

e:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{81.8}{14} = 5.842857$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{40.9}{14} = 2.92143$$

Quindi:

$$\begin{aligned} b_0 &= \bar{Y} - b_1 \bar{X} \\ &= 5.842857 - (1.6699)(2.92143) \\ &= 0.9645 \end{aligned}$$

Esercizi al Paragrafo 12.2

Apprendere i concetti di base

- 12.1** Stimando un modello di regressione per un insieme di dati, è stata ottenuta la seguente retta di regressione:

$$\hat{Y}_i = 4 + 8X_i$$

- Interpretare il significato dell'intercetta b_0 .
 - Interpretare il significato del coefficiente angolare b_1 .
 - Stimare il valore della Y per $X = 2$.
- 12.2** Se i valori della X dell'Esercizio 12.1 variano nell'intervallo da 2 a 25, si può usare il modello di regressione per predire i valori di Y quando X è uguale a:
- 3?
 - 3?
 - 0?
 - 24?

- 12.3** Stimando un modello di regressione per un insieme di dati, è stata ottenuta la seguente retta di regressione:

$$\hat{Y}_i = 24 - 0.8X_i$$

- Interpretare il significato dell'intercetta b_0 .
- Interpretare il significato del coefficiente angolare b_1 .
- Stimare il valore della Y per $X = 5$.

Applicazioni

- 12.4** Il manager di una catena di supermercati intende verificare se lo spazio destinato a una certa tipologia di prodotto influisce sulle vendite. A tale scopo su un campione casuale di 12 supermercati si ottengono i seguenti risultati:

Supermercato	Spazio sullo scaffale (X) (in metri)	Vendite settimanali (Y) (\$)
1	5	160
2	5	220
3	5	140
4	10	190
5	10	240
6	10	260
7	15	230
8	15	270
9	15	280
10	20	260
11	20	290
12	20	310

- Costruire un diagramma di dispersione. Per questi dati si ha $b_0 = 145$ e $b_1 = 7.4$.
- Interpretare il significato del coefficiente angolare b_1 per il problema in questione.
- Stimare quali potrebbero essere le vendite settimanali del prodotto nel caso in cui lo spazio X sullo scaffale è pari a 12 metri.

12.5 È noto che, per una rivista, la capacità di attrarre inserzionisti dipende dalla sua diffusione. Recentemente l'autorità garante delle comunicazioni negli Stati Uniti ha riscontrato una sostanziale differenza tra quelle che sono le vendite dichiarate direttamente dalle riviste e quelle riportate nei dati ufficiali. Tale strategia ha lo scopo di attrarre più inserzionisti ma non è un'operazione legittima. La tabella seguente riporta le vendite annuali, in migliaia di dollari, per 10 tra le più diffuse riviste americane.

Rivista	Vendite dichiarate (X)	Vendite ufficiali (Y)
YM	621.0	299.6
CosmoGirl	359.7	207.7
Rosie	530.0	325.0
Playboy	492.1	336.3
Esquire	70.5	48.6
TeenPeople	567.0	400.3
More	125.5	91.2
Spin	50.6	39.1
Vogue	353.3	268.6
Elle	263.6	214.3

Fonte: M. Rose, In Fight for Ads, Publishers Often Overstate Their Sales, *The Wall Street Journal*, 6 agosto 2003, pp. A1, A10.

- Costruire un diagramma di dispersione. Per questi dati si ha $b_0 = 26.724$ e $b_1 = 0.5719$.
 - Interpretare il significato del coefficiente angolare b_1 per il problema in questione.
 - Predire le vendite ufficiali di una rivista che dichiara di avere un volume di vendite annuali pari a \$300000.
- 12.6** Un aspetto cruciale per la qualità del servizio ai clienti in un supermercato è il cosiddetto tempo di *check-out* (ovvero il tempo che il cliente impiega dal momento in cui si mette in fila alla cassa fino a quando non viene emesso lo scontrino). Questo tempo di attesa potrebbe dipendere dal numero di clienti presenti all'interno della struttura. In un grande supermercato vengono raccolte le informazioni relative al numero di clienti presenti nel supermercato e il tempo di attesa in fila alle casse (in minuti) (i dati sono contenuti nel file *Supermarket*).
- Costruire un diagramma di dispersione.
 - Assumendo una relazione lineare tra il numero di clienti e il tempo di attesa, usare il metodo dei minimi quadrati per stimare i valori dei coefficienti b_0 e b_1 .
 - Interpretare il significato del coefficiente angolare b_1 .
- 12.7** Un agente immobiliare di una grande città vor-

rebbe stimare l'affitto mensile di un appartamento, in un quartiere residenziale, sulla base della sua grandezza (in piedi quadrati). Egli seleziona un campione casuale di 25 appartamenti del quartiere e su di essi rileva i dati relativi all'affitto mensile e all'ampiezza (i dati sono contenuti nel file *Rent*).

- Costruire un diagramma di dispersione.
- Usare il metodo dei minimi quadrati per stimare i valori dei coefficienti b_0 e b_1 .
- Interpretare il significato di b_0 e b_1 .
- Stimare l'affitto mensile quando si considera un appartamento di 1000 piedi quadrati.
- Motivare perché non sarebbe appropriato predire quale sia l'affitto mensile di un appartamento grande 500 piedi quadrati?
- Una giovane coppia sta pensando di firmare un contratto per affittare un appartamento nel quartiere residenziale oggetto dell'indagine. La coppia sta decidendo tra due appartamenti, uno di 1000 piedi quadrati, che costerebbe mensilmente \$1275, e l'altro, di 1200 piedi quadrati che costerebbe \$1425. Sulla base dei risultati ottenuti ai punti precedenti, quale appartamento sarebbe più conveniente?

- 12.8** Una compagnia che distribuisce DVD ha condotto un'indagine per stabilire quale sia la relazione tra le vendite dei DVD e la spesa in pubblicità effettuata per ognuno di essi. I dati riguardano un campione di 30 DVD relativi a film da poco usciti nelle sale cinematografiche, su cui sono rilevate le vendite (in migliaia) e il budget destinato alla pubblicità (in milioni di dollari) (i dati sono contenuti nel file *Movie*).

- Costruire un diagramma di dispersione.
- Usare il metodo dei minimi quadrati per stimare i valori dei coefficienti b_0 e b_1 .
- Interpretare il significato del coefficiente angolare b_1 per questo problema.
- Predire le vendite dei DVD quando il budget pubblicitario ammonta a 75 milioni di dollari.

12.3 La scomposizione della devianza totale e l'indice di determinazione

Quando si utilizza il metodo dei minimi quadrati per stimare i coefficienti di regressione su un insieme di dati, al fine di valutare quanto il modello si adatti ai dati, è necessario calcolare tre importanti indici di variabilità. Il primo di questi indici è la **devianza totale** (o somma complessiva degli scarti), denotata con SST , che è una misura della variabilità complessiva delle Y_i rispetto alla media generale $\bar{Y}$. La devianza totale si sud-

istribuzione rispetto alla media. In altri termini, l'errore standard delle stime rappresenta lo scarto quadratico medio dei dati attorno alla retta di regressione. L'Equazione (12.13) definisce l'errore standard delle stime, denotato con il simbolo S_{YX} .

ERRORE STANDARD DELLE STIME

$$S_{YX} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}} \quad (12.13)$$

dove:

Y_i = valore osservato di Y per una data osservazione X_i
 $\hat{Y}_i$ = valore stimato di Y per una data osservazione X_i
 SSE = devianza residua

Per i dati della Sunflowers conosciamo la devianza residua $SSE = 11.2067$; è pertanto possibile calcolare agevolmente l'errore standard delle stime:

$$S_{YX} = \sqrt{\frac{11.2067}{14-2}} = 0.9664$$

Questo errore standard delle stime, pari a 0.9664 milioni di dollari (cioè \$966 400), è indicato come Errore Standard nell'output di Excel. Questo indice ha la stessa unità di misura della variabile dipendente Y e la sua interpretazione è analoga a quella dello scarto quadratico medio. Per i dati della Sunflowers, quindi, si può concludere che la deviazione media tra le vendite annuali osservate e quelle stimate dal modello è approssimativamente di \$966 400.

Esercizi al Paragrafo 12.3

Apprendere i concetti di base

- 12.9** Come si interpreta un coefficiente di determinazione $R^2 = 0.85$?
- 12.10** Se $SSR = 36$ e $SSE = 4$, trovare SST , calcolare il coefficiente di determinazione e interpretarne il significato.
- 12.11** Se $SSR = 77$ e $SST = 110$, calcolare il coefficiente di determinazione e interpretarne il significato.
- 12.12** Se $SSE = 10$ e $SSR = 30$, calcolare il coefficiente di determinazione e interpretarne il significato.
- 12.13** Se $SSR = 150$, perché è impossibile che SST sia uguale a 140?

Applicazioni

- 12.14** Nell'Esercizio 12.4, il manager di un supermercato ha costruito un modello di previsione delle vendite di un prodotto a partire dallo spazio destinato negli

scaffali per il prodotto stesso. Per questi dati si ricava che $SSR = 20 535$ e $SST = 30 025$.

a. Calcolare il coefficiente di determinazione R^2 e interpretarne il significato.

b. Calcolare l'errore standard delle stime.

c. A partire da questi due indici, quanto risulta utile il modello di regressione per predire le vendite?

- 12.15** Nell'Esercizio 12.5, è stato costruito un modello di regressione per stimare le vendite ufficiali di 10 riviste americane a partire da quelle dichiarate (i dati sono riportati nel file **Circulation**). Per questi dati si ricava che $SSR = 130 301.41$ e $SST = 144 538.64$.

a. Calcolare il coefficiente di determinazione R^2 e interpretarne il significato.

b. Calcolare l'errore standard delle stime.

c. A partire da questi due indici, quanto risulta utile il modello di regressione per stimare le vendite effettive?