are not as large as those for the group effect, and this is reflected in the smaller value of the F statistic.

The analysis indicates that a complete description of the average heart rates requires consideration of the interaction in addition to the main effects. The two lines in the plot are not parallel. This interaction can be described in two ways. The female-male difference in average heart rates is greater for the controls than for the runners. Alternatively, the difference in average heart rates between controls and runners is greater for women than for men. As the plot suggests, the interaction is not large. It is statistically significant because there were 800 subjects in the study.

Two-way ANOVA output for other software is similar to that given by SAS. Figure 13.7 gives the analysis of the heart rate data using Excel and Minitab.

Excel

	F	G	H	I	J	K	L	M
1	Anova Two Factor With Replication							
2								
3	SUMMARY	Control	Runners	Total				
4	*Female*							
5	Count	200	200	400				
6	Sum	29600	23197	52797				
7	Average	148	115.985	131.9925				
8	Variance	264.7437	255.0902	516.1478				
9								
10	*Male*							
11	Count	200	200	400				
12	Sum	26000	20795	46795				
13	Average	130	103.975	116.9875				
14	Variance	292.4221	156.2356	393.5161				
15								
16	*Total*							
17	Count	400	400					
18	Sum	55600	43992					
19	Average	139	109.98					
20	Variance	359.0877	241.2978					
21								
22								
23	ANOVA							
24	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
25	Sample	45030.01	1	45030.01	185.9799	3.29E-38	3.85316	
26	Columns	168432.1	1	168432.1	695.6471	1.10E-110	3.85316	
27	Interaction	1794.005	1	1794.005	7.409481	0.00663	3.85316	
28	Within	192729.8	796	242.1229				
29								
30	Total	407985.9	799					

Minitab

```
Analysis of Variance for HR
Source          DF        SS        MS
Group            1    168432    168432
Gender           1     45030     45030
Interaction      1      1794      1794
Error          796    192730       242
Total          799    407986
```

FIGURE 13.7 Excel and Minitab two-way ANOVA output for the heart rate study, for Example 13.10.

CHAPTER 13 Summary

Two-way analysis of variance is used to compare population means when populations are classified according to two factors.

ANOVA assumes that the populations are Normal with possibly different means and the same standard deviation and that independent SRSs are drawn from each population.

As with one-way ANOVA, preliminary analysis includes examination of means, standard deviations, and Normal quantile plots. **Marginal means** are calculated by taking averages of the cell means across rows and columns. Pooling is used to estimate the within-group variance.

ANOVA separates the total variation into parts for the **model** and **error.** The model variation is separated into parts for each of the **main effects** and the **interaction.**

The calculations are organized into an **ANOVA table.** F statistics and P-values are used to test hypotheses about the main effects and the interaction.

Careful inspection of the means is necessary to interpret significant main effects and interactions. Plots are a useful aid.

CHAPTER 13 Exercises

For Exercises 13.1 and 13.2, see pages 679–680.

13.3 What's wrong? For each of the following, explain what is wrong and why.

(a) You should reject the null hypothesis that there is no interaction in a two-way ANOVA when the AB F statistic is small.

(b) Sums of squares are equal to mean squares divided by degrees of freedom.

(c) The test statistics for the main effects in a two-way ANOVA have a chi-square distribution when the null hypothesis is true.

(d) The sums of squares always add in two-way ANOVA.

13.4 Is there an interaction? Each of the following tables gives means for a two-way ANOVA. Make a plot of the means with the levels of factor A on the x axis. State whether or not there is an interaction, and if there is, describe it.

(a)

	Factor A		
Factor B	**1**	**2**	**3**
1	11	16	21
2	6	16	26

(b)

	Factor A		
Factor B	**1**	**2**	**3**
1	10	15	20
2	40	45	50

(c)

	Factor A		
Factor B	**1**	**2**	**3**
1	10	15	20
2	50	45	40

(d)

	Factor A		
Factor B	**1**	**2**	**3**
1	10	15	20
2	50	55	52

13.5 Describing a two-way ANOVA model. A 3×2 ANOVA was run with 5 observations per cell.

(a) Give the degrees of freedom for the F statistic that is used to test for interaction in this analysis and the entries from Table E that correspond to this distribution.

(b) Sketch a picture of this distribution with the information from the table included.

(c) The calculated value of this F statistic is 3.72. How would you report the P-value?

(d) Would you expect a plot similar to Figure 13.1 to have mean profiles that look parallel? Explain your answer.

13.6 Determining the critical value of *F*. For each of the following situations, state how large the F statistic needs to be for rejection of the null hypothesis at the 5% level. Sketch each distribution and indicate the region where you would reject.

(a) The main effect for the first factor in a 2×4 ANOVA with 3 observations per cell.

(b) The interaction in a 4×4 ANOVA with 2 observations per cell.

(c) The interaction in a 2×2 ANOVA with 51 observations per cell.

13.7 Identifying the factors of a two-way ANOVA model. For each of the following situations, identify both factors and the response variable. Also, state the number of levels for each factor (I and J) and the total number of observations (N).

(a) A child psychologist is interested in studying how a child's percent of pretend play differs with gender and age (4, 8, and 12 months). There are 11 infants assigned to each cell of the experiment.

(b) Brewers malt is produced from germinating barley. A homebrewer wants to determine the best conditions to germinate the barley. A total of 30 lots of barley seed were equally and randomly assigned to 10 germination conditions. The conditions are combinations of the week after harvest (1, 3, 6, 9, or 12 weeks) and the amount of water used in the process (4 or 8 milliliters). The percent of seeds germinating is the outcome variable.

(c) The strength of concrete depends upon the formula used to prepare it. An experiment compares six different mixtures. Nine specimens of concrete are poured from each mixture. Three of these specimens are subjected to 0 cycles of freezing and thawing, three are subjected to 100 cycles, and three are subjected to 500 cycles. The strength of each specimen is then measured.

(d) A company wants to compare four different training programs for its new employees. Each of these programs takes 6 hours to complete. The training can be given for 6 hours on one day or for 3 hours on two consecutive days. The next 80 employees hired by the company will be the subjects for this study.

13.8 Determining the degrees of freedom. For each part in Exercise 13.7, outline the ANOVA table, giving the sources of variation and the degrees of freedom.

13.9 The influences of transaction history and a thank-you statement. A service failure is defined as any service-related problem (real or perceived) that transpires during a customer's experience with a firm. In the hotel industry, there is a high human component so these sorts of failures commonly occur regardless of extensive training and established policies. As a result, hotel firms must learn to effectively react to these failures. A recent study investigated the relationship between a consumer's transaction history (levels: long and short) and an employee thank-you statement (levels: yes and no) on a consumer's repurchase intent.[8] Each subject was randomly assigned to one of the four treatment groups and asked to read some service failure/resolution scenarios and respond accordingly. Repurchase intent was measured using a 9-point scale. Here is a summary of the means:

	Thank-you	
History	**No**	**Yes**
Short	5.69	6.80
Long	7.53	7.37

(a) Plot the means. Do you think there is an interaction? If yes, describe the interaction in terms of the two factors.

(b) Find the marginal means. Are they useful for understanding the results of this study? Explain your answer.

13.10 Transaction history and a thank-you statement, cont. Refer to the previous exercise. The number of subjects in each cell were not equal so the researchers used linear regression to analyze the data. This was done by creating an indicator variable for each factor and the interaction. Below is a partial ANOVA table. Complete it and state your conclusions regarding the main effects and interaction described in the previous exercise.

Source	DF	SS	MS	F	P-value
Transaction history		61.445			
Thank-you statement		21.810			
Interaction		15.404			
Error	160	759.904			

13.11 The effects of proximity and visibility on food intake. A recent study investigated the influence that proximity and visibility of food have on food intake.[9] A total of 40 secretaries from the University of Illinois participated in the study. A candy dish full of individually wrapped chocolates was placed either at the desk of the participant or at a location 2 meters from the participant. The candy dish was either a clear (candy visible) or opaque (candy not visible) covered bowl. After a week, the researchers noted not only the number of candies consumed per day but also the self-reported number of candies consumed by each participant. The following table summarizes the mean difference between these two values (reported minus actual):

	Visibility	
Proximity	**Clear**	**Opaque**
Proximate	−1.2	−0.8
Less proximate	0.5	0.4

(a) Make a plot of the means and describe the patterns that you see. Does the plot suggest an interaction between visibility and proximity?

(b) This study actually took four weeks, with each participant being observed at each treatment combination in a random order. Explain why a "repeated-measures" design like this may be beneficial.

13.12 Hypotension and endurance exercise. In sedentary individuals, low blood pressure (hypotension) often occurs after a single bout of aerobic exercise and lasts nearly two hours. This can cause dizziness, light-headedness, and possibly fainting upon standing. It is thought that endurance exercise training can reduce the degree of postexercise hypotension. To test this, researchers studied 16 endurance-trained and 16 sedentary men and women.[10] The following table summarizes the postexercise systolic arterial pressure (mmHg) after 60 minutes of upright cycling:

Group	n	$\bar{x}$	**Std Err**
Women, sedentary	8	100.7	3.4
Women, endurance	8	105.3	3.6
Men, sedentary	8	114.2	3.8
Men, endurance	8	110.2	2.3

(a) Make a plot similar to Figure 13.3 with the systolic blood pressure on the y axis and training level on the x axis. Describe the pattern you see.

(b) From the table, one can show that SSA = 677.12, SSB = 0.72, SSAB = 147.92, and SSE = 2478 where A is the gender effect and B is the training level. Construct the ANOVA table with F statistics and degrees of freedom, and state your conclusions regarding main effects and interaction.

(c) The researchers also measured the before-exercise systolic blood pressure of the participants and looked at a model that incorporated both the pre- and postexercise values. Explain why it is likely beneficial to incorporate both measurements in the study.

13.13 The effect of humor. In advertising, humor is often used to overcome sales resistance and stimulate customer purchase behavior. A recent experiment looked at the use of humor as an approach to offset negative feelings often associated with web site encounters.[11] The setting of their experiment was an online travel agency and they used a three-factor design, each factor with two levels. They were humor (used, not used), process (favorable, unfavorable), and outcome (favorable, unfavorable). For the humor condition, cartoons and jokes of the day about skiing were used on the site. For the no humor condition, standard pictures of ski sites were used. Two hundred and forty-one business students from a large Dutch university participated in the experiment. Each was randomly assigned to one of the eight treatment conditions. The students were asked to book a skiing holiday and then rate their perceived enjoyment and satisfaction with the process. All responses were measured on a 7-point Likert scale. A summary of the results for satisfaction follows:

Treatment	n	$\bar{x}$	s
No humor - favorable process - unfavorable outcome	27	3.04	0.79
No humor - favorable process - favorable outcome	29	5.36	0.47
No humor - unfavorable process - unfavorable outcome	26	2.84	0.59
No humor - unfavorable process - favorable outcome	31	3.08	0.59
Humor - favorable process - unfavorable outcome	32	5.06	0.59
Humor - favorable process - favorable outcome	30	5.55	0.65
Humor - unfavorable process - unfavorable outcome	36	1.95	0.52
Humor - unfavorable process - favorable outcome	30	3.27	0.71

(a) Plot the means of the four treatments without humor. Do you think there is an interaction? If yes, describe the interaction in terms of the process and outcome factors.

(b) Plot the means of the four treatments that used humor. Do you think there is an interaction? If yes, describe the interaction in terms of the process and outcome factors.

(c) The three-factor interaction can be assessed by looking at the two interaction plots created in parts (a) and (b). If the relationship between process and outcome is different across the two humor conditions, there is evidence of an interaction among all three factors. Do you think there is a three-factor interaction? Explain your answer.

13.14 Pooling the standard deviations. Refer to the previous exercise. Find the pooled estimate of the standard deviation for these data. What are its degrees of freedom? Using the rule from Chapter 12 (page 633), is it

(a) What can you conclude from the information given?

(b) What additional information would you need to write a summary of the results for this study?

13.25 Conspicuous consumption and men's testosterone levels. It is argued that conspicuous consumption is a means by which men communicate their social status to prospective mates. One study looked at changes in a male's testosterone level in response to fluctuations in his status created by the consumption of a product.[14] The products considered were a new and luxurious sports car, and an old family sedan. Participants were asked to drive either on an isolated highway or a busy urban street. A table of cell means and standard deviations for the change (post-pre) in testosterone follows:

	Location			
	Highway		City	
Car	$\bar{x}$	s	$\bar{x}$	s
Old sedan	0.03	0.12	−0.03	0.12
New sports car	0.15	0.14	0.13	0.13

(a) Make a plot of the means and describe the patterns that you see. Does the plot suggest an interaction between location and type of car?

(b) Compute the pooled standard error, s_p, assuming equal sample sizes.

(c) The researchers wanted to test the following hypotheses:

(1) Testosterone levels will increase more in men who drive the new car

(2) For men driving the new car, testosterone levels will increase more in men who drive in the city

(3) For men driving the old car, testosterone levels will decrease less in men who drive the old car on the highway

Write out the contrasts for each of these hypotheses.

(d) This study actually involved each male participating in all four combinations. Half of them drove the sedan first and the other half drove the sports car first. Explain why a "repeated-measures" design like this may be beneficial.

13.26 The effects of peer pressure on mathematics achievement. Researchers were interested in comparing the relationship between high achievement in mathematics and peer pressure across several countries.[15] They hypothesized that in countries where high achievement is not valued highly, considerable peer pressure may exist. A questionnaire was distributed to 14-year-olds from three countries (Germany, Canada, and Israel). One of the questions asked students to rate how often they fear being called a nerd or teacher's pet on a 4-point scale (1 = never, 4 = frequently). The following table summarizes the response:

Country	Gender	n	$\bar{x}$
Germany	Female	336	1.62
Germany	Male	305	1.39
Israel	Female	205	1.87
Israel	Male	214	1.63
Canada	Female	301	1.91
Canada	Male	304	1.88

(a) The P-values for the interaction and the main effects of country and gender are 0.016, 0.068, and 0.108, respectively. Using the table and P-values, summarize the results both graphically and numerically.

(b) The researchers contend that Germany does not value achievement as highly as Canada and Israel. Do the results from part (a) allow you to address their primary hypothesis? Explain.

(c) The students were also asked to indicate their current grade in mathematics on a 6-point scale (1 = excellent, 6 = insufficient). How might both responses be used to address the researchers' primary hypothesis?

13.27 The effect of chromium on insulin metabolism. The amount of chromium in the diet has an effect on the way the body processes insulin. In an experiment designed to study this phenomenon, four diets were fed to male rats. There were two factors. Chromium had two levels: low (L) and normal (N). The rats were allowed to eat as much as they wanted (M) or the total amount that they could eat was restricted (R). We call the second factor Eat. One of the variables measured was the amount of an enzyme called GITH.[16] The means for this response variable appear in the following table:

	Eat	
Chromium	M	R
L	4.545	5.175
N	4.425	5.317

(a) Make a plot of the mean GITH for these diets, with the factor Chromium on the x axis and GITH on the y axis. For each Eat group, connect the points for the two Chromium means.

(b) Describe the patterns you see. Does the amount of chromium in the diet appear to affect the GITH mean? Does restricting the diet rather than letting the rats eat as much as they want appear to have an effect? Is there an interaction?

(c) Compute the marginal means. Compute the differences between the M and R diets for each level of Chromium. Use this information to summarize numerically the patterns in the plot.

13.28 Changing your major. A study of undergraduate computer science students examined changes in major after the first year.[17] The study examined the fates of 256 students who enrolled as first-year computer science students in the same fall semester. The students were classified according to gender and their declared major at the beginning of the second year. For convenience we use the labels CS for computer science majors, EO for engineering and other science majors, and O for other majors. The explanatory variables included several high school grade summaries coded as 10 = A, 9 = A−, etc. Here are the mean high school mathematics grades for these students:

	Major		
Gender	CS	EO	O
Males	8.68	8.35	7.65
Females	9.11	9.36	8.04

Describe the main effects and interaction using appropriate graphs and calculations.

13.29 More on changing your major. The mean SAT Mathematics scores for the students in the previous exercise are summarized in the following table:

	Major		
Gender	CS	EO	O
Males	628	618	589
Females	582	631	543

Summarize the results of this study using appropriate plots and calculations to describe the main effects and interaction.

13.30 Designing a study. The students studied in the previous two exercises were enrolled at a large Midwestern university more than two decades ago. Discuss how you would conduct a similar study at a college or university of your choice today. Include a description of all variables that you would collect for your study.

13.31 Trust of individuals and groups. Trust is an essential element in any exchange of goods or services. The following trust game is often used to study trust experimentally:

A *sender* starts with \$X and can transfer any amount $x \le X$ to a *responder*. The responder then gets $\$3x$ and can transfer any amount $y \le 3x$ back to the sender. The game ends with final amounts $X - x + y$ and $3x - y$ for the sender and responder respectively.

The value x is taken as a measure of the sender's trust and the value $y/3x$ indicates the responder's trustworthiness. A recent study used this game to study the dynamics between individuals and groups of three.[18] The following table summarizes the average amount x sent by senders starting with \$100.

Sender	Responder	n	$\bar{x}$	s
Individual	Individual	32	65.5	36.4
Individual	Group	25	76.3	31.2
Group	Individual	25	54.0	41.6
Group	Group	27	43.7	42.4

(a) Find the pooled estimate of the standard deviation for this study and its degrees of freedom.

(b) Is it reasonable to use a pooled standard deviation for the analysis? Explain your answer.

(c) Compute the marginal means.

(d) Plot the means. Do you think there is an interaction? If yes, then describe it.

(e) The F statistics for sender, responder, and interaction are 9.05, 0.001, and 2.08, respectively. Compute the P-values and state your conclusions.

13.32 Does the type of cooking pot affect iron content? Iron-deficiency anemia is the most common form of malnutrition in developing countries, affecting about 50% of children and women and 25% of men. Iron pots for cooking foods had traditionally been used in many of these countries, but they have been largely replaced by aluminum pots, which are cheaper and lighter. Some research has suggested that food cooked in iron pots will contain more iron than food cooked in other types of pots. One study designed to investigate this issue compared the iron content of some Ethiopian foods cooked in aluminum, clay, and iron pots.[19] In Exercise 12.49 (page 665), we analyzed the iron content of *yesiga wet'*, beef cut into small pieces and prepared with several Ethiopian spices. The researchers who conducted this study also examined the iron content of *shiro wet'*, a legume-based mixture of chickpea flour and Ethiopian spiced pepper, and *ye-atkilt allych'a*, a lightly spiced vegetable casserole. In the following table, these three foods are labeled meat, legumes, and vegetables. Four samples of each food were cooked in each type of pot. The iron in the food is measured in milligrams of iron per 100 grams of cooked food. The data are shown in Table 13.2 (page 692). IRONCONTENT

(a) Make a table giving the sample size, mean, and standard deviation for each type of pot. Is it reasonable to pool the variances? Although the standard deviations

TABLE 13.2

Iron content

Type of pot	Meat				Legumes				Vegetables			
Aluminum	1.77	2.36	1.96	2.14	2.40	2.17	2.41	2.34	1.03	1.53	1.07	1.30
Clay	2.27	1.28	2.48	2.68	2.41	2.43	2.57	2.48	1.55	0.79	1.68	1.82
Iron	5.27	5.17	4.06	4.22	3.69	3.43	3.84	3.72	2.45	2.99	2.80	2.92

vary more than we would like, this is partially due to the small sample sizes and we will proceed with the analysis of variance.

(b) Plot the means. Give a short summary of how the iron content of foods depends upon the cooking pot.

(c) Run the analysis of variance. Give the ANOVA table, the F statistics with degrees of freedom and P-values, and your conclusions regarding the hypotheses about main effects and interactions.

13.33 Interpreting the results. Refer to the previous exercise. Although there is a statistically significant interaction, do you think that these data support the conclusion that foods cooked in iron pots contain more iron than foods cooked in aluminum or clay pots? Discuss.

13.34 Analysis using a one-way ANOVA. Refer to Exercise 13.32. Rerun the analysis as a one-way ANOVA with 9 groups and 4 observations per group. Report the results of the F test. Examine differences in means using a multiple-comparisons procedure. Summarize your results and compare them with those you obtained in Exercise 13.32.

13.35 Examination of a drilling process. One step in the manufacture of large engines requires that holes of very precise dimensions be drilled. The tools that do the drilling are regularly examined and are adjusted to ensure that the holes meet the required specifications. Part of the examination involves measurement of the diameter of the drilling tool. A team studying the variation in the sizes of the drilled holes selected this measurement procedure as a possible cause of variation in the drilled holes. They decided to use a designed experiment as one part of this examination. Some of the data are given in Table 13.3. The diameters in millimeters (mm) of five tools were measured by the same operator at three times (8:00 A.M., 11:00 A.M., and 3:00 P.M.). Three measurements were taken on each tool at each time. The person taking the measurements could not tell which tool was being measured, and the measurements were taken in random order.[20] DRILLING

(a) Make a table of means and standard deviations for each of the 5 × 3 combinations of the two factors.

TABLE 13.3

Tool diameter data

Tool	Time	Diameter (mm)		
1	1	25.030	25.030	25.032
1	2	25.028	25.028	25.028
1	3	25.026	25.026	25.026
2	1	25.016	25.018	25.016
2	2	25.022	25.020	25.018
2	3	25.016	25.016	25.016
3	1	25.005	25.008	25.006
3	2	25.012	25.012	25.014
3	3	25.010	25.010	25.008
4	1	25.012	25.012	25.012
4	2	25.018	25.020	25.020
4	3	25.010	25.014	25.018
5	1	24.996	24.998	24.998
5	2	25.006	25.006	25.006
5	3	25.000	25.002	24.999

(b) Plot the means and describe how the means vary with tool and time. Note that we expect the tools to have slightly different diameters. These will be adjusted as needed. It is the process of measuring the diameters that is important.

(c) Use a two-way ANOVA to analyze these data. Report the test statistics, degrees of freedom, and P-values for the significance tests.

(d) Write a short report summarizing your results.

13.36 Examination of a drilling process, continued. Refer to the previous exercise. Multiply each measurement by 0.04 to convert from millimeters to inches. Redo the plots and rerun the ANOVA using the transformed measurements. Summarize what parts of the analysis have changed and what parts have remained the same. DRILLING

13.37 Do left-handed people live shorter lives than right-handed people? A study of this question examined a sample of 949 death records and contacted next of kin to determine handedness.[21] Note that there are many possible definitions of "left-handed." The researchers

examined the effects of different definitions on the results of their analysis and found that their conclusions were not sensitive to the exact definition used. For the results presented here, people were defined to be right-handed if they wrote, drew, and threw a ball with the right hand. All others were defined to be left-handed. People were classified by gender (female or male) and handedness (left or right), and a 2 × 2 ANOVA was run with the age at death as the response variable. The F statistics were 22.36 (handedness), 37.44 (gender), and 2.10 (interaction). The following marginal mean ages at death (in years) were reported: 77.39 (females), 71.32 (males), 75.00 (right-handed), and 66.03 (left-handed).

(a) For each of the F statistics given, find the degrees of freedom and an approximate P-value. Summarize the results of these tests.

(b) Using the information given, write a short summary of the results of the study.

13.38 A radon exposure study. Scientists believe that exposure to the radioactive gas radon is associated with some types of cancers in the respiratory system. Radon from natural sources is present in many homes in the United States. A group of researchers decided to study the problem in dogs because dogs get similar types of cancers and are exposed to environments similar to those of their owners. Radon detectors are available for home monitoring, but the researchers wanted to obtain actual measures of the exposure of a sample of dogs. To do this, they placed the detectors in holders and attached them to the collars of the dogs. One problem was that the holders might in some way affect the radon readings. The researchers therefore devised a laboratory experiment to study the effects of the holders. Detectors from four series of production were available, so they used a two-way ANOVA design (series with 4 levels and holder with 2, representing the presence or absence of a holder). All detectors were exposed to the same radon source and the radon measure in picocuries per liter was recorded.[22] The F statistic for the effect of series is 7.02, for holder it is 1.96, and for the interaction it is 1.24.

(a) Using Table E or statistical software, find approximate P-values for the three test statistics. Summarize the results of these three significance tests.

(b) The mean radon readings for the four series were 330, 303, 302, and 295. The results of the significance test for series were of great concern to the researchers. Explain why.

13.39 A comparison of plant species under low water conditions. The PLANTS1 data set gives the percent of nitrogen in four different species of plants grown in a laboratory. The species are *Leucaena leucocephala*, *Acacia saligna*, *Prosopis juliflora*, and *Eucalyptus citriodora*. The researchers who collected these data were interested in commercially growing these plants in parts of the country of Jordan where there is very little rainfall. To examine the effect of water, they varied the amount per day from 50 millimeters (mm) to 650 mm in 100 mm increments. There were nine plants per species-by-water combination. Because the plants are to be used primarily for animal food, with some parts that can be consumed by people, a high nitrogen content is very desirable. PLANTS1

(a) Find the means for each species-by-water combination. Plot these means versus water for the four species, connecting the means for each species by lines. Describe the overall pattern.

(b) Find the standard deviations for each species-by-water combination. Is it reasonable to pool the standard deviations for this problem? Note that with sample sizes of size 9, we expect these standard deviations to be quite variable.

(c) Run the two-way analysis of variance. Give the results of the hypothesis tests for the main effects and the interaction.

13.40 Examination of the residuals. Refer to the previous exercise. Examine the residuals. Are there any unusual patterns or outliers? If you think that there are one or more points that are somewhat extreme, rerun the two-way analysis without these observations. Does this change the results in any substantial way? PLANTS1

13.41 Analysis using multiple one-way ANOVAs. Refer to Exercise 13.39. Run a separate one-way analysis of variance for each water level. If there is evidence that the species are not all the same, use a multiple-comparisons procedure to determine which pairs of species are significantly different. In what way, if any, do the differences appear to vary by water level? Write a short summary of your conclusions. PLANTS1

13.42 More on the analysis using multiple one-way ANOVAs. Refer to Exercise 13.39. Run a separate one-way analysis of variance for each species and summarize the results. Since the amount of water is a quantitative factor, we can also analyze these data using regression. Run simple linear regressions separately for each species to predict nitrogen percent from water. Use plots to determine whether or not a line is a good way to approximate this relationship. Summarize the regression results and compare them with the one-way ANOVA results. PLANTS1

13.43 Another comparison of plant species under low water conditions. Refer to Exercise 13.39. Additional data collected by the same researchers according to a