Chapter 9

# Applying the Concepts: Excel Instructions

## Perform a Two-Sample T-Test

1. Open the carseats\_sales\_us.csv data set.
2. Click “Data” tab
3. Click “Data Analysis”
4. Click “t-Test: Two-Sample Assuming Unequal Variances”
5. Click “OK”
6. Variable 1 Range → *select the data in the sales\_us\_yes column* A2:A259
7. Variable 2 Range → *select the data in the sales\_us\_no column* B2:B143
8. Click “OK”

## Connecting Linear Regression and T-Tests

Note: Because of the way t-tests and regression are implemented in Excel, the data need to be in a format specific to each. This is why there is a different file for each procedure. In other words, carsteats\_sales\_us.csv is the same data as carseats.csv but formatted differently.

Perform the Two-Sample T-Test:

1. Open the carseats\_sales\_us.csv data set.
2. Click “Data” tab
3. Click “Data Analysis”
4. Click “t-Test: Two-Sample Assuming Equal Variances”
5. Click “OK”
6. Variable 1 Range → *select the data in the sales\_us\_yes column* A2:A259
7. Variable 2 Range → *select the data in the sales\_us\_no column* B2:B143
8. Click “OK”

Fit a Regression Line:

1. Open the carseats.csv data set.
2. Create Indicator Variables for US **first**.
   1. In cell M1, type “us\_yes”
   2. In cell M2, type:

=IF(K2=”Yes”,1,0)

* 1. Drag this formula to M401 to apply to all cells.

1. Click “Data” tab
2. Click “Data Analysis”
3. Click “Regression”
4. Click “OK”
5. Input Y Range → *select all data in the sales column*, A2:A401
6. Input X Range → *select all data in the us\_yes column*, M2:M401
   1. Be sure that your columns do not contain any missing values!
7. Check the box for “Labels”
8. Click “OK”

## Compute a Correlation and Fit a Regression Line

Compute a Correlation:

1. Open the carseats.csv data set.
2. Click “Data” tab
3. Click “Data Analysis”
4. Click “Correlation”
5. Click “OK”
6. Input Range → *select all data in the sales and advertising columns* (A1:D401)
   1. Your data selection can include more than these two columns as long as it contains at least these two.
   2. Be sure that your data selection does not include any non-numeric values.
7. Check the box for “Labels in First Row”
8. Click “OK”

## Extrapolate Weight

Fit regression model

1. Enter the age and weight data from the text into two columns in a new spreadsheet.
   1. Age in column A
   2. Weight in column B
2. Click “Data” tab
3. Click “Data Analysis”
4. Click “Regression”
5. Click “OK”
6. Input Y Range → *select all data in the weight column* B1:B12
7. Input X Range → *select all data in the age column* A1:A12
   1. Be sure that your columns do not contain any missing values!
8. Check the box for “Labels”
9. Click “OK”

Make prediction for age 40

1. In an empty cell type:

=B17+B18\*40

## Exploring R-Squared in Multiple Linear Regression

Fit Regression Model:

1. Open the carseats.csv data set.
2. Create Indicator Variables for US and Shelf\_Location **first**
   1. In cell M1, type “us\_yes”
   2. In cell M2, type:

=IF(K2=”Yes”,1,0)

* 1. Drag this formula to M401 to apply to all cells.
  2. In cell N1, type “shelf\_good”
  3. In cell N2, type:

=IF(G2=”Good”,1,0)

* 1. Drag this formula to N401 to apply to all cells.
  2. In cell O1, type “shelf\_medium”
  3. In cell O2, type:

=IF(G2=”Medium”,1,0)

* 1. Drag this formula to O401 to apply to all cells.

1. Click “Data” tab
2. Click “Data Analysis”
3. Click “Regression”
4. Click “OK”
5. Input Y Range → *select all data in the Sales column* A2:A401
6. Input X Range → *select all data in the columns for the variables you want to include in your model* (e.g., if your model includes Competitor\_Price then you would include B2:B401)
   1. Be sure that your columns do not contain any missing values!
   2. Be sure not to include Sales in your X Range!
   3. If your model includes US, then include the *us\_yes* variable
   4. If your model includes Shelf\_Location, then include both the *shelf\_good* and the *shelf\_medium* variables
   5. Be sure that all variables you want to include are in columns next to each other. This may require copying existing columns to new empty columns in column P or farther to the right.
7. Check the box for “Labels”
8. Click “OK”
9. Record R-squared value in a new column (T), with the name “R-squared” at the top
10. Record the number of variables in your model in a new column (U), with the name “Number of Variables” at the top (e.g., U2 = 1 because there is one variable in the regression, and T2 = .004 as the R-squared value for it).
11. Repeat steps 2-11 for each model.

Create Summary Plot of Models:

1. Select the data.
   1. Click on the column headers for both R-square (column T) and number of variables (Column U) to select all the data in those columns.
2. Insert a scatter plot.
   1. With Columns T and U highlighted, go to the Insert tab on the Excel ribbon.
   2. Click on the Scatter (X,Y) icon in the Charts group.
   3. Choose the first scatter chart option, a simple scatter chart with no lines.

## Fit, Evaluate, and Interpret Decision Trees

Currently not available in Excel.

## Fit, Evaluate, and Interpret Random Forests

Currently not available in Excel.

## Fit and Evaluate k-Nearest Neighbors

Currently not available in Excel.

## Compare Regression Models

Currently not available in Excel.

## Fit, Evaluate, and Interpret Logistic Regression

Currently not available in Excel.

## Compare Classification Models

Currently not available in Excel.