

quindi l'intervallo di predizione sarebbe dato da

$$69.86 \pm 8.36 \quad \square$$

Problemi

1. I dati seguenti mettono in relazione la percentuale di acqua x , contenuta in un certo materiale in una delle fasi di lavorazione, con la densità Y del prodotto finito.

x	5	6	7	10	12	15	18	29
Y	7.4	9.3	10.6	15.4	18.1	22.2	24.1	24.8

(a) Traccia il diagramma di dispersione.

(b) Trova la retta di regressione che interpola questi dati.

2. I dati seguenti illustrano la relazione esistente tra il prezzo unitario di un certo bene in 6 luoghi differenti e il numero di unità dello stesso bene che sono state ordinate.

Pezzi ordinati	88	112	123	136	158	172
Prezzo	50	40	35	30	20	15

Secondo te quante unità verrebbero ordinate se il prezzo fosse 25? \square

3. Si studia il livello di corrosione di una certa sostanza metallica esponendola ad una atmosfera di ossigeno puro, ad una temperatura di 500 gradi Celsius. L'aumento relativo di massa della sostanza viene utilizzato come indicatore della quantità di ossigeno che ha reagito. I dati raccolti sono i seguenti:

Ore di esposizione	1.0	2.0	2.5	3.0	3.5	4.0
Incremento percentuale	0.02	0.03	0.035	0.042	0.05	0.054

(a) Traccia il diagramma di dispersione.

(b) Trova la relazione lineare che interpola meglio i dati.

(c) Fornisci una previsione dell'incremento di massa dopo 3.2 ore di esposizione. \square

4. I dati che seguono mostrano la relazione tra la densità x di certi campioni di legname e Y , la massima resistenza alla compressione opposta dal legno nella direzione della fibra (misurata in psi).

x	0.41	0.46	0.44	0.47	0.42	0.39	0.41	0.44	0.43	0.44
Y	1850	2620	2340	2690	2160	1760	2500	2750	2730	3120

(a) Traccia il diagramma di dispersione. Pensi che sussista una relazione lineare?

(b) Stima i coefficienti di regressione.

(c) Predici la resistenza alla compressione per un campione di legname con una densità di 0.43.

5. I dati seguenti mostrano l'incremento nella velocità di lettura (misurata in parole al minuto) dopo un numero diverso di settimane per 10 individui iscritti ad un corso di lettura veloce.

Numero di settimane	2	3	8	11	4	5	9	7	5	7
Aumento di velocità	21	42	102	130	52	57	105	85	62	90

(a) Traccia il diagramma di dispersione per capire se può sussistere una relazione lineare.

(b) Trova le stime dei minimi quadrati dei coefficienti di regressione.

(c) Stima il guadagno nel quale può mediamente sperare un iscritto che intenda seguire il corso per 7 settimane.

6. La spettroscopia infrarossa è spesso impiegata per determinare la percentuale di gomma naturale in misture di gomma naturale e sintetica. Per esemplari di composizione nota, lo strumento ha fornito le letture seguenti:

Percentuale	0	20	40	60	80	100
Letture	0.734	0.885	1.050	1.191	1.314	1.432

Se una nuova miscela dà una lettura di 1.15 allo spettroscopio, qual è la percentuale di gomma naturale stimata?

7. La tabella che segue fornisce i punteggi medi per le parti linguistica e matematica del SAT⁷ del 1996, in ciascuno degli stati americani. Viene anche riportata la percentuale di studenti diplomati che hanno sostenuto il test.

(a) Usa i dati dei primi 20 stati (da Alabama a Maine) per ottenere una predizione del punteggio medio in matematica in funzione della percentuale di studenti che sostengono il test.

(b) Confronta i valori predetti con quelli riscontrati nei 5 stati successivi.

Punteggi medi del SAT, ordinati per stato, 1996 (scala ricentrata)

	Linguistico	Matematico	Percentuale di partecipazione
Alabama	565	558	8
Alaska	521	513	47
Arizona	525	521	28
Arkansas	566	550	6
California	495	511	45
Colorado	536	538	30
Connecticut	507	504	79
Delaware	508	495	66
Dist. of Columbia	489	473	50
Florida	498	496	48
Georgia	484	477	63
Hawaii	485	510	54
Idaho	543	536	15
Illinois	564	575	14
Indiana	494	494	57
Iowa	590	600	5
Kansas	579	571	9
Kentucky	549	544	12
Louisiana	559	550	9
Maine	504	498	68
Maryland	507	504	64
Massachusetts	507	504	80
Michigan	557	565	11
Minnesota	582	593	9
Mississippi	569	557	4
Missouri	570	569	9
Montana	546	547	21
Nebraska	567	568	9
Nevada	508	507	31
New Hampshire	520	514	70
New Jersey	498	505	69
New Mexico	554	548	12
New York	497	499	73
North Carolina	490	486	59
North Dakota	596	599	5
Ohio	536	535	24
Oklahoma	566	557	8
Oregon	523	521	50
Pennsylvania			1

8. Verifica l'Equazione (9.3.3) che afferma che

$$\text{Var}(A) = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

9. Considera nuovamente il Problema 4.

- (a) Stima la varianza di una singola risposta.
- (b) Determina un intervallo di confidenza al 90% per tale parametro.

10. Verifica l'identità seguente.

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$$

11. I dati seguenti riguardano 12 studenti di uno stesso corso di studi in legge. Tutti riportarono punteggi simili nella prova finale, inoltre ciascuno di essi entrò a lavorare in uno studio legale. La tabella mette a confronto i loro redditi in migliaia di dollari con le loro stature in pollici.

Statura	64	65	66	67	69	70	72	72	74	74	75	76
Reddito	91	94	88	103	77	96	105	88	122	102	90	114

- (a) Ad un livello di significatività del 5%, questi dati confermano che vi sia un legame tra salario e altezza?
 - (b) Qual è stata la tua scelta per l'ipotesi nulla nel punto (a)? Motiva la risposta.
12. I dati che seguono rappresentano il numero di macchie solari apparse e il numero di vittime di incidenti stradali che si sono verificati negli anni dal 1970 al 1983. Verifica l'ipotesi che il numero delle vittime della strada non sia influenzato dalle macchie solari.

Anno	Macchie solari	Vittime di incidenti stradali (1 000)
1970	165	54.6
1971	89	53.3
1972	55	56.3
1973	34	49.6
1974	9	47.1
1975	30	45.9
1976	59	48.5
1977	83	50.1
1978	109	52.4
1979	127	52.5
1980	153	53.2
1981	112	51.4
1982	80	46.0
1983	45	44.6

Fonte: per le macchie solari, Jastrow and Thompson, Fundamentals and frontiers of astronomy; per gli incidenti, General Statistics of the U.S. 1985.

13. Considera il modello di regressione lineare semplice

$$Y = \alpha + \beta x + e$$

e supponi che $0 < \beta < 1$.

(a) Dimostra che se $x < \frac{\alpha}{1-\beta}$, allora

$$x < E[Y] < \frac{\alpha}{1-\beta}$$

(b) Dimostra che se $x > \frac{\alpha}{1-\beta}$, allora

$$x > E[Y] > \frac{\alpha}{1-\beta}$$

E concludi che $E[Y]$ è sempre compresa tra x e $\frac{\alpha}{1-\beta}$.

16. I dati seguenti rappresentano la relazione esistente tra il numero di errori di allineamento Y e numero di rivetti mancanti x , per 10 differenti aeromobili.

Rivetti mancanti	13	15	10	22	30	7	25	16	20	15
Errori di allineamento	7	7	5	12	15	2	13	9	11	8

- (a) Disegna il diagramma di dispersione.
- (b) Stima i coefficienti di regressione.
- (c) Verifica l'ipotesi che $\alpha = 1$.
- (d) Stima il numero medio di errori di allineamento per un aeroplano cui manchino 24 rivetti.
- (e) Calcola un intervallo di confidenza al 90% per la quantità del punto (d).

17. Le cifre che seguono sono le medie annuali dei prezzi di tutti i libri recensiti dalla rivista *Science*, dal 1990 al 1996. Dai un intervallo che con il 95% di confidenza contenga la media dei prezzi di tutti i libri che sono stati recensiti nel 1997.

Anno	Prezzo medio (dollari)
1990	54.43
1991	54.08
1992	57.58
1993	51.21
1994	59.96
1995	60.52
1996	62.13

I Problemi dal 18 al 22 si riferiscono alla tabella di pagina seguente, che mette in relazione il livello di fumo con i tassi di morte per 4 tipi di tumore in 14 stati americani.

- 18. (a) Disegna il diagramma di dispersione dei decessi per tumore alla vescica rispetto ai consumi di sigarette.
- (b) Diresti che è possibile che vi sia una relazione lineare?
- (c) Trova il miglior fit lineare.
- (d) Se il consumo medio pro capite in un certo stato fosse di x ...

23. Disegna i residui standardizzati per i dati del Problema 22. Discuti i risultati e spiega il significato di regressione lineare riguardo alla nostra assunzione che sia valido un modello di regressione lineare.
24. Misurare direttamente il contenuto di proteine nei campioni di fegato richiede un procedimento lungo e difficile. Per questo motivo i laboratori di medicina fanno spesso uso della spettrofotometria, grazie al fatto che la luce assorbita dal campione è legata alla quantità di proteine presenti. La procedura di misurazione consiste nel preparare una sospensione del campione in acqua e registrarne l'assorbimento luminoso tramite uno spettrofotometro; essa è stata effettuata su 5 campioni con un contenuto di proteine noto, ottenendo i risultati seguenti.

Luce assorbita	0.44	0.82	1.20	1.61	1.83
Contenuto di proteine	2	16	30	46	55

- (a) Calcola il coefficiente di determinazione.
- (b) Ti sembra che questo sia un modo ragionevole di misurare le proteine nei campioni di fegato?
- (c) Qual è la stima del contenuto di proteine se l'assorbimento di luce è 1.5?
- (d) Determina un intervallo di predizione al 90% per la stima del punto (c).

25. Determinare la sollecitazione di taglio di un punto di saldatura è relativamente difficile: misurare il diametro è molto più semplice. Sarebbe molto vantaggioso perciò se la prima grandezza potesse essere predetta da una misurazione della seconda. I dati trovati in una sperimentazione sono i seguenti.

Sollecitazione di taglio (psi)	Diametro della saldatura (10^{-4} pollici)
370	400
780	800
1 210	1 250
1 560	1 600
1 980	2 000
2 450	2 500
3 070	3 100
3 550	3 600
3 940	4 000
3 950	4 000

- (a) Traccia il diagramma di dispersione.
- (b) Determina gli stimatori dei minimi quadrati dei coefficienti di regressione.
- (c) Verifica al 5% di significatività l'ipotesi che il coefficiente angolare della retta di regressione sia 1.
- (d) Stima il valore atteso della sollecitazione di taglio quando il diametro è di 0.25 pollici.
- (e) Trova un intervallo di predizione che contenga con il 95% di confidenza la sollecitazione di taglio di un punto di saldatura del diametro di 0.225 pollici.

- (f) Traccia il grafico dei residui standardizzati.
- (g) Il grafico ottenuto al punto (f) è in accordo con le assunzioni del modello?
26. Un produttore di viti vuole fornire ai suoi clienti dei dati sulla relazione tra lunghezze nominali ed effettive dei suoi prodotti. Vengono osservati i dati (in pollici) che sono riportati nella tabella alla fine del problema.
- (a) Stima i coefficienti di regressione.
- (b) Stima la varianza che risulta nella fabbricazione di una vite.
- (c) Trova un intervallo di confidenza al 90% per la lunghezza media di un elevato numero di viti di 1 pollice nominale.
- (d) Determina un intervallo di predizione al 90% per la lunghezza di una singola vite di 1 pollice nominale.
- (e) Traccia il grafico dei residui standardizzati.
- (f) Il grafico ottenuto al punto (e) fa sorgere qualche dubbio sul modello di regressione?
- (g) Calcola il coefficiente di correlazione lineare.

Lunghezza nominale x Lunghezza effettiva y

$\frac{1}{4}$	0.262	0.262	0.245
$\frac{1}{2}$	0.496	0.512	0.490
$\frac{3}{4}$	0.743	0.744	0.751
1	0.976	1.010	1.004
$1\frac{1}{4}$	1.265	1.254	1.252
$1\frac{1}{2}$	1.498	1.518	1.504
$1\frac{3}{4}$	1.738	1.759	1.750
2	2.005	1.992	1.992

27. Il vetro gioca un ruolo importante nelle indagini criminali, infatti l'attività criminale finisce spesso col causare la rottura di finestre e altri oggetti di vetro, e siccome piccoli frammenti tendono a rimanere attaccati ai vestiti del colpevole, è fondamentale riuscire a identificare i diversi tipi di vetro e collegarli con il luogo del delitto. Due proprietà fisiche del vetro che sono utili per l'identificazione sono l'indice di rifrazione e la densità. Il primo è di facile misurazione, mentre il secondo è molto più complicato; siccome inoltre la misurazione esatta della densità è molto facilitata se si possiede almeno una sua buona stima prima di approntare l'esperimento, sarebbe piuttosto utile se si potesse impiegare l'indice di rifrazione per stimare l'altro parametro.

I dati seguenti mettono in relazione l'indice di rifrazione di 18 tipi di vetro con la loro densità.

Indice di rifrazione	Densità	Indice di rifrazione	Densità
1.5139	2.4801	1.5161	2.4843
1.5153	2.4819	1.5165	2.4858
1.5155	2.4791	1.5178	2.4950
1.5155	2.4796	1.5181	2.4922
1.5156	2.4773	1.5191	2.5035
1.5157	2.4811	1.5227	2.5086
1.5158	2.4765	1.5227	2.5117
1.5159	2.4781	1.5232	2.5146
1.5160	2.4909	1.5233	2.5187

- (a) Predici la densità di un frammento di vetro che abbia un indice di rifrazione di 1.52.
- (b) Determina un intervallo che con il 95% di confidenza contenga la densità cercata al punto (a).

3. Il modello di regressione

$$Y = \beta x + e, \quad e \sim \mathcal{N}(0, \sigma^2)$$

è detto regressione attraverso l'origine, perché suppone che la risposta media quando il livello di ingresso è $x = 0$ sia nulla. Supponi che (x_i, Y_i) , per $i = 1, 2, \dots, n$ sia un campione di coppie di dati provenienti da questo modello.

- (a) Determina lo stimatore dei minimi quadrati B , di β .
- (b) Qual è la distribuzione di B ?
- (c) Definisci SS_R e trova la sua distribuzione.
- (d) Costruisci un test per verificare H_0 .

1.5150	2.4811
1.5157	2.4765
1.5158	2.4781
1.5159	2.4909
1.5160	

2.5167

2.5167

2.5167

- (a) Predici la densità di un frammento di vetro che abbia un indice di rifrazione di 1.52.
- (b) Determina un intervallo che con il 95% di confidenza contenga la densità cercata al punto (a).

28. Il modello di regressione

$$Y = \beta x + e, \quad e \sim \mathcal{N}(0, \sigma^2)$$

è detto regressione attraverso l'origine, perché suppone che la risposta media quando il livello di ingresso è $x = 0$ sia nulla. Supponi che (x_i, Y_i) , per $i = 1, 2, \dots, n$ sia un campione di coppie di dati provenienti da questo modello.

(a) Determina lo stimatore dei minimi quadrati B , di β .

(b) Qual è la distribuzione di B ?

(c) Definisci SS_R e trova la sua distribuzione.

(d) Costruisci un test per verificare $H_0 : \beta = \beta_0$ di contro a $H_1 : \beta \neq \beta_0$.

(e) Determina un intervallo di predizione con un livello di confidenza di $1 - \gamma$ per $Y(x_0)$, la risposta al livello di ingresso x_0 .

29. Dimostra l'identità seguente:

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

30. La tabella che segue riporta il peso e la pressione sistolica per un campione casuale di 20 uomini americani di età compresa tra i 25 e 30 anni.

Soggetto	Peso (libbre)	Pressione	Soggetto	Peso (libbre)	Pressione
1	165	130	11	172	153
2	167	133	12	159	128
3	180	150	13	168	132
4	155	128	14	174	149
5	212	151	15	183	158
6	175	146	16	215	150
7	190	150	17	195	163
8	210	140	18	180	156
9	200	148	19	143	124
10	149	125	20	240	170

- (a) Stima i coefficienti di regressione.
- (b) Ti sembra che i dati supportino la tesi che la pressione del sangue non dipenda dal peso corporeo?
- (c) Scelto un campione numeroso di soggetti del peso di 182 libbre, trova un intervallo che con il 95% di confidenza contenga la media delle loro pressioni sistoliche.
- (d) Analizza i residui standardizzati.
- (e) Determina il coefficiente di correlazione campionaria.

31. Si è determinato che la relazione tra la tensione T e il numero di cicli N prima di una rottura, per una particolare lega metallica è dato da

$$T = \frac{A}{N^m}$$

dove A e m sono costanti da determinare. Stimale, sapendo che una sperimentazione che è stata effettuata ha ottenuto i dati seguenti.

Tensione (1 000 psi)	55.0	50.5	43.5	42.5	42.0	41.0	35.7	34.5	33.0	32.0
N (milioni di cicli)	.223	.925	6.75	18.1	29.1	50.5	126	215	445	420

32. Nel 1957 l'ingegnere olandese J. R. DeJong propose un modello per il tempo necessario per svolgere una semplice operazione manuale, in funzione del numero di volte che era stata praticata. La formula era

$$T \approx ts^{-n}$$

dove T è il tempo necessario, n è il numero di volte che si è praticata l'operazione e t e s sono parametri che dipendono dal tipo di lavoro e dalla persona coinvolta. Stimati t e s per il