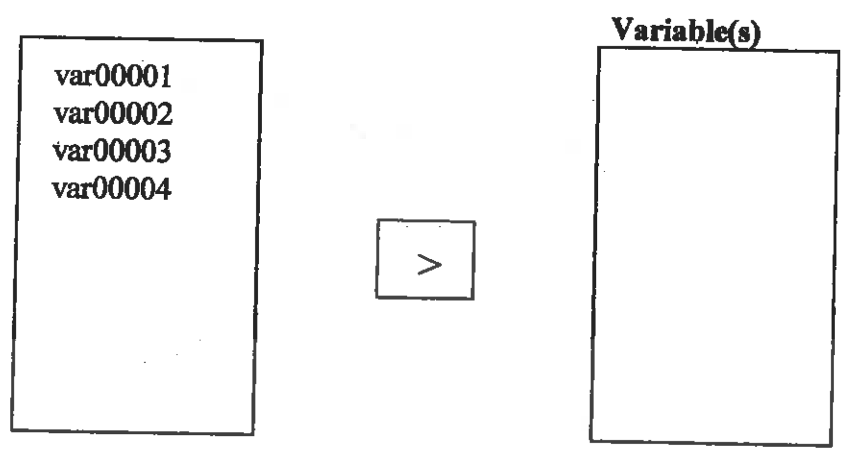
SPSS

The Statistical Package for the Social Sciences, commonly known as SPSS, is a computer program that performs statistical calculations, and is widely available on college campuses. SPSS makes it possible for researchers to compute basic descriptive statistics (such as means and standard deviations) and to perform hypothesis tests relatively quickly and easily.

SPSS consists of two basic components: A Data Matrix and a set of Statistical Commands. The **data matrix** is a huge matrix of numbered rows and columns. To begin any analysis, you must type your data into the matrix. Typically, the scores are entered into columns of the matrix. You may also enter names for the columns (for example, 'xscores" or "selfestm") or you can simply remember that the X scores are in column 1 and the self-esteem scores are in column 2. The **statistical commands** are listed in menus that are made available by clicking on the Analyze box that is located on the tool bar at the top of the screen. When you select a statistical command, SPSS will typically ask you to identify exactly where the scores are located and exactly what other options you want to use. This is accomplished by identifying the column(s) in the data matrix that contain the needed information. Typically, you will be presented with a display similar to the figure below. On the left is a box that lists all of the columns in the data matrix that contain information. In this example, we have typed values into columns 1, 2, 3, and 4. On the right is an empty box that is waiting for you to identify the correct column. For example, suppose that you wanted to do a statistical calculation using the scores in column 3. You should highlight var00003 by clicking on it in the left-hand box, then click the arrow to move the column label into the right hand box. (If you make a mistake, you can highlight the variable in the right-hand box and the arrow will reverse so that you can move the variable back to the left-hand box.)



If you want to name a column (instead of using var00001), click on the **Variable View** tab at the bottom of the data matrix. You will get a description of each variable in the matrix, including a box for the name. You may type in a new name using up to 8 lower-case characters (no spaces, no hyphens). Click the **Data View** tab to go back to the data matrix.

Frequency Distributions (Chapter 2)

A frequency distribution is an organized tabulation showing how many individuals have scores in each category on the scale of measurement. A frequency distribution can be presented either as a table or a graph.

To construct a frequency distribution table with SPSS, first enter the scores in one column of the data matrix, then click **Analyze** on the tool bar. Then, select **Descriptive Statistics** and select **Frequencies.** Move the column label for the set of scores into the Variable box. Be sure that the option to Display Frequency Table is selected and then click **OK.** The frequency distribution table will list the score values in a column from smallest to largest, with the percentage and cumulative percentage also listed for each score. Score values that do not occur (zero frequency) are not included in the table, and the program does not group scores into class intervals (all values are listed).

To draw a histogram or bar graph for the scores in a column, click **Analyze** on the tool bar, then select **Descriptive Statistics** and select **Frequencies.** Enter the column label for the set of scores into the Variable box and click **Charts.** Select either Bar Graph or Histogram depending on the scale of measurement for your data, then click **Continue** and click **OK.** After a brief delay, SPSS will display a frequency distribution table and a graph. Note that the graphing program will automatically group scores into class intervals if the range of scores is too wide.

Example: The following set of scores produce a frequency distribution (either a table or a graph) showing that three people had scores of X = 1, five people had X = 2, six people had X = 3, four had X = 4, and two had X = 5. Scores: 1, 2, 4, 2, 3, 3, 5, 1, 3, 4, 2, 4, 3, 2, 4, 3, 1, 3, 2, 5

Means and Standard Deviations (Chapters 3 and 4)

The mean and standard deviation are probably the two most commonly used statistics for describing a set of scores. The mean describes the center of the set of scores and the standard deviation describes how the scores are scattered around the mean. In simple terms, the standard deviation provides a measure of the average distance from the mean.

To calculate the mean and the standard deviation for a sample of scores, enter the scores in one column of the data matrix and click **Analyze** on the tool bar. Then, select **Descriptive Statistics** and select **Descriptives**. Move the column label for the set of scores into the Variable box and click **OK**. SPSS will produce a summary table listing the maximum score, the minimum score, the mean, and the standard deviation. Note that SPSS computes the *sample* standard deviation using n - 1. If your scores are intended to be a population, you can multiply the reported sample standard deviation by the square root of (n-1)/n to obtain the population standard deviation.

You can also obtain the mean and standard deviation for a sample if you use SPSS to display the scores in a frequency distribution histogram (see the preceding section on frequency distributions). The mean and standard deviation are displayed beside the graph.

Example: The following scores produce a mean of M = 2.85 and a standard deviation of SD = 1.23. Scores: 1, 2, 4, 2, 3, 3, 5, 1, 3, 4, 2, 4, 3, 2, 4, 3, 1, 3, 2, 5

The Independent-Measures t Test (Chapter 10)

The independent-measures t test is used to compare two means from an independent-measures (between-subjects) research design: that is, the test evaluates the mean difference between two separate samples that represent two separate treatment conditions or two separate populations. A *significant difference* indicates that there appears to be a consistent, systematic difference between the two treatments and that the obtained mean difference is very unlikely ($p < .05$) to have occurred by chance alone. The significance is determined by the p value that is reported as part of the computer output.

To conduct an independent-measures t hypothesis test, the data are entered in what is called a *stacked format*, which means that all the scores from *both samples* are entered in one column of the data matrix. Values are then entered into a second column to designate the sample or treatment condition corresponding to each of the scores. For example, enter all of the scores in the first column. In the second column, enter a 1 beside each score from sample #1 and enter a 2 beside each score from sample #2. Then, click **Analyze** on the tool bar, select **Compare Means,** and click on **Independent-Samples T Test.** Move the label for the column containing the scores into the Test Variable box. Then move the label for the column that contains the group number for each score into the Group Variable box. Next, click on **Define Groups.** Assuming that you used a 1 to identify group #1 and a 2 to identify group #2, enter the values 1 and 2 into the appropriate group boxes, then click **Continue** and click **OK.** SPSS will produce a summary table showing the number of scores, the mean, the standard deviation, and the standard error for each of the two samples. SPSS also conducts a test for homogeneity of variance, using Levene's test (similar to an F-max test). Homogeneity of variance is an assumption for the t test and requires that the two populations from which the samples were obtained have equal variances. This test should *not* be significant (you do not want the two variance to be different), so you want the reported Sig. value to be greater than .05. Next, the results of the hypothesis test are presented using two different assumptions; we will focus on the top row where equal variance are assumed. The test results include the calculated t value, the degrees of freedom, the level of significance (probability of a Type I error), and the size of the mean difference. Finally, the output includes a report of the standard error for the mean difference and a 95% Confidence Interval that provides a range of values estimating how much difference exists between the two treatment conditions.

Example: The following two samples produce a t statistic of $t = 3.834$, with degrees of freedom equal to $df = 6$, and a significance level of $p = 0.009$.

Treatment 1 (sample 1)	Treatment 2 (sample 2)
3	12
5	10
7	8
1	14

The Repeated-Measures t Test (Chapter 11)

The repeated-measures t test is used to compare two means from a repeated-measures (within-subjects) research design: that is, the test evaluates the mean difference between two treatment conditions where the same set of individuals is measured in both treatments. A *significant difference* indicates that there appears to be a consistent, systematic difference between the two treatments and that the obtained mean difference is very unlikely ($p < .05$) to have occurred by chance alone. The significance is determined by the p value that is reported as part of the computer output.

To conduct a repeated-measures t hypothesis test, the data are entered into two columns in the data matrix with the first score for each participant in the first column and the second score in the second column. Then, click **Analyze** on the tool bar, select **Compare Means**, and click on **Paired-Samples T Test.** Next you must enter the labels for the two data columns into the Paired Variables box. (Highlight the first-score column as Variable 1 and highlight the second-score column as Variable 2, then click the arrow to move you selections into the box.) Then click **OK.** SPSS will produce a summary table showing descriptive statistics for each of the two sets of scores, and a table showing the correlation between the first and second score. Finally, SPSS conducts the t test for the difference scores. The output shows the mean difference, the standard deviation and the standard error for the difference scores, as well as the value for t, the value for df, and the level of significance (p value). The output also includes a 95% Confidence Interval that provides a range of values estimating how much difference exists between the two treatment conditions.

Note: If you have already computed the difference score for each participant (instead of pairs of scores), you can do the repeated-measures t test by entering the difference scores in one column and using the **One-Sample T Test** option. Click **Analyze** on the tool bar, select **Compare Means**, and click on **One-Sample T Test.** Enter the column label for the set of difference scores into the Test Variable box, and enter a value of zero in the Test Value box.

Example: The following data show a mean difference of 5 points between the two treatments and produce $t = 2.50$, with $df = 3$, and a significance level of $p = 0.088$.

Participant	1st Treatment	2nd Treatment	Difference
A	19	12	-7
B	35	36	+1
C	20	13	-7
D	31	24	-7

Single-Factor, Independent-Measures Analysis of Variance (Chapter 13)

The single-factor, independent-measures analysis of variance is used to compare the means from an independent-measures (between-subjects) research study using two or more separate samples to compare two or more separate treatment conditions or populations. A *significant difference* indicates that there appears to be a consistent, systematic difference between at least two of the treatments and that the obtained mean differences are very unlikely ($p < .05$) to have occurred by chance alone. The significance is determined by the p value that is reported as part of the computer output.

To conduct a single-factor, independent-measures analysis of variance, the data are entered in a *stacked format* in the data matrix, which means that all the scores are entered in a single column and a second column is used to designate the treatment condition for each score. For example, enter all the scores in the first column and, in the second column enter a 1 beside each score from the first treatment, enter a 2 beside each score from the second treatment, and so on. Then, click **Analyze** on the tool bar, select **Compare Means**, and click on **One-Way ANOVA**. Move the column label for the scores into the Dependent List box, and move the column label for the treatment numbers into the Factor box. Before you conduct the ANOVA, you can click on the **Options** box and select Descriptives to get descriptive statistics for each sample. Then click **OK**. If you selected the Descriptives Option, SPSS will produce a table showing descriptive statistics for each of the samples along with a summary table showing the results from the analysis of variance.

Example: For the following data, the 1st treatment has $M = 1.00$ with $SD = 1.73$, the 2nd treatment has $M = 5.00$ with $SD = 2.24$, and the 3rd treatment has $M = 6.00$ with $SD = 1.87$. The analysis produces an F-ratio of $F = 9.13$, with $df = 2, 12$, and a significance level of $p = 0.004$.

1st Treatment	2nd Treatment	3rd Treatment
0	6	6
4	8	5
0	5	9
1	4	4
0	2	6

Single-Factor, Repeated-Measures Analysis of Variance (Chapter 14)

The single-factor, repeated-measures analysis of variance is used to compare the means from a repeated-measures (within-subjects) research study using one sample to compare two or more separate treatment conditions (each individual is measured in each of the treatment conditions). A *significant difference* indicates that there appears to be a consistent, systematic difference between at least two of the treatments and that the obtained mean differences are very unlikely ($p < .05$) to have occurred by chance alone. The significance is determined by the p value that is reported as part of the computer output.

To conduct a single-factor, repeated-measures analysis of variance, the data are entered in the data matrix with the scores for each treatment condition in a separate column (with the scores for each individual in the same row). Then, click **Analyze** on the tool bar, select **General Linear Model**, and click on **Repeated Measures**. SPSS will present a box titled Repeated-Measures Define Factors. Within the box, the Within-Subjects Factor Name should already contain Factor1. If not, type in Factor 1. Then, enter the number of levels (number of treatment conditions) in the next box, click on **Add**, then click **Define**. One by one, move the column labels for your treatment conditions into the Within Subjects Variables box. (Highlight the column label on the left and click the arrow to move it into the box.) Before you conduct the ANOVA, you can click on the **Options** box and select Descriptives (to get descriptive statistics for each treatment). Then click **OK**. If you selected the Descriptives Option, SPSS will produce a table showing the mean and standard deviation for each treatment condition. The rest of the SPSS output is relatively complex and includes a lot of statistical information that goes well beyond the scope of this book. However, if you focus on the table showing *Test of Within-Subjects Effects*, the top line of the FACTOR1 box and the top line of the Error(FACTOR1) box will show the sum of squares, the degrees of freedom, and the mean square for the numerator and denominator of the F-ratio, as well as the value of the F-ratio and the level of significance (p value).

Example: For the following data, the 1st treatment has $M = 5.00$ with $SD = 1.87$, the 2nd treatment has $M = 4.00$ with $SD = 1.58$, and the 3rd treatment has $M = 3.00$ with $SD = 1.58$. The analysis produces an F-ratio of $F = 10.00$, with $df = 2, 8$, and a significance level of $p = 0.007$.

Participant	1st Treatment	2nd Treatment	3rd Treatment
A	5	4	3
B	3	2	1
C	4	3	2
D	5	6	4
E	8	5	5

Two-Factor, Independent-Measures Analysis of Variance (Chapter 15)

The two-factor, independent-measures analysis of variance is used to compare the means from an independent-measures (between-subjects) research study using two independent variables (or quasi-independent variables). The structure of a two-factor study can be represented as a matrix with the levels of one independent variable defining the rows and the levels of the second independent variable defining the columns. Each cell in the matrix corresponds to a unique treatment condition, and there is a separate sample for each cell. The two-factor ANOVA consists of three separate tests for mean differences: (1) The main effect for one factor consists of the mean differences between the rows of the matrix; (2) the main effect for the second factor consists of the mean differences between the columns of the matrix; (3) the interaction consists of any additional mean differences that are not accounted for by the two main effects. For each of the three tests, a *significant difference* indicates that there appears to be a consistent, systematic difference between at least two of the treatments and that the obtained mean differences are very unlikely ($p < .05$) to have occurred by chance alone. The significance is determined by the p value that is reported as part of the computer output.

To conduct a two-factor, independent-measures analysis of variance, the scores are entered into the SPSS data matrix in a *stacked format*, which means that all the scores from all the different treatment conditions are entered in a single column. A second column is used to designate the level of factor A for each score, and a third column is used to designate the level of factor B. For example, enter all the scores in the first column. In the second column, enter a 1 beside each score from A1, enter a 2 beside each score from A2, and so on. In the third column, enter a 1 beside each score from B1, a 2 beside each score from B2, and so on. Then, click **Analyze** on the tool bar, select **General Linear Model**, and click on **Univariant**. Move the column label for the scores into the Dependent Variable box, and move the two column labels for the two factors into the Fixed Factors box. Before you conduct the ANOVA, you can click on the **Options** box and select Descriptives (to get descriptive statistics for each sample). Then click **OK**. If you selected the Descriptives Option, SPSS will produce a table showing the means and standard deviations for each treatment condition. The results of the ANOVA are shown in a summary table in which each factor is identified by its column label. (Note that the summary table includes some extra values, such as *Corrected Model* and *Intercept*, that are beyond the scope of this text.)

Example: The following data produce an F-ratio for the main effect of factor A of F = 8.167 with df = 1, 24 and a significance level of p = 0.009. The F-ratio for the main effect of factor B is F = 3.167 with df = 2, 24 and a significance level of p = 0.060 (not significant). The AxB interaction has F = 1.167 with df = 2, 24 and a significance level of p = 0.328 (not significant). The means and standard deviations are shown with the individual samples.

		Factor B		
		B1	B2	B3
Factor A	A1	3 1 1 6 4 M = 3 SD = 2.12	1 4 8 6 6 M = 5 SD = 2.24	5 5 9 2 4 M = 5 SD = 2.54
	A2	0 2 0 0 3 M = 1 SD = 1.41	3 8 3 3 3 M = 4 SD = 2.24	0 0 0 5 0 M = 1 SD = 2.24

The Pearson or Spearman Correlation (Chapter 16)

The Pearson correlation measures and describes the direction and degree of linear relationship between two variables. The data are numerical scores, with two separate scores, representing two different variables, for each individual. The two scores are identified as X and Y. A positive correlation indicates that X and Y tend to vary in the same direction (as X increases, Y also increases), and a negative correlation indicates that X and Y vary in opposite directions (as X increases, Y decreases). A correlation of 1.00 (or -1.00) indicates that the data points fit perfectly on a straight line. A correlation of 0.00 indicates that there is no linear relationship whatsoever. Values between 0 and 1.00, indicate intermediate degrees of relationship. It is also possible to evaluate the statistical significance of a correlation by determining the probability that the sample correlation was obtained, just by chance, from a population in which there is a zero correlation.

To compute correlations and to determine the statistical significance of correlations, the data are entered into two columns in the data matrix with the X values in one column and the Y values in a second column. Then, click **Analyze** on the tool bar, select **Correlate**, and click on **Bivariate**. Next you must move the labels for the two data columns into the Variables box. The Pearson box should be checked, but you can click the Spearman box if you want to compute a Spearman correlation (SPSS will convert the scores to ranks). Then click **OK**. SPSS will produce a correlation matrix showing all the possible correlations. You want the correlation of X and Y which is contained in the upper right corner (or the lower left). The output includes the significance level (p value) for the correlation.

Example: The following data produce a Pearson correlation of 0.875 with a significance level of p = 0.052 (not significant).

X	Y
0	1
10	3
4	1
8	2
8	3

The Chi-Square Test for Independence (Chapter 17)

The chi-square test for independence evaluates the relationship between two variables. Instead of measuring numerical scores, each individual is simply classified into a category for each of the two variables; for example, each individual could be classified by gender (male/female) and by personality (introvert/extrovert). The data are usually organized in a matrix with the categories of one variable defining the rows and the categories of the second variable defining the columns. The actual data (called *observed frequencies*) consist of the number of individuals from the sample who are in each cell of the matrix; for example, how many introverted males, how many introverted females, how many extroverted males, and how many extroverted females.

Although SPSS can be used to perform the chi-square test for independence, the program requires that the scores for each participant be entered individually. This system works well for preexisting data sheets (for example, data from a large survey that have already been entered into the computer). However, it can be very tedious if you simply want to use SPSS for a single set of observed frequencies .

Suppose for example, that you are using a chi-square test to examine the relationship between academic performance and self-esteem for a sample of n = 150 students. Each student's academic performance is classified as either high or low, and each student is classified as either high, medium, or low in terms of self-esteem. To perform the chi-square test with SPSS, you would need to enter the academic performance score for each of the 150 participants in one column (coded as high = 1 and low = 2), and enter the same participant's self-esteem score (coded as high = 1, medium = 2, low = 3) in a second column. Note that you are entering a total of 300 scores. Then click **Analyze** on the tool bar, select **Descriptive Statistics**, and click on **Crosstabs**. Move the column label for the academic scores into the Rows box, and move the column label for the self-esteem scores into the Columns box. Then, click on **Statistics**. Select **Chi-Square**, click **Continue**, and click **OK**. The SPSS output will include a Crosstabulation table showing the matrix of observed frequencies, and a table of Chi-Square Tests in which you should focus on the Pearson Chi-Square. The table includes the calculated chi-square value, the degrees of freedom, and the level of significance (p value).

Example: The following data represent the observed frequencies for a sample of 50 students who have been classified by gender (male/female) and by self-esteem (high, medium, low). The data produce a chi-square statistic of 2.91 with df = 2 and a significance level of p = 0.233 (no significant relationship).

	Self-Esteem		
	High	Medium	Low
males	10	6	4
females	8	12	10