIETRO 12 - 20121 MILANO
COPERTINA: ANONIMA S.p.A. - 20121 MILANO



44 Psicologia
Madri pentite

56 Fotografia
Diario di un'anoressica

Interpretare l'associazione

← → ↻ 🏠 Sicuro | <https://www.ncbi.nlm.nih.gov/pubmed/14738551>

App WordReference.com Google Nuova scheda 7.2 - Diagnosing Logi

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

Format: Abstract ▾

[Paediatr Perinat Epidemiol.](#) 2004 Jan;18(1):88-92.

New evidence for the theory of the stork.


[Höfer T](#)¹, [Przyrembel H](#), [Verleger S](#).

Author information

Abstract

Data from Berlin (Germany) show a significant correlation between the increase in the stork population around the city and the increase in deliveries outside city hospitals (out-of-hospital deliveries). However, there is no correlation between deliveries in hospital buildings (clinical deliveries) and the stork population. The decline in the number of pairs of storks in the German state of Lower Saxony between 1970 and 1985 correlated with the decrease of deliveries in that area. The nearly constant number of deliveries from 1985 to 1995 was associated with an unchanged stork population (no statistical significance). However, the relevance of the stork for the birth rate in that part of Germany remains unclear, because the number of out-of-hospital deliveries in this area is not well documented. A lack of statistical information on out-of-hospital deliveries in general is a severe handicap for further proof for the Theory of the Stork.

PMID: 14738551
[Indexed for MEDLINE]



Interpretare l'associazione

88

For the classroom

New evidence for the Theory of the Stork

Thomas Höfer^a, Hildegard Przyrembel^b and Silvia Verleger^c

^aFederal Institute for Risk Assessment, Berlin, ^bOffice of the National Breast Feeding Committee at BfR, Berlin, and ^cIndependent Midwife, Berlin, Germany

Summary

Correspondence:

Dr Thomas Höfer, Federal
Institute for Risk Assessment,
Thielallee 88-92, D-14195
Berlin, Germany.
E-mail:
thomas.hoefler@bfr.bund.de

Data from Berlin (Germany) show a significant correlation between the increase in the stork population around the city and the increase in deliveries outside city hospitals (out-of-hospital deliveries). However, there is no correlation between deliveries in hospital buildings (clinical deliveries) and the stork population. The decline in the number of pairs of storks in the German state of Lower Saxony between 1970 and 1985 correlated with the decrease of deliveries in that area. The nearly constant number of deliveries from 1985 to 1995 was associated with an unchanged stork population (no statistical significance). However, the relevance of the stork for the birth rate in that part of Germany remains unclear, because the number of out-of-hospital deliveries in this area is not well documented. A lack of statistical information on out-of-hospital deliveries in general is a severe handicap for further proof for the Theory of the Stork.

The intended value (disclaimer): This article is not intended to disprove the value of serious epidemiological investigations. It is an example of how studies based on popular belief and unsubstantiated theory, seconded by low quality references and supported by coincidental statistical association could lead to apparent scientific endorsement. Insofar it is a humorous case study for education in perinatal epidemiology.

Interpretare l'associazione

88

For the classroom

New evidence for the Theory of the Stork

Thomas Höfer^a, Hildegard Przyrembel^b and Silvia Verleger^c

^aFederal Institute for Risk Assessment, Berlin, ^bOffice of the National Breast Feeding Committee at BfR, Berlin, and ^cIndependent Midwife, Berlin, Germany

Summary

Correspondence:
Dr Thomas Höfer, Federal Institute for Risk Assessment, Thielallee 88-92, D-14195 Berlin, Germany.
E-mail: thomas.hoefler@bfr.bund.de

Data from Berlin (Germany) show a significant correlation between the increase in the stork population around the city and the increase in deliveries outside city hospitals (out-of-hospital deliveries). However, there is no correlation between deliveries in hospital buildings (clinical deliveries) and the stork population. The decline in the number of pairs of storks in the German state of Lower Saxony between 1970 and 1985 correlated with the decrease of deliveries in that area. The nearly constant number of deliveries from 1985 to 1995 was associated with an unchanged stork population (no statistical significance). However, the relevance of the stork for the birth rate in that part of Germany remains unclear, because the number of out-of-hospital deliveries in this area is not well documented. A lack of statistical information on out-of-hospital deliveries in general is a severe handicap for further proof for the Theory of the Stork.

The intended value (disclaimer): This article is not intended to disprove the value of serious epidemiological investigations. It is an example of how studies based on popular belief and unsubstantiated theory, seconded by low quality references and supported by coincidental statistical association could lead to apparent scientific endorsement. Insofar it is a humorous case study for education in perinatal epidemiology.

Introduction

Background

Two different theories exist concerning the origin of children: the Theory of Sexual Reproduction (ThoSR) and the Theory of the Stork (ThoS).

For many people the stork has been and still is the 'bringer of new life' or the 'baby-carrier'. During pregnancy people say, 'the stork has been visiting'. Old German scientists therefore named the stork *odebero* (in dutch *oievaar*), which means 'bringing luck'.

Nowadays, many people believe in the theory of reproduction, simply because they have been taught this theory in school, although it is a scientific theory, not a truth (Leisti T, personal communication via leisti@cc.helsinki.fi, 2001). A number of the world's leading scientists are still in favour of ThoS. Some recent scientific work (1997-2001) has shown new evidence for ThoS.

Status of ThoS

A search conducted on the internet revealed that more than 20 internet domains are communicating about aspects of ThoS (Table 1). This is evidently important for the future of the ThoS's relevance.

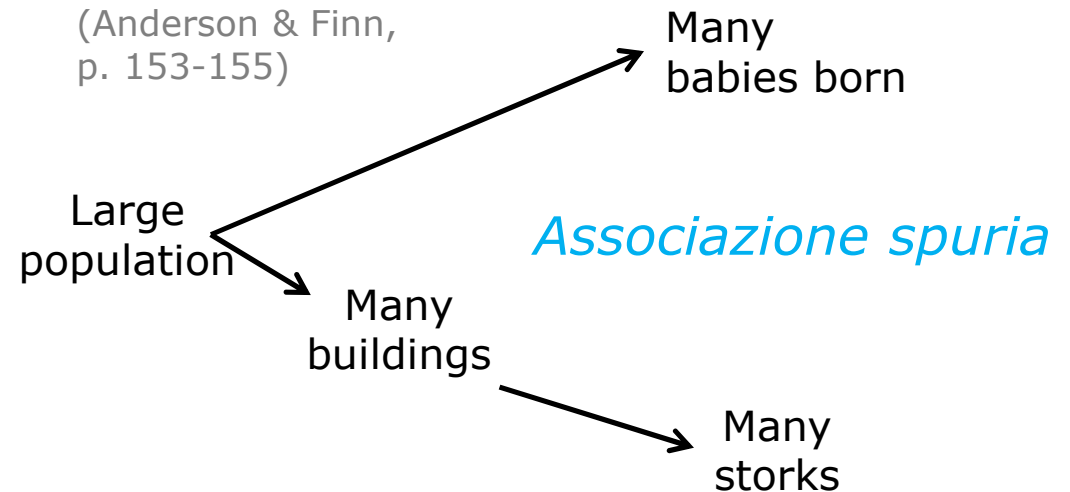
According to work of Prof Erkki Aalto¹ from the University of Helsinki the evidence supporting ThoS is based on six facts:

- (i) storks exist
- (ii) unexplained features of fetal development
- (iii) a newborn is new-born
- (iv) sexual intercourse without delivery
- (v) positive correlation of birth rate to stork population
- (vi) scientific studies.

Aalto's work focussed not only on aspects in the field of gynaecology and obstetrics:

Aspect no. (i): It is a scientifically established fact that the stork exists, which is confirmed by ornithologists.

(Anderson & Finn, p. 153-155)



According to work of Prof Erkki Aalto¹ from the University of Helsinki the evidence supporting ThoS is based on six facts:

- (i) storks exist
- (ii) unexplained features of fetal development
- (iii) a newborn is new-born
- (iv) sexual intercourse without delivery
- (v) positive correlation of birth rate to stork population
- (vi) scientific studies.

Interpretare l'associazione

88

For the classroom

New evidence for the Theory of the Stork

Thomas Höfer^a, Hildegard Przyrembel^b and Silvia Verleger^c

^aFederal Institute for Risk Assessment, Berlin, ^bOffice of the National Breast Feeding Committee at BfR, Berlin, and ^cIndependent Midwife, Berlin, Germany

Summary

Correspondence:

Dr Thomas Höfer, Federal Institute for Risk Assessment, Thielallee 88-92, D-14195 Berlin, Germany.
E-mail: thomas.hoefer@bfr.bund.de

Data from Berlin (Germany) show a significant correlation between the increase in the stork population around the city and the increase in deliveries outside city hospitals (out-of-hospital deliveries). However, there is no correlation between deliveries in hospital buildings (clinical deliveries) and the stork population. The decline in the number of pairs of storks in the German state of Lower Saxony between 1970 and 1985 correlated with the decrease of deliveries in that area. The nearly constant number of deliveries from 1985 to 1995 was associated with an unchanged stork population (no statistical significance). However, the relevance of the stork for the birth rate in that part of Germany remains unclear, because the number of out-of-hospital deliveries in this area is not well documented. A lack of statistical information on out-of-hospital deliveries in general is a severe handicap for further proof for the Theory of the Stork.

The intended value (disclaimer): This article is not intended to disprove the value of serious epidemiological investigations. It is an example of how studies based on popular belief and unsubstantiated theory, seconded by low quality references and supported by coincidental statistical association could lead to apparent scientific endorsement. Insofar it is a humorous case study for education in perinatal epidemiology.

Introduction

Background

Two different theories exist concerning the origin of children: the Theory of Sexual Reproduction (ThoSR) and the Theory of the Stork (ThoS).

For many people the stork has been and still is the 'bringer of new life' or the 'baby-carrier'. During pregnancy people say, 'the stork has been visiting'. Old German scientists therefore named the stork *odebero* (in dutch *oievaar*), which means 'bringing luck'.

Nowadays, many people believe in the theory of reproduction, simply because they have been taught this theory in school, although it is a scientific theory, not a truth (Leisti T, personal communication via leisti@cc.helsinki.fi, 2001). A number of the world's leading scientists are still in favour of ThoS. Some recent scientific work (1997-2001) has shown new evidence for ThoS.

Status of ThoS

A search conducted on the internet revealed that more than 20 internet domains are communicating about aspects of ThoS (Table 1). This is evidently important for the future of the ThoS's relevance.

According to work of Prof Erkki Aalto¹ from the University of Helsinki the evidence supporting ThoS is based on six facts:

- (i) storks exist
- (ii) unexplained features of fetal development
- (iii) a newborn is new-born
- (iv) sexual intercourse without delivery
- (v) positive correlation of birth rate to stork population
- (vi) scientific studies.

Aalto's work focussed not only on aspects in the field of gynaecology and obstetrics:

Aspect no. (i): It is a scientifically established fact that the stork exists, which is confirmed by ornithologists.

La correlazione da sola non implica una diretta relazione di causa-effetto.

L'associazione può essere causata da una terza variabile non osservata.

According to work of Prof Erkki Aalto¹ from the University of Helsinki the evidence supporting ThoS is based on six facts:

- (i) storks exist
- (ii) unexplained features of fetal development
- (iii) a newborn is new-born
- (iv) sexual intercourse without delivery
- (v) positive correlation of birth rate to stork population
- (vi) scientific studies.

Interpretare l'associazione

The screenshot shows a web browser window displaying the Airc website. The URL is www.airc.it/cancro/disinformazione/pillola-anticoncezionale-tumore?ets_cmmk=3598&utm_source=newsletter&utm_medium=ema. The page title is "È vero che la pillola anticoncezionale aumenta il rischio di cancro?". The article text states: "Dipende dal tipo di cancro. In alcuni casi può aumentare il rischio, in altri ridurlo. In ogni caso la scelta di farne uso deve comprendere una valutazione generale dei rischi e dei benefici che vanno oltre la sola valutazione del rischio oncologico." Below the text is an image of a blister pack of white pills. The browser's taskbar at the bottom shows the time as 10:13 and the language as ENG.

La correlazione da sola non implica una diretta relazione di causa-effetto.

L'associazione può essere causata da una terza variabile non osservata.

Interpretare l'associazione

La correlazione da sola non implica una diretta relazione di causa-effetto.

L'associazione può essere causata da una terza variabile non osservata.

Parla con noi X

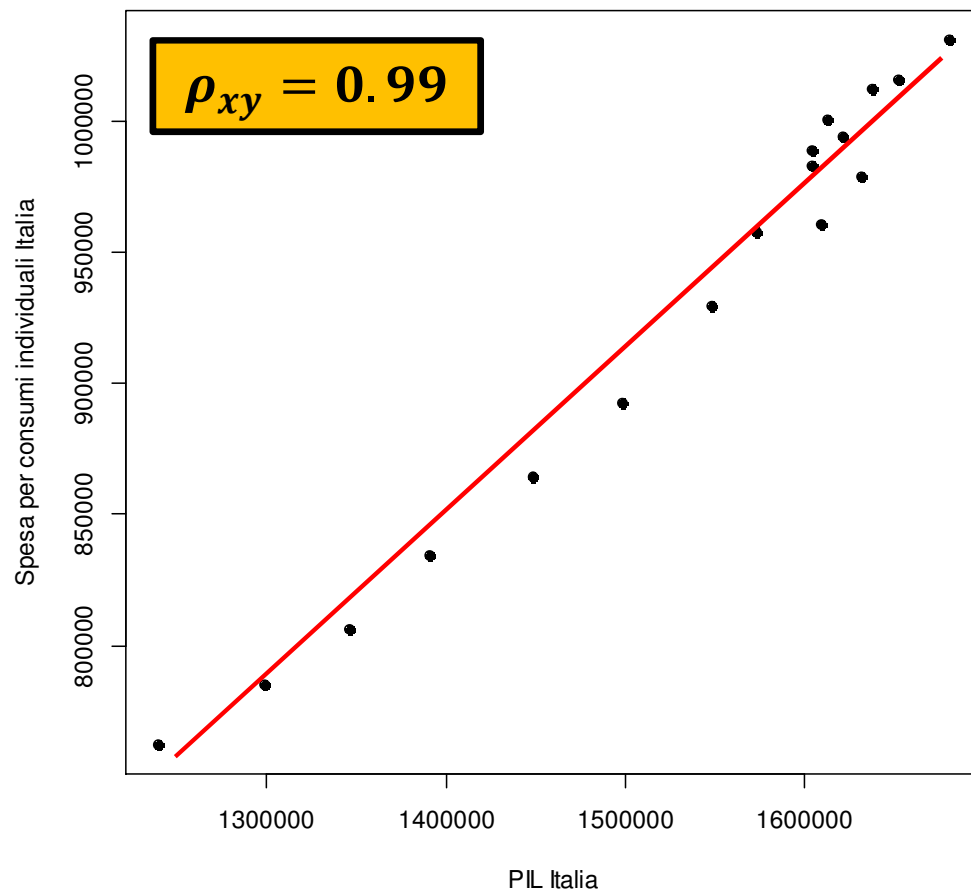
Pillola e cancro della cervice uterina

Un uso prolungato di contraccettivi orali è stato associato a un aumento del rischio di cancro della cervice uterina che tende a diminuire alla cessazione. Nel 2002, però, uno studio dell'Organizzazione mondiale della sanità ha dimostrato che **l'aumento di rischio è legato alla presenza di un'infezione da virus del Papilloma umano (HPV)**.

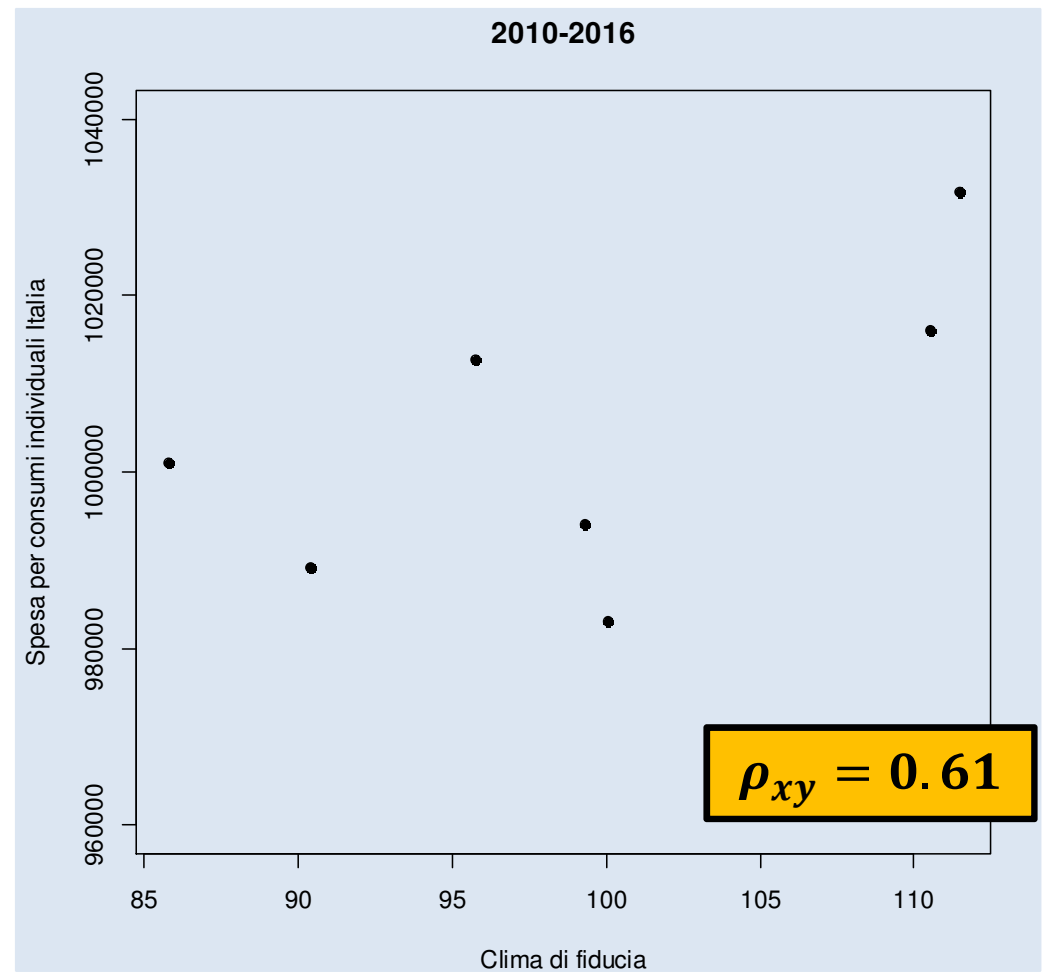
Di conseguenza è probabile che fra la pillola e il tumore vi sia un legame indiretto: le donne che usano questo metodo contraccettivo in genere non usano il preservativo e sono quindi a maggior rischio di contrarre l'HPV, [il vero responsabile del tumore della cervice uterina](#). Questo problema dovrebbe ridursi man mano che aumenta la percentuale di persone (femmine e maschi) vaccinate contro l'HPV.

La retta di regressione

2000-2016

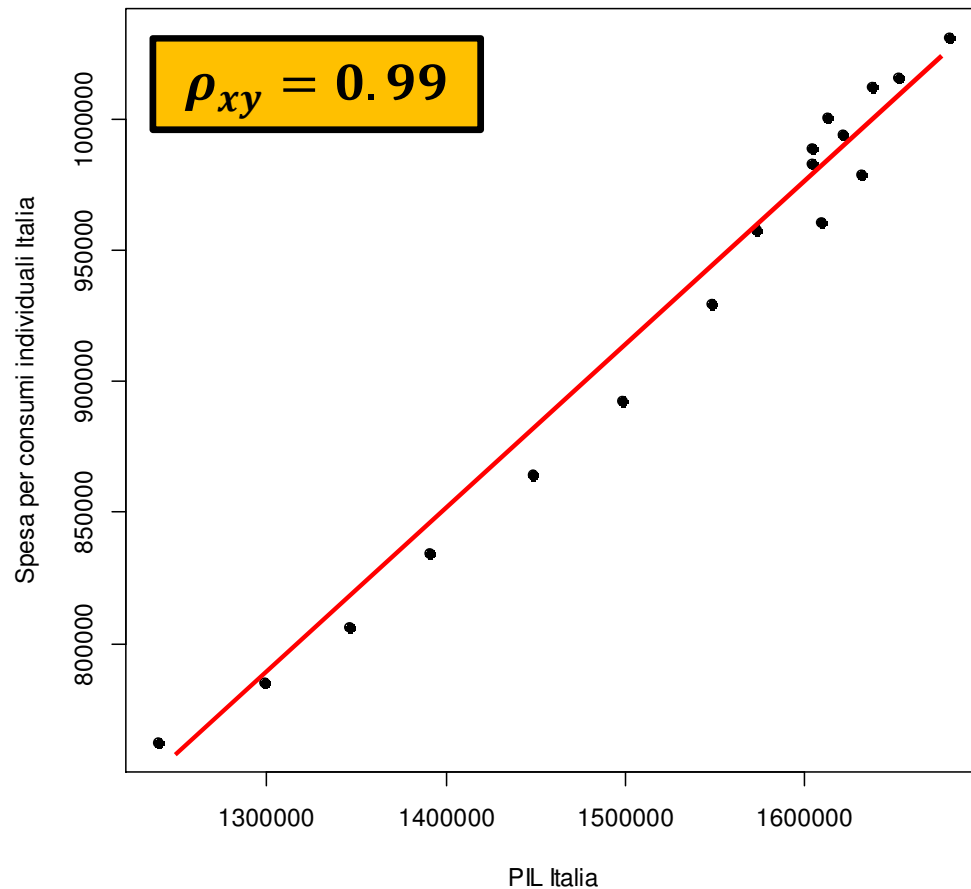


2010-2016



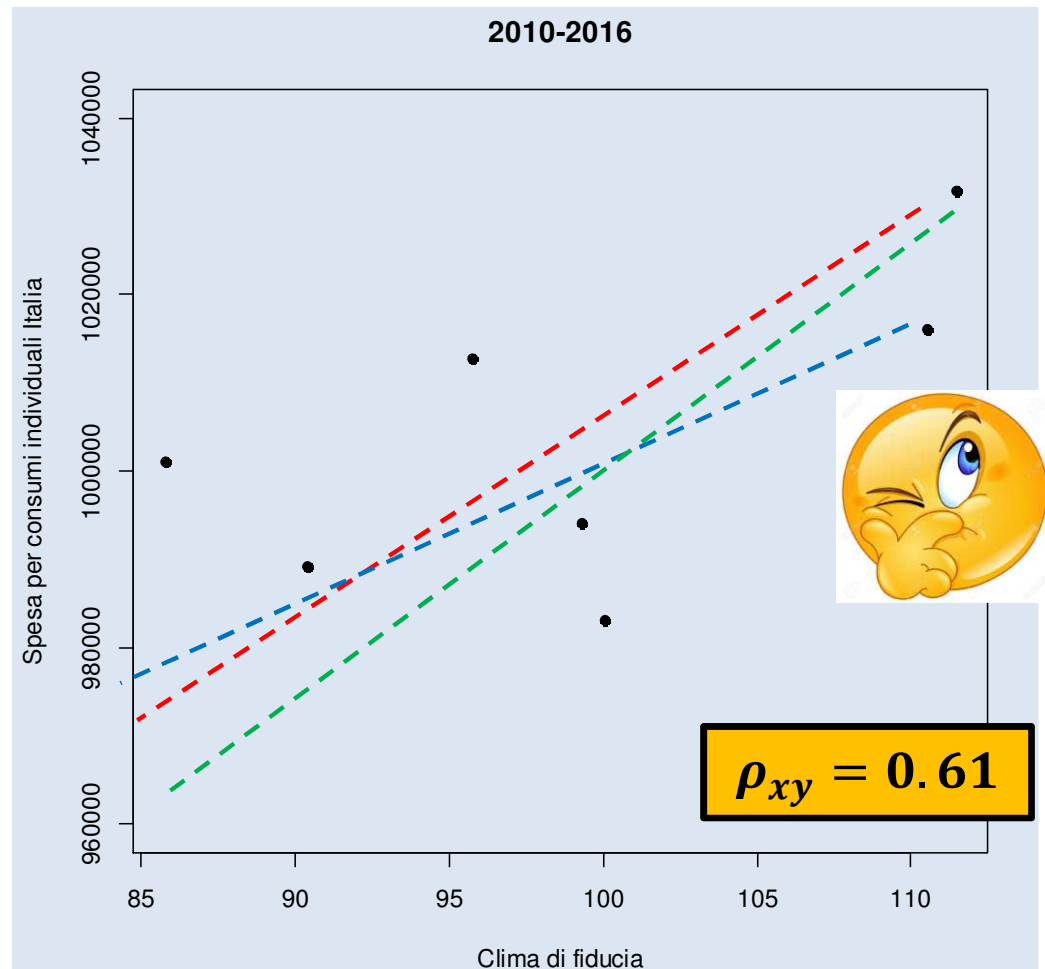
La retta di regressione

2000-2016



Qual è la retta che passa più vicino a tutti i punti, più o meno?

2010-2016



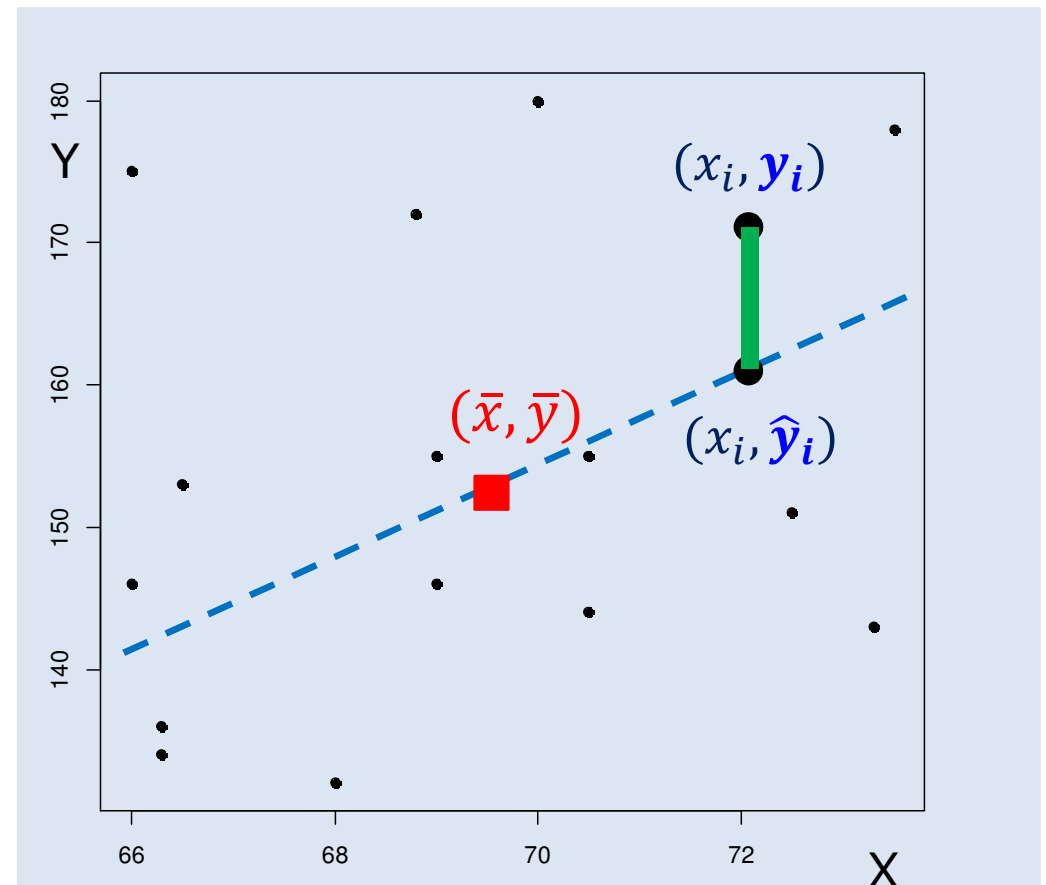
La retta di regressione

Di tutte le rette $y = a + bx$ quella che passa più vicino a tutti i punti, nel senso dei *minimi quadrati*, è quella con coeff. \hat{a} e \hat{b} che rendono minima la quantità:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

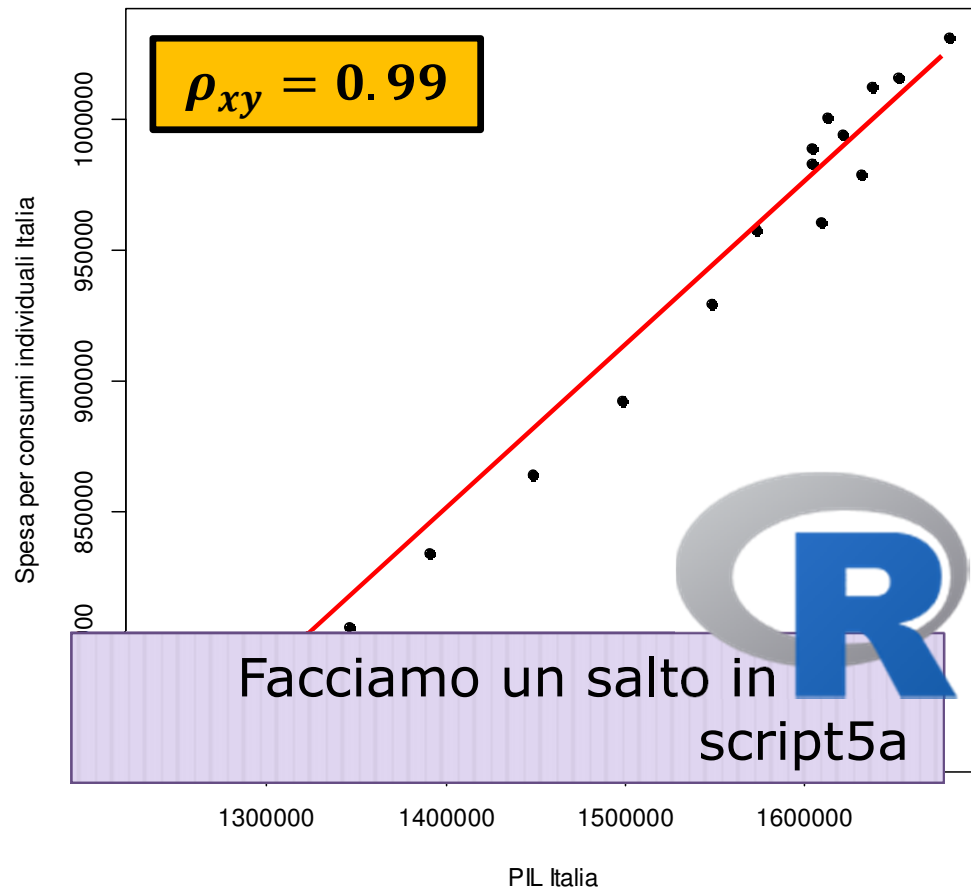
$$\hat{b} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

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



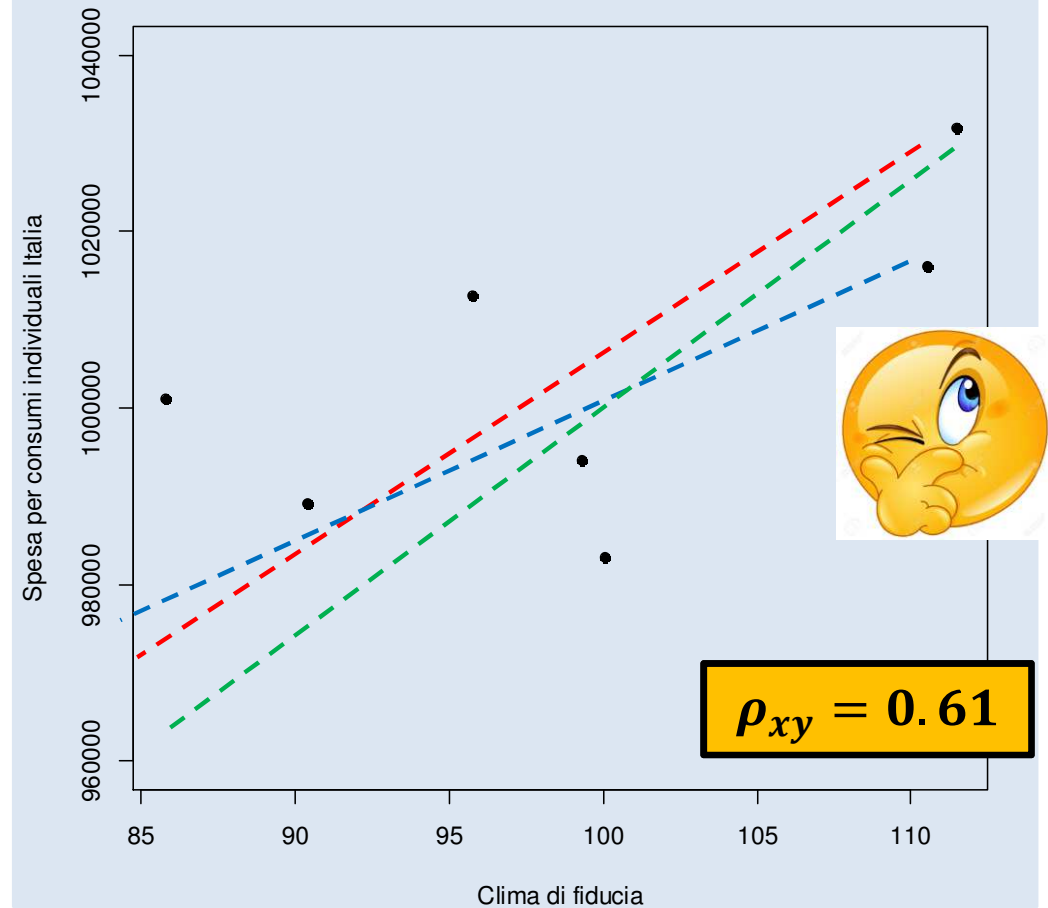
La retta di regressione

2000-2016



Qual è la retta che passa più vicino a tutti i punti, più o meno?

2010-2016





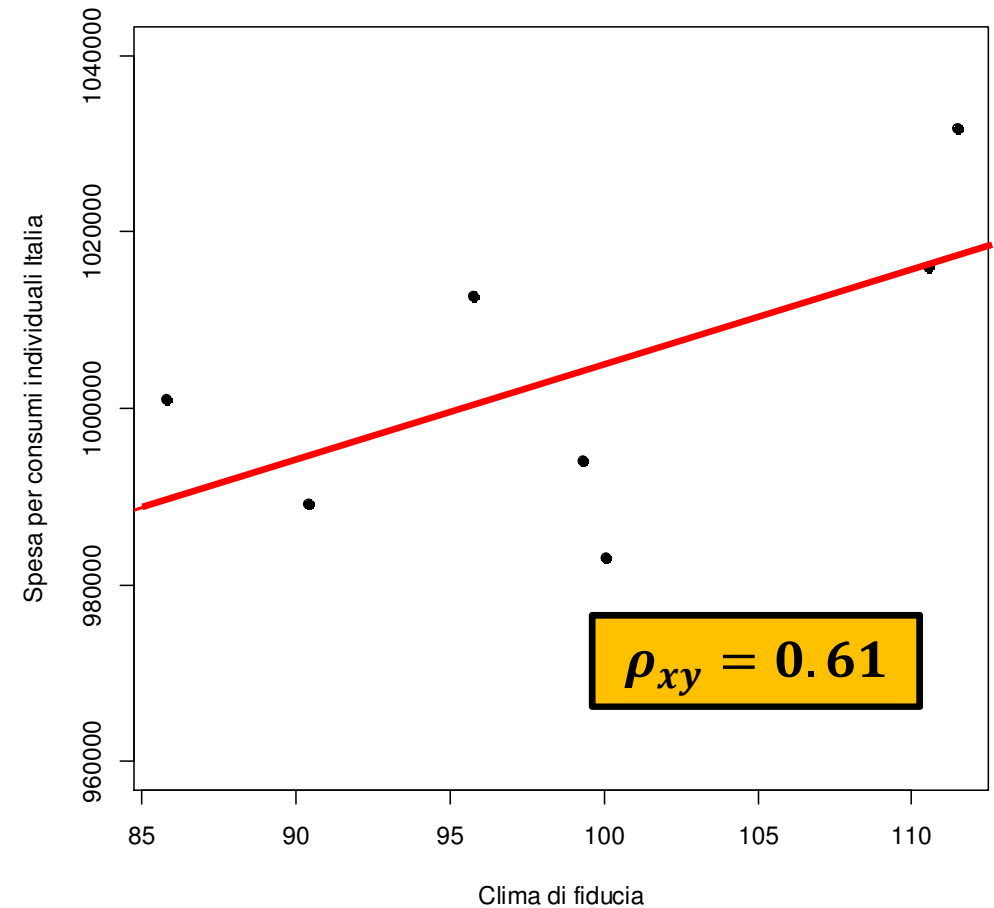
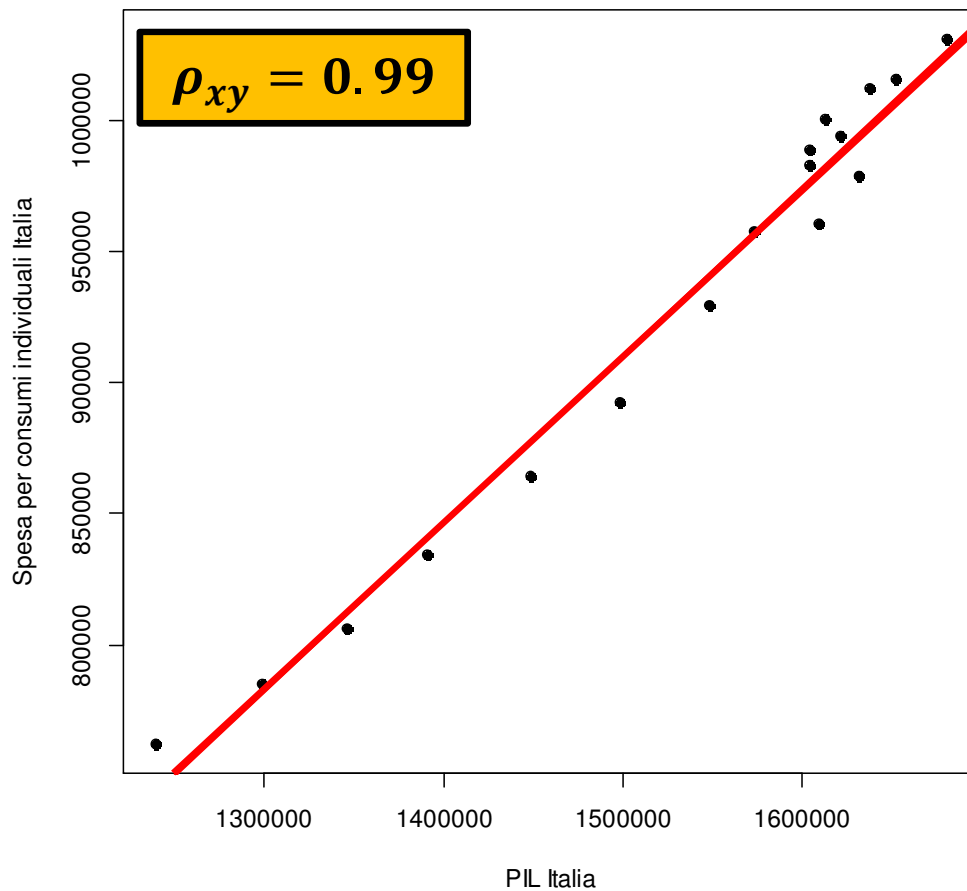
La retta di regressione

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.272e+04  3.665e+04  -1.438   0.171
PIL          6.422e-01  2.388e-02  26.894 4.16e-14 ***
---
```

2000-2016

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 896307.8  63146.5  14.194 3.13e-05 ***
CFC         1087.0    634.9   1.712  0.148
---
```

2010-2016





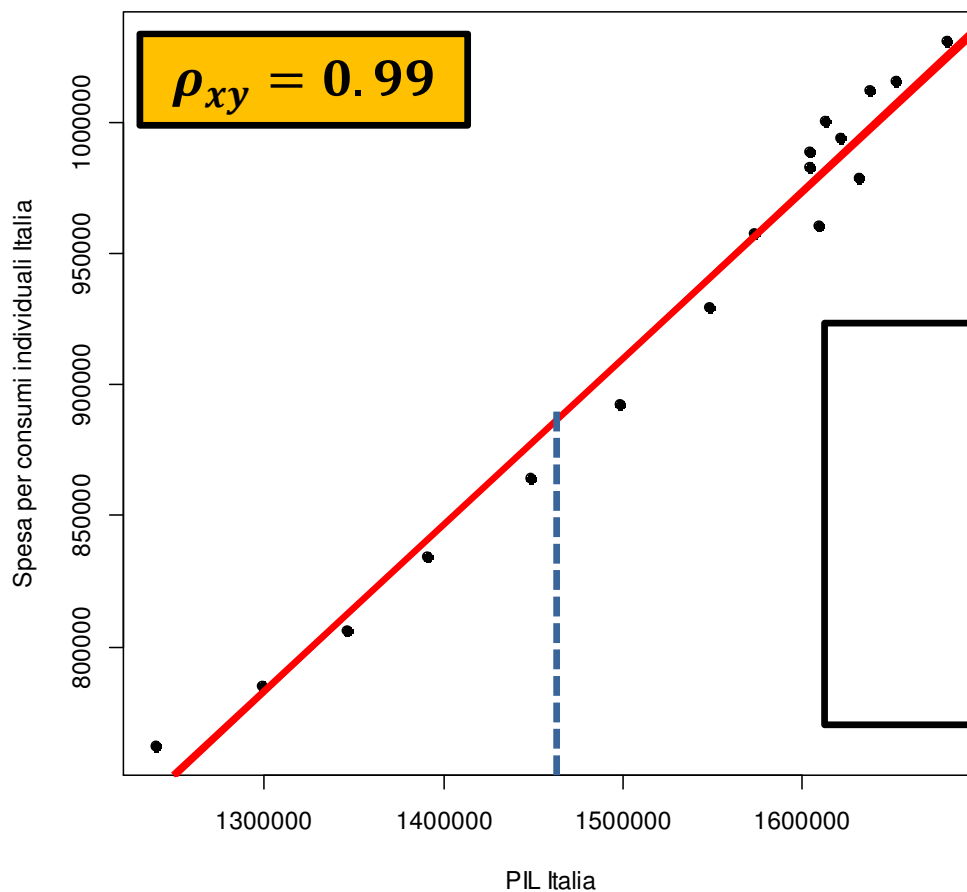
La retta di regressione

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.272e+04  3.665e+04  -1.438   0.171
PIL          6.422e-01  2.388e-02  26.894 4.16e-14 ***
---
```

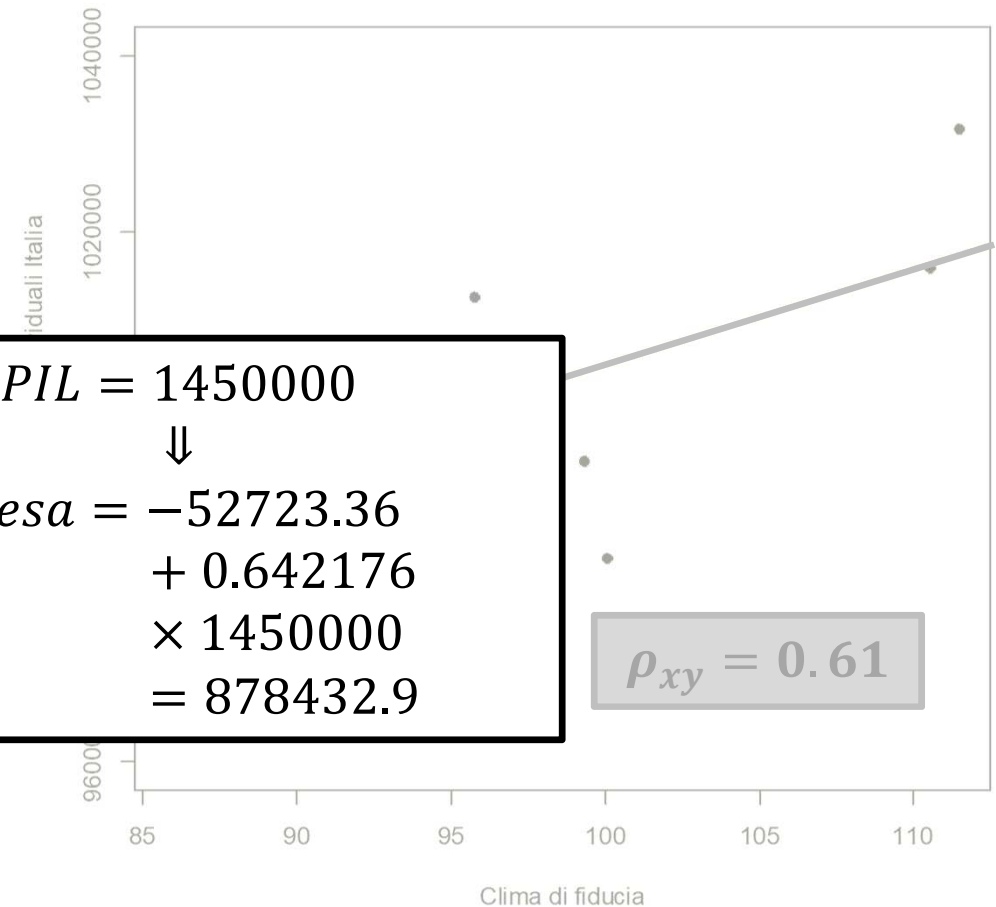
2000-2016

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  896307.8   63146.5   14.194 3.13e-05 ***
CFC          1087.0     634.9     1.712  0.148
---
```

2010-2016



$$\begin{aligned} \text{PIL} &= 1450000 \\ &\Downarrow \\ \text{Spesa} &= -52723.36 \\ &\quad + 0.642176 \\ &\quad \times 1450000 \\ &= 878432.9 \end{aligned}$$

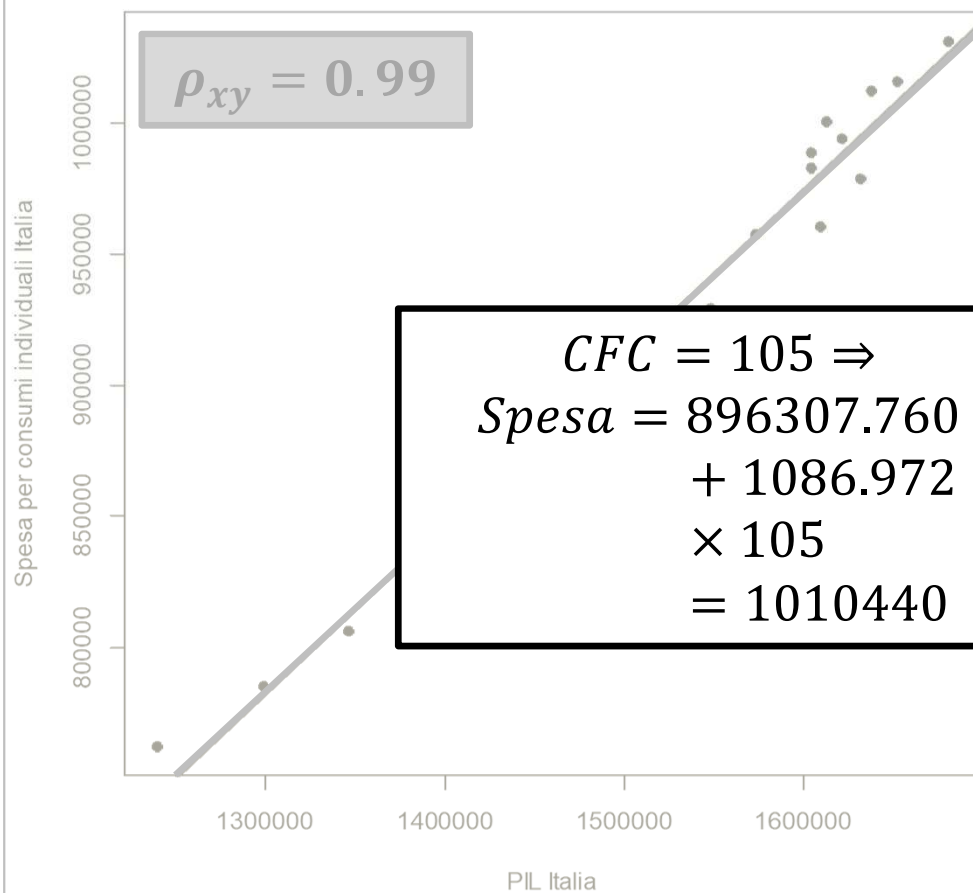




La retta di regressione

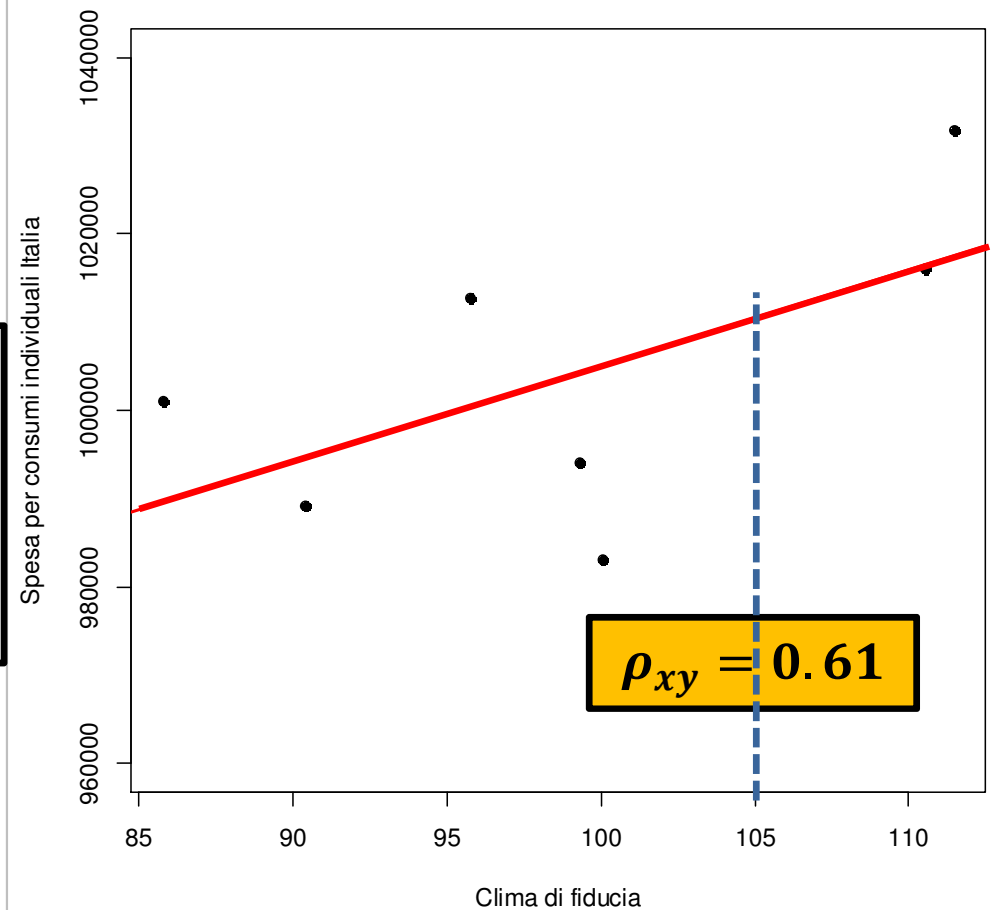
```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.272e+04  3.665e+04  -1.438   0.171
PIL          6.422e-01  2.388e-02  26.894 4.16e-14 ***
---
```

2000-2016

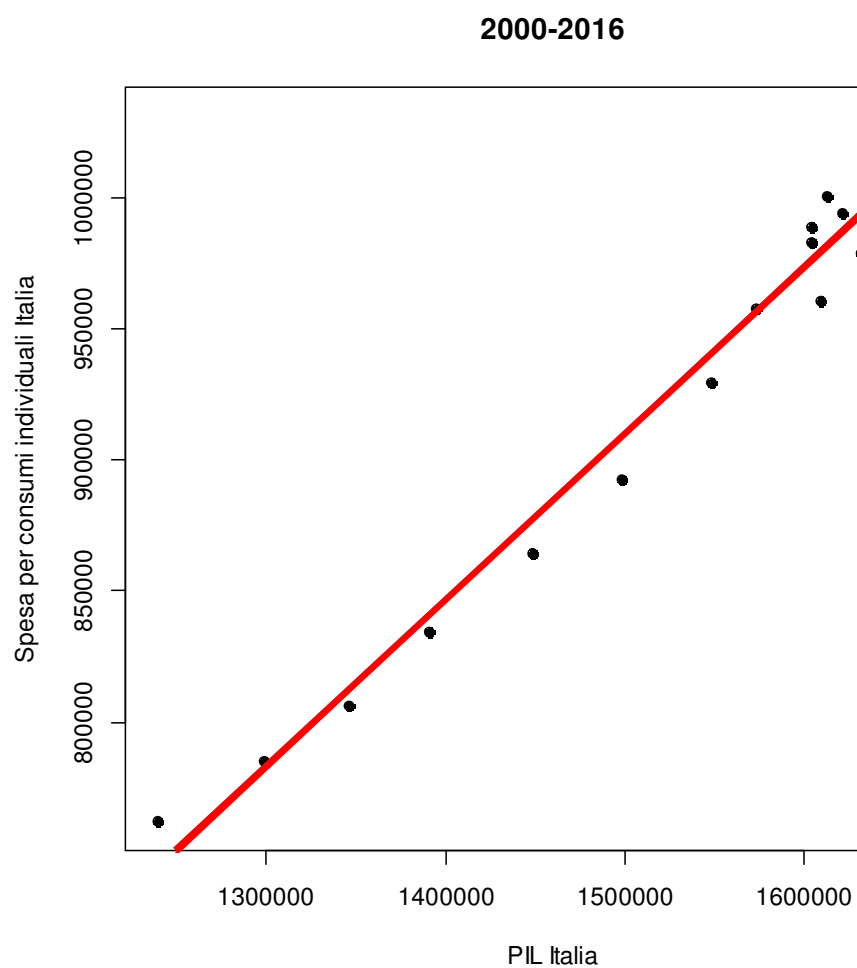


```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  896307.8   63146.5  14.194 3.13e-05 ***
CFC          1087.0     634.9   1.712   0.148
---
```

2010-2016



La retta di regressione



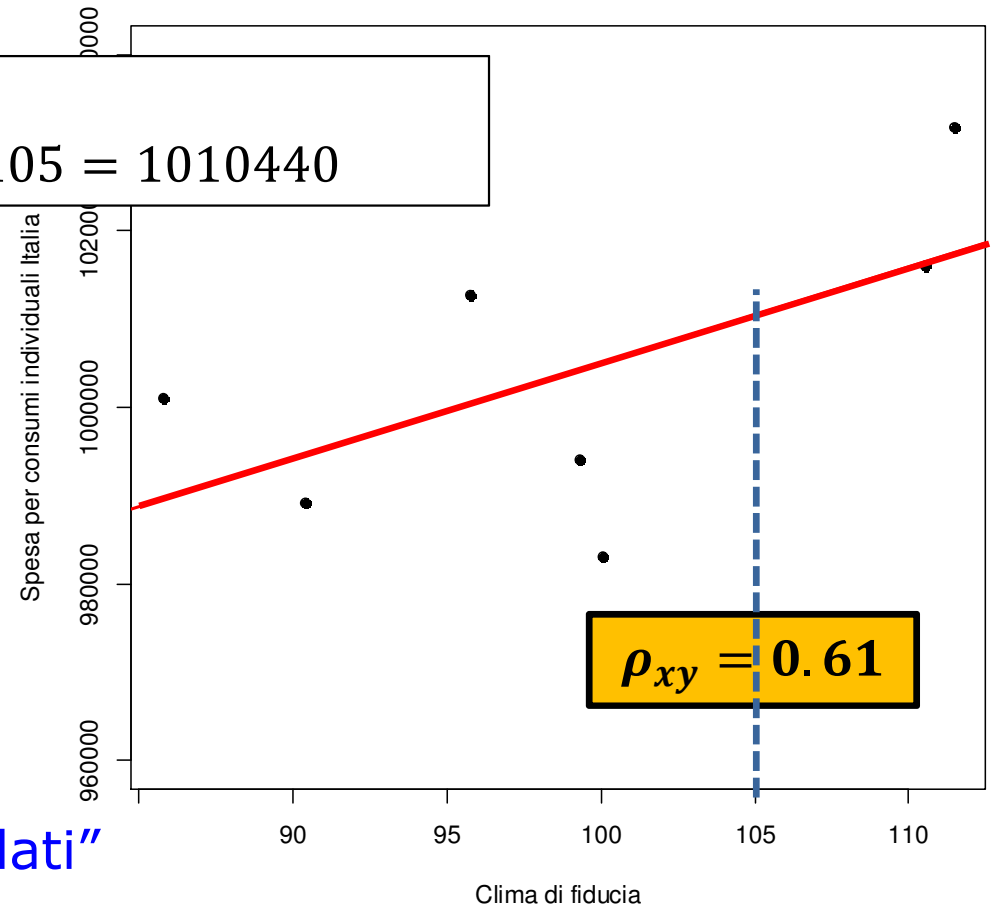
La retta di regressione

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 896307.8    63146.5  14.194 3.13e-05 ***
CFC          1087.0      634.9    1.712  0.148
---

```

2010-2016

$$CFC = 105 \Rightarrow$$
$$Spesa = 896307.760 + 1086.972 \times 105 = 1010440$$



La previsione è attendibile?

solo se "il modello si adatta bene ai dati"

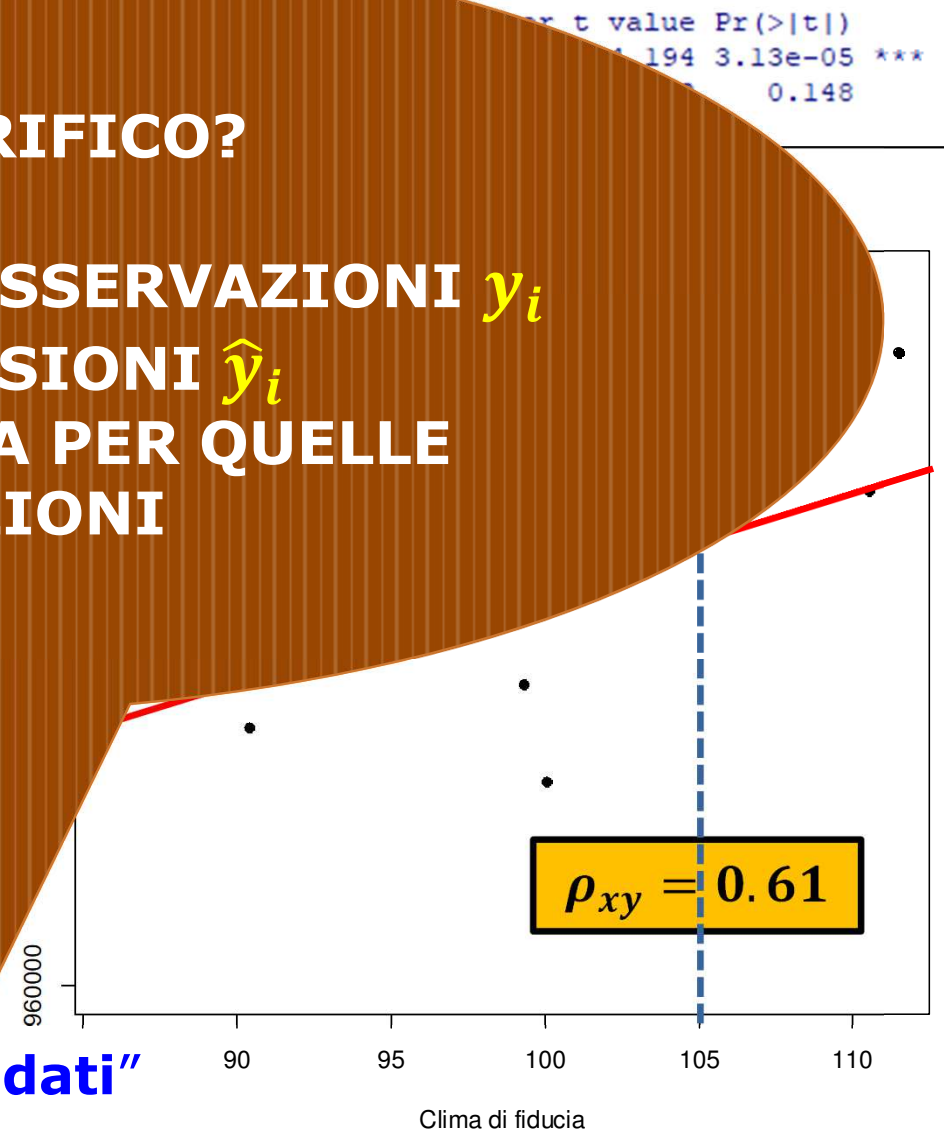
La retta di regressione

COME LO VERIFICO?

CONFRONTANDO LE OSSERVAZIONI y_i
CON LE PREVISIONI \hat{y}_i
CHE IL MODELLO FA PER QUELLE
OSSERVAZIONI

La previsione è
attendibile?

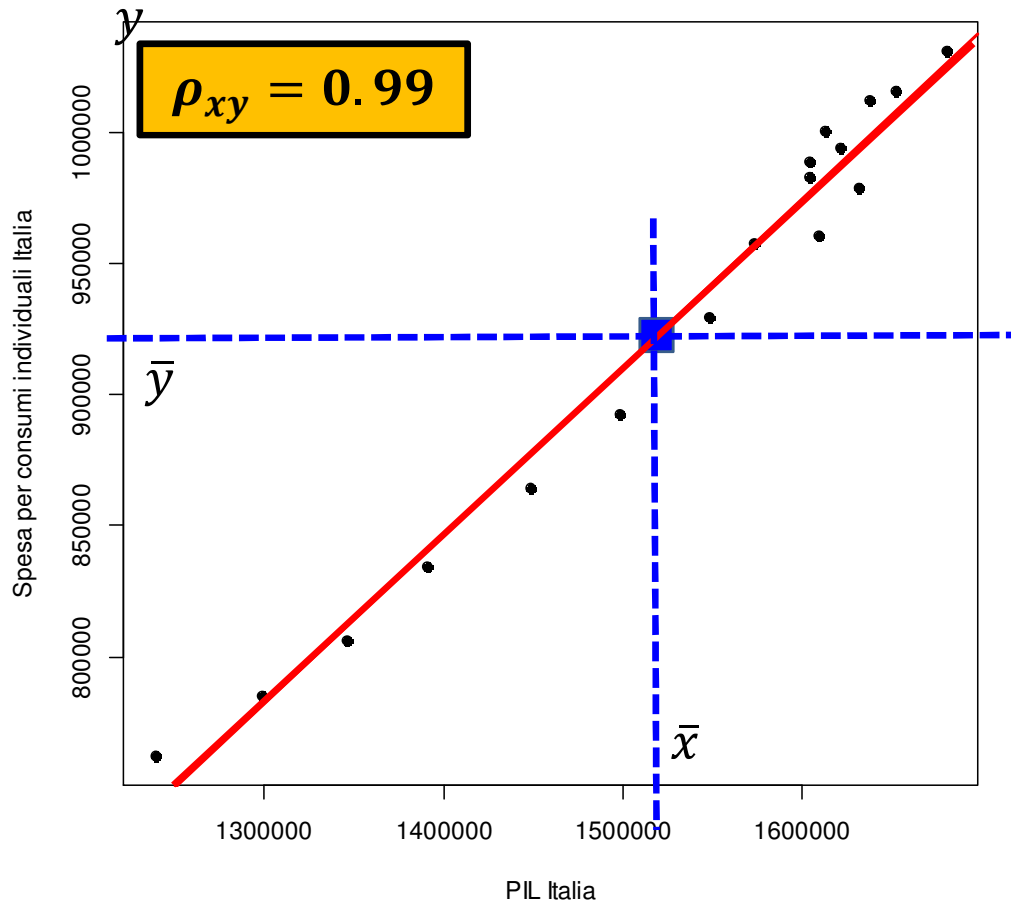
solo se "il modello si adatta bene ai dati"



La varianza spiegata dalla retta

$$\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}$$

2000-2016

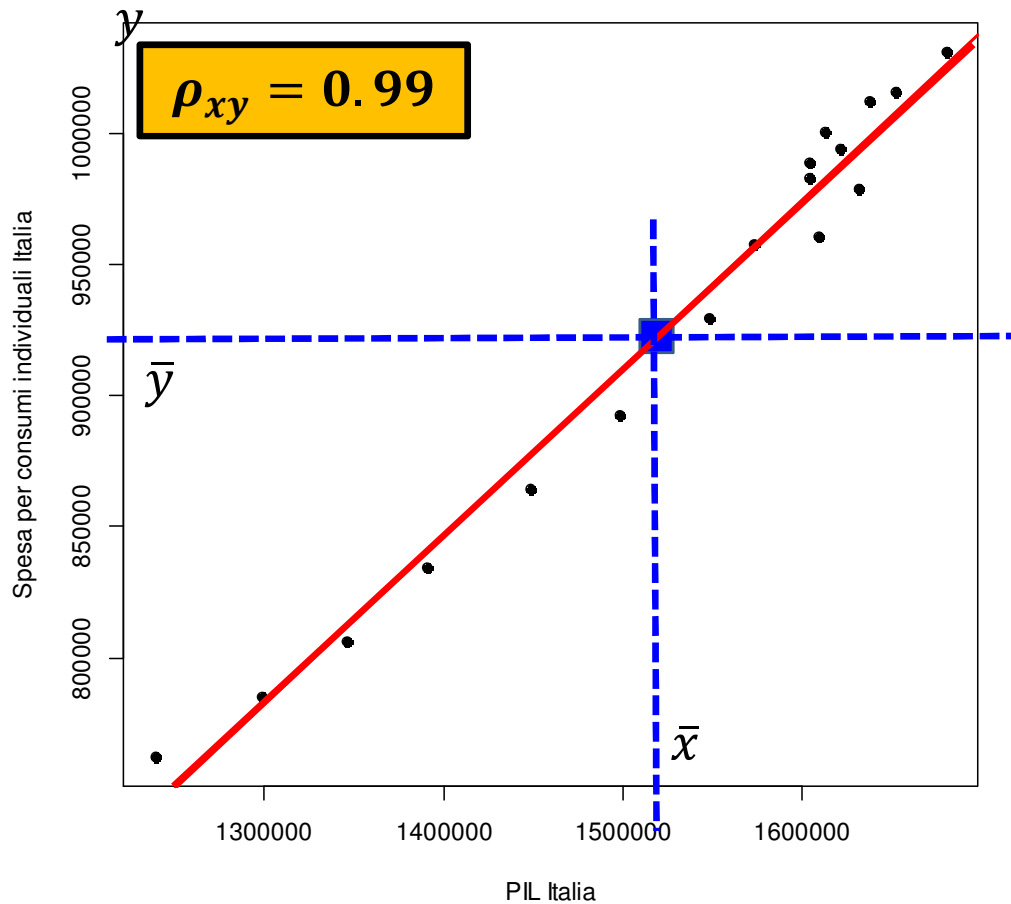


$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2 \approx \sigma_{\hat{y}}^2 = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$$

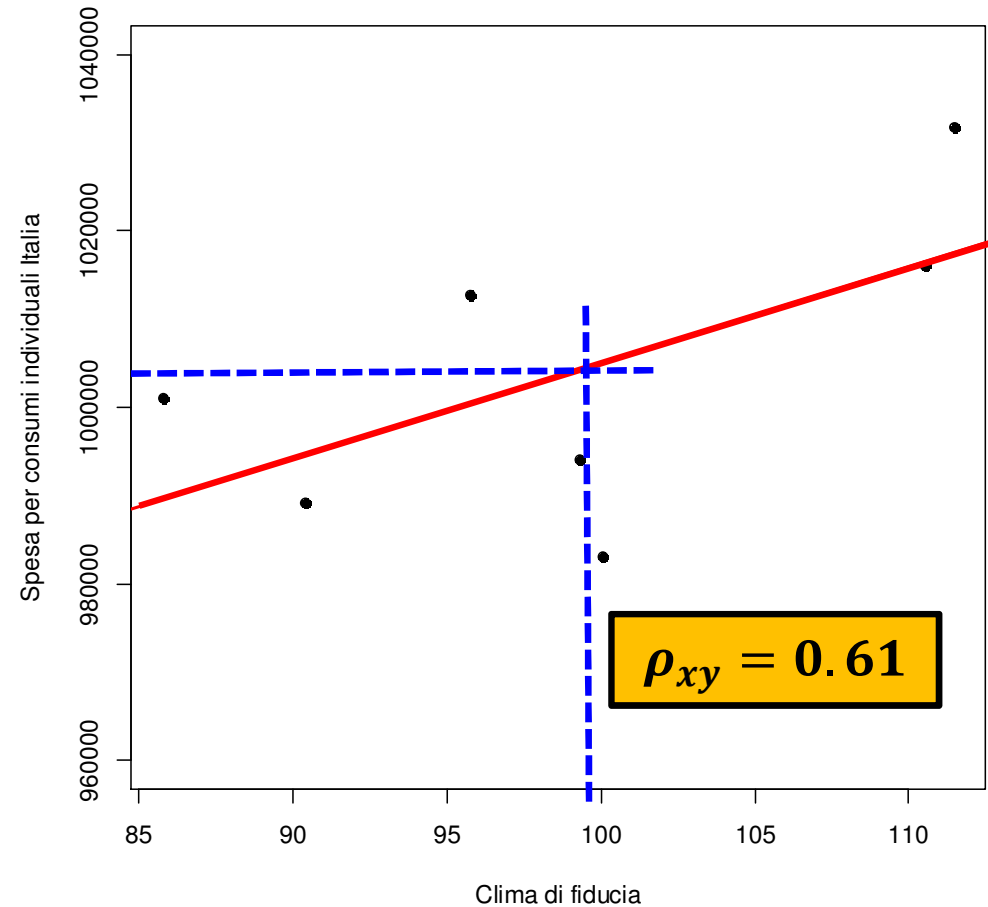
La varianza spiegata dalla retta

$$\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}$$

2000-2016



2010-2016



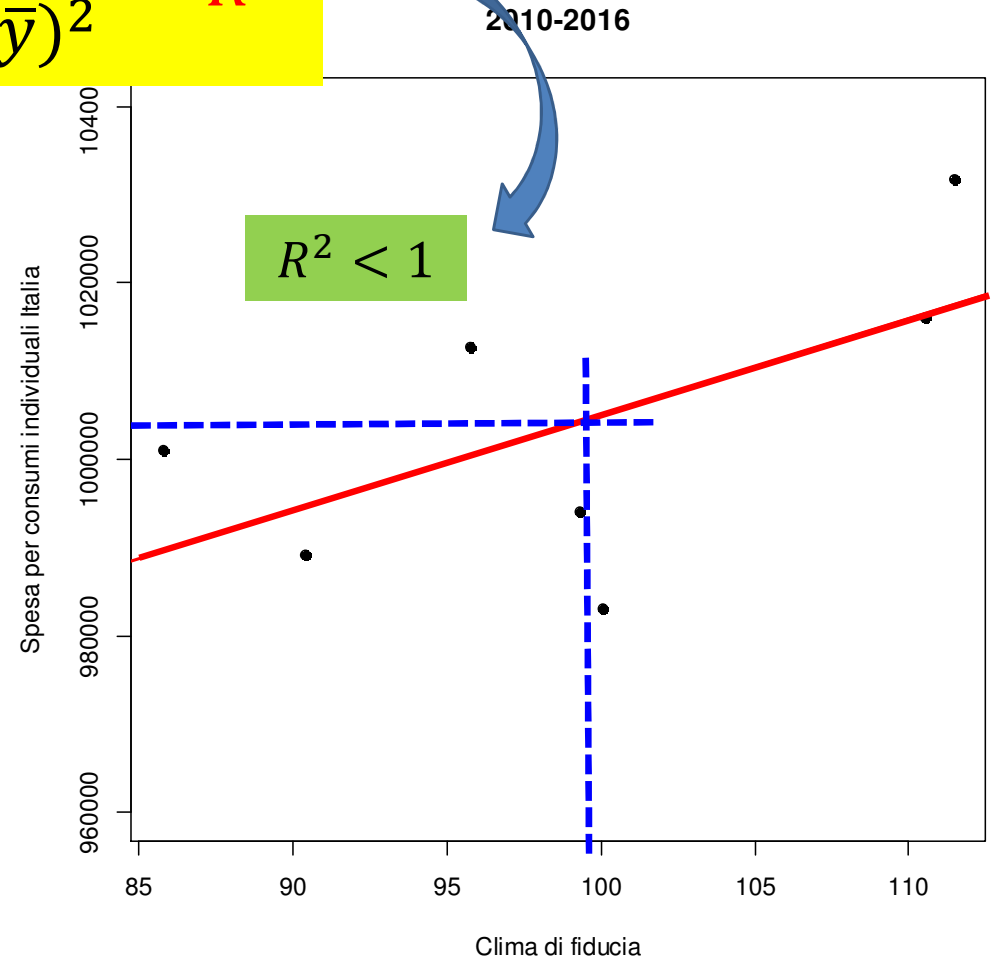
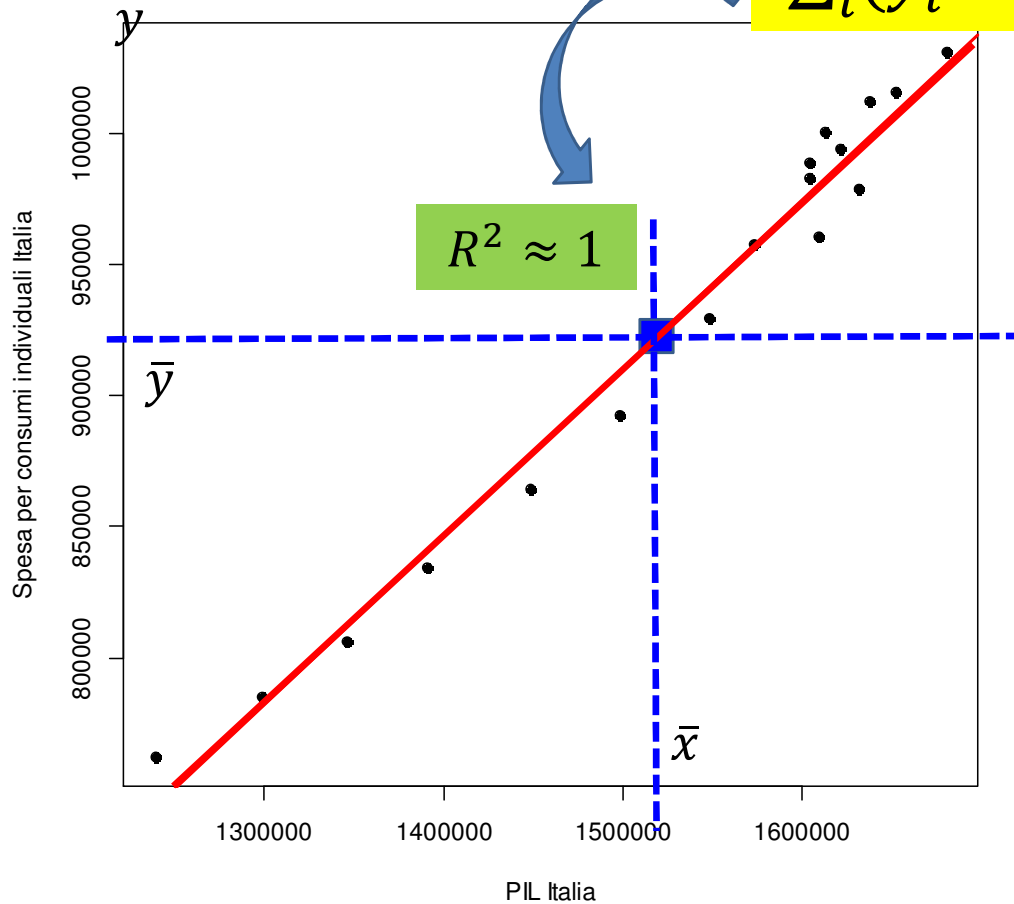
$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2 \approx \sigma_{\hat{y}}^2 = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$$

e qui???

La varianza spiegata dalla retta

$$\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}$$

$$\frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = R^2$$



$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2 \approx \sigma_{\hat{y}}^2 = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$$

e qui???

La bontà della regressione

$$\frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = R^2$$

coeff. di determinazione

$$0 \leq R^2 \leq 1$$

$$R^2 = \rho_{xy}^2$$



$$R^2 > 0.7 \Leftrightarrow \rho_{xy} > 0.837 \text{ o } \rho_{xy} < -0.837$$

(per tendenze
crescenti)

(per tendenze
decrescenti)

Analisi della Varianza

Fonte di variabilità	Gradi di libertà (gl)	SS (Sum of Square)	Mean Square (SS/gl)
Retta di regressione	1	$\sum_i (\hat{y}_i - \bar{y})^2$	
Attorno alla retta	$n - 2$	$\sum_i (y_i - \hat{y}_i)^2$	$\frac{1}{n - 2} \sum_{i=1}^n e_i^2$
Totale	$n - 1$	$\sum_i (y_i - \bar{y})^2$	

varianza **spiegata**

varianza **totale**

num. di parametri stimati (a e b)

La bontà della regressione

Per fare un buon modello lineare serve:

- ✓ una **correlazione alta**, che dice che i dati stanno vicini alla retta
- ✓ alcune **ipotesi** che dicano che il meccanismo che genera i dati è (ragionevolmente) lineare