NOTE: Original exercises and their questions are in *italics*, while the suggested answers are in **bold face** type.

## Module 2: *Statistics Review for Psychological Measurement*

***PROLOGUE***: *The Equal Employment Opportunity Commission (EEOC) has received a complaint about our current Mechanical Comprehension (MC) test from a former job applicant (a female minority) who applied, but was rejected, for our engineering assistant position. As you know, we are in the process of replacing our current MC test with a new one. The EEOC analyst assigned to our case will be here to meet with us in 1 hour so we better have some answers by then! Use the data set "Mechanical Comprehension.sav" to complete the following exercises.*

### EXERCISE 2.1: *COMPUTING DESCRIPTIVE STATISTICS*

***OBJECTIVE****: To practice computing and interpreting descriptive statistics on test data.*

1. *What descriptive information can we provide to the EEOC regarding the current MC test being used? How about the proposed one?*

**Basic information such as the sample sizes (N=474) and frequencies for nominal level data such as demographic data (e.g., sex and race) would be useful. These can be displayed in frequency tables. For interval level data, descriptive statistics such as means, standard deviations, skew, and so on would be useful. See example output below.**

**Statistics**

| | AGE Age of employee | MECH1 Current Mechanical Aptitude Test | MECH2 Proposed Mechanical Aptitude Test | PERF Job Performance Rating | SEXRACE Sex and race classification | WORK Work Experience in years |
|---|---|---|---|---|---|---|
| N | 474 | 474 | 474 | 474 | 474 | 474 |
| Mean | 37.34 | 137.6783 | 68.0643 | 5.0063 | 2.13 | 7.9886 |
| Median | 32.00 | 115.5000 | 60.0000 | 5.0000 | 2.00 | 4.5800 |
| Mode | 30 | 123.00 | 60.00 | 5.00 | 1 | .00 |
| Std. Deviation | 11.931 | 68.30265 | 31.48255 | 1.37709 | 1.050 | 8.71541 |
| Skewness | .849 | 2.125 | 2.853 | -.358 | .176 | 1.510 |
| Std. Error of Skewness | .112 | .112 | .112 | .112 | .112 | .112 |
| Minimum | 23 | 63.00 | 36.00 | 1.00 | 1 | .00 |
| Maximum | 65 | 540.00 | 319.92 | 7.00 | 4 | 39.67 |

**EDLEVEL Education level**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 Less than HS | 53 | 11.2 | 11.2 | 11.2 |
| | 2 HS Diploma or GED | 190 | 40.1 | 40.1 | 51.3 |
| | 3 Some College | 116 | 24.5 | 24.5 | 75.7 |
| | 4 Associates Degree | 50 | 10.5 | 10.5 | 86.3 |
| | 5 Bachelors Degree | 6 | 1.3 | 1.3 | 87.6 |
| | 6 Graduate or Professional Degree | 59 | 12.4 | 12.4 | 100.0 |
| | Total | 474 | 100.0 | 100.0 | |

**JOBCAT Job category**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 CLERICAL | 227 | 47.9 | 47.9 | 47.9 |
| | 2 OFFICE TRAINEE | 136 | 28.7 | 28.7 | 76.6 |
| | 3 SECURITY OFFICER | 27 | 5.7 | 5.7 | 82.3 |
| | 4 COLLEGE TRAINEE | 41 | 8.6 | 8.6 | 90.9 |
| | 5 EXEMPT EMPLOYEE | 32 | 6.8 | 6.8 | 97.7 |
| | 6 MBA TRAINEE | 5 | 1.1 | 1.1 | 98.7 |
| | 7 TECHNICAL | 6 | 1.3 | 1.3 | 100.0 |
| | Total | 474 | 100.0 | 100.0 | |

**MINORITY Minority classification**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 WHITE | 370 | 78.1 | 78.1 | 78.1 |
| | 1 NONWHITE | 104 | 21.9 | 21.9 | 100.0 |
| | Total | 474 | 100.0 | 100.0 | |

**SEX Sex of employee**

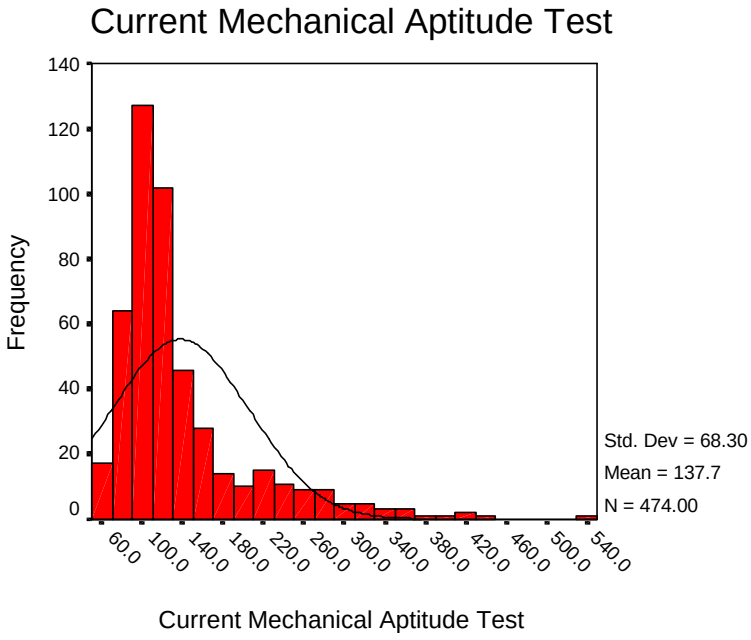
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 MALES | 258 | 54.4 | 54.4 | 54.4 |
| | 1 FEMALES | 216 | 45.6 | 45.6 | 100.0 |
| | Total | 474 | 100.0 | 100.0 | |

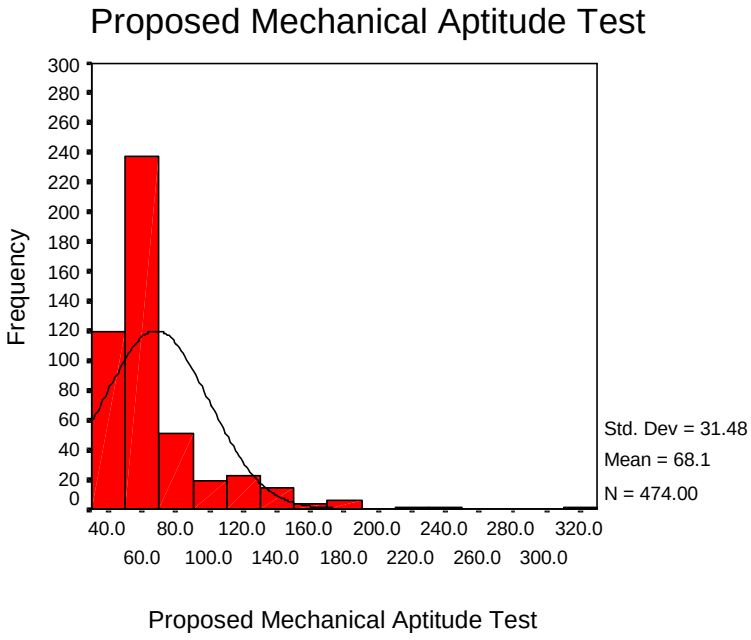
**SEXRACE Sex and race classification**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 WHITE MALES | 194 | 40.9 | 40.9 | 40.9 |
| | 2 MINORITY MALES | 64 | 13.5 | 13.5 | 54.4 |
| | 3 WHITE FEMALES | 176 | 37.1 | 37.1 | 91.6 |
| | 4 MINORITY FEMALES | 40 | 8.4 | 8.4 | 100.0 |
| | Total | 474 | 100.0 | 100.0 | |

2. *Create appropriate graphs to describe the current and proposed MC tests.*

**Frequency histograms would probably work best in depicting the test scores. Superimposing a normal curve (such as in the example output below) tends to be helpful in examining this type of test data. Clearly both the current and proposed mechanical comprehension tests are strongly positively skewed (2+), as confirmed in the first summary table above.**




3. *Compute appropriate measures of central tendency, variability, and shape for the current and proposed MC tests.*

**The table below was created using the "explore" procedure in SPSS instead of "descriptives" or "frequencies" (which was used for the tables above), because it includes additional descriptive statistics not found in those procedures such as trimmed means, inter-quartile ranges, and kurtosis. Explore also computes the standard errors for the mean, skewness, and kurtosis.**

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| MECH1 Current Mechanical Aptitude Test | Mean | | 137.6783 | 3.13724 |
| | 95% Confidence Interval for Mean | Lower Bound | 131.5136 | |
| | | Upper Bound | 143.8429 | |
| | 5% Trimmed Mean | | 129.8208 | |
| | Median | | 115.5000 | |
| | Variance | | 4665.251 | |
| | Std. Deviation | | 68.30265 | |
| | Minimum | | 63.00 | |
| | Maximum | | 540.00 | |
| | Range | | 477.00 | |
| | Interquartile Range | | 52.6500 | |
| | Skewness | | 2.125 | .112 |
| | Kurtosis | | 5.378 | .224 |
| MECH2 Proposed Mechanical Aptitude Test | Mean | | 68.0643 | 1.44604 |
| | 95% Confidence Interval for Mean | Lower Bound | 65.2229 | |
| | | Upper Bound | 70.9058 | |
| | 5% Trimmed Mean | | 64.1669 | |
| | Median | | 60.0000 | |
| | Variance | | 991.151 | |
| | Std. Deviation | | 31.48255 | |
| | Minimum | | 36.00 | |
| | Maximum | | 319.92 | |
| | Range | | 283.92 | |
| | Interquartile Range | | 20.6700 | |
| | Skewness | | 2.853 | .112 |
| | Kurtosis | | 12.390 | .224 |

## EXERCISE 2.2: *COMPUTING BIVARIATE STATISTICS*

**OBJECTIVE**: *To practice computing and interpreting inferential statistics.*

1. *Is the current test related to any other demographic information such as age, education level, or work experience? How about the proposed test?*

**As can be seen in the example output below, both the current and proposed mechanical comprehension tests have a strong, positive correlation with education. In addition, both are negatively correlated with sex (i.e., men scoring significantly higher on both tests than women). In addition, while the correlation between minority classification and both tests scores are statistically significant (i.e., majority scoring higher than minorities), the effect sizes are rather small (i.e., minority classification is only associated with about 2% to 3% of variance in test scores). Work experience and age seem to have little relationship with the test scores.**

**Correlations**

| | | MECH1 Current Mechanical Aptitude Test | MECH2 Proposed Mechanical Aptitude Test |
|---|---|---|---|
| AGE Age of employee | Pearson Correlation | -.132** | -.005 |
| | Sig. (2-tailed) | .004 | .914 |
| | N | 474 | 474 |
| EDLEVEL Education level | Pearson Correlation | .576** | .526** |
| | Sig. (2-tailed) | .000 | .000 |
| | N | 474 | 474 |
| WORK Work Experience in years | Pearson Correlation | -.097* | .045 |
| | Sig. (2-tailed) | .034 | .327 |
| | N | 474 | 474 |
| SEX Sex of employee | Pearson Correlation | -.450** | -.457** |
| | Sig. (2-tailed) | .000 | .000 |
| | N | 474 | 474 |
| MINORITY Minority classification | Pearson Correlation | -.177** | -.158** |
| | Sig. (2-tailed) | .000 | .001 |
| | N | 474 | 474 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

2. *The complainant (with ID #450) is suggesting that the test is biased/unfair. What was her score? What is your best guess of her "true" score? How does her score compare to the scores of other applicants? To the scores of other female applicants? To the scores of other minority applicants? (Look at this in terms of both the current and the proposed test).*

**The person with ID#450 has a score on the current MC test of 65.4, and the proposed MC test of 40.8. In order to get an estimate of her "true" score, you would have to compute the standard error of measurement (SEM). In order to compute the SEM, you would have to have some estimate of reliability. No estimate of reliability is provided however. Given the answers to individual items are not provided, we will be unable to compute measures of internal consistency (e.g., alpha or split half). Therefore, about our only option is to correlate the two versions of the test and use that as an estimate of parallel or equivalent forms reliability. The correlation between the current and proposed MC test is r = .88. We would also need to know the standard deviation of the test. For the current MC test, $S_c$=68.30 and for the proposed test, $S_p$=31.48.**

$$Current\ MC - SEM = S_x\sqrt{1 - r_{xx}} = 68.3\sqrt{1 - .88} = 68.3 * .3464 = 23.66$$

**95% CI for Current MC test true score: 65.4±1.96(23.66)=65.4±46.37=19.03 ≤ $T_c$ ≤ 111.77**

$$\text{Pr}\,oposed\ MC - SEM = S_x\sqrt{1 - r_{xx}} = 31.48\sqrt{1 - .88} = 31.48 * .3464 = 10.90$$

**95% CI for Proposed MC test true score: 40.8±1.96(10.90)=40.8±21.36=19.44 ≤ $T_c$ ≤ 62.16**

**Thus, even with a relatively strong reliability value, the test still shows that the estimated true score is in a very wide range, particularly for the current MC test.**

**In comparing applicant #450 to all other applicants, women, and minorities, she scored well below the mean for each of those subgroups on both the current (her score=65.4) and proposed MC (her score=40.8) tests.**

**Descriptive Statistics - All Applicants**

| | N | Mean | Std. Deviation |
|---|---|---|---|
| MECH1 Current Mechanical Aptitude Test | 474 | 137.6783 | 68.30265 |
| MECH2 Proposed Mechanical Aptitude Test | 474 | 68.0643 | 31.48255 |
| Valid N (listwise) | 474 | | |

**Descriptive Statistics - Female Applicants**

| | N | Mean | Std. Deviation |
|---|---|---|---|
| MECH1 Current Mechanical Aptitude Test | 216 | 104.1277 | 30.23209 |
| MECH2 Proposed Mechanical Aptitude Test | 216 | 52.3679 | 11.74240 |
| Valid N (listwise) | 216 | | |

**Descriptive Statistics - Minority Applicants**

| | N | Mean | Std. Deviation |
|---|---|---|---|
| MECH1 Current Mechanical Aptitude Test | 104 | 114.8558 | 45.68655 |
| MECH2 Proposed Mechanical Aptitude Test | 104 | 58.7158 | 20.03296 |
| Valid N (listwise) | 104 | | |

---

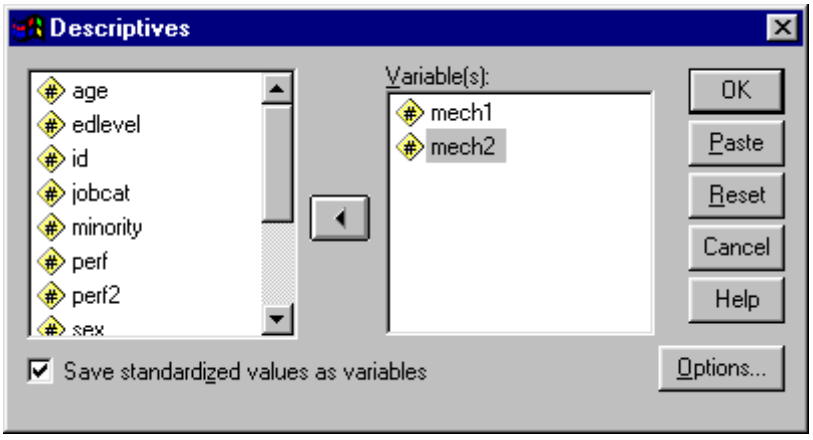
# Additional Exercises

**EXERCISE 2.E1: *COMPUTING STANDARDIZED SCORES*** (Note: *You will have to refer to the further readings for information on how to compute stanine and T-scores, which were not covered in this module because of space limitations.*)

**OBJECTIVE: *To practice creating standardized scores.***

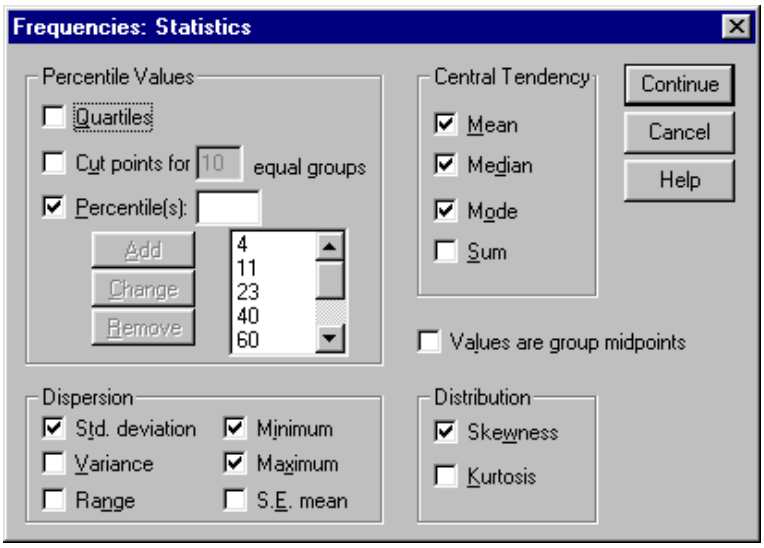
1. *Create standardized Z-scores for both the current and the proposed MC tests.*

**The easiest way to create Z-scores in SPSS is to use the "descriptives" procedure and simply check the box that says, "Save standardized values as variables" (see screen shot below). These new variables will be added as additional columns of data at the end of the data set and labeled "zmech1" and "zmech2," respectively. In addition, the variable labels from the original variables will automatically be duplicated for these variables with the words, "Zscore: " added in front of them.**



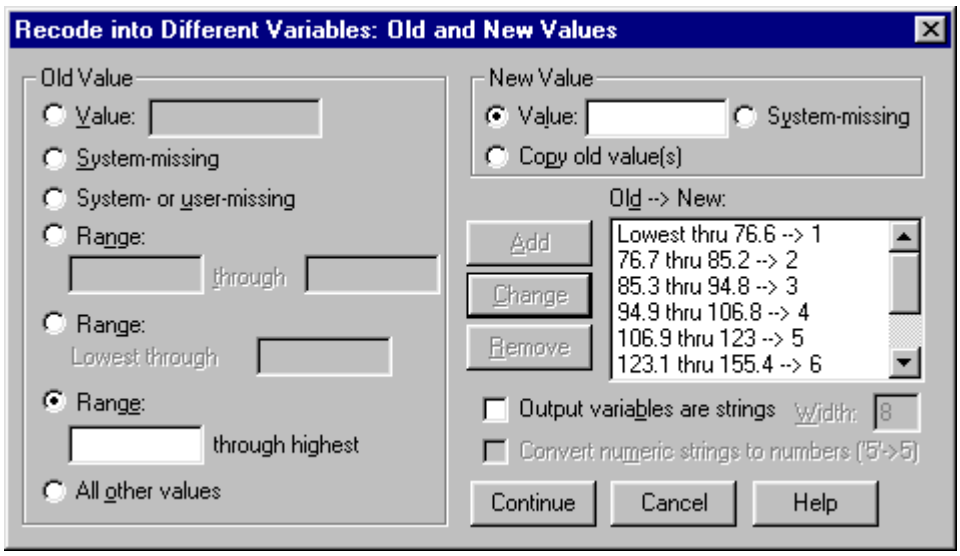
2. *Create stanine scores for both the current and the proposed MC tests.*

**The easiest way to do this in SPSS is to use the "frequencies" procedure and click on "statistics" and request the specific percentiles associated with stanine scores (see screen shot below). Thus, scores below 4% would be recoded to give a stanine score of 1, between 4% and 11% a 2, between 11% and 23% a 3, between 23% and 40% a 4, between 40% and 60% a 5, between 60% and 77% a 6, between 77% and 89% a 7, between 89% and 96% an 8, and those above 96% a 9.**




**Statistics**

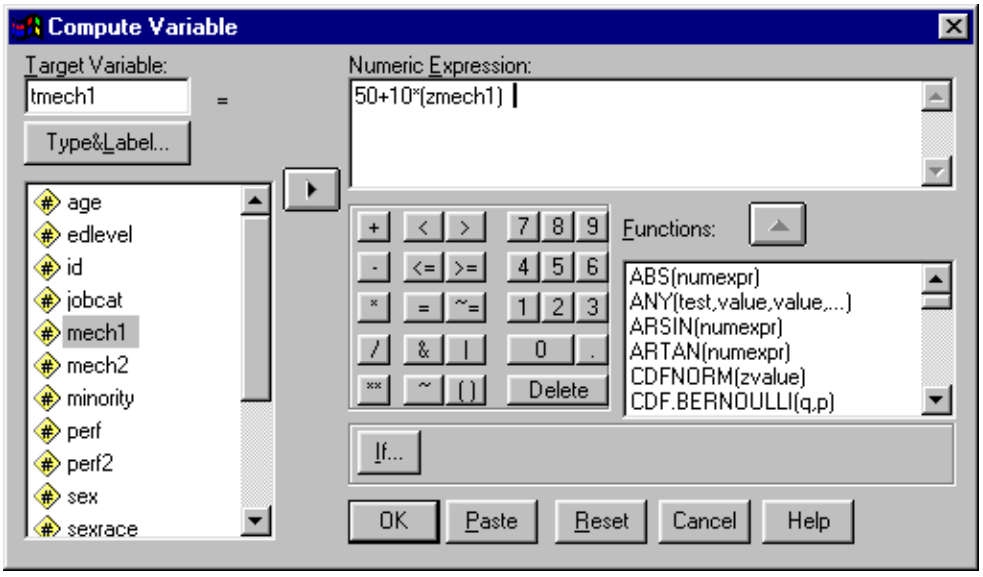
| | | MECH1 Current Mechanical Aptitude Test | MECH2 Proposed Mechanical Aptitude Test |
|---|---|---|---|
| N | Valid | 474 | 474 |
| Percentiles | 4 | 72.6000 | 40.8000 |
| | 11 | 85.2000 | 43.8000 |
| | 23 | 94.8000 | 48.0000 |
| | 40 | 106.8000 | 57.0000 |
| | 60 | 123.0000 | 63.0000 |
| | 77 | 155.4000 | 72.0000 |
| | 89 | 227.7500 | 109.9200 |
| | 96 | 300.0000 | 139.9200 |




3. *Create standardized T-scores (with a mean of 50 and a standard deviation of 10) for both the current and the proposed MC tests.*

   **If students have already saved the z-scores for Mech1 and Mech2 then they can do a simple compute statement using the following formula.**

$$T = 50 + 10(z)$$




**Otherwise students must use this formula which is much more cumbersome.**

$$T = \left(\frac{S_{new}}{S_{old}}\right) X_{old} + \overline{X}_{new} - \left(\frac{S_{new}}{S_{old}}\right) \overline{X}_{old}$$

4. *How do the Z-scores, stanine, and T-scores compare?*

**Below is an abbreviated output comparing the three standardized variables for Mech1 with the original variable. We have only printed the first 10 cases in this example output.**

**Case Summaries[a]**

| | MECH1 Current Mechanical Aptitude Test | ZMECH1 Zscore: Current Mechanical Aptitude Test | SMECH1 Stanine: Current Mechanical Aptitude Test | TMECH1 Tscore: Current Mechanical Aptitude Test |
|---|---|---|---|---|
| 1 | 160.80 | .33852 | 7.00 | 53.39 |
| 2 | 414.00 | 4.04555 | 9.00 | 90.46 |
| 3 | 160.80 | .33852 | 7.00 | 53.39 |
| 4 | 141.00 | .04863 | 6.00 | 50.49 |
| 5 | 124.20 | -.19733 | 6.00 | 48.03 |
| 6 | 123.00 | -.21490 | 5.00 | 47.85 |
| 7 | 220.00 | 1.20525 | 7.00 | 62.05 |
| 8 | 228.00 | 1.32238 | 8.00 | 63.22 |
| 9 | 123.00 | -.21490 | 5.00 | 47.85 |
| 10 | 222.00 | 1.23453 | 7.00 | 62.35 |

a. Limited to first 10 cases.

## EXERCISE 2.E2: *ENTERING AND COMPUTING STATISTICS*

**OBJECTIVE***: To practice entering and computing statistics.*

*Then the data in Tables 2.1 and 2.2 into a statistical analysis program. You can even use a common spreadsheet program to conduct most of the analyses. Try to replicate the findings presented in the module overview. Compute measures of central tendency, variability, and shape. In addition, try to recreate the figures presented in the module. Finally, create your own figures that you think best represent the data.*

**The tables and figures that the student produces should look vary similar to those displayed in the Module 2 overview. Actual statistics, such as measures of central tendency, variability, and shape, should be the same within rounding error. If students do obtain statistics with different values check to see how the program and/or procedure handles missing data. For example, in SPSS the *Descriptives* or *Frequencies* procedures, by default, uses casewise or pairwise deletion of variables. However, the default for the *Explore* procedure is to use listwise deletion. Hence, students may obtain different values than those displayed in the module overview due to the treatment of missing data (i.e., because they have different sample sizes).**