This chapter also presented ***analysis of variance*** (ANOVA) methods for comparing several groups according to their means on a quantitative response variable. The groups are categories of categorical explanatory variables.

- Analysis of variance methods are a special case of multiple regression analyses. ***Dummy variables*** in the regression model are indicators that represent the groups. Each dummy variable equals 1 for a particular group and 0 otherwise.
- ***Multiple comparison*** methods provide confidence intervals for differences between pairs of means, while controlling the overall error probability. The ***Bonferroni*** method does this using an error probability for each comparison that equals the desired overall error probability divided by the number of comparisons.
- ***One-way ANOVA*** methods compare means for categories of a single factor. ***Two-way ANOVA*** methods compare means across categories of each of two factors. Assuming no interaction, the main effects describe the effect of each factor while controlling for the other one. Ordinary ANOVA methods compare groups with *independent* random samples from the groups.
- For longitudinal and repeated-measures studies, different samples have the same subjects, and are *dependent*. Methods for ***repeated-measures ANOVA*** result from regression models with ***random effects*** that represent the effects of the random sample of observed subjects. Such methods treat ***within-subjects*** effects (for repeated measurements on subjects) differently from ***between-subjects*** effects (for independent samples of subjects).

Exercises

Practicing the Basics

12.1. For GSS data comparing the reported number of good friends for those who are (married, widowed, divorced, separated, never married), an ANOVA table reports $F = 0.80$.

(a) Specify the null and alternative hypotheses for the test.

(b) Software reports a P-value of 0.53. Explain how to interpret it.

(c) State the hypotheses tested in terms of parameters of a regression model. Define variables in that model.

12.2. A General Social Survey asked subjects how many good friends they have. Is this associated with the respondent's astrological sign (the 12 symbols of the zodiac)? The ANOVA table for the GSS data reports $F = 0.61$ based on $df_1 = 11, df_2 = 813$.

(a) Specify the null and alternative hypotheses for the analysis.

(b) Software reports a P-value of 0.82. Explain how to interpret it.

(c) State a regression model and corresponding null hypothesis that can yield these results. Define variables in that model.

12.3. A recent General Social Survey asked, "What is the ideal number of kids for a family?" Show how to define dummy variables, and formulate a model for this response with explanatory variable religious affiliation (Christian, Muslim, Jewish, Other or none).

12.4. Refer to the previous exercise. Table 12.22 shows an ANOVA table for the model.

(a) Specify the hypotheses tested in this table.

(b) Report the F test statistic value and the P-value. Interpret the P-value.

(c) Based on (b), can you conclude that *every* pair of religious affiliations has different population means for ideal family size? Explain.

TABLE 12.22

Source	SS	df	Mean Square	F	Sig (Prob>F)
Religion	11.72	3	3.91	5.48	0.001
Residual (Error)	922.82	1295	0.71		
Total	934.54	1298			

12.5. A recent GSS asked, "How often do you go to a bar or tavern?" Table 12.23 shows descriptive statistics and an ANOVA table for comparing the mean reported number of good friends at three levels of this variable.

(a) State the (i) hypotheses, (ii) test statistic value, (iii) P-value, (iv) decision for an $\alpha = 0.05$-level test.

(b) Does any aspect of the summary here suggest that an assumption for the F test may be badly violated? Explain.

(c) Set up dummy variables, and show the prediction equation you would obtain, based on the results shown in the table.

TABLE 12.23

	How often go to bar or tavern?		
	Very often	Occasional	Never
Mean no. good friends	12.1	6.4	6.2
Standard deviation	21.3	10.2	14.0
Sample size	41	166	215

Source	Sum of Squares	df	Mean Square	F	Prob>F
Group	1116.8	2	558.4	3.03	0.049
Residual (Error)	77171.8	419	184.2		
Total	78288.5	421			

12.6. Table 12.24 shows scores on the first quiz (maximum score 10 points) in a beginning French course. Students in the course are grouped as follows:

Group A: Never studied foreign language before, but have good English skills.
Group B: Never studied foreign language before; have poor English skills.
Group C: Studied other foreign language.

Using software for regression or ANOVA, conduct a test comparing the means. Report the assumptions, hypotheses, test statistic, and P-value. Interpret the P-value.

TABLE 12.24

Group A	Group B	Group C
4	1	9
6	5	10
8		5

12.7. Refer to the previous exercise.

(a) Suppose that the first observation in the second group was actually 9, not 1. Then, the standard deviations are the same, but the sample means are 6, 7, and 8 rather than 6, 3, and 8. Do you think the F test statistic would be larger, the same, or smaller? Explain your reasoning, without doing any calculations.

(b) Suppose you had the same means as these data, but the sample standard deviations were 1.0, 1.8, and 1.6, instead of the actual 2.0, 2.8, and 2.6. Do you think the F test statistic would be larger, the same, or smaller? Explain your reasoning.

(c) Suppose you had the same means and standard deviations as these data, but the sample sizes were 30, 20, and 30, instead of 3, 2, and 3. Do you think the F test statistic would be larger, the same, or smaller? Explain your reasoning.

(d) In (a), (b), and (c), would the P-value be larger, the same, or smaller? Why?

12.8. In a study to compare customer satisfaction at service centers for PC technical support in San Jose (California), Toronto (Canada), and Bangalore (India), each center randomly sampled 100 people who called during a two-week period. Callers rated their satisfaction on a scale of 0 to 10, with higher scores representing greater satisfaction. The sample means were 7.6 for San Jose, 7.8 for Toronto, and 7.1 for Bangalore. Table 12.25 shows an ANOVA table.

(a) Explain how to obtain the F test statistic value reported in the table from the mean square values shown. Report the df_1 and df_2 values for the F distribution, and report and interpret the P-value.

TABLE 12.25

Source	SS	df	MS	F	Prob>F
Group	26.00	2	13.00	27.6	0.000
Residual (Error)	140.00	297	0.47		
Total	60.00	299			

(b) Explain why the margin of error for separate 95% confidence intervals is the same (0.19) for comparing the population means for each pair of cities. Construct and interpret the three intervals.

(c) The margin of error for Bonferroni or for Tukey 95% multiple comparison confidence intervals is 0.23. Why is it different from that in (b), and what is an advantage of this approach?

(d) With dummy variables to represent the service centers, the prediction equation is $\hat{y} = 7.1 + 0.5z_1 + 0.7z_2$. Show how the terms in this equation relate to the sample means of 7.6, 7.8, and 7.1.

12.9. For g groups with $n = 100$ each, we plan to compare all pairs of population means. We want the probability to equal at least 0.80 that the entire set of confidence intervals contains the true differences. For the Bonferroni method, which t-score multiple of the standard error should we use for each interval if **(a)** $g = 10$, **(b)** $g = 5$? Describe how the t-score depends on g, and explain the implication regarding width of the intervals.

12.10. A recent GSS asked, "Would you say that you are very happy, pretty happy, or not too happy?" and "About how many good friends do you have?" Table 12.26 summarizes results, with number of friends as the response variable.

(a) State a research question you could answer with these data.

(b) Interpret the result of the F test, but indicate one assumption of the test that is clearly violated.

(c) Software reports Tukey 95% confidence intervals of (0.3, 5.7) comparing very happy and pretty happy, (−2.3, 6.5) comparing very happy and not too happy, and (−5.1, 3.3) comparing pretty happy and not too happy. Interpret.

12.11. When we use the GSS to evaluate how the mean number of hours a day watching TV depends on sex and race, for subjects of age 18–25, we get the results shown in Table 12.27. The sample means were 2.66 for white females, 2.62 for white males, 3.48 for black females, and 3.14 for black males. Explain how these results seem to be compatible with the results of the tests shown.

12.12. A recent GSS asked, "What is the ideal number of kids for a family?" Table 12.28 shows results of evaluating the effects of gender and race.

(a) Explain how to interpret the results of the F tests.

(b) Let $s = 1$ for females and 0 for males, and let $r = 1$ for blacks and 0 for whites. The no interaction model has $\hat{y} = 2.42 + 0.04s + 0.37r$. Find the estimated mean for each combination of gender and race. Explain how these means satisfy "no interaction."

TABLE 12.26

	Very happy	Pretty happy	Not too happy
Mean	10.4	7.4	8.3
Standard deviation	17.8	13.6	15.6
Sample size	276	468	87

Source	Sum of Squares	df	MS	F	Prob>F
Group	1626.8	2	813.4	3.47	0.032
Residual (Error)	193900.9	828	234.2		
Total	195527.7	830			

TABLE 12.27

Source	SS	df	MS	F	Prob>F
Sex	2.22	1	2.22	0.35	0.555
Race	489.65	1	489.65	76.62	0.000
Residual (Error)	11094.16	1737	6.39		
Total	11583.81	1739			

TABLE 12.28

Source	SS	df	MS	F	P-value
Gender	0.25	1	0.25	0.36	0.550
Race	16.98	1	16.98	24.36	0.000
Residual	868.67	1246	0.70		
Total	886.12	1248			