MICROSOFT® EXCEL® TECHNOLOGY STEP-BY-STEP

STATISTICS: INFORMED DECISIONS USING DATA

SEVENTH EDITION

Michael Sullivan, III

Joliet Junior College
and
Florida SouthWestern State College

Pearson

MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAKE NO REPRESENTATIONS ABOUT THE SUITABILITY OF THE INFORMATION CONTAINED IN THE DOCUMENTS AND RELATED GRAPHICS PUBLISHED AS PART OF THE SERVICES FOR ANY PURPOSE. ALL SUCH DOCUMENTS AND RELATED GRAPHICS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS HEREBY DISCLAIM ALL WARRANTIES AND CONDITIONS WITH REGARD TO THIS INFORMATION, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. IN NO EVENT SHALL MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF INFORMATION AVAILABLE FROM THE SERVICES. THE DOCUMENTS AND RELATED GRAPHICS CONTAINED HEREIN COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED HEREIN AT ANY TIME. PARTIAL SCREEN SHOTS MAY BE VIEWED IN FULL WITHIN THE SOFTWARE VERSION SPECIFIED. MICROSOFT®, WINDOWS®, and MICROSOFT OFFICE® ARE REGISTERED TRADEMARKS OF THE MICROSOFT CORPORATION IN THE U.S.A. AND OTHER COUNTRIES. THIS BOOK IS NOT SPONSORED OR ENDORSED BY OR AFFILIATED WITH THE MICROSOFT CORPORATION.

PPID: 0000000000000

Table of Contents

Microsoft Excel Technology Step-by-Step

Section 1.3 Obtaining a Simple Random Sample

1. Be sure Data Analysis Tool Pak is activated.
2. Data > Data Analysis > Random Number Generation
3. Fill in boxes, setting Distribution to Uniform, and type a random seed.
4. Ignore digits to the right of the decimal point.

Section 2.1 Organizing Qualitative Data: The Popular Displays

Frequency Distributions

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab:
4. Qualitative data: Check the box and highlight the data cell range to be analyzed (such as A1:A31).
5. Range/Sheet/Workbook: Select the Sheet option.
6. Variable labels: Check the box if the first row of data contains a label.
7. Outputs tab: Under "Qualitative Data" select Categories and Frequency per category.
8. Charts (2) tab: Under "Values used" choose Frequencies.
9. Click OK.

Relative Frequency Distributions

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab:
4. Qualitative data: Check the box and highlight the data cell range to be analyzed (such as A1:A31).
5. Range/Sheet/Workbook: Select the Sheet option.
6. Variable labels: Check the box if the first row of data contains a label.
7. Outputs tab: Under "Qualitative Data" select Categories and Frequency per category.
8. Charts (2) tab: Under "Values used" choose Relative frequencies.
9. Click OK.

Frequency and Relative Frequency Bar Graphs

Microsoft Excel

1. Enter the summarized data into the spreadsheet. Highlight the categories and their frequencies (or relative frequencies).
2. Insert > Column Charts > Clustered Column Chart

3. Edit as needed.

XLSTAT

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab:
4. Qualitative data: Check the box and highlight the data cell range to be analyzed (such as A1:A31).
5. Range/Sheet/Workbook: Select the Sheet option.
6. Variable labels: Check the box if the first row of data contains a label.
7. Outputs tab: Under "Qualitative Data" select Categories and Frequency per category [or Rel. frequency per category (%)].
8. Charts (2) tab: Under "Values used" choose Frequencies (or Relative frequencies).
9. Click OK.

Side-by-Side Bar Graphs

1. Enter the categories in column A, including a title in cell A1.
2. Enter the respective frequencies in columns B and C, with titles in cells B1 and C1.
3. In a cell below column B, find the sum of the column using "=SUM(frequency array)" and press Enter.
   Repeat for column C.
4. Copy the categories in column D.
5. In cell E1, type the title "Relative Frequency 1."
   Type "Relative Frequency 2" in cell F1.
6. In cell E2, type the formula "=FrequencyCell/$Total$Cell" and press Enter. (Example: =B2/B8)
7. Drag the formula down column E.
8. Repeat to find the relative frequencies for column C in column F.
9. After constructing the side-by-side relative frequency distribution, highlight the categories and relative frequencies.
10. Insert > 2-D Column > Clustered Column Chart
11. Manipulate the chart options as desired.

Pie Charts

Microsoft Excel

1. Enter the summarized data into the spreadsheet.
2. Highlight the categories and their respective frequencies.
3. Insert > Pie Charts > 2-D Pie Chart
4. Edit as needed.

XLSTAT

1. Enter the raw data into the spreadsheet.
2. Highlight the raw data.
3. XLSTAT > Describing Data > Descriptive Statistics
4. Uncheck the Quantitative data box.
   Check the Qualitative data box.
   The sheet describing the raw data should be in the cell.
5. Click Outputs.
6. Be sure Categories, Frequency per category, and Rel. frequency per category (%) are selected.
7. Click Charts (2).
8. Check the box "Pie charts."
9. Choose either Frequencies or Relative frequencies under "Values used."
10. Click OK.

Section 2.2 Organizing Quantitative Data: The Popular Displays

Frequency and Relative Frequency Distributions for Discrete Data

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab:
   Uncheck the Quantitative data box.
   Check the Qualitative data box.
   Highlight the data cell range to be analyzed.
4. Range/Sheet/Workbook: Select Sheet.
5. Variable labels: Check the box if the first row of data contains a label.
6. Outputs tab: Under Qualitative data select Categories, Frequency per category, and Rel. frequency per category (%).
7. Charts (2) tab: Uncheck the boxes for Bar charts and Pie charts.
8. Choose either Frequencies or Relative frequencies under "Values used."
9. Click OK.

Histograms for Discrete Data

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Histograms
3. General tab:
   Data: Highlight cell range to be analyzed (such as A1:A40).
   Data type: Select Discrete or Continuous
   Variable labels: Check the box if the first row of data contains a label.

4. Options tab: Select either Number and enter the desired number of classes or Range and enter the desired class width.
5. Check the Minimum box and enter a value for the lower limit of the first class, if desired.
6. For example, if the smallest observation is 1, then enter 0.5 for Minimum and 1 for Range.
7. Outputs tab: Check the Descriptive statistics box.
8. Charts tab: Check the Histograms box and select Bars.
9. Under "Ordinate of the histograms" select either Frequency or Relative frequency.
10. Click OK.
11. Edit the chart options as desired.

Frequency and Relative Frequency Distributions for Continuous Data

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Histograms
3. General tab:
   Data: Highlight cell range to be analyzed (such as A1:A40).
   Data type: Select Continuous
   Variable labels: Check the box if the first row of data contains a label.
4. Options tab: Select either Number and enter the desired number of classes or Range and enter the desired class width.
5. Check the Minimum box and enter a value for the lower limit of the first class, if desired.
6. If the lower class limit of the first class and class width are both given, enter these values for the Minimum and Range, respectively.
7. Outputs tab: Check the Descriptive statistics box.
8. Charts tab: Uncheck the Histograms box.
9. Click OK.
10. Edit the chart options as desired.

Histograms for Continuous Data

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Histograms
3. General tab:
   Data: Highlight cell range to be analyzed (such as A1:A40).
   Data type: Select Discrete or Continuous.
   Variable labels: Check the box if the first row of data contains a label.
4. Options tab: Select either Number and enter the desired number of classes or Range and enter the desired class width.
5. Check the Minimum box and enter a value for the lower limit of the first class, if desired.
6. Outputs tab: Check the Descriptive statistics box.
7. Charts tab: Check the Histograms box and select Bars.

8. Under "Ordinate of the histograms" select either Frequency or Relative frequency.
9. Click OK.
10. Edit the chart options as desired.

Section 2.3 Additional Displays of Quantitative Data

Stem-and-Leaf Plots

1. Enter the data into a spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab:
   Quantitative Data: Check the box and highlight the data cell range to be analyzed (such as A1:A40).
   Range/Sheet/Workbook: Select Sheet.
   Variable Labels: Check the box if the first row of data contains a label.
4. Charts (1) | Chart types tab: Check the Stem-and-leaf plots box.
   Unit 10^: Fill in the box with 0 if the stem unit is the ones digit, 1 if it's the tens digit, etc.
5. Click OK.

Frequency Polygons

1. Construct a frequency (or relative frequency) distribution, including a "Class Midpoints" column as well as classes before the first class and after the last class using the appropriate midpoints and frequencies of 0.
2. Highlight the column containing the frequencies (or relative frequencies).
3. Insert > Line Charts > Line Chart with Markers
4. Design > Select Data
   In the Select Source dialog box, select Horizontal (Category) Axis Labels | Edit.
5. In the Axis Labels dialog box, highlight the Class Midpoint values and click OK.
6. In the Select Data Source dialog box, click OK.
7. Edit the chart options as desired.

Ogives

1. Type the upper class limits in column A and the appropriate cumulative frequencies in column B.
   (Include a value of 0 for the class that would come before the first class.)
2. Highlight the column containing the cumulative frequencies or cumulative relative frequencies.
3. Insert > Line Charts > Lines with Markers
4. Design > Select Data
   In the Select Data Source dialog box, select Horizontal (Category) Axis Labels > Edit.
5. In the Axis Labels dialog box highlight the Upper Class Limit values and click OK.
6. In the Select Data Source dialog box click OK.

7. Edit the chart options as desired.

Time-Series Plots

1. Enter the data in a spreadsheet.
2. Highlight the column containing the data values (but NOT the time column).
3. Insert > Line Charts > Lines with Markers
4. Design > Select Data
   In the Select Data Source dialog box, select Horizontal (Category) Axis Labels > Edit.
5. Choose the time data.
6. Click the + button to access Chart Elements.
   Select the elements you want visible.
7. Edit the chart options as desired.

Section 3.1 Measures of Central Tendency

Determining the Mean

Microsoft Excel

1. Enter the raw data in one column.
2. Type **=Average(** in the cell you want to include the mean.
3. Highlight the raw data and hit Enter.

XLSTAT

1. Enter the raw data in one column.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General Tab:
   Quantitative data: Check the box and highlight the data cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check the box if the first row contains a label.
4. Outputs tab: Under Quantitative Data, select **Mean**. Click OK.

Determining the Median

Microsoft Excel

1. Enter the raw data in one column.
2. Type **=Median(** in the cell you want to include the mean.
3. Highlight the raw data and hit Enter.

XLSTAT

1. Enter the raw data in one column.
2. XLSTAT > Describing Data > Descriptive Statistics

3. General Tab:
   Quantitative data: Check the box and highlight the data cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check the box if the first row contains a label.
4. Outputs tab: Under Quantitative Data, select **Median**. Click OK.

Section 3.2 Measures of Dispersion

Determining the Range

Microsoft Excel

1. Enter the raw data in one column.
2. Type **=MAX(***data range***)-MIN(***data range***)** in the cell you want to include the range, where *data range* represents the cells containing the raw data (such as A1:A40).
3. Hit Enter.

XLSTAT

1. Enter the raw data in one column.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General Tab:
   Quantitative data: Check the box and highlight the data cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check the box if the first row contains a label.
4. Outputs tab: Under Quantitative Data, select **Range**. Click OK.

Determining the Variance

Microsoft Excel

1. Enter the raw data in one column.
2. Sample Variance: Type **=VAR.S** in the cell you want to include the standard deviation.
   Highlight the raw data and hit Enter.

   Population Variance: Type **=VAR.P** in the cell you want to include the standard deviation.
   Highlight the raw data and hit Enter.

XLSTAT

1. Enter the raw data in one column.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General Tab:
   Quantitative data: Check the box and highlight the data cell range to be analyzed.

Range/Sheet/Workbook: Select the Sheet option.
Variable labels: Check the box if the first row contains a label.

4. Outputs tab: Under Quantitative Data, select **Variance (n-1)** for the sample variance and **Variance (n)** for the population variance. Click OK.

Determining the Standard Deviation

Microsoft Excel

1. Enter the raw data in one column.
2. Sample Standard Deviation: Type **=STDEV.S** in the cell you want to include the standard deviation.
   Highlight the raw data and hit Enter.

   Population Standard Deviation: Type **=STDEV.P** in the cell you want to include the standard deviation.
   Highlight the raw data and hit Enter.

XLSTAT

1. Enter the raw data in one column.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General Tab:
   Quantitative data: Check the box and highlight the data cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check the box if the first row contains a label.
4. Outputs tab: Under Quantitative Data, select **Standard deviation (n-1)** for the sample standard deviation and **Standard deviation (n)** for the population standard deviation. Click OK.

Section 3.3 Measures of Central Tendency and Dispersion from Grouped Data

Determining the Mean from Grouped Data

1. Right-click on Column A and select Format Cells. In the Number Tab select Text and click OK. Enter the classes in Column A, including a title in cell A1.
2. Enter the class midpoints in Column B, including a title in cell B1.
3. Enter the class frequencies in Column C, including a title in cell C1.
4. In cell G1, type the formula "**=SUM(**frequency array**)**" and press Enter.
   For example, =SUM(C2:C13).
5. In cell G2, type the formula "**=SUMPRODUCT**(midpoint array, frequency array**)**" and press Enter.
   For example, =SUMPRODUCT(B2:B13, C2:C13).
6. In cell F3, type the title "Mean".

7. In cell G3, type the formula "**=G2/G1**" and press Enter.

Determining the Weighted Mean

1. Enter the values of the variable in Column A, including a title in cell A1.
2. Enter the weights in Column B, including a title in cell B1.
3. In cell E1, type the formula "=**SUM(**frequency array**)**" and press Enter to find the sum of the weights.
   For example, =SUM(B2:B6).
4. In cell E2, type the formula "**=SUMPRODUCT(**midpoint array, frequency array**)**" and press Enter.
   For example, =SUMPRODUCT(A2:A6,B2:B6).
5. In cell D3, type the title "Weighted Mean".
6. In cell E3, type the formula "**=E2/E1**" and press Enter to compute the weighted mean.

Determining the Standard Deviation from Grouped Data

1. Right-click on Column A and select Format Cells. In the Number Tab select Text and click OK. Enter the classes in Column A, including a title in cell A1.
2. Enter the class midpoints in Column B, including a title in cell B1.
3. Enter the class frequencies in Column C, including a title in cell C1.
4. In cell G1, type the formula "**=SUM(**frequency array**)**" and press Enter.
   For example, =SUM(C2:C13).
5. In cell G2, type the formula "**=SUMPRODUCT(**midpoint array, frequency array**)**" and press Enter.
   For example, =SUMPRODUCT(B2:B13, C2:C13).
6. In cell F3, type the title "Mean".
7. In cell G3, type the formula "**=G2/G1**" and press Enter.
8. In cell D1, type the title "Deviation^2".
9. In cell D2, type the formula "**=(B2-G3)^2**" and press Enter.
   Drag the formula down the column to the bottom row of the frequency distribution.
10. In cell G4, type the formula "**=SUMPRODUCT(**deviation^2 array, frequency array**)**" and press Enter.
    For example, =SUMPRODUCT(D2:D13,C2:C13)
11. In cell F5, type the title "Standard Deviation".
12. In cell G5, type the formula "**=SQRT(G4/(G1-1))**" and press Enter.
    This is the sample standard deviation.
    If the population standard deviation is desired, change the formula to "**=SQRT(G4/G1)**"

Section 3.4 Measures of Position and Outliers

Determining Quartiles

1. Type all the values into column A of the spreadsheet.

2. Type the formula "**=QUARTILE(**array, quartile desired**)**" and press Enter.
   For example, =QUARTILE(A2:A19,1).
3. For "quartile desired," use 1 for Q1, 2 for the median, and 3 for Q3.

Section 3.5 The Five-Number Summary and Boxplots

Determining the Five-Number Summary

1. Enter the raw data into column A of the spreadsheet.
2. Minimum: "**=MIN(**array**)**"
   For example, =MIN(A2:A22)
3. Quartiles: "**=QUARTILE(**array, quartile desired**)**"
   For example, =QUARTILE(A2:A22,1)
4. Maximum: "**=MAX(**array**)**"
   For example, =MAX(A2:A22)

Drawing Boxplots

1. Enter the raw data into the spreadsheet.
2. XLSTAT > Describing Data > Descriptive Statistics
3. General tab
   Quantitative data: Check the box and highlight the cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable Labels: Check the box if the first row contains a label.
4. Charts(1) > Chart types tab: Check the Box plots box.
5. Charts(1) > Options tab: Check the boxes for Horizontal, Group Plots, Variables, Minimum/Maximum, and Outliers.
6. Click OK.

Section 4.1 Scatter Diagrams and Correlation

Scatter Diagrams

Microsoft Excel

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. Highlight the data.
3. Insert > Charts
   Select the Scatter Diagram.
4. Add labels.

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Visualizing Data > Scatter plots
3. General tab
   X: Highlight the explanatory data cell range.
   Y: Highlight the response data cell range.
   Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
   Click OK

Correlation Coefficient

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. In cell D1, type "**=CORREL(**explanatory variable array, response variable array**)**".
   For example, =CORREL(A2:A9,B2:B9)

Section 4.2 Least-Squares Regression

Determining the Least-Squares Regression Line

Data Analysis Toolpak

1. If you have not done so already, load the Data Analysis Toolpak.
2. Data > Data Analysis.
   Highlight Regression in the dialogue box and click OK.
3. Place the cursor in the Input Y Range cell and highlight the data containing the response variable.
4. Place the cursor in the Input X Range cell and highlight the data containing the explanatory variable.
5. Check the boxes of any additional output desired and click OK.

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression

3. General tab
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
4. Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
5. Outputs | General Tab: Check the Correlations and the Predictions and residuals boxes.
   Click OK

Section 4.3 Diagnostics on the Least-Squares Regression Line

The Coefficient of Determinations, R^2

Microsoft Excel

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. In cell D1, type "**=RSQ(**response variable range, explanatory variable range**)**" and press Enter.
   For example, =RSQ(B2:B9,A2:A9)

Data Analysis Toolpak

1. If you have not done so already, load the Data Analysis Toolpak.
2. Data > Data Analysis.
   Highlight Regression in the dialogue box and click OK.
3. Place the cursor in the Input Y Range cell and highlight the data containing the response variable.
4. Place the cursor in the Input X Range cell and highlight the data containing the explanatory variable.
5. Check the boxes of any additional output desired and click OK.

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression
3. General tab
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
4. Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.

5. Outputs | General Tab: Check the Correlations and the Predictions and residuals boxes.
   Click OK

Residual Plots

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression
3. General tab
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
4. Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
5. Outputs | General Tab: Check the Correlations and the Predictions and residuals boxes.
   Click OK

Section 4.4 Diagnostics on the Least-Squares Regression Line

Determining Frequency Marginal Distributions, Relative Frequency Marginal Distributions, and Conditional Distributions

1. Enter the given contingency table into a worksheet, including column and row labels.
2. XLSTAT > Correlation/Association tests > Tests on contingency tables (Chi-square…)
3. General tab:
   Contingency table: Highlight the contingency table cell range
   Data format: Select the contingency table option
   Range/Sheet/Workbook: Select the Sheet option.
   Labels included: Check this box.
4. Outputs tab:
   Check the boxes for
   - Observed frequencies
   - Proportions/Row
   - Proportions/Column
   - Proportions/Total

   and select Proportions.
   Click OK.

Drawing a Bar Graph of a Conditional Distributions

1. Type the conditional distribution in consecutive columns.
2. Highlight the table, click Insert.
3. Select 2-D Column.

Section 5.5 Counting Techniques

Factorials

1. Click on a new cell and type the formula "**=FACT(**n**)**" where n is the number you want to compute the factorial of: $n!$.
2. Press Enter.

Permutations

3. Click on a new cell and type the formula "**=PERMUT(**n,x**)**" where n is the number of objects and x is the number being chosen.
4. Press Enter.

Combinations

1. Click on a new cell and type the formula "**=COMBIN(**n,x**)**" where n is the number of objects and x is the number being chosen.
2. Press Enter.

Section 5.6 Simulating Probability Experiments

Random Integers

1. Click on a new cell and type the formula "**=RANDBETWEEN(**bottom,top**)**" to select a random integer from "bottom" through "top".
2. Press Enter.
3. Copy and paste the formula to a range of cells to generate several random numbers.

Section 6.1 Discrete Random Variables

Finding the Mean of a Discrete Random Variable

1. Enter the values of the random variable in column A, and the corresponding probabilities in column B.
2. Multiply each value from column A by its corresponding probability in column B. Place the product in column C. For example, in cell C1 enter "= A1*B1". Copy the contents of cell C1 for the remaining values.
3. Find the sum of the entries in column C using the sum command. For example, if there are entries in cells C1 through C10, enter "=sum(C1:C10)" into cell C11.

Finding the Standard Deviation of a Discrete Random Variable

1. Enter the values of the random variable in column A, and their corresponding probabilities in column B.

2. Multiply each value from column A by its corresponding probability in column B, placing the product in column C. For example, in cell C1 enter "= A1*B1". Copy the contents of cell C1 for the remaining values.
3. Find the sum of the entries in column C using the sum command. For example, if there are entries in cells C1 through C10, enter "=sum(C1:C10)" into cell C11. This value is the mean.
4. Subtract the mean from each value of the random variable in column A. Enter these values in column D.
5. Square the entries in column D. Store the results in column E.
6. Find the sum of the entries in column E. This represents the variance of the random variable.
7. Find the square root of the sum of the entries in column E using the "=SQRT" command.

Section 6.2 The Binomial Probability Distribution

Computing $P(x)$

- In a new cell, type the formula
  "**=BINOM.DIST(***x*,*n*,*p*,**FALSE)**"
  where x represents the number of successes,
  n represents the number of trials, and
  p represents the probability of any trial's success.
- Press ENTER.

Computing $P(X \leq x)$

- In a new cell, type the formula
  "**=BINOM.DIST(***x*,*n*,*p*,**TRUE)**"
  where x represents the number of successes,
  n represents the number of trials, and
  p represents the probability of any trial's success.
- Press ENTER.

Computing $P(X > x)$

- In a new cell, type the formula
  "**=1 – BINOM.DIST(***x*,*n*,*p*,**TRUE)**"
  where x represents the number of successes,
  n represents the number of trials, and
  p represents the probability of any trial's success.
- Press ENTER.

Section 6.3 The Poisson Probability Distribution

Computing $P(x)$

1. In a new cell, type the formula
   "**=POISSON.DIST(***x*,mean,**FALSE)**"

where x represents the number of successes in an interval of fixed length t, and mean (given by λt) represents the average number of occurrences of the event in a given time t.

2. Press ENTER.

Computing $P(X \leq x)$

1. In a new cell, type the formula
   "**=POISSON.DIST(***x*,mean,**TRUE)**"
   where x represents the number of successes in an interval of fixed length t, and mean (given by λt) represents the average number of occurrences of the event in a given time t.
2. Press ENTER.

Computing $P(X > x)$

1. In a new cell, type the formula
   "**=1 – POISSON.DIST(***x*,mean,**TRUE)**"
   where x represents the number of successes in an interval of fixed length t, and mean (given by λt) represents the average number of occurrences of the event in a given time t.
2. Press ENTER.

Section 7.2 Applications of the Normal Distribution

Computing *P(X < a)*

1. In a new cell type "**=NORM.DIST(***a*, mean, st.dev., **TRUE)**"
   where a is the value desired for the distribution, mean and st.dev. are the mean and standard deviation values for the distribution, and TRUE implies calculating cumulative area up to the given value (from the left side).
2. Press Enter.

Computing *P(X > b)*

1. In a new cell type "**=1 – NORM.DIST(***b*, mean, st.dev., **TRUE)**"
   where b is the value desired for the distribution, mean and st.dev. are the mean and standard deviation values for the distribution, and TRUE implies calculating cumulative area up to the given value (from the left side).
2. Press Enter.

Computing *P(a < X < b)*

1. In a new cell type
   "**=NORM.DIST(***b*, mean, st.dev., **TRUE) – NORM.DIST(**a, mean, st.dev., **TRUE)**"
   where a is the smaller value and b is the larger value desired for the distribution, mean and st.dev. are the mean and standard deviation values for the distribution, and TRUE implies calculating cumulative area up to the given value (from the left side).
2. Press Enter.

Finding Normal Values Corresponding to an Area

1. In a new cell, type "=NORM.INV(left-tail area, mean, st.dev)"
2. Press Enter.

Section 7.3 Assessing Normality

Normal Probability Plots

1. Enter the raw data into the spreadsheet.
2. XLSTAT > Describing data > Normality tests
3. General tab
   Data: Highlight the data cell range to be analyzed.
   Range/Sheet/Workbook: Select the Sheet option.
   Shapiro-Wilk Test: Check this box.
   Sample Labels: Check the box if the data set contains a label in row 1.
   (***Note:*** *Sample Labels may be changed to Variable Labels is future versions of XLSTAT.)*
4. Charts tab: Check both the Normal P-P plots and Normal Q-Q plots boxes. Click OK.

Section 9.1 Estimating a Population Proportion

Confidence Interval for a Population Proportion

1. XLSTAT > Parametric tests > Test for one proportion
2. General Tab
   Frequency: Enter the number of successes, *x*
   Sample size: Enter the sample size, *n*
   Test proportion: Enter any value between 0 and 1
   Data format: Select Frequency
   Range/Sheet/Workbook: Select Sheet
3. Options Tab
   Alternative Hypothesis: Select the ≠ option
   Hypothesized difference (D): Enter 0
   Significance level (%):

Enter the value of $\alpha = 1 -$ confidence level as a whole number
Variance (confidence interval): Select Sample
Confidence Interval: Select Wald

4. Click OK.

Determining Sample Size to Estimate a Population Proportion

With a Prior Estimate

1. In an open cell enter the formula "= $\hat{p}$ * (1 – $\hat{p}$) * **NORM.S.INV**(1 – α/2)/Error)^2" where $\hat{p}$ is the prior estimate of p, $\alpha = 1 -$ confidence level, and *Error* is the desired margin of error.

With No Prior Estimate

1. In an open cell enter the formula "=0.25 * **NORM.S.INV**(1 – α/2)/Error)^2" where $\alpha = 1 -$ confidence level, and *Error* is the desired margin of error.

Section 9.2 Estimating a Population Mean

Confidence Interval for a Population Mean

From Summary Statistics

1. Given $\bar{x}$ is the sample mean, s is the sample standard deviation, n is the sample size, and $\alpha = 1 -$ confidence level:
2. *Critical Value*: In cell A1 enter the formula **"=T.INV.2T**(a, n – 1)"
3. *Standard Error*: In cell A2 enter the formula **"=**s**/SQRT**(n)**"**
4. *Lower Confidence Interval Limit for Mean*: In cell A3 enter the formula: **"=** $\bar{x}$ –A1*A2"
5. *Upper Confidence Interval Limit for Mean*: In cell A4 enter the formula: **"=** $\bar{x}$ +A1*A2"

From Raw Data

1. Enter the data in the worksheet.
2. XLSTAT > Parametric tests > One-sample t-test and z-test
3. General Tab
   Data: Highlight the cell range to be analyzed
   Data format: Select One column per sample
   Range/Sheet/Workbook: Select Sheet
   Column labels: Check this box if the first row of data set contains a label
   Student's t test: Check this box

4. Options Tab
   Alternative Hypothesis: Select the ≠ option
   Theoretical mean: Enter 0
   Significance level (%):
   Enter the value of $\alpha = 1 -$ confidence level as a whole number
5. Click OK.

Determining Sample Size to Estimate a Population Mean

1. In any cell enter the formula: "=(**NORM.S.INV**(1–α/2)*s/Error)^2"
   where s is the sample standard deviation,
   $\alpha = 1 -$ confidence level, and
   Error is the desired margin of error.

Section 10.2 Hypothesis Tests for a Population Proportion

1. XLSTAT > Parametric Tests > Test for one proportion
2. General Tab:
   Frequency: Enter x
   Sample size: Enter n
   Test proportion: Enter the value of p_0 from the null hypothesis
   Data format: Select *Frequency*
   Range/Sheet/Workbook: Select *Sheet*
3. Options Tab:
   Alternative Hypothesis: Select the appropriate alternative hypothesis.
   Hypothesized difference (D): Enter 0
   Significance level (%): Enter the value of α as a whole number
   Variance (confidence interval): Select *Test proportion*
4. Click OK

Section 10.3 Hypothesis Tests for a Population Mean

From Summary Statistics

1. *T-Test Statistic*: In cell A1 enter the formula
   =STANDARDIZE($\bar{x}$, μ_0, s/**SQRT**(n))
2. *P-value*: In cell A2 enter the appropriate formula, as follows.
   Left-tailed Test: **=T.DIST**(A1, $n - 1$, TRUE)
   Right-tailed Test: **=T.DIST.RT**(A1, $n - 1$)
   Two-tailed Test: **=T.DIST.2T**(**ABS**(A1), $n - 1$)

From Raw Data

1. Enter the raw data in a worksheet.
2. XLSTAT > Parametric tests > One sample t-test and z-test
3. **General Tab:**
   Data: Highlight the cell range to be analyzed
   Data Format: Select One column per sample
   Range/Sheet/Workbook: Select Sheet
   Column labels: Check the box if the first row contains a label
   Student's t test: Check this box
4. **Options Tab:**
   Alternative Hypothesis: Select the appropriate alternative hypothesis
   Theoretical mean: Enter μ_0 from the null hypothesis
   Significance level (%): Enter the value of α as a whole number
5. Click OK

Section 11.1 Inference about Two Population Proportions

Hypothesis Tests

1. XLSTAT > Parametric tests > Test for two proportions
2. **General Tab:**
   Frequency 1: Enter *x1*
   Sample size 1: Enter *n1*
   Frequency 2: Enter *x2*
   Sample size 2: Enter *n2*
   Data format: Select Frequencies
   Range/Sheet/Workbook: Select Sheet
   z-test: Check this box
3. **Options Tab:**
   Alternative Hypothesis: Select the appropriate alternative hypothesis.
   Hypothesized difference (D): Enter 0
   Significance level (%): Enter the value of α as a whole number
   Variance: Select pq(1/n1+1/n2)
4. Click OK

Confidence Intervals

1. XLSTAT > Parametric tests > Test for two proportions
2. **General Tab:**
   Frequency 1: Enter *x1*
   Sample size 1: Enter *n1*
   Frequency 2: Enter *x2*

Sample size 2: Enter *n2*
Data format: Select Frequencies
Range/Sheet/Workbook: Select Sheet
z-test: Check this box

3. **Options Tab:**
   Alternative Hypothesis: Select the ≠ option.
   Hypothesized difference (D): Enter 0
   Significance level (%): Enter the value of $\alpha = 1 -$ confidence level as a whole number
   Variance: Select pq(1/n1+1/n2)
4. Click OK

Determining Sample Size to Estimate the Difference Between Two Population Proportions

With a Prior Estimate

1. In an open cell enter the formula:
   "= $\hat{p}_1$ * (1 – $\hat{p}_1$) * $\hat{p}_2$ * (1 – $\hat{p}_2$) * **NORM.S.INV**(1 – α/2)/Error)^2"
   where $\hat{p}_1$ and $\hat{p}_2$ are the prior estimates, $\alpha = 1 -$ confidence level, and *Error* is the desired margin of error.

With No Prior Estimate

2. In an open cell enter the formula "=0.5 * **NORM.S.INV**(1 – α/2)/Error)^2"
   where $\alpha = 1 -$ confidence level, and *Error* is the desired margin of error.

Section 11.2 Inference about Two Means: Dependent Samples

Hypothesis Tests

1. Enter the matched pairs data into columns A and B.
2. XLSTAT > Parametric Tests > Two-sample t-test and z-test
3. **General Tab:**
   Sample 1: Highlight the data cell range to be analyzed as group 1
   Sample 2: Highlight the data cell range to be analyzed as group 2
   Data Format: Select Paired Samples
   Range/Sheet/Workbook: Select Sheet
   Column labels: Check the box if the first row of the data set contains a label
   Student's t test: Check this box
4. **Options Tab:**
   Alternative Hypothesis: Select the appropriate alternative hypothesis
   Hypothesized mean: Enter **0** from the null hypothesis
   Significance level (%): Enter the value of α as a whole number
5. Click OK

Confidence Intervals

1. Enter the matched pairs data into columns A and B.
2. XLSTAT > Parametric Tests > Two-sample t-test and z-test
3. **General Tab:**
   Sample 1: Highlight the data cell range to be analyzed as group 1
   Sample 2: Highlight the data cell range to be analyzed as group 2
   Data Format: Select Paired Samples
   Range/Sheet/Workbook: Select Sheet
   Column labels: Check the box if the first row of the data set contains a label
   Student's t test: Check this box
4. **Options Tab:**
   Alternative Hypothesis: Select the ≠ option
   Hypothesized mean: Enter **0** from the null hypothesis
   Significance level (%): Enter the value of $\alpha = 1$ – confidence level as a whole number
5. Click OK

Section 11.3 Inference about Two Means: Independent Samples

Hypothesis Tests

1. Enter the independent data into columns A and B, respectively.
2. XLSTAT > Parametric Tests > Two-sample t-test and z-test
3. **General Tab:**
   Sample 1: Highlight the data cell range to be analyzed as group 1
   Sample 2: Highlight the data cell range to be analyzed as group 2
   Data Format: Select One column per sample
   Range/Sheet/Workbook: Select Sheet
   Column labels: Check the box if the first row of the data set contains a label
   Student's t test: Check this box
4. **Options Tab:**
   Alternative Hypothesis: Select the appropriate alternative hypothesis
   Hypothesized mean: Enter **0** from the null hypothesis
   Significance level (%): Enter the value of α as a whole number
5. Click OK

Confidence Intervals

1. Given, for samples 1 and 2, that $\bar{x}_1$ and $\bar{x}_2$ are the respective sample means, s_1 and s_2 are the respective sample standard deviations, and n_1 and n_2 are the respective sample sizes, and $\alpha = 1$ – confidence level:
2. *Degrees of Freedom*: In cell A1 enter either $n_1 - 1$ or $n_2 - 1$, whichever is smaller. (Note: This is a conservative approximation of *degrees of freedom*)

3. *Critical Value*: In cell A2 enter the formula: "**=T.INV.2T**(α, A1)"
4. *Standard Error*: In cell A3 enter the formula "**=SQRT**(s_1^2/n_1+s_2^2/n_2)"
5. *Lower Confidence Interval Limit for Mean*:
   In cell A4 enter the formula "=$\left(\bar{x}_1 - \bar{x}_2\right)$ – A2*A3"
6. *Upper Confidence Interval Limit for Mean*:
   In cell A5 enter the formula "=$\left(\bar{x}_1 - \bar{x}_2\right)$ + A2*A3"

Section 12.1 Goodness-of-Fit Test

1. Enter the data into a worksheet, including the categories in column A, observed counts (frequency) in column B, and the expected proportions (in decimal form) or expected counts in column C.
2. XLSTAT > Parametric tests > Multinomial goodness of fit test
3. *Frequencies*: Highlight the observed counts
4. *Expected proportions*: Highlight the expected proportions or counts
5. *Data format*: Select Frequencies if using expected counts or Proportions if using expected proportions (in decimal form)
6. *Range/Sheet/Workbook*: Select Sheet
7. *Column labels*: Check the box if the first row of data set contains a label
8. *Chi-Square test*: Check this box
9. Click OK

Section 12.2 Tests for Independence and the Homogeneity of Proportions

Chi-Square Tests

1. Enter the given contingency table into a worksheet, including column and row labels.
2. XLSTAT > Correlation/Association tests > Tests on contingency tables (Chi-square…)
3. **General Tab**
   Contingency table: Highlight the contingency table range.
   Data format: Select the contingency table option.
   Range/Sheet/Workbook: Select the Sheet option.
   Labels included: Check this box.
4. **Options Tab**: Check the Chi-square test box.
5. **Outputs Tab:**
   Check the boxes for Theoretical frequencies, Proportions/Row, and Proportions/Column. Also select Proportions.
   Click OK.

Section 12.3 Inference about Two Population Proportions: Dependent Samples

McNemar's Test

1. Enter the contingency table into a worksheet, including column and row labels as well as success/failure labels. The numerical data values will be in cells C3:D4.
2. χ^2 *Test Statistic*: In cell A6 enter the formula: "=(D3-C4)^2/(D3+C4)"
3. *P-value*: In cell A7 enter the formula: "**=CHISQ.DIST.RT**(A6,1)"

Section 13.1 Comparing Three or More Means (One-Way Analysis of Variance)

ANOVA

1. In column A enter the group labels in a repeating, non-alternating pattern. Starting in column B enter the respective counts from the table. Include appropriate labels in row 1.
2. XLSTAT > Modeling Data > ANOVA
3. **General Tab:**
   Y/Dependent variables: Highlight the numerical cell range to be analyzed.
   X/Explanatory variables: Check the Qualitative box and highlight the group labels.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check this box.
4. **Outputs | General Tab:**
   Check the boxes for Descriptive statistics, Analysis of Variance, and Predictions and residuals.

Section 13.2 Post Hoc Tests on One-Way Analysis of Variance

Performing Tukey's Test

1. In column A enter the group labels in a repeating, non-alternating pattern. Starting in column B enter the respective counts from the table. Include appropriate labels in row 1.
2. XLSTAT > Modeling Data > ANOVA
3. **General Tab:**
   Y/Dependent variables: Highlight the numerical cell range to be analyzed.
   X/Explanatory variables: Check the Qualitative box and highlight the group labels.
   Range/Sheet/Workbook: Select the Sheet option.
   Variable labels: Check this box.
4. **Outputs | General Tab:**
   Check the boxes for Descriptive statistics, Analysis of Variance, and Predictions and residuals.
5. **Outputs | Means Tab:**
   Check the boxes for Multiple comparisons, Pairwise comparisons, and Tukey (HSD) in the *Pairwise comparisons* menu.

Section 14.1 Testing the Significance of the Least-Squares Regression Model

Hypothesis Test on the Slope

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression
3. **General Tab:**
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
   Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
4. **Outputs | General Tab:**
   Check the Correlations and the Predictions and residuals boxes. Click OK

Data Analysis Toolpak

1. If you have not already done so, load the Data Analysis ToolPak.
2. Data > Data Analysis, highlight Regression and click OK.
3. Place the cursor in the Input Y Range cell. Highlight the data containing the response variable.
4. Place the cursor in the Input X Range cell. Highlight the data containing the explanatory variable.
5. Check the boxes of any additional output desired.
6. Click OK.

Confidence Interval on the Slope

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression
3. **General Tab:**
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
   Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
4. **Outputs | General Tab:**
   Check the Correlations and the Predictions and residuals boxes. Click OK

Data Analysis Toolpak

1. If you have not already done so, load the Data Analysis ToolPak.
2. Data > Data Analysis, highlight Regression and click OK.
3. Place the cursor in the Input Y Range cell. Highlight the data containing the response variable.
4. Place the cursor in the Input X Range cell. Highlight the data containing the explanatory variable.
5. Check the boxes of any additional output desired.
6. Click OK.

Section 14.3 Multiple Regression

XLSTAT

1. Enter the data into the spreadsheet with the explanatory variable in column A and the response variable in column B.
   Include labels in row 1.
2. XLSTAT > Modeling Data > Linear Regression
3. **General Tab:**
   Y/Dependent variables: Highlight the response data cell range.
   X/Explanatory variables: Highlight the explanatory data cell range.
   Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
4. **Outputs | General Tab:**
   Check the Correlations and the Predictions and residuals boxes. Click OK

Section 14.5 Polynomial Regression

Data Analysis ToolPak

1. Enter the explanatory variable in column A. In column B, determine the squared values of the entries in column A. Enter the response variable in column C.
2. Select the Data menu; then select Data Analysis. Select Regression and click OK.
3. With the cursor in the "Input Y Range:" cell, highlight the data for the response variable. With the cursor in the "Input X Range:" cell, highlight the data in columns A and B. Check the boxes for Residuals, Residual Plots, and Normal Probability Plots. Click OK. The residuals are stored in the spreadsheet. Draw a boxplot of the residuals as indicated in Section 3.5. NOTE: XLSTAT may also be used to find a polynomial regression.

Section 15.2 Runs Test for Randomness

XLSTAT

1. Enter the data in column A, if necessary.
2. XLSTAT > Test a hypothesis > Nonparametric tests > One-sample runs test
3. **General Tab:**
   Data: Highlight the data cell range.
   Data type: Choose either Quantitative or Qualitative
   Range/Sheet/Workbook: Select Sheet.
   Variable labels: Check the box if the first row contains a label.
4. **Outputs | General Tab:**
   Check the Descriptive statistics box. Click OK.

Section 15.4 Wilcoxon Matched-Pairs Signed-Rank Test

XLSTAT

1. Enter the raw paired data in Columns A and B.
2. XLSTAT > Test a hypothesis > Nonparametric tests > Comparison of two samples
3. **General Tab:**
   Sample 1: Highlight the data cell range in column A.
   Sample 2: Highlight the data cell range in column B.
   Data format: Check the box "Paired samples".
   Data type: Choose either Quantitative or Qualitative
   Range/Sheet/Workbook: Select Sheet.
   Column labels: Check the box if the first row contains a label.
   Check the box "Wilcoxon signed-rank test".
4. **Options Tab:**
   Alternative hypothesis: Choose the appropriate alternative hypothesis.
5. **Outputs Tab:**
   Check the Descriptive statistics, Detailed results, and Summary table boxes. Click OK.

XLSTAT Wilcoxon Signed-Rank Test on a Single Sample

1. Enter the raw data in Column A.
2. XLSTAT > Test a hypothesis > Nonparametric tests > One sample Wilcoxon signed-rank test
3. **General Tab:**
   Samples: Highlight the data cell range in column A.
   Range/Sheet/Workbook: Select Sheet.
   Column labels: Check the box if the first row contains a label.
4. **Options Tab:**

Alternative hypothesis: Choose the appropriate alternative hypothesis.

5. **Outputs Tab:**
   Check the Descriptive statistics, Detailed results, and Summary table boxes. Click OK.

Section 15.5 Mann-Whitney Test

XLSTAT

1. Enter the raw paired data in Columns A and B.
2. XLSTAT > Test a hypothesis > Nonparametric tests > Comparison of two samples
3. **General Tab:**
   Sample 1: Highlight the data cell range in column A.
   Sample 2: Highlight the data cell range in column B.
   Data format: Check the box "One column per sample".
   Range/Sheet/Workbook: Select Sheet.
   Column labels: Check the box if the first row contains a label.
   Check the box "Mann-Whitney test".
4. **Options Tab:**
   Alternative hypothesis: Choose the appropriate alternative hypothesis.
   Hypothesized difference (D): Enter 0
5. **Outputs Tab:**
   Check the Descriptive statistics, Detailed results, and Summary table boxes. Click OK.

Section 15.7 Mann-Whitney Test

XLSTAT

1. Enter the level of the factor in Column A. Enter the value of the response variable in Column B.
2. XLSTAT > Test a hypothesis > Nonparametric tests > Comparison of *k* samples
3. **General Tab:**
   Samples: Highlight the data cell range in column B.
   Sample identifiers: Highlight the data cell range in column A.
   Data format: Check the box "One column per variable".
   Range/Sheet/Workbook: Select Sheet.
   Column labels: Check the box if the first row contains a label.
   Check the box "Kruskal-Wallis test".
4. **Outputs Tab:**
   Check the Descriptive statistics, Detailed results, and Summary table boxes. Click OK.