STATISTICA A – K (63 ore)

Marco Riani

mriani@unipr.it

http://www.riani.it

Richiami sulla regressione

MODELLO DI REGRESSIONE

- $y_i = a + bx_i + e_i$ $\qquad\qquad i = 1, \ldots, n$

dove:

- $a + bx_i$ *rappresenta una retta:*
- a = *ordinata all'origine* $\rightarrow$ *intercetta*
- b = *coeff. angolare* $\rightarrow$ *coeff. di regressione*
- e_i *è un termine di errore (accidentale)*

RETTA DI REGRESSIONE

$$\hat{y}_i = a + bx_i$$

- $i = 1, \ldots, n$

$\hat{y}_i$ = valore *teorico* (valore *stimato*) di $y_i \rightarrow$ funzione *lineare* di $i = 1, \ldots, n$

Residui

$$e_i = y_i - \hat{y}_i$$

Come si calcolano i parametri *a* e *b*?

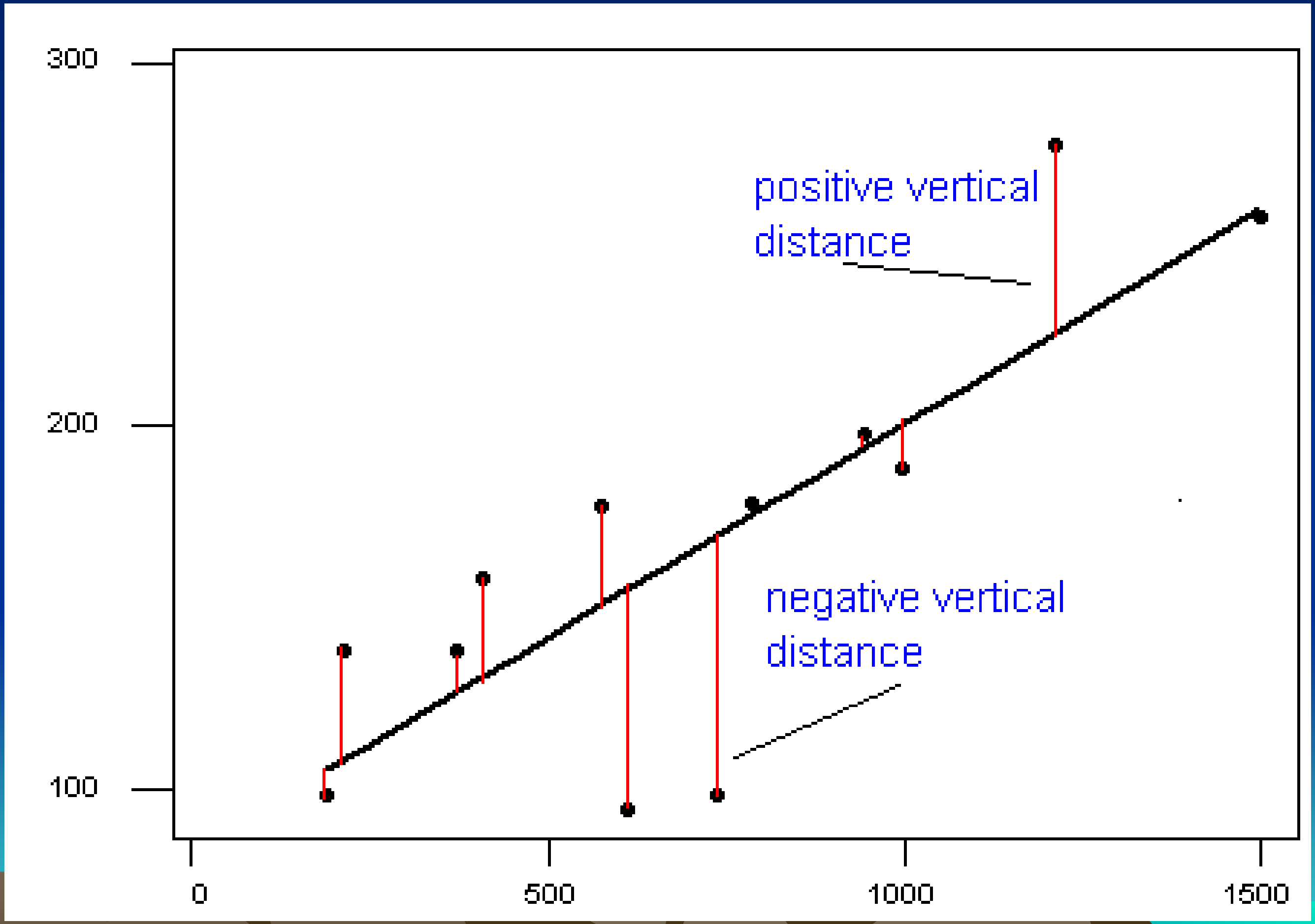
- **METODO DEI MINIMI QUADRATI**

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min$$

Le incognite sono i parametri della retta

$$\hat{y}_i = a + bx_i$$

Visualizzazione grafica dei residui



Sistema di equazioni normali

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} x_i e_i = 0$$

Formule per il calcolo di a e b

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} = \bar{y} - b\bar{x}$$

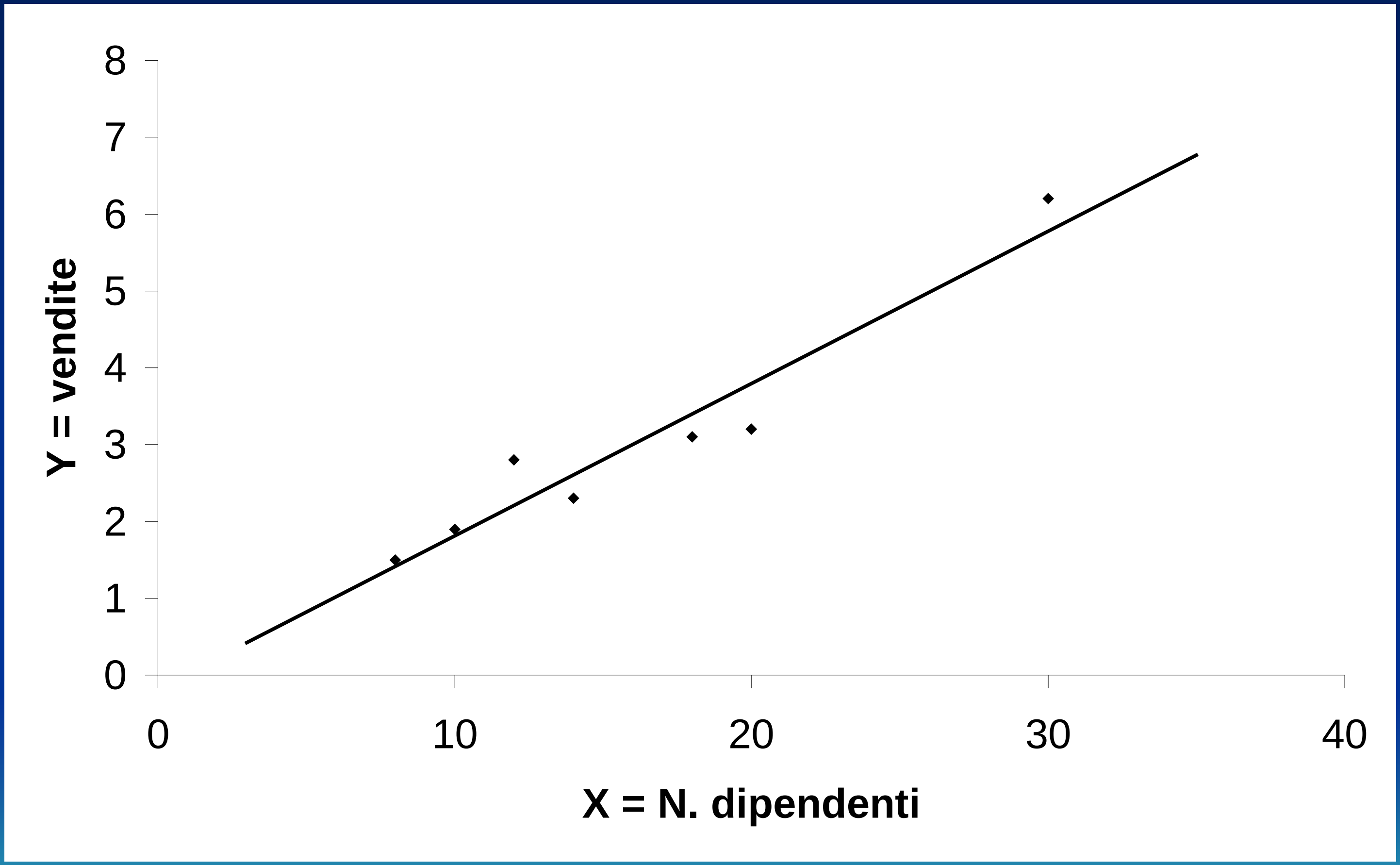
$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

a e *b* sono funzioni lineari delle osservazioni y_i

ESEMPIO (7 supermercati)

	N. dipendenti (X)	Fatturato in milioni di € (Y)
A	10	1,9
B	18	3,1
C	20	3,2
D	8	1,5
E	30	6,2
F	12	2,8
G	14	2,3

Scatter con retta di regressione



- **Come variano le vendite in funzione del numero di dipendenti?**

Calcolo di a e b

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
A	10	1,9	100	3,61	19
B	18	3,1	324	9,61	55,8
C	20	3,2	400	10,24	64
D	8	1,5	…	…	…
E	30	6,2	…	…	…
F	12	2,8	…	…	…
G	14	2,3	…	…	…
Tot.	112	21	2128	77,28	402,6

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a = \frac{21 \cdot 2.128 - 112 \cdot 402,6}{7 \cdot 2.128 - 112^2} = -\frac{403,2}{2.352} = -0,17$$

Calcolo di a e b

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
A	10	1,9	100	3,61	19
B	18	3,1	324	9,61	55,8
C	20	3,2	400	10,24	64
D	8	1,5	…	…	…
E	30	6,2	…	…	…
F	12	2,8	…	…	…
G	14	2,3	…	…	…
Tot.	112	21	2128	77,28	402,6

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{7 \cdot 402{,}6 - 112 \cdot 21}{7 \cdot 2.128 - 112^2} = \frac{466{,}2}{2.352} = 0{,}198$$

BONTA' DI ADATTAMENTO

- **Retta di regressione:** $\hat{y}_i = a + bx_i$

DEVIANZA TOTALE

$$DEV(Y) = \sum_{i=1}^{n} (y_i - M_y)^2$$

DEVIANZA DI REGRESSIONE

$$DEV(\hat{Y}) = \sum_{i=1}^{n} (\hat{y}_i - M_y)^2$$

DEVIANZA RESIDUA

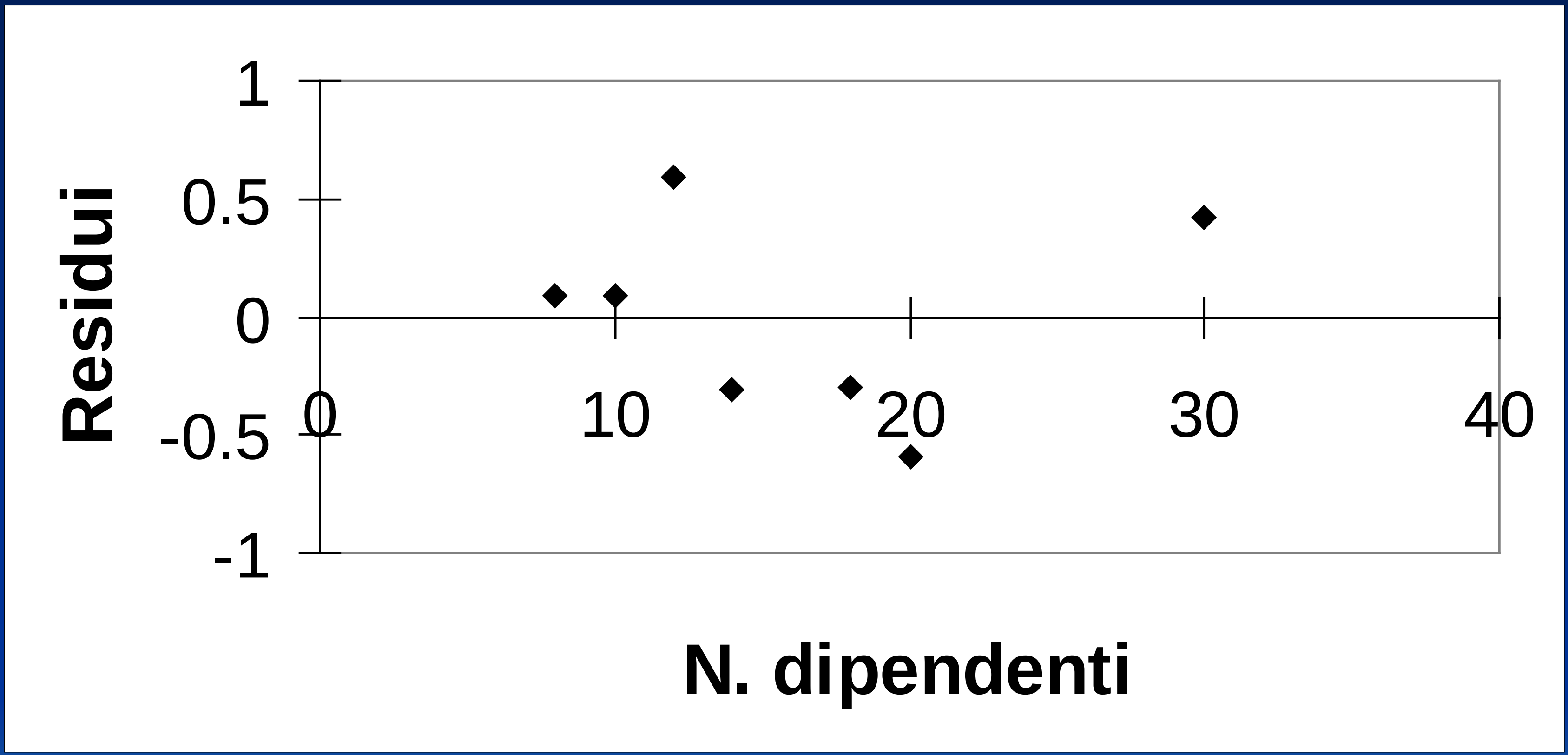
$$DEV(E) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

Indice di determinazione lineare (R^2)

$$\delta = \frac{DEV(\hat{Y})}{DEV(Y)} = 1 - \frac{DEV(E)}{DEV(Y)}$$

- $\delta = 1$ se $\sum (y_i - \hat{y}_i)^2 = 0$
- $\delta = 0$ se $\sum (\hat{y}_i - M_y)^2 = 0$

Grafico dei residui



- **Modello soddisfacente: distribuzione casuale dei residui → componente erratica**

ESTRAPOLAZIONE

- **Si tenta di valutare in maniera attendibile il valore che assumerà la variabile dipendente in corrispondenza di un valore noto della variabile esplicativa.**
- ***CONDIZIONI***
  - **Validità della retta di regressione (δ prossimo ad 1)**
  - **valore noto della variabile esplicativa non lontano dai valori utilizzati nel calcolo della retta**

Regressione inferenziale

Introduzione agli elementi aleatori

Introduzione agli elementi aleatori

	N. dipendent i (X)	Vendite in milioni di € (Y)
A	10	1,9
B	18	3,1
C	20	3,2
D	8	1,5
E	30	6,2
F	12	2,8
G	14	2,3

	Prezzi in Euro (x)	Vendite (Y)
A	1.55	410
B	1.60	380
C	1.65	350
D	1.60	400
E	1.50	440
F	1.65	380
G	1.45	450
H	1.50	420

Introduzione agli elementi aleatori

- Le vendite sono dovute in parte ai prezzi e in parte a fattori di natura aleatoria e perciò sono esse stesse delle v.c.

- Al contrario i dipendenti e/o i prezzi non sono v.c. poiché sono del tutto prevedibili dalla compagnia che li stabilisce

Introduzione agli elementi aleatori

- Una successione di valori fissi
- $x_1, x_2, \ldots x_n$
- a cui sono associate n v.c. indipendenti
- $Y_1, Y_2, \ldots Y_n$
- Il punto cruciale consiste nel descrivere in modo appropriato tali v.c.
- $E(Y_i)$? $var(Y_i)$? Distribuzione di Y_i?

Assunzioni su Y_i

- Tutte le osservazioni sono caratterizzate dallo stesso grado di incertezza
- $var(Y_i) = \sigma^2$ $\qquad$ $i=1, 2, \ldots, n$
- σ^2 è un parametro incognito da stimare
- $cov(Y_i, Y_j)=0$ $\qquad$ $i \neq j$

Assunzioni su Y_i

- $E(Y_i) = \mu_i$ $\quad\quad$ $i=1, 2, \ldots, n$
- i valori osservati della variabili dipendente provengono da *n* distribuzioni di probabilità con medie incognite

- Ip. le medie delle distribuzioni variano linearmente con la variabili indipendente
- $\mu_i = E(Y_i) = \alpha+\beta\, x_i$

Assunzioni su Y_i (continua)

- Ip: $\mu_i = E(Y_i) = \alpha+\beta x_i$
- Questa ipotesi equivale ad affermare che i punti

  $(x_1, \mu_1), (x_2, \mu_2), \dots, (x_n, \mu_n)$

  stiano tutti su una retta con parametri α, β
- Oss: questa assunzione non implica che tutti i punti (x_i, y_i) stiano sulla retta ma che i valori medi delle distribuzioni da cui i punti provengono verificano l'equazione della retta

Interpretazione di α e β

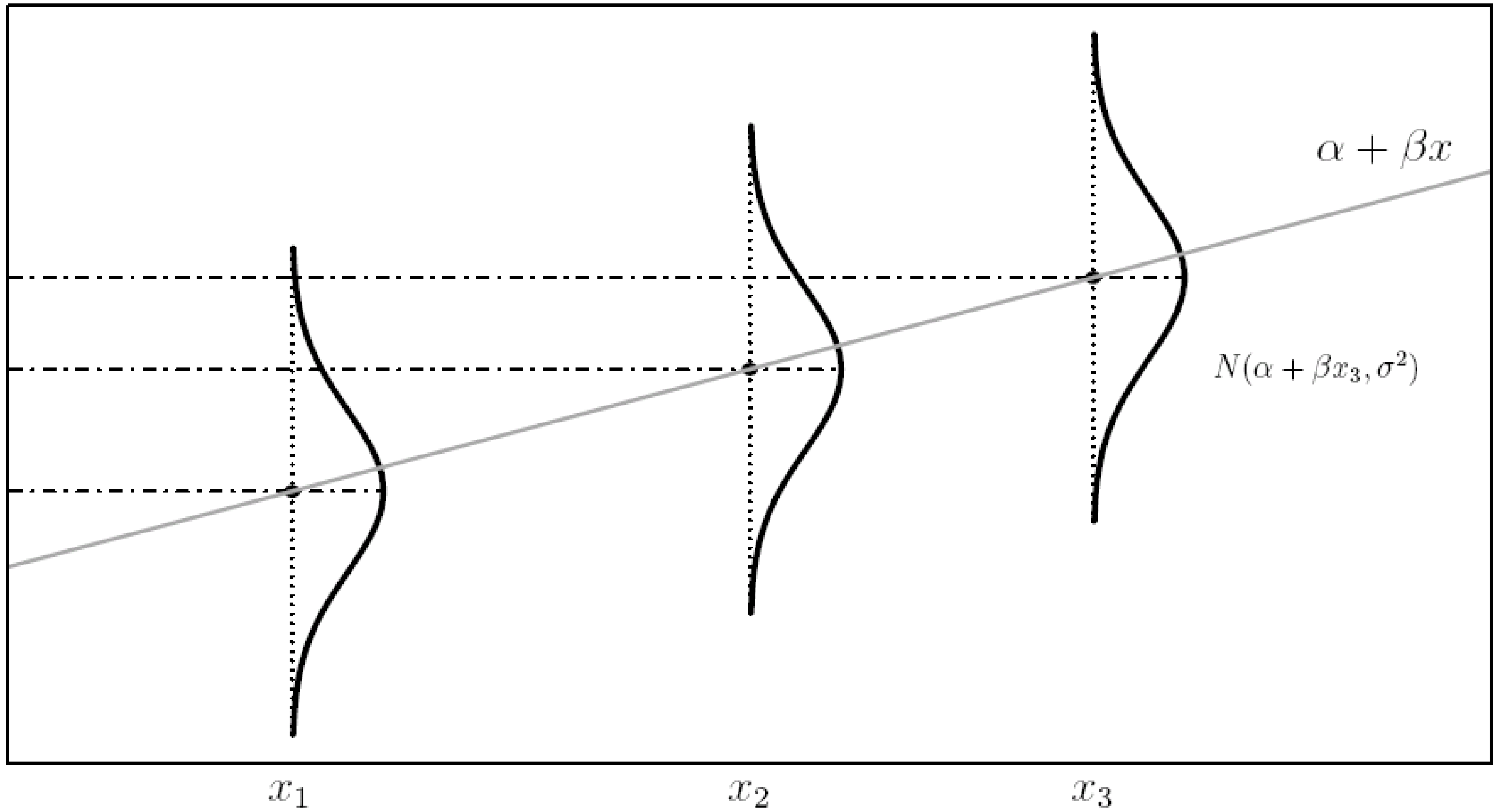
- I parametri α e β rappresentano l'intercetta ed il coeff. angolare della retta sulla quale giacciono le medie incognite delle distribuzioni $Y_1, \ldots, Y_n$

Interpretazione di α e β

$E(Y_3)=\alpha+\beta x_3$

$E(Y_2)=\alpha+\beta x_2$

$E(Y_1)=\alpha+\beta x_1$



Osservazione sulla notazione

- In queste slide utilizziamo la notazione
- α per indicare l'intercetta nella popolazione β per indicare la pendenza nella popolazione

$\hat{\alpha}$ per indicare lo stimatore e la stima di α

$\hat{\beta}$ per indicare lo stimatore e la stima di β

Notazione utilizzata nel testo

- Nel testo si usa la notazione β_0 per indicare l'intercetta nella popolazione B_0 per indicare lo stimatore dell'intercetta e b_0 per indicare la stima dell'intercetta

- Nel testo si usa la notazione β_1 per indicare la pendenza nella popolazione B_1 per indicare lo stimatore della pendenza e b_1 per indicare la stima della pendenza

Osservazione

- Dato il modello di regressione
- $Y_i = \alpha + \beta x_i + \varepsilon_i$

- L'ip: $\mu_i = E(Y_i) = \alpha + \beta x_i$

- equivale ad affermare che

- $E(\varepsilon_i)=0$

Stima dei parametri

- I parametri da stimare sono
- $\alpha, \beta, \mu_1, \mu_2, \ldots, \mu_n, \sigma^2$

- La conoscenza di α, β consente di ricostruire tutte le medie incognite $\mu_1, \mu_2, \ldots, \mu_n$

- $\mu_i = E(Y_i) = \alpha + \beta x_i$

Riepilogo ipotesi

- $Y_i = \alpha + \beta x_i + \varepsilon_i$
- $\mu_i = E(Y_i) = \alpha + \beta x_i$
- $var(Y_i) = \sigma^2$
- $cov(Y_i, Y_j) = 0$
- $var(\varepsilon_i) = ?$
- $var(\varepsilon_i) = \sigma^2$ (Dato che var(Y+k)=var(Y) con k costante)
- $cov(\varepsilon_i, \varepsilon_j) = ?$
- $cov(\varepsilon_i, \varepsilon_j) = 0$ (Dato che $cov(Y_i+k, Y_j+c) = cov(Y_i, Y_j)$ con k e c costanti)

Stime di α e β

- Pensando di ripetere più volte l'esperimento che ha generato le osservazioni $y_1, \ldots, y_n$, per valori fissi di $x_1, \ldots, x_n$ si ottiene una distribuzione campionaria di valori

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x}$$

$$\hat{\beta} = \frac{\sum (x_i - \overline{x}) Y_i}{\sum (x_i - \overline{x})^2}$$

Coeff. di regressione campionari e nella popolazione

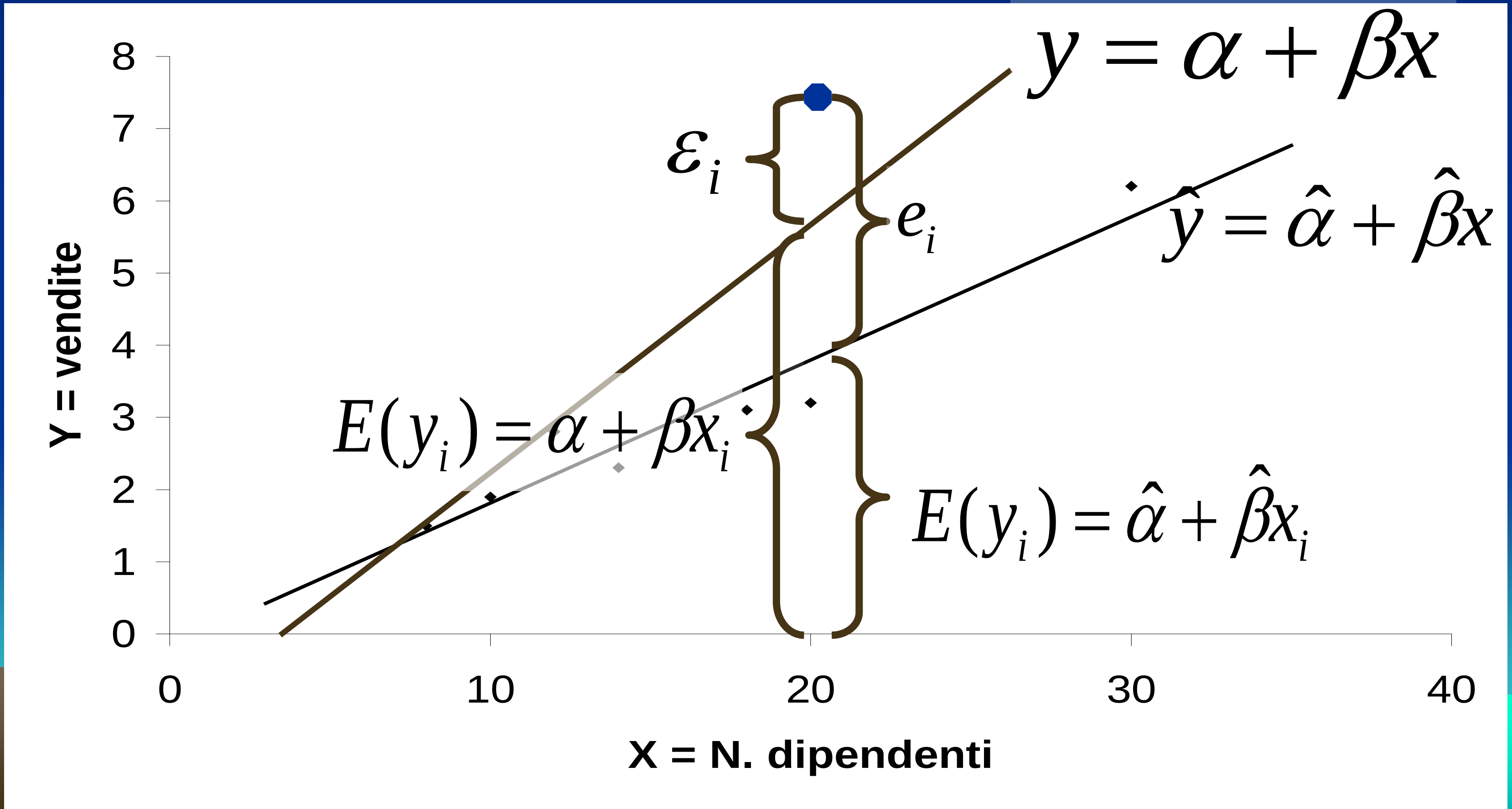
$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$y_i = \hat{\alpha} + \hat{\beta} x_i + e_i$$

Coeff. di regressione campionari e nella popolazione

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$y_i = \hat{\alpha} + \hat{\beta} x_i + e_i$$



Stima di σ^2

- $var(Y_i)=var(\varepsilon_i)=\sigma^2$= dispersione verticale attorno alla retta che unisce i valori medi delle popolazioni
- Dato che $\sigma^2=E(\varepsilon_i^2)-[E(\varepsilon_i)]^2= E(\varepsilon_i^2)$
- Dato che e_i è una stima di ε_i sembra naturale utilizzare come stimatore di σ^2 la seguente espressione

$$s^2 = \frac{\sum (y_i - \hat{\mu}_i)^2}{n-2} = \frac{\sum e_i^2}{n-2}$$

$$E(s^2) = \sigma^2$$

Stima di σ

- s="errore standard nella stima di Y"

$$s = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Ip. aggiuntiva

- Le distribuzioni Y_i sono normali
- y_1 è una realizzazione di $Y_1 \sim N(\mu_1, \sigma^2)$
- y_2 è una realizzazione di $Y_2 \sim N(\mu_2, \sigma^2)$
- …
- y_n è una realizzazione di $Y_n \sim N(\mu_n, \sigma^2)$

- $Y_1, Y_2, \ldots, Y_n$ sono indipendenti

Obiettivo

Costruire intervalli di confidenza e test di verifica d’ipotesi sul coeff. angolare

$$\hat{\beta}$$

Studio della distribuzione di $\hat{\beta}$

$$\hat{\beta} = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}$$

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$y_i = \hat{\alpha} + \hat{\beta} x_i + e_i$$

Studio della distribuzione di $\hat{\beta}$

$$\hat{\beta} = \frac{\sum (x_i - \overline{x}) Y_i}{\sum (x_i - \overline{x})^2}$$

$$E(\hat{\beta}) = ?$$

$$E(\hat{\beta}) = \beta$$

$$\operatorname{var}(\hat{\beta}) = ?$$

$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2}{\sum (x_i - \overline{x})^2}$$

Varianza di beta cappello

$$\mathrm{var}(\hat{\beta}) = \mathrm{var}\left(\frac{\sum_{i=1}^{n} (x_i - \overline{x}) Y_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right)$$

$$\mathrm{var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \right]^2} \mathrm{var}\left(\sum_{i=1}^{n} (x_i - \overline{x}) Y_i \right)$$

$$\mathrm{var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \right]^2} \left(\sum_{i=1}^{n} \mathrm{var}(x_i - \overline{x}) Y_i \right)$$

Varianza di beta cappello

$$\operatorname{var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^2}\left(\sum_{i=1}^{n}\operatorname{var}(x_i - \overline{x})Y_i\right)$$

$$\operatorname{var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^2}\left(\sum_{i=1}^{n}(x_i - \overline{x})^2 \operatorname{var}Y_i\right)$$

$$\operatorname{var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^2}\left(\sum_{i=1}^{n}(x_i - \overline{x})^2 \sigma^2\right)$$

$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Al posto di σ^2 sostituiamo il suo stimatore

$$Stima(\mathrm{var}(\hat{\beta})) = s^2(\hat{\beta}) = \frac{s^2}{\sum (x_i - \bar{x})^2}$$

- La radice quadrata della stima della varianza di uno stimatore è l'errore standard (standard error, SE) dello stimatore

$$s_{\hat{\beta}} = SE(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Interpretazione dello standard error di beta cappello

- Rappresenta l'errore quadratico medio che si commette quando si stima il coefficiente di regressione con le formule dei minimi quadrati

Studio della distribuzione di $\hat{\alpha}$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$$

$$E(\hat{\alpha}) = ?$$

$$E(\hat{\alpha}) = \alpha$$

$$\mathrm{var}(\hat{\alpha}) = ?$$

$$\mathrm{var}(\hat{\alpha}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]$$

Esercizio: nell'esempio dei 7 supermercati calcolare lo standard error di beta cappello e alpha cappello

$$s_{\hat{\beta}} = SE(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = 0.025$$

$$s_{\hat{\alpha}} = SE(\hat{\alpha}) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} = 0.44$$

Costruzione di intervalli di confidenza dei parametri

Punto di partenza

$$\hat{\beta} = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}$$

- beta cappello è funzione lineare di v.c. Normali e quindi è distribuito in maniera normale

Lo scostamento standardizzato di beta capello ha una distribuzione N(0,1)

$$\Pr\left(-Z_{\alpha/2} < \frac{\hat{\beta} - E(\hat{\beta})}{\sqrt{\operatorname{var}(\hat{\beta})}} < Z_{\alpha/2}\right) = 1 - \alpha$$

- Se il livello di confidenza $1\text{-}\alpha=0.95$

$$\Pr(-1.96 < \frac{\hat{\beta} - E(\hat{\beta})}{\sqrt{\operatorname{var}(\hat{\beta})}} < 1.96) = 0.95$$

$$\Pr\left(-1.96 < \frac{\hat{\beta} - E(\hat{\beta})}{\sqrt{\operatorname{var}(\hat{\beta})}} < 1.96\right) = 0.95$$

$$\Pr\left(-1.96 < \frac{\hat{\beta} - \beta}{\sqrt{\dfrac{\sigma^2}{\sum (x_i - \bar{x})^2}}} < 1.96 \right) = 0.95$$

Problema: σ^2 è ignoto (occorre sostituire il suo stimatore s^2)

Sostituendo al posto di σ^2 il suo stimatore

$$\frac{\hat{\beta}-\beta}{\sqrt{\dfrac{\sigma^2}{\sum(x_i-\bar{x})^2}}} \sim N(0,1)$$

$$\frac{\hat{\beta}-\beta}{\sqrt{\dfrac{s^2}{\sum(x_i-\bar{x})^2}}} \sim T(n-2)$$

$$P\left\{-t(\alpha/2) \le \frac{\hat{\beta}-\beta}{\sqrt{\dfrac{s^2}{\sum(x_i-\bar{x})^2}}} \le t(\alpha/2)\right\} = 1-\alpha$$

$$P\left\{-t(\alpha/2) \le \frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \le t(\alpha/2)\right\} = 1-\alpha$$

Costruzione di un intervallo di confidenza per il coeff. angolare

$$P\left\{\hat{\beta} - t(\alpha/2)SE(\hat{\beta}) \le \beta_1 \le \hat{\beta} + t(\alpha/2)SE(\hat{\beta})\right\} = 1 - \alpha$$

- Dove $t(\alpha/2)$ è valore critico associato alla distribuzione T di student con n-2 gradi di libertà (T(n-2)) tale che

$F(-t(\alpha/2))= \quad P(T<-t(\alpha/2))=\alpha/2$

Costruzione di intervalli di confidenza dei parametri

Esercizio: nell'esempio dei 7 supermercati costruire un intervallo di confidenza al 95% per β ed interpretare i risultati ottenuti

ESEMPIO (7 supermercati)

	N. dipendenti (X)	Fatturato in milioni di € (Y)
A	10	1,9
B	18	3,1
C	20	3,2
D	8	1,5
E	30	6,2
F	12	2,8
G	14	2,3

$$\hat{\alpha} = \frac{21 \cdot 2.128 - 112 \cdot 402{,}6}{7 \cdot 2.128 - 112^2} = -\frac{403{,}2}{2.352} = -0{,}17$$

$$\hat{\beta} = \frac{7 \cdot 402{,}6 - 112 \cdot 21}{7 \cdot 2.128 - 112^2} = \frac{466{,}2}{2.352} = 0{,}198$$

Costruzione di un intervallo di confidenza al 95% per il coeff. angolare

$$P\left\{\hat{\beta} - t(\alpha/2)SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t(\alpha/2)SE(\hat{\beta})\right\} = 1 - \alpha$$

$$\hat{\beta} = 0{,}198$$

$$s_{\hat{\beta}} = SE(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \overline{x})^2}} = 0.025$$

- t(0.025)=+2.5706 (=INV.T(0.05;5))
- (Oss: Pr.(T>2.5706)=0.025)

Costruzione di un intervallo di confidenza al 95% per il coeff. angolare

$$P\left\{\hat{\beta} - t(\alpha/2)SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t(\alpha/2)SE(\hat{\beta})\right\} = 1 - \alpha$$

- Pr(0.198-2.5706×0.0253<β< 0.198+2.5706× 0.0253)=0.95
- Pr(0.133<β< 0.263)=0.95

Interpretazione

- L'intervallo di confidenza del coefficiente di regressione, con probabilità uguale a 0.95, va da 0,133 a 0,263.
- Questo significa che nell'universo di riferimento, all'aumento di un dipendente può corrispondere un aumento delle vendite compreso tra 133 mila Euro e 263 mila Euro circa (con probabilità del 95%).
- Oss: l'intervallo è piuttosto ampio e questo dipende dalla ridotta numerosità campionaria (solo 7 supermercati).

Intervallo di confidenza per l’intercetta

Costruzione di un intervallo di confidenza al 95% per l'intercetta

$$P\left\{\hat{\alpha} - 2.571 SE(\hat{\alpha}) \leq \alpha \leq \hat{\alpha} + 2.571 SE(\hat{\alpha})\right\} = 0.95$$

- t(0.05)=+2.5706   (=INV.T(0.05;5))
- (Oss: Pr.(T>2.5706)=0.025)

- Pr(1.31<α< 0.96)=0.95

Costruzione di test di ipotesi per

$\alpha \quad \beta \quad \sigma^2$

Dato che

$$\frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim T_{n-2}$$

Sotto H_0: β =0

$$\frac{\hat{\beta}}{SE(\hat{\beta})} \sim T_{n-2}$$

$$s_{\hat{\beta}} = SE(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = 0.025$$

Esercizio: nell'esempio dei 7 supermercati testare H_0:β=0 (H_1:$\beta \neq 0$), trovare il relativo p-value ed interpretare il risultato del test

$t_\beta = 0.1982/0.02534 = 7.82$ p-value <0.001 (dalle tavole)

DISTRIB.T(7.82;5;2)= 0.000548

Interpretazione : rifiuto decisamente l'ipotesi nulla

Esercizio: nell’esempio dei 7 supermercati testare $H_0: \alpha=0$, ($H_1: \alpha \neq 0$), trovare il relativo p-value ed interpretare il risultato del test

$t_\alpha = 0.39$ p-value = >0.1 (dalle tavole)

DISTRIB.T(0.39;5;2)=0.714

Interpretazione : non posso rifiutare l’ipotesi nulla

Funzione regr.lin

- Ordine in cui vengono restituite le statistiche aggiuntive di regressione dalla funzione di Excel REGR.LIN

	A	B	C	D	E	F	G
1	b_k	b_{k-1}	…	b_2	b_1	b_0	a
2	s_k	s_{k-1}	…	s_2	s_1	s_0	s_a
3	R2	se_y					
4	F	d_f					
5	SS_{reg}	SS_{resid}					
6							

Intervallo di confidenza delle previsioni con il metodo dei minimi quadrati

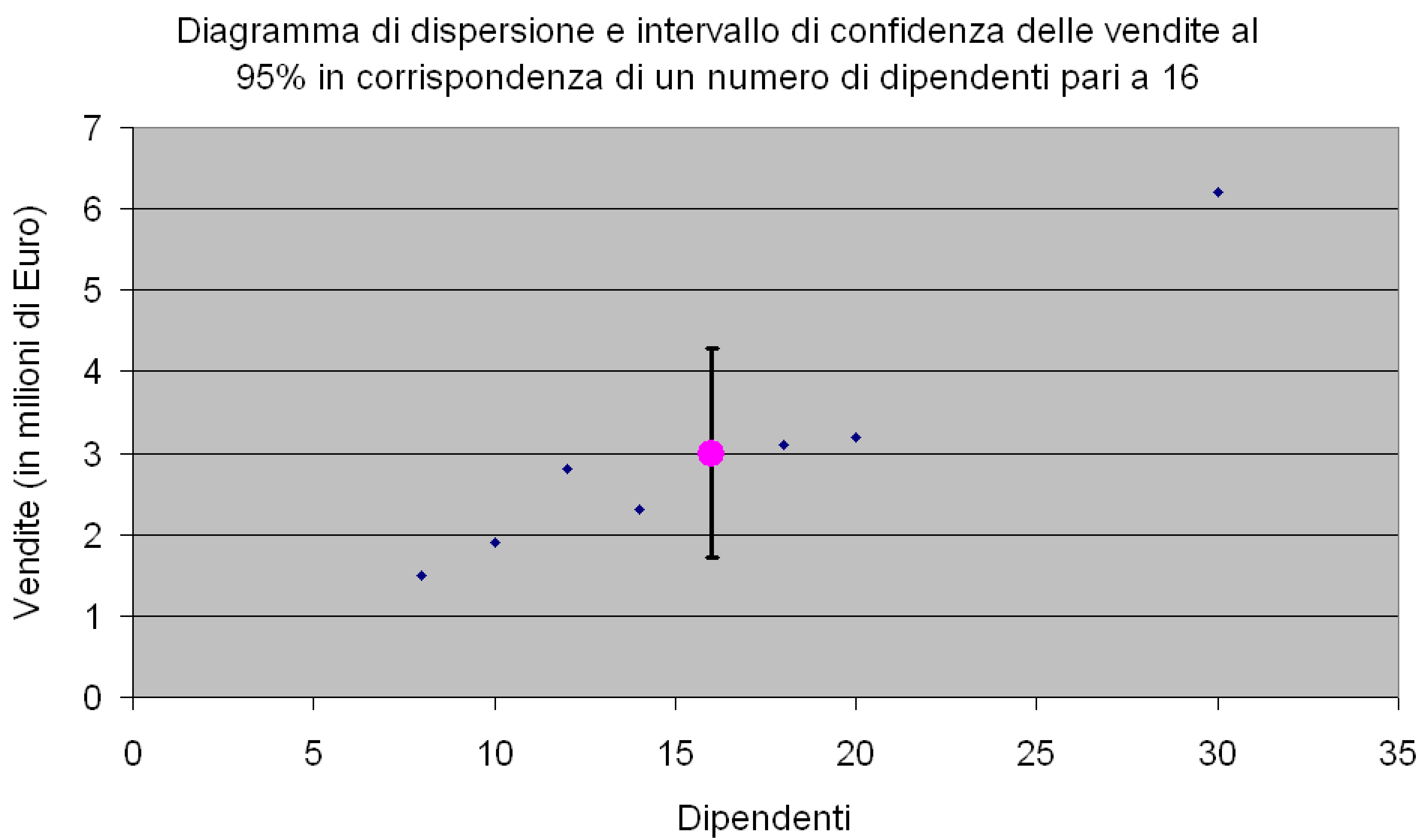


Diagramma di dispersione e intervallo di confidenza delle vendite al 95% in corrispondenza di un numero di dipendenti pari a 16
Vendite (in milioni di Euro)
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
Dipendenti

Approccio inferenziale al δ (coefficiente di determinazione lineare)

$$\delta = \frac{DEV(\hat{Y})}{DEV(Y)} = 1 - \frac{DEV(E)}{DEV(Y)}$$

- H_0: $\delta=0$ (non vi è alcuna relazione tra X e Y)

- La statistica test da utilizzare è la seguente:

$$F = \frac{\delta}{(1-\delta)/(n-2)} = \frac{DEV(\hat{Y})}{DEV(E)/(n-2)}$$

- Nel modello di regressione lineare semplice

$$F = \left[\frac{\hat{\beta}}{SE(\hat{\beta})}\right]^2$$

Le quantità necessarie per il calcolo della statistica F sono riportate nella Tabella di analisi della varianza

Modello	*Gradi di libertà*	*Somme dei quadrati*	*Medie dei quadrati*	*F*	*P-value*
Regressione	1	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 /1$	$\dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} e_i^2 /(n-2)}$	$P\{F \geq$ del valore osservato nel campione$\}$
Residuo	$n-2$	$\sum_{i=1}^{n} e_i^2$	$\sum_{i=1}^{n} e_i^2 /(n-2)$		
Totale	$n-1$	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$\sum_{i=1}^{n}(y_i - \bar{y})^2 /(n-1)$		

- v. p. 142

Tutti i precedenti calcoli sono stati implementati in Excel

- Pagina http://www.riani.it/SV/sv.htm

Indirizzo del file da scaricare

http://www.riani.it/SV/inputfiles/SV-regr1(out).xls