

is employed. We first convert our estimator,  $r$ , to  $z_r$ , construct a confidence interval about  $z_\rho$ , and then reconvert the limits to obtain a  $100(1 - \alpha)$  percent confidence interval about  $\rho$ . The general formula then becomes

$$z_r \pm z \left( 1/\sqrt{n-3} \right) \quad (9.7.10)$$

For our present example the 95 percent confidence interval for  $z_\rho$  is given by

$$1.24726 \pm 1.96(1/\sqrt{155-3})$$

$$(1.08828, 1.40624)$$

Converting these limits (by interpolation in Appendix Table I), which are values of  $z_r$ , into values of  $r$  gives

$z_r$	$r$
1.08828	.7962
1.40624	.8866

We are 95 percent confident, then, that  $\rho$  is contained in the interval .7962 to .88866. Because of the limited entries in the table, these limits must be considered as only approximate.

## EXERCISES

In each of the following exercises:

- Prepare a scatter diagram.
- Compute the sample correlation coefficient.
- Test  $H_0 : \rho = 0$  at the .05 level of significance and state your conclusions.
- Determine the  $p$  value for the test.
- Construct the 95 percent confidence interval for  $\rho$ .

- 9.7.1** The purpose of a study by Brown and Persley (A-8) was to characterize acute hepatitis A in patients more than 40 years old. They performed a retrospective chart review of 20 subjects who were diagnosed with acute hepatitis A, but were not hospitalized. Of interest was the use of age (years) to predict bilirubin levels (mg/dl). The following data were collected.

Age (Years)	Bilirubin (mg/dl)	Age (Years)	Bilirubin (mg/dl)
78	7.5	44	7.0
72	12.9	42	1.8
81	14.3	45	.8
59	8.0	78	3.8
64	14.1	47	3.5
48	10.9	50	5.1
46	12.3	57	16.5

(Continued)

Age (Years)	Bilirubin (mg/dl)	Age (Years)	Bilirubin (mg/dl)
42	1.0	52	3.5
58	5.2	58	5.6
52	5.1	45	1.9

Source: Data provided courtesy of Geri R. Brown, M.D.

**9.7.2** Another variable of interest in the study by Reiss et al. (A-3) (see Exercise 9.3.4) was partial thromboplastin (aPTT), the standard test used to monitor heparin anticoagulation. Use the data in the following table to examine the correlation between aPTT levels as measured by the CoaguCheck point-of-care assay and standard laboratory hospital assay in 90 subjects receiving heparin alone, heparin with warfarin, and warfarin and exoenoxaparin.

Heparin		Warfarin		Warfarin and Exoenoxaparin	
CoaguCheck aPTT	Hospital aPTT	CoaguCheck aPTT	Hospital aPTT	CoaguCheck aPTT	Hospital aPTT
49.3	71.4	18.0	77.0	56.5	46.5
57.9	86.4	31.2	62.2	50.7	34.9
59.0	75.6	58.7	53.2	37.3	28.0
77.3	54.5	75.2	53.0	64.8	52.3
42.3	57.7	18.0	45.7	41.2	37.5
44.3	59.5	82.6	81.1	90.1	47.1
90.0	77.2	29.6	40.9	23.1	27.1
55.4	63.3	82.9	75.4	53.2	40.6
20.3	27.6	58.7	55.7	27.3	37.8
28.7	52.6	64.8	54.0	67.5	50.4
64.3	101.6	37.9	79.4	33.6	34.2
90.4	89.4	81.2	62.5	45.1	34.8
64.3	66.2	18.0	36.5	56.2	44.2
89.8	69.8	38.8	32.8	26.0	28.2
74.7	91.3	95.4	68.9	67.8	46.3
150.0	118.8	53.7	71.3	40.7	41.0
32.4	30.9	128.3	111.1	36.2	35.7
20.9	65.2	60.5	80.5	60.8	47.2
89.5	77.9	150.0	150.0	30.2	39.7
44.7	91.5	38.5	46.5	18.0	31.3
61.0	90.5	58.9	89.1	55.6	53.0
36.4	33.6	112.8	66.7	18.0	27.4
52.9	88.0	26.7	29.5	18.0	35.7
57.5	69.9	49.7	47.8	78.3	62.0
39.1	41.0	85.6	63.3	75.3	36.7
74.8	81.7	68.8	43.5	73.2	85.3
32.5	33.3	18.0	54.0	42.0	38.3
125.7	142.9	92.6	100.5	49.3	39.8
77.1	98.2	46.2	52.4	22.8	42.3
143.8	108.3	60.5	93.7	35.8	36.0

Source: Data provided courtesy of Curtis E. Haas, Pharm.D.

## EXERCISES

Obtain the regression equation for each of the following data sets.

- 10.3.1** Machiel Naeije (A-2) studied the relationship between maximum mouth opening and measurements of the lower jaw (mandible). He measured the dependent variable, maximum mouth opening (MMO, measured in mm), as well as predictor variables, mandibular length (ML, measured in mm) and angle of rotation of the mandible (RA, measured in degrees) of 35 subjects.

MMO ( $Y$ )	ML ( $X_1$ )	RA ( $X_2$ )	MMO ( $Y$ )	ML ( $X_1$ )	RA ( $X_2$ )
52.34	100.85	32.08	50.82	90.65	38.33
51.90	93.08	39.21	40.48	92.99	25.93
52.80	98.43	33.74	59.68	108.97	36.78
50.29	102.95	34.19	54.35	91.85	42.02
57.79	108.24	35.13	47.00	104.30	27.20
49.41	98.34	30.92	47.23	93.16	31.37
53.28	95.57	37.71	41.19	94.18	27.87
59.71	98.85	44.71	42.76	89.56	28.69
53.32	98.32	33.17	51.88	105.85	31.04
48.53	92.70	31.74	42.77	89.29	32.78
51.59	88.89	37.07	52.34	92.58	37.82
58.52	104.06	38.71	50.45	98.64	33.36
62.93	98.18	43.89	43.18	83.70	31.93
57.62	91.01	41.06	41.99	88.46	28.32
65.64	96.98	41.92	39.45	94.93	24.82
52.85	97.85	35.25	38.91	96.81	23.88
64.43	96.89	45.11	49.10	93.13	36.17
57.25	98.35	39.44			

Source: Data provided courtesy of M. Naeije, D.D.S.

- 10.3.2** Family caregiving of older adults is more common in Korea than in the United States. Son et al. (A-3) studied 100 caregivers of older adults with dementia in Seoul, South Korea. The dependent variable was caregiver burden as measured by the Korean Burden Inventory (KBI). Scores ranged from 28 to 140, with higher scores indicating higher burden. Explanatory variables were indexes that measured the following:

ADL: total activities of daily living (low scores indicate that the elderly perform activities independently).

MEM: memory and behavioral problems (higher scores indicate more problems).

COG: cognitive impairment (lower scores indicate a greater degree of cognitive impairment).

The reported data are as follows:

KBI ( $Y$ )	ADL ( $X_1$ )	MEM ( $X_2$ )	COG ( $X_3$ )	KBI ( $Y$ )	ADL ( $X_1$ )	MEM ( $X_2$ )	COG ( $X_3$ )
28	39	4	18	88	76	50	5
68	52	33	9	54	79	44	11
59	89	17	3	73	48	57	9
91	57	31	7	87	90	33	6

(Continued)

KBI ( $Y$ )	ADL ( $X_1$ )	MEM ( $X_2$ )	COG ( $X_3$ )	KBI ( $Y$ )	ADL ( $X_1$ )	MEM ( $X_2$ )	COG ( $X_3$ )
70	28	35	19	47	55	11	20
38	34	3	25	60	83	24	11
46	42	16	17	65	50	21	25
57	52	6	26	57	44	31	18
89	88	41	13	85	79	30	20
48	90	24	3	28	24	5	22
74	38	22	13	40	40	20	17
78	83	41	11	87	35	15	27
43	30	9	24	80	55	9	21
76	45	33	14	49	45	28	17
72	47	36	18	57	46	19	17
61	90	17	0	32	37	4	21
63	63	14	16	52	47	29	3
77	34	35	22	42	28	23	21
85	76	33	23	49	61	8	7
31	26	13	18	63	35	31	26
79	68	34	26	89	68	65	6
92	85	28	10	67	80	29	10
76	22	12	16	43	43	8	13
91	82	57	3	47	53	14	18
78	80	51	3	70	60	30	16
103	80	20	18	99	63	22	18
99	81	20	1	53	28	9	27
73	30	7	17	78	35	18	14
88	27	27	27	112	37	33	17
64	72	9	0	52	82	25	13
52	46	15	22	68	88	16	0
71	63	52	13	63	52	15	0
41	45	26	18	49	30	16	18
85	77	57	0	42	69	49	12
52	42	10	19	56	52	17	20
68	60	34	11	46	59	38	17
57	33	14	14	72	53	22	21
84	49	30	15	95	65	56	2
91	89	64	0	57	90	12	0
83	72	31	3	88	88	42	6
73	45	24	19	81	66	12	23
57	73	13	3	104	60	21	7
69	58	16	15	88	48	14	13
81	33	17	21	115	82	41	13
71	34	13	18	66	88	24	14
91	90	42	6	92	63	49	5
48	48	7	23	97	79	34	3
94	47	17	18	69	71	38	17
57	32	13	15	112	66	48	13
49	63	32	15	88	81	66	1

Source: Data provided courtesy of Gwi-Ryung Son, R.N., Ph.D.

Another problem sometimes encountered in the application of multiple regression is an apparent incompatibility in the results of the various tests of significance that one may perform. In a given problem for a given level of significance, one or the other of the following situations may be observed.

1.  $R^2$  and all  $\hat{\beta}_i$  significant
2.  $R^2$  and some but not all  $\hat{\beta}_i$  significant
3.  $R^2$  significant but none of the  $\hat{\beta}_i$  significant
4. All  $\hat{\beta}_i$  significant but not  $R^2$
5. Some  $\hat{\beta}_i$  significant, but not all nor  $R^2$
6. Neither  $R^2$  nor any  $\hat{\beta}_i$  significant

Notice that situation 1 exists in our illustrative example, where we have a significant  $R^2$  and two significant regression coefficients. This situation does not occur in all cases. In fact, situation 2 is very common, especially when a large number of independent variables have been included in the regression equation.

## EXERCISES

---

- 10.4.1** Refer to Exercise 10.3.1. (a) Calculate the coefficient of multiple determination; (b) perform an analysis of variance; (c) test the significance of each  $\hat{\beta}_i (i > 0)$ . Let  $\alpha = .05$  for all tests of significance and determine the  $p$  value for all tests; (d) construct a 95 percent confidence interval for each significant sample slope.
- 10.4.2** Refer to Exercise 10.3.2. Do the analysis suggested in Exercise 10.4.1.
- 10.4.3** Refer to Exercise 10.3.3. Do the analysis suggested in Exercise 10.4.1.
- 10.4.4** Refer to Exercise 10.3.4. Do the analysis suggested in Exercise 10.4.1.
- 10.4.5** Refer to Exercise 10.3.5. Do the analysis suggested in Exercise 10.4.1.
- 10.4.6** Refer to Exercise 10.3.6. Do the analysis suggested in Exercise 10.4.1.

## 10.5 USING THE MULTIPLE REGRESSION EQUATION

---

As we learned in the previous chapter, a regression equation may be used to obtain a computed value of  $Y$ ,  $\hat{y}$ , when a particular value of  $X$  is given. Similarly, we may use our multiple regression equation to obtain a  $\hat{y}$  value when we are given particular values of the two or more  $X$  variables present in the equation.

Just as was the case in simple linear regression, we may, in multiple regression, interpret a  $\hat{y}$  value in one of two ways. First we may interpret  $\hat{y}$  as an estimate of the mean of the subpopulation of  $Y$  values assumed to exist for particular combinations of  $X_i$  values. Under this interpretation  $\hat{y}$  is called an *estimate*, and when it is used for this purpose, the equation is thought of as an *estimating equation*. The second interpretation of  $\hat{y}$  is that it is the value  $Y$  is most likely to assume for given values of the  $X_i$ . In this case  $\hat{y}$