

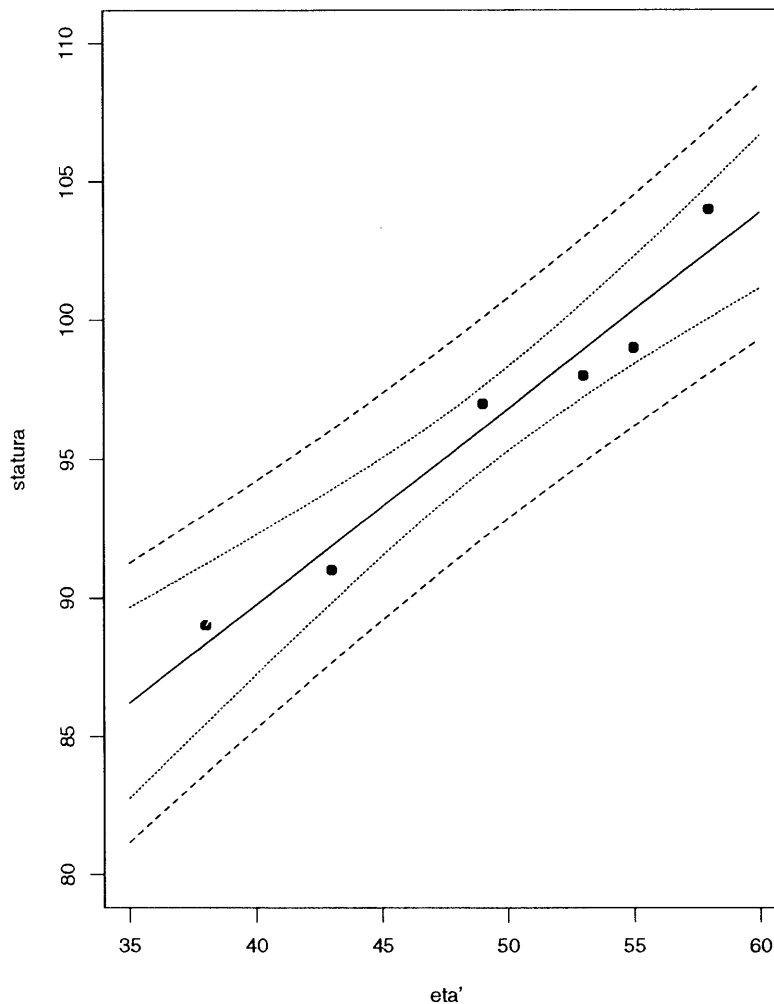


Figura 5.4: *Dati dell'Esercizio 5.13 con la retta di regressione, linea continua. La linea punteggiata e la linea tratteggiata rappresentano, rispettivamente, gli intervalli di confidenza di livello 95% per $E[Y_q]$ e gli intervalli di previsione per Y_q per ogni x_q nell'intervallo $(35,60)$. L'ampiezza di entrambi aumenta verso gli estremi dell'intervallo $(35,60)$.*

③ TROVARE I LIMITI DI CONFIDENZA AL 95% PER Y_0 QUANDO $X_0 = 12$

$X_0 = 12$ $Y_0 = 8,18$ (ottenuto sostituendo X_0 nella equazione di regressione)

Y_p = valore di Y previsto nella popolazione

$$Y_p = Y_0 \pm \frac{t}{\sqrt{N-2}} \cdot S_{yx} \sqrt{N+1 + \frac{(X_0 - \bar{X})^2}{(S_x)^2}}$$

per $\alpha = 0,05$ (test BIDIREZIONALE) $\rightarrow t_{crit.} = 2,447$

$$Y_p = 8,18 \pm \frac{2,447}{\sqrt{6}} \cdot 0,56 \sqrt{8+1 + \frac{(12-7)^2}{(4,06)^2}}$$

$Y_p = 8,18 \pm 1,81$ ossia siamo confidenti al 95% che $6,37 < Y_p < 9,99$

④ STABILIRE LA SIGNIFICATIVITÀ STATISTICA DEL COEFFICIENTE DI REGRESSIONE b_{yx} AL LIVELLO DI SIGNIFICATIVITÀ $\alpha = 0,05$

b_{yx} È STATISTICAMENTE SIGNIFICATIVO QUANDO $W \neq \pi$. IN QUESTO CASO ABBIAMO CALCOLATO $r = 0,977$.

Per $\alpha = 0,05$ r è statisticamente significativo (ossia $p \neq 0$) quando $r > r_{critico}$.

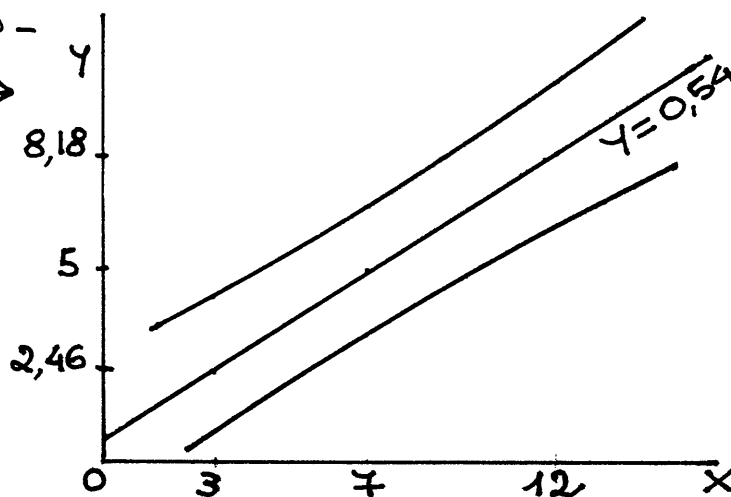
$r_{crit} = \pm 0,6215$

H_1	H_0	H_1
-0,6215	+0,6215	0,977

Quindi anche b_{yx} è statisticamente significativo.

IL TEST N°3 →
CONSENTE DI COSTRUIRE IL GRAFICO DELLA FASCE DI FIDUCIA AL 95% INTORNO ALLA RETTA DI REGRESSIONE

cf. BLALOCK, pp. 525-526 e, in particolare, fig. 18.2



H. BLALOCK, *Statistica per la ricerca sociale*, Bologna, Il Mulino, 1984, pp. 524-526

A volte può essere necessario porre un intervallo di fiducia attorno al valore di b , oppure trovare una fascia entro la quale ci si può aspettare che giaccia la vera equazione di regressione. In tutti e due i casi possiamo utilizzare la distribuzione t in modo abbastanza diretto. La stima dell'errore standard di b è data da:

$$\hat{\sigma}_{b|x} = \frac{\hat{\sigma}_{y|x}}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2}} \quad (18.3)$$

in cui, come si ricorderà,

$$\hat{\sigma}_{y|x} = \sqrt{\sum_{i=1}^N \frac{(Y_i - Y_p)^2}{N-2}}$$

Una formula di calcolo che si può dimostrare dà risultati uguali a quelli della formula precedente pur facilitando i calcoli rispetto a quelli della formula precedente, è la seguente:

$$\hat{\sigma}_{y|x} = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2 - b \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N-2}} \quad (18.4)$$

Possiamo, ora, utilizzare i calcoli già fatti per i dati sulla discriminazione riportati alla tabella 17.1, e otteniamo:

$$\hat{\sigma}_{y|x} = \sqrt{\frac{560.024 - 19,931(2.553,77)}{11}} = \sqrt{46.284} = 215,1$$

¹ E. A. Haggard, *Intraclass Correlation and the Analysis of Variance*, New York, The Dryden Press, 1958.

² L. A. Goodman e W. H. Kruskal, *Measures of Association for Cross Classifications II: Further Discussion and Reference*, in «Journal of the American Statistical Association», LIV (1959), pp. 123-163, e degli stessi autori, *Measures of Association for Cross Classifications III: Approximate Sampling Theory*, in «Journal of the American Statistical Association», LVIII (1963), pp. 310-364.

e

$$\hat{\sigma}_b = \frac{215,1}{\sqrt{128,131}} = \frac{215,1}{11,32} = 19,00$$

Se vogliamo costruire un intervallo di fiducia del 99 per cento, possiamo utilizzare direttamente la tabella della distribuzione t usando $N-2$, e cioè, nel nostro caso, 11 gradi di libertà. Otteniamo pertanto:

$$b \pm (3,106)(19,00) = 19,931 \pm 59,014$$

* Nella stima dell'equazione di regressione, abbiamo visto che la miglior stima singola («puntuale») che ne possiamo fare è data dalla retta ottenuta col metodo dei minimi quadrati. Dal momento che la quantità che vogliamo stimare non è più un singolo valore, ma piuttosto una intera retta, la nostra «stima a intervallo» non sarà più un vero e proprio intervallo, ma piuttosto una fascia posta attorno alla retta ottenuta col metodo dei minimi quadrati. A prima vista ci si potrebbe aspettare che la fascia fosse compresa tra due rette parallele alla retta ottenuta col metodo dei minimi quadrati, ma potremmo tracciare una fascia del genere solo nel caso che conoscessimo già l'inclinazione della vera retta di regressione e, quindi, l'unico errore possibile fosse nella stima di α . Dobbiamo invece ricordarci che sono due le quantità che dobbiamo stimare questa volta (α e β) e che, quindi, vi sono due potenziali fonti di errore. Si può dimostrare che, dal momento che la stima dell'inclinazione può essere errata, la possibilità di errore aumenta man mano che ci si allontana dal singolo punto $(\bar{X}, \bar{Y})$. La fascia di fiducia assume, quindi, in generale la forma indicata dalla figura 18.2.

Per tracciare una fascia di fiducia del genere, sarà necessario calcolare l'errore standard di Y_p per i vari valori di X . La stima dell'errore standard è data dalla formula

$$\hat{\sigma}_{y_p} = \hat{\sigma}_{y|x} \sqrt{\frac{1}{N} + \frac{(X - \bar{X})^2}{\sum_{i=1}^N (X_i - \bar{X})^2}} \quad (18.5)$$

in cui il particolare valore di X , che deve essere usato per il

calcolo di $(X - \bar{X})^2$, può essere scelto in qualsiasi punto dell'asse della X . Si noti, tra l'altro, che tanto più X è lontano da $\bar{X}$, tanto maggiore è il valore numerico dell'errore standard. Supponiamo di voler trovare la stima dell'errore standard nel caso in cui X sia 10,0. Se $\bar{X}=4,837$ abbiamo:

$$\hat{\sigma}_{y_p} = 215,1 \sqrt{\frac{1}{13} + \frac{(10,0 - 4,837)^2}{128,131}} = 215,1 \sqrt{0,28496} = 114,86$$

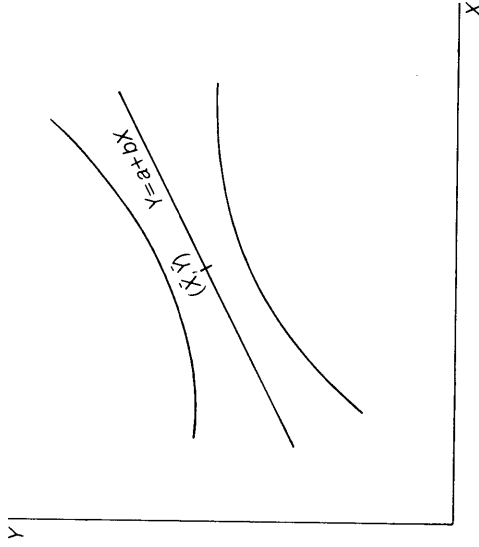


FIG. 18.2. Fascia di fiducia attorno alla retta ottenuta col metodo dei minimi quadrati

*Usando di nuovo la tabella della distribuzione di t per un intervallo di fiducia del 99 per cento attorno all' Y_p relativo a questo determinato valore di X , avremo:

$$Y_p \pm (3,106)(114,86) = Y_p \pm 356,8$$

Dopo aver ottenuto diversi intervalli attorno a Y_p per altri particolari valori di X , possiamo tracciare tutta la fascia; questo procedimento è, peraltro, lungo e di notevole complicazione se non si può ricorrere ad un calcolatore.