since the t ratio of 4.46 for testing H_0: $\beta_3 = 0$ is also greater than the critical t of 2.0423, we can conclude (at the .05 level of significance) that the y-intercept of the population regression line for treatment B is also different from the y-intercept of the population regression line for treatment C. (See the y-intercept of Equation 11.2.6.)

Now let us consider the slopes. We see by Equation 11.2.5 that the slope of the regression line for treatment A is equal to $\hat{\beta}_1$ (the slope of the line for treatment C) $+\hat{\beta}_4$. Since the t ratio of -6.45 for testing H_0: $\beta_4 = 0$ is less than the critical t of -2.0423, we can conclude (for $\alpha = .05$) that the slopes of the population regression lines for treatments A and C are different. Similarly, since the computed t ratio for testing H_0: $\beta_5 = 0$ is also less than -2.0423, we conclude (for $\alpha = .05$) that the population regression lines for treatments B and C have different slopes (see the slope of Equation 11.2.6). Thus, we conclude that there is interaction between age and type of treatment. This is reflected by a lack of parallelism among the regression lines in Figure 11.2.6. ■

Another question of interest is this: Is the slope of the population regression line for treatment A different from the slope of the population regression line for treatment B? To answer this question requires computational techniques beyond the scope of this text. The interested reader is referred to books devoted specifically to regression analysis.

In Section 10.4 the reader was warned that there are problems involved in making multiple inferences from the same sample data. Again, books on regression analysis are available that may be consulted for procedures to be followed when multiple inferences, such as those discussed in this section, are desired.

We have discussed only two situations in which the use of dummy variables is appropriate. More complex models involving the use of one or more qualitative independent variables in the presence of two or more quantitative variables may be appropriate in certain circumstances. More complex models are discussed in the many books devoted to the subject of multiple regression analysis.

At this point it may be evident that there are many similarities between the use of a linear regression model using dummy variables and the basic ANOVA approach. In both cases, one is attempting to model the relationship between predictor variables and an outcome variable. In the case of linear regression, we are generally most interested in prediction, and in ANOVA, we are generally most interested in comparing means. If the desire is to compare means using regression, one could develop a model to predict mean response, say μ_i, instead of an outcome, y_i. Modeling the mean response using regression with dummy variables is equivalent to ANOVA. For the interested student, we suggest the book by Bowerman and O'Connell (1), who provide an example of using both approaches for the same data.

EXERCISES

For each exercise do the following:

(a) Draw a scatter diagram of the data using different symbols for the different categorical variables.

(b) Use dummy variable coding and regression to analyze the data.

(c) Perform appropriate hypothesis tests and construct appropriate confidence intervals using your choice of significance and confidence levels.

(d) Find the p value for each test that you perform.

11.2.1 For subjects undergoing stem cell transplants, dendritic cells (DCs) are antigen-presenting cells that are critical to the generation of immunologic tumor responses. Bolwell et al. (A-2) studied lymphoid DCs in 44 subjects who underwent autologous stem cell transplantation. The outcome variable is the concentration of DC2 cells as measured by flow cytometry. One of the independent variables is the age of the subject (years), and the second independent variable is the mobilization method. During chemotherapy, 11 subjects received granulocyte colony-stimulating factor (G-CSF) mobilizer (μg/kg/day) and 33 received etoposide (2 g/m^2). The mobilizer is a kind of blood progenitor cell that triggers the formation of the DC cells. The results were as follows:

G-CSF		Etoposide					
DC	**Age**	**DC**	**Age**	**DC**	**Age**	**DC**	**Age**
6.16	65	3.18	70	4.24	60	4.09	36
6.14	55	2.58	64	4.86	40	2.86	51
5.66	57	1.69	65	4.05	48	2.25	54
8.28	47	2.16	55	5.07	50	0.70	50
2.99	66	3.26	51	4.26	23	0.23	62
8.99	24	1.61	53	11.95	26	1.31	56
4.04	59	6.34	24	1.88	59	1.06	31
6.02	60	2.43	53	6.10	24	3.14	48
10.14	66	2.86	37	0.64	52	1.87	69
27.25	63	7.74	65	2.21	54	8.21	62
8.86	69	11.33	19	6.26	43	1.44	60

Source: Data provided courtesy of Lisa Rybicki, M.S.

11.2.2 According to Pandey et al. (A-3) carcinoma of the gallbladder is not infrequent. One of the primary risk factors for gallbladder cancer is cholelithiasis, the asymptomatic presence of stones in the gallbladder. The researchers performed a case-control study of 50 subjects with gallbladder cancer and 50 subjects with cholelithiasis. Of interest was the concentration of lipid peroxidation products in gallbladder bile, a condition that may give rise to gallbladder cancer. The lipid peroxidation product melonaldehyde (MDA, μg/mg) was used to measure lipid peroxidation. One of the independent variables considered was the cytochrome P-450 concentration (CYTO, nmol/mg). Researchers used disease status (gallbladder cancer vs. cholelithiasis) and cytochrome P-450 concentration to predict MDA. The following data were collected.

Cholelithiasis				Gallbladder Cancer			
MDA	**CYTO**	**MDA**	**CYTO**	**MDA**	**CYTO**	**MDA**	**CYTO**
0.68	12.60	11.62	4.83	1.60	22.74	9.20	8.99
0.16	4.72	2.71	3.25	4.00	4.63	0.69	5.86
0.34	3.08	3.39	7.03	4.50	9.83	10.20	28.32
3.86	5.23	6.10	9.64	0.77	8.03	3.80	4.76
0.98	4.29	1.95	9.02	2.79	9.11	1.90	8.09
3.31	21.46	3.80	7.76	8.78	7.50	2.00	21.05
1.11	10.07	1.72	3.68	2.69	18.05	7.80	20.22
4.46	5.03	9.31	11.56	0.80	3.92	16.10	9.06
1.16	11.60	3.25	10.33	3.43	22.20	0.98	35.07

(*Continued*)

Cholelithiasis				Gallbladder Cancer			
MDA	**CYTO**	**MDA**	**CYTO**	**MDA**	**CYTO**	**MDA**	**CYTO**
1.27	9.00	0.62	5.72	2.73	11.68	2.85	29.50
1.38	6.13	2.46	4.01	1.41	19.10	3.50	45.06
3.83	6.06	7.63	6.09	6.08	36.70	4.80	8.99
0.16	6.45	4.60	4.53	5.44	48.30	1.89	48.15
0.56	4.78	12.21	19.01	4.25	4.47	2.90	10.12
1.95	34.76	1.03	9.62	1.76	8.83	0.87	17.98
0.08	15.53	1.25	7.59	8.39	5.49	4.25	37.18
2.17	12.23	2.13	12.33	2.82	3.48	1.43	19.09
0.00	0.93	0.98	5.26	5.03	7.98	6.75	6.05
1.35	3.81	1.53	5.69	7.30	27.04	4.30	17.05
3.22	6.39	3.91	7.72	4.97	16.02	0.59	7.79
1.69	14.15	2.25	7.61	1.11	6.14	5.30	6.78
4.90	5.67	1.67	4.32	13.27	13.31	1.80	16.03
1.33	8.49	5.23	17.79	7.73	10.03	3.50	5.07
0.64	2.27	2.79	15.51	3.69	17.23	4.98	16.60
5.21	12.35	1.43	12.43	9.26	9.29	6.98	19.89

Source: Data provided courtesy of Manoj Pandey, M.D.

11.2.3 The purpose of a study by Krantz et al. (A-4) was to investigate dose-related effects of methadone in subjects with *torsades de pointes*, a polymorphic ventricular tachycardia. In the study of 17 subjects, 10 were men (sex = 0) and seven were women (sex = 1). The outcome variable, is the QTc interval, a measure of arrhythmia risk. The other independent variable, in addition to sex, was methadone dose (mg/day). Measurements on these variables for the 17 subjects were as follows.

Sex	Dose (mg/day)	QTc (msec)
0	1000	600
0	550	625
0	97	560
1	90	585
1	85	590
1	126	500
0	300	700
0	110	570
1	65	540
1	650	785
1	600	765
1	660	611
1	270	600
1	680	625
0	540	650
0	600	635
1	330	522

Source: Data provided courtesy of Mori J. Krantz, M.D.

For each exercise in Sec 11.4 do the following:

(a) Compute the logistic regression equation
(b) Draw a scatter diagram of the data with the logistic equation
(c) Perform appropriate hypothesis tests and construct appropriate confidence intervals using your choice of significance and confidence levels.
(d) Find the p value for each test that you perform.

Further Reading We have discussed only the basic concepts and applications of logistic regression. The technique has much wider application. Stepwise regression analysis may be used with logistic regression. There are also techniques available for constructing confidence intervals for odds ratios. The reader who wishes to learn more about logistic regression may consult the books by Hosmer and Lemeshow (2) and Kleinbaum (3).

EXERCISES

11.4.1 In a study of violent victimization of women and men, Porcerelli et al. (A-11) collected information from 679 women and 345 men ages 18 to 64 years at several family-practice centers in the metropolitan Detroit area. Patients filled out a health history questionnaire that included a question about victimization. The following table shows the sample subjects cross-classified by gender and whether the subject self-identified as being "hit, kicked, punched, or otherwise hurt by someone within the past year." Subjects answering yes to that question are classified "violently victimized." Use logistic regression analysis to find the regression coefficients and the estimate of the odds ratio. Write an interpretation of your results.

Victimization	Women	Men	Total
No victimization	611	308	919
Violently victimized	68	37	105
Total	679	345	1024

Source: John H. Porcerelli, Rosemary Cogan, Patricia P. West, Edward A. Rose, Dawn Lambrecht, Karen E. Wilson, Richard K. Severson, and Dunia Karana, "Violent Victimization of Women and Men: Physical and Psychiatric Symptoms," *Journal of the American Board of Family Practice, 16* (2003), 32–39.

11.4.2 Refer to the research of Gallagher et al. (A-10) discussed in Example 11.4.2. Another covariate of interest was a score using the Hospital Anxiety and Depression Index. A higher value for this score indicates a higher level of anxiety and depression. Use the following data to predict whether a woman in the study participated in a cardiac rehabilitation program.

Hospital Anxiety and Depression Index Scores for Nonparticipating Women				Hospital Anxiety and Depression Index Scores for Participating Women	
17	14	19	16	23	25
7	21	6	9	3	6
19	13	8	22	24	29
16	15	13	17	13	22
23	21	4	14	26	11
27	12	15	14	19	12
23	9	23	5	25	20
18	29	19	5	15	18
21	4	14	14	22	24
27	18	19	20	13	18

(*Continued*)

Hospital Anxiety and Depression Index Scores for Nonparticipating Women					Hospital Anxiety and Depression Index Scores for Participating Women
14	22	17	21	21	8
25	5	13	17	15	10
19	27	14	17	12	17
23	16	14	10	25	14
6	11	17	13	29	21
8	19	26	10	17	25
15	23	15	20	21	25
30	22	19	3	8	16
18	25	16	18	19	23
10	11	10	9	16	19
29	20	15	10	24	24
8	11	22	5	17	11
12	28	8	15	26	17
27	12	15	13	12	19
12	19	20	16	19	20
9	18	12		13	17
16	13	2		23	31
6	12	6		11	0
22	7	14		17	18
10	12	19		29	18
9	14	14		6	15
11	13	19		20	

Source: Data provided courtesy of Robyn Gallagher, R.N., Ph.D.

11.5 SUMMARY

This chapter is included for the benefit of those who wish to extend their understanding of regression analysis and their ability to apply techniques to models that are more complex than those covered in Chapters 9 and 10. In this chapter we present some additional topics from regression analysis. We discuss the analysis that is appropriate when one or more of the independent variables is dichotomous. In this discussion the concept of dummy variable coding is presented. A second topic that we discuss is how to select the most useful independent variables when we have a long list of potential candidates. The technique we illustrate for the purpose is stepwise regression analysis. Finally, we present the basic concepts and procedures that are involved in logistic regression analysis. We cover two situations: the case in which the independent variable is dichotomous, and the case in which the independent variable is continuous.

Since the calculations involved in obtaining useful results from data that are appropriate for analysis by means of the techniques presented in this chapter are complicated and time-consuming when attempted by hand, it is recommended that a computer be used to work the exercises.