Chapter 14

ANOVA: Comparisons of Several Populations

In Chapter 11, we demonstrate the use of *t*-tests for comparing the means of two populations (such as males and females). The analysis of variance (ANOVA) allows us to compare the means of a larger number of populations (i.e., three or more). Through analysis of variance, variability among the group means is compared to the variability within the groups. If between-group variability is substantially greater than within-group variability, the means are declared to be significantly different. The ANOVA allows you to answer questions such as:

- Is political affiliation (Democrat, Republican, Independent) related to the number of elections in which people vote?
- Do three different techniques for memorizing words have different impacts on the mean number of words recalled?
- Is annual movie income from movie sales related to the type of movie (e.g., comedy, drama, horror, action, family)?
- Is heart rate after a step-workout affected by the height of the step (low or high) and the frequency of stepping (slow, medium, fast)?

In this chapter, we will use the One-Way ANOVA procedure in SPSS. We will also discuss techniques for obtaining follow-up tests to determine which group means differ from which others, and effect sizes to determine the magnitude of the differences between specific groups. We will also explore the Analysis of Variance of Ranks, which is appropriate when the assumptions required for performing ANOVA are violated. Finally, we will use the General Linear

Model (Univariate) procedure to conduct an ANOVA when there is more than one independent variable.

14.1 ONE-WAY ANALYSIS OF VARIANCE

We illustrate the ANOVA procedure for one independent variable by comparing the relationship between three ways of organizing information and children's ability to memorize words. The data are contained in the "words.sav" file. One of the two variables in this file denotes the organizational method provided to the child (1 = no information, 2 = words divided into three categories, 3 = words divided into six categories), and the other variable denotes the number of words memorized by the child. We shall use SPSS to test the null hypothesis that the mean number of words memorized by the three groups are equal, that is, H_0: $\mu_1 = \mu_2 = \mu_3$. In this example, the number of observations in the three conditions is equal (6).

Examining the Data

It is important to examine data visually prior to conducting an ANOVA test. One method for doing so is to visually inspect a scatter plot of the data. Another is by creating a box-and-whisker plot for the dependent variable (here, number of words memorized) separately for each level of the independent variable (here, the type of organizational method used). A third method available through SPSS is producing an error bar chart. Scatter plots are described in Chapters 5 and 13 and box-and-whisker plots in Chapter 5. Therefore, we will illustrate the error bar chart in this section. After opening the "words.sav" data file:

1. Click on **Graphs** from the menu bar.
2. Click on **Error bar** from the pull down menu.
3. Click on **Simple** and **Summaries for groups of cases** in the Error Bar dialog box.
4. Click on **Define** to open the Define Simple Error Bar: Summaries for Groups of Cases dialog box (Fig. 14.1).
5. Click on and move the dependent variable ("words") into the Variable box using the **top right arrow button.**
6. Click on and move the independent variable ("info_set") into the Category Axis variable box using the **bottom right arrow button.**
7. Select the representation for the bars in the Bars Represent box. You have option of selecting a specified confidence interval for the mean or a speci-

fied multiplier for the standard error of the mean. In this example, we maintain the default, which is a 95% confidence interval for the mean.

8. Click on **OK.**

The output is displayed in Figure 14.2.

The figure shows the 95% confidence interval for the average number of words memorized by children in each of the three information set groups. The circle represents the mean, and the horizontal lines the endpoints of the confidence interval. For instance, we see that children in the *no information* group memorized, on average, 4 words. Further, the 95% confidence interval for the mean is approximately 2.5 words to 5.5 words.

From inspection, it appears that the children in the *3 categories* group had the most success in memorizing words. To determine whether the groups differ significantly, however, we must conduct the one-way ANOVA test of significance.

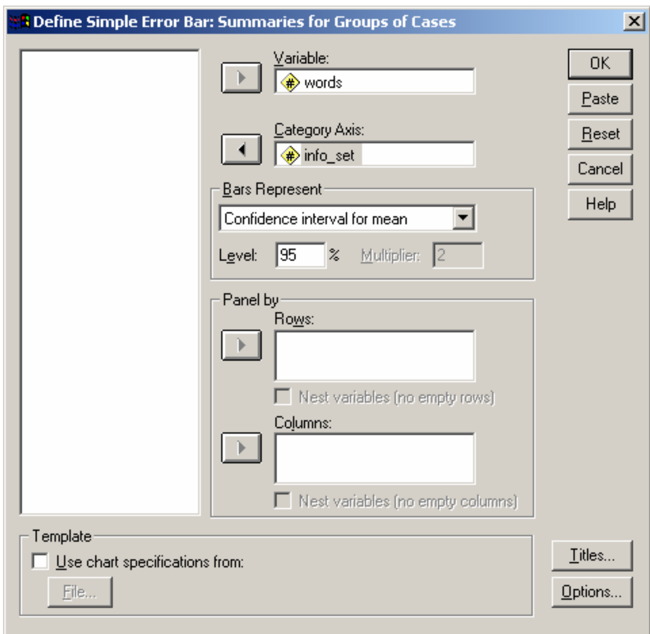

Figure 14.1 Define Simple Error Bar: Summaries for Groups of Cases Dialog Box

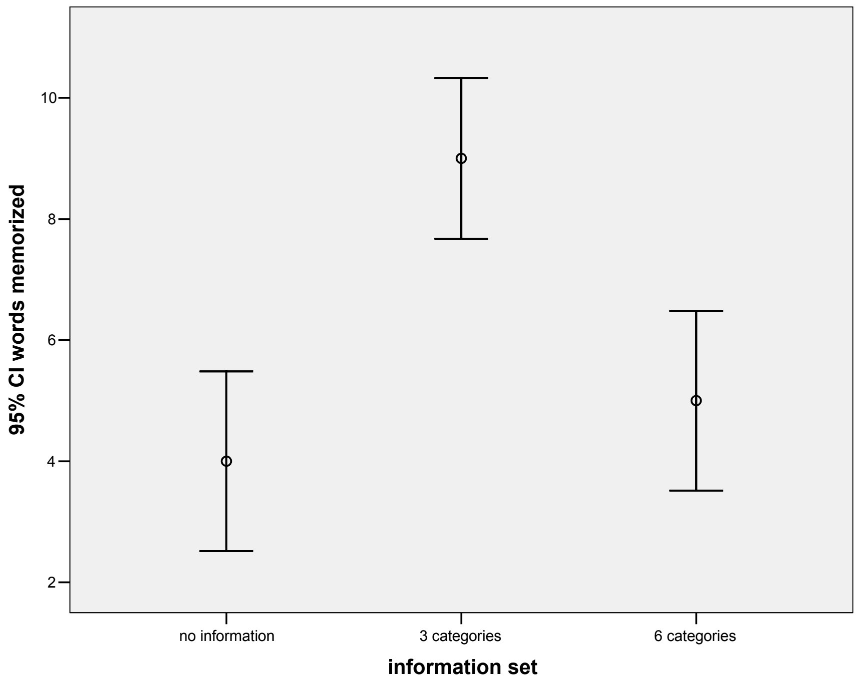

Figure 14.2 Error Bar Chart of Words Memorized by Information Set

Running the One-Way Procedure

To direct SPSS to perform the One-Way procedure:

1. Click on **Analyze** from the menu bar.
2. Click on **Compare Means** from the pull-down menu.
3. Click on **One-Way ANOVA** from the pull-down menu to open the One-Way ANOVA dialog box (Fig. 14.3).
4. Click on and move the "words" variable to the Dependent List box using the **top right arrow button.**
5. Click on and move the "info_set" variable to the Factor box using the **bottom right arrow button.**
6. Click on the **Options** button in the lower right corner of the One-Way ANOVA dialog box.
7. Click on the **Descriptives** option of the Statistics box. (This option provides mean, standard deviation, confidence interval, standard error, and the

minimum and maximum for each information category separately.) Also click on the **Means Plot** option to produce a graphical display of the means.

8. Click on **Continue** to close the One-way ANOVA: Option dialog box.
9. Click on **OK** to run the procedure.

The output containing descriptive statistics, the ANOVA table, and the means plot are displayed in Figure 14.4. The Sig. (.000) in the ANOVA table represents the *P* value corresponding to the *F*-ratio of 22.5 with 2 and 15 degrees of freedom. The null hypothesis that the population means are equal is rejected for any α greater than or equal to .0005. Thus, we conclude that there are differences among the three groups in the mean numbers of words memorized based on the information set.

From the Descriptives table and the Means plot, we see that the average number of words remembered by the no information group was 4; by the three categories group it was 9; and by the six categories group it was 5. SPSS also lists the range of words remembered for each group and computes a 95% confidence interval for each of the means. (This is similar to the information obtained in Section 14.1.)

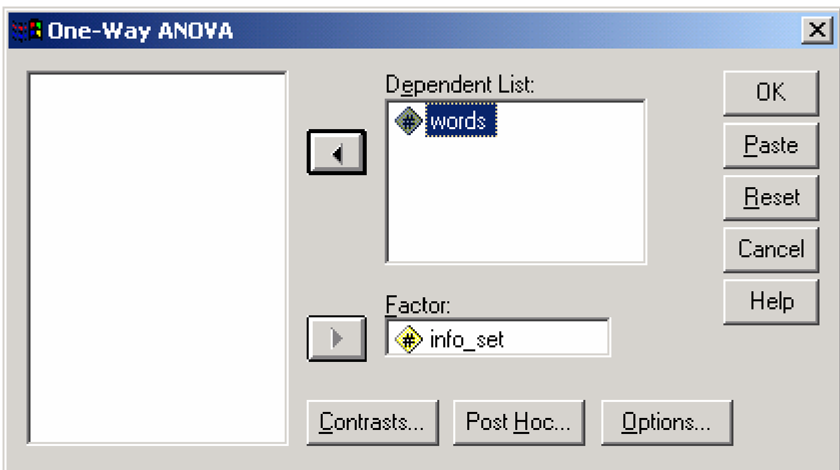

Figure 14.3 One-Way ANOVA Dialog Box

Descriptives

words memorized

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
no information	6	4.0000	1.41421	.57735	2.5159	5.4841	2.00	6.00
3 categories	6	9.0000	1.26491	.51640	7.6726	10.3274	7.00	10.00
6 categories	6	5.0000	1.41421	.57735	3.5159	6.4841	3.00	7.00
Total	18	6.0000	2.56676	.60499	4.7236	7.2764	2.00	10.00

ANOVA

words memorized

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	84.000	2	42.000	22.500	.000
Within Groups	28.000	15	1.867		
Total	112.000	17			

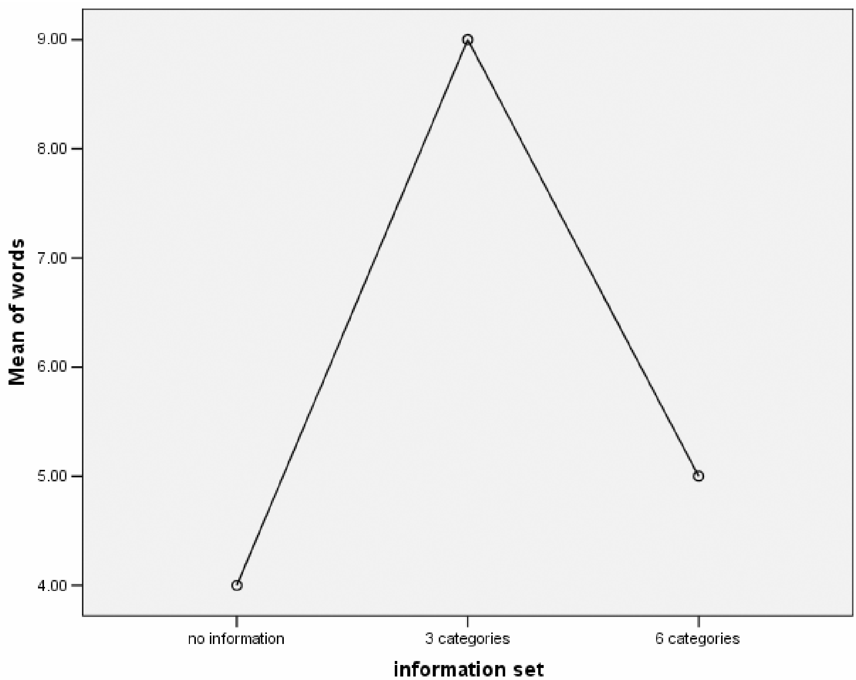

Figure 14.4 One-way ANOVA Listing with Descriptives and Means Plot

14.2 WHICH GROUPS DIFFER FROM WHICH, AND BY HOW MUCH?

Post-Hoc Comparisons of Specific Differences

When three or more populations are compared through analysis of variance, it is impossible to tell from the overall *F*-ratio which means differ significantly from which other means simply by inspection. You may wish to conduct tests for particular pairs of means. Conducting multiple tests inflates the probability of making a Type I error, however. There are several ways to compensate for this, many offered by SPSS. One approach is to use the Bonferroni correction procedure for computing the tests of significance for all possible pairs of information set (e.g., no information compared with three categories, no information compared with six categories, and three categories compared with six categories).

To illustrate, conduct the One-way procedure again, by following steps 1–8 as described in Section 14.1 under and then:

1. Click on the **Post Hoc** button to open the One-Way ANOVA: Post Hoc Multiple Comparisons dialog box (Fig. 14.5).
2. Click on the **Bonferroni** option; leave the significance level at the default .05.
3. Click on **Continue** to close the dialog box.
4. Click on **OK** to run the procedure.

The results should look similar to those displayed in Figure 14.4, with the addition of the Multiple Comparisons Table, contained in Figure 14.6. This table lists the results of tests of significance for all possible pairwise comparisons among the three information sets. For instance, the mean difference between number of words memorized by children in the no information group (I column) minus that of the 3 categories group (J column) was –5 words. Thus, the 3 categories group had a higher mean. (Note that the same information can be obtained in the pairing of no information in the J column and 3 categories in the I column.)

The Sig. column contains the P value for the test of significance. Here, $P < .0005$, indicating that the difference is statistically significant. The difference for no information minus 6 categories is –1, which is not statistically significant at the .05 level ($P < .673$). The difference between 3 categories minus 6 categories (4 words) is significant ($P < .0005$).

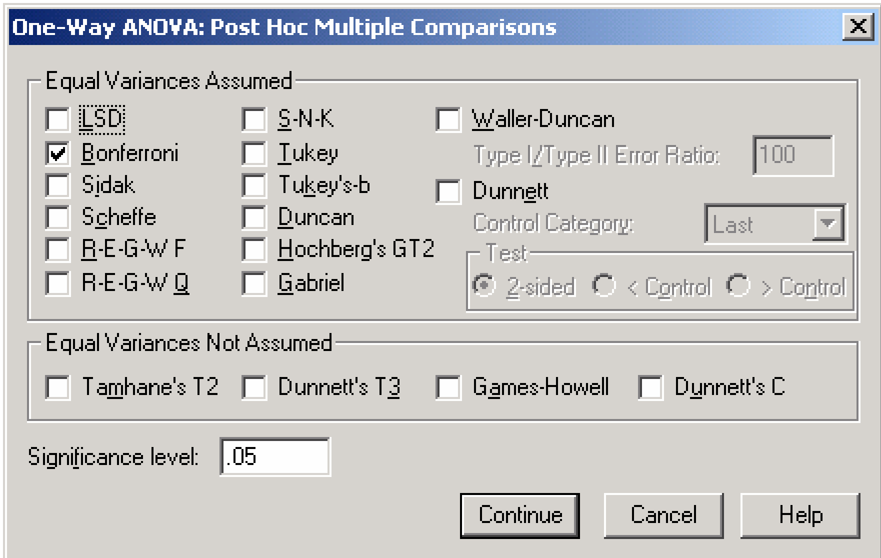

Figure 14.5 One-Way ANOVA: Post Hoc Multiple Comparisons Dialog Box

Multiple Comparisons

Dependent Variable: words memorized

Bonferroni

(I) information set	(J) information set	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
no information	3 categories	-5.0000*	.78881	.000	-7.1249	-2.8751
	6 categories	-1.0000	.78881	.673	-3.1249	1.1249
3 categories	no information	5.0000*	.78881	.000	2.8751	7.1249
	6 categories	4.0000*	.78881	.000	1.8751	6.1249
6 categories	no information	1.0000	.78881	.673	-1.1249	3.1249
	3 categories	-4.0000*	.78881	.000	-6.1249	-1.8751

*. The mean difference is significant at the .05 level.

Figure 14.6 Multiple Comparisons Table from the Post Hoc Option of the One-Way ANOVA

Therefore, we summarize the findings by stating that the 3 categories information set appears superior for word memorization to both the no information and the 6 categories sets, but that there is no difference between the no information and 6 categories sets. (Refer back to the graphs in Figure 4.2 and Figure 14.4; is our conclusion supported by the graphs?)

Effect Sizes

The *t*-tests allow us to determine which means differ significantly from which other means, but do not provide a clear indication of the *magnitude* of the differences. In some cases, the difference itself is a meaningful way to describe the strength of the effect. For example, students in the three categories condition remembered, on average, 9 – 4 = 5 more words than those in the no information condition.

There are instances when the scale of the dependent variable is not familiar, as with many educational and psychological tests. In such situations, it is helpful to express the mean difference in standard deviation units; the result is called the effect size. For example, the effect size for the 5-word difference between children in the three-category group and those in the no information group is

$$\frac{9-4}{\sqrt{1.867}} = 3.66,$$

where the pooled within-group standard deviation is the square root of the mean square within group (found on the ANOVA table in Figure 14.4). The mean difference (3.66) is over 3 standard deviations, representing a very large effect size. On average, children given information grouped into three sets were 3.66 standard deviations above children in the no information group.

14.3 ANALYSIS OF VARIANCE OF RANKS

As with other statistical procedures, it is important to evaluate whether the data meet the assumptions underlying the ANOVA. These assumptions include (1) independence of observations, (2) a normal distribution of subgroup means, and (3) homogeneity of population variances.

While the analysis of variance is fairly robust with regard to violation of conditions (2) and (3), these assumptions should be examined routinely, especially if sample sizes are small to moderate. You can, for instance, inspect histograms of the dependent variable to assess departures from normality, as illustrated in Section 10.2. When there are more than two populations involved, the sample variances should also be examined to see if they are similar. In Figure 14.4, for example, the standard deviations of the three groups appear to be in the same general range. (There are also formal tests for equal variances, but they are not discussed in this manual.)

When the assumptions of normality and/or homogeneity of variance are severely violated, the ANOVA results may be misleading. In this case, an alternative procedure, the Kruskal-Wallis analysis of variance of ranks, is preferable.

The data file "bottle.sav" contains the daily output for three bottle capping machines. Using this data file, the procedure for using SPSS to perform the analysis of variance of ranks is:

1. Click on **Analyze** from the menu bar.
2. Click on **Nonparametric Tests** from the pull-down menu.
3. Click on **K Independent samples** from the pull-down menu. This opens the Tests for Several Independent Samples dialog box (Fig. 14.7).
4. Click on and move the "output" variable to the Test Variable List box using the **upper right arrow** button.
5. Click on and move the "machine" variable to the Grouping Variable box using the **lower right arrow** button.
6. Click on the **Define Range** button to open the Several Independent Samples: Define Range dialog box.
7. The machine variable is coded 1 through 3, so enter **1** in the minimum box and **3** in the maximum box.
8. Click on **Continue** to close the dialog box.
9. Notice that the Kruskal-Wallis H test is the default option in the Test Type box. Therefore, click on **OK** to run the procedure.

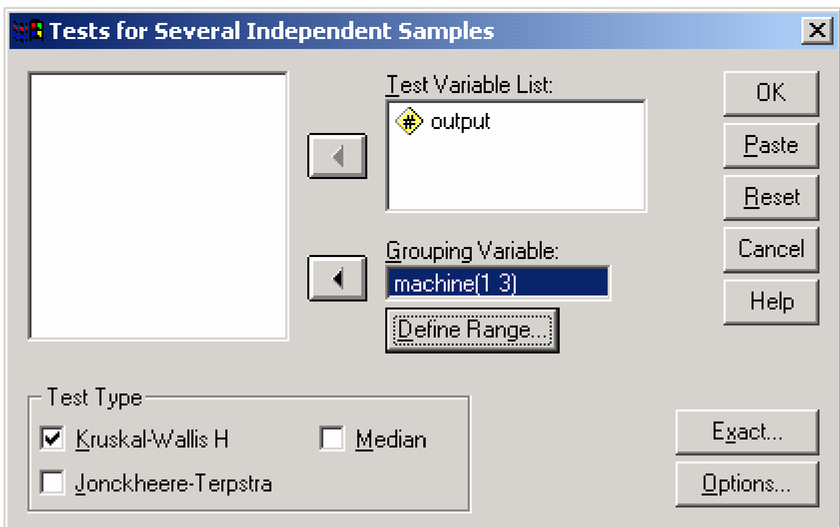

Figure 14.7 Test for Several Independent Samples Dialog Box

Ranks

	machine	N	Mean Rank
bottle cap output	Machine A	5	4.80
	Machine B	3	4.67
	Machine C	4	10.00
	Total	12	

Test Statistics[a,b]

	bottle cap output
Chi-Square	5.656
df	2
Asymp. Sig.	.059

a. Kruskal Wallis Test

b. Grouping Variable: machine

Figure 14.8 Kruskal-Wallis One-Way ANOVA Listing

The output from this test in contained in Figure 14.8. It shows the mean rank and sample size for each machine. For example, there are 4 machines of type C, and their mean rank is 10.00. The listing also reports the chi-square statistic (5.656) and the *P* value (.059). Using an α level of .05, the null hypothesis is accepted and we conclude that the machines do not differ with respect to bottle cap output.

14.4 TWO-FACTOR ANALYSIS OF VARIANCE

In many instances, the data analyst will be interested in the relationship between two or more independent categorical variables and one dependent numerical variable. For instance, the "stepping.sav" data file contains information on an experiment conducted at The Ohio State University to explore the relationship between heart rate after a stepping exercise and the frequency of stepping (slow, medium, or fast) and the height of the step (low or high). Thirty individuals took part in the study.

Because there are two independent variables of interest, we conduct a two-factor or two-way (rather than a one-way) ANOVA. This allows us to look not only at the *main effect* of height and frequency, but also the *interaction* of the two variables. That is, we can ask if the relationship between frequency of stepping and heart rate after exercise is the same for different step heights.

We illustrate by opening the "stepping.sav" data file and doing the following:

1. Click on **Analyze** from the menu bar.
2. Click on **General Linear Model** from the pull-down menu and on **Univariate** from the supplementary pull-down menu. This will open the Univariate dialog box (Fig. 14.9).
3. Click on the variable "hr" (heart rate after exercise) and move it to the Dependent Variable box with the **top right arrow** button.
4. Click on "height" and move it to the Fixed Factor(s) box with the **second right arrow** button.
5. Repeat step 4 with the "frequenc" variable.
6. Click on the **Model** button to open the Univariate: Model dialog box. This is the box that allows you to construct your model. For instance, you can decide whether you want to include the interaction term (height-by-frequency) in the model. The default option, **Full factorial** includes both main effects and the interaction. We will maintain this option.

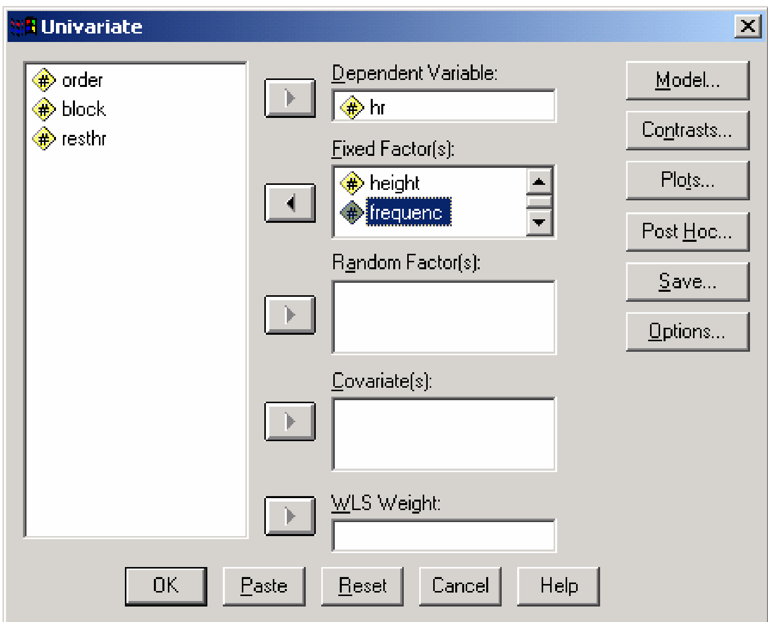

Figure 14.9 Univariate Dialog Box

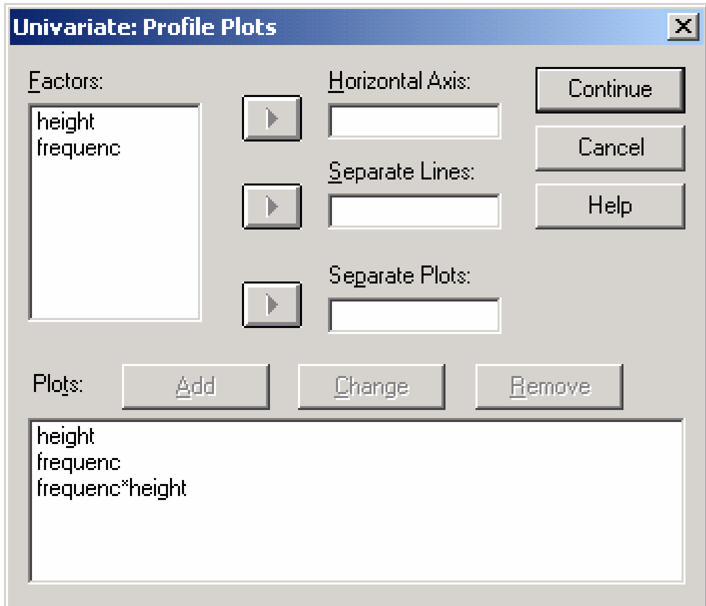

Figure 14.10 Univariate: Profile Plots Dialog Box

7. Click on **Continue.**
8. Click on **Plots** to open the Univariate: Profile Plots dialog box (Fig. 14.10). This option is similar to the means plots option in the one-way procedure. We can obtain one-way and two-way means graphs, however.
9. To obtain a graph of mean heart rate after exercise based on step height, click on the "height" variable and move it to the Horizontal Axis box using the **top right arrow** button.
10. Click on the **Add** button to move it into the Plots area.
11. Repeat steps 9-10 with the "frequenc" variable.
12. To obtain the interaction graph, click on and move the "frequenc" variable to the Horizontal Axis box with the **top right arrow** button and click on and move the "height" variable to the Separate Lines box with the **middle right arrow** button.
13. Click on **Add** to add this plot to the Plots area.
14. Click on **Continue** to close the dialog box.
15. Click on the **Options** button to open the Univariate: Options dialog box.
16. Click on the **Descriptive Statistics** option to produce means and standard deviation for the groups.

17. Click on **Continue**.
18. Click on **OK** to run the procedure.

The output of this procedure is displayed in Figure 14.11. The first table lists the between-subjects factors — the independent variables height of step and frequency of stepping. We see that there were 15 people in each of the height groups and 10 in each of the frequency groups.

The Descriptives table breaks down the groups, and displays the average heart rate after stepping for each cross-classification of the values of the independent variables. For instance, average heart rate after stepping for the 5 people in the low height, slow frequency group was 87.6; average in the high fast group was 136.8. This table also displays the marginal means — those by one variable. For instance, average heart rate for the 15 people in the low step group was 96.6, and 104.1 for the 10 people in the medium frequency group.

The ANOVA table is contained in the Tests of Between-Subjects Effects table. The Corrected Model row (with 5 degrees of freedom) refers to the full model — with two main effects and the interaction. The Sig. column displays the P value ($< .0005$), and indicates that model is significantly related to heart rate after stepping.

When there is an interaction effect in the model, it is important to interpret it before looking that the main effects. Thus, we examine the HEIGHT * FREQUENC effect first. The F-statistic is .540, and $P = .590$. We conclude that there is not a significant interaction between height and frequency of stepping on the heart rate. In other words, the effect of frequency of stepping on heart rate is the same whether the step height is low or high. (We can see this graphically in the last plot displayed in Figure 14.11. The blue line displays the relationship between heart rate and step frequency for low steps and the green line displays the relationship for high steps. The fact that the lines follow the same pattern and are almost parallel indicates the lack of a significant interaction.)

Because there is not a significant interaction, it is appropriate to interpret the main effects. The ANOVA table indicates that the height of the step is significantly related to heart rate ($F = 17.931$, $P < .0005$). In addition, frequency of stepping is also significantly related to heart rate ($F = 9.551$, $P < .0005$).

The descriptive statistics and the graphs indicate that heart rate increases with the height of the step (in the sample, mean is 96.6 for a low step and 118.2 for a high step). Because there are only two levels to this variable, we know where the significant difference lies.

Between-Subjects Factors

		Value Label	N
Height of step	.00	low	15
	1.00	high	15
Frequency of stepping	.00	slow	10
	1.00	medium	10
	2.00	fast	10

Descriptive Statistics

Dependent Variable: heart rate after stepping

Height of step	Frequency of stepping	Mean	Std. Deviation	N
low	slow	87.6000	9.09945	5
	medium	94.2000	8.89944	5
	fast	108.0000	16.70329	5
	Total	96.6000	14.26184	15
high	slow	103.8000	10.52141	5
	medium	114.0000	21.00000	5
	fast	136.8000	13.34916	5
	Total	118.2000	20.30904	15
Total	slow	95.7000	12.60555	10
	medium	104.1000	18.44180	10
	fast	122.4000	20.82306	10
	Total	107.4000	20.44437	30

Tests of Between-Subjects Effects

Dependent Variable: heart rate after stepping

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7437.600[a]	5	1487.520	7.622	.000
Intercept	346042.800	1	346042.800	1773.214	.000
HEIGHT	3499.200	1	3499.200	17.931	.000
FREQUENC	3727.800	2	1863.900	9.551	.001
HEIGHT * FREQUENC	210.600	2	105.300	.540	.590
Error	4683.600	24	195.150		
Total	358164.000	30			
Corrected Total	12121.200	29			

a. R Squared = .614 (Adjusted R Squared = .533)

Figure 14.11 Results of Two-Way ANOVA

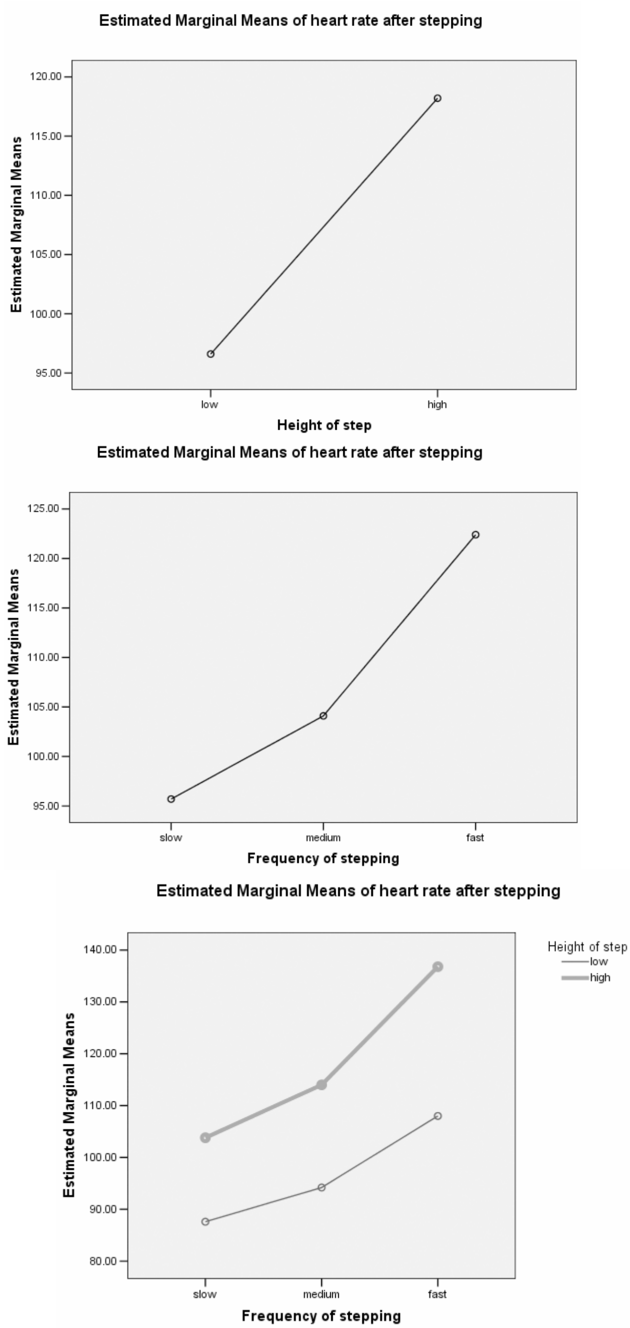

Figure 14.11 *Continued*

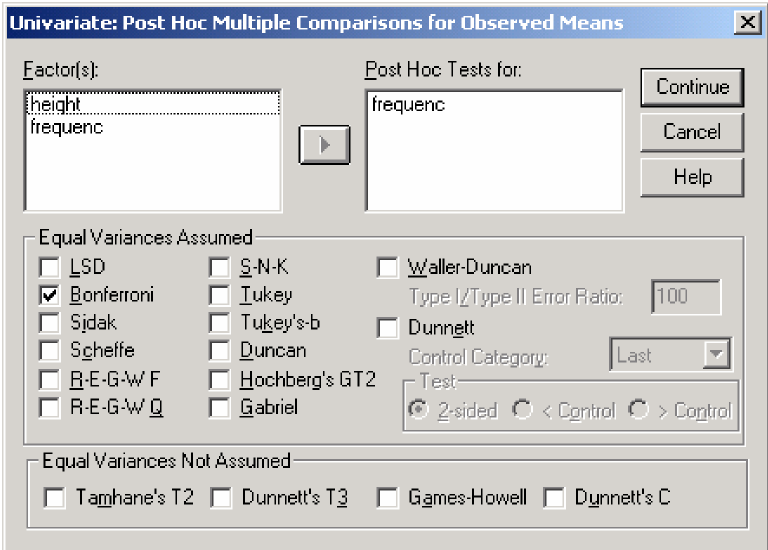

Figure 14.12 Univariate: Post Hoc Multiple Comparisons for Observed Means Dialog Box

The descriptive statistics and graphs also suggest that heart rate increases as frequency of stepping increases (the sample means are 95.7, 104.1, and 122.4 for slow, medium, and fast stepping, respectively). However, because there are three levels, we cannot say which levels differ statistically from one another without performing a follow-up test. To do so, re-run the ANOVA, adding the following steps before clicking **OK** to run the procedure.

1. From the Univariate dialog box, click on the **Post Hoc** button to open the Univariate: Post Hoc Multiple Comparisons for Observed Means dialog box (Fig. 14.12).
2. Click on and move the "frequenc" variable to the Post Hoc Tests for box with the **right arrow** button.
3. Click on the **Bonferroni** option.
4. Click on **Continue**.

The Multiple Comparisons table is listed in Figure 14.13. We read this table the same as in the one-way procedure. Results indicate that fast stepping is associated with a heart rate that is significantly higher than both slow stepping (P = .001) and medium stepping (P = .022), but that there is no significant difference in heart rate between slow and medium stepping (P = .574).

Multiple Comparisons

Dependent Variable: heart rate after stepping
Bonferroni

(I) Frequency of stepping	(J) Frequency of stepping	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
slow	medium	-8.4000	6.24740	.574	-24.4786	7.6786
	fast	-26.7000*	6.24740	.001	-42.7786	-10.6214
medium	slow	8.4000	6.24740	.574	-7.6786	24.4786
	fast	-18.3000*	6.24740	.022	-34.3786	-2.2214
fast	slow	26.7000*	6.24740	.001	10.6214	42.7786
	medium	18.3000*	6.24740	.022	2.2214	34.3786

Based on observed means.

*. The mean difference is significant at the .05 level.

Figure 14.13 Bonferroni Multiple Comparisons of Frequency of Stepping

Chapter Exercises

14.1 Using the "hotdog.sav" data file:

a. Test the hypothesis that there are no differences, on average, in calories based on type of hot dog. Use $\alpha = .05$ and state your conclusions in words.

b. If you found a significant difference in (a), conduct the Bonferroni post hoc tests to determine where the differences lie. State your conclusions in one or two sentences.

c. Compute an effect size for any of the comparisons that were significant in (b).

d. Repeat steps (a) through (c) for sodium content.

14.2 The data file "movies.sav" contains information on number of weeks the movies for the year 2001 were in the Top 60.

a. Perform an ANOVA to test whether weeks in Top 60 is independent of genre (type) of movie. Use $\alpha = .05$ and state your conclusions in words.

b. If you found a significant difference in (a), conduct the Bonferroni post hoc tests to determine where the differences lie. State your conclusions in one or two sentences.

c. Compute an effect size for any of the comparisons that were significant in (b).