Chapter 9

# Applying the Concepts: StatCrunch Instructions

## Perform a Two-Sample T-Test

1. Open the carseats\_sales\_us.csv data set.
2. Stat → T Stats → Two Sample → With Data
3. Sample 1 Values in → sales\_us\_yes
4. Sample 2 Values in → sales\_us\_no
5. Optional graphs and tables → Summary statistics
6. Click Compute!

## Connecting Linear Regression and T-Tests

Note: Because of the way t-tests and regression are implemented in StatCrunch, 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 carsteats\_sales\_us.csv data set.
2. Stat → T Stats → Two Sample → With Data
3. Sample 1 Values in → sales\_us\_yes
4. Sample 2 Values in → sales\_us\_no
5. Calculation options → Pool variances
6. Click Compute!

Fit a Regression Line:

Create Indicator Variables for US first:

1. Open the carseats.csv data set.
2. Data → Indicator
3. Select columns → US

Regression Line second:

1. Stat → Regression → Simple Linear
2. X variable → US = Yes
3. Y variable → Sales
4. Click Compute!

## Compute a Correlation and Fit a Regression Line

Compute a Correlation:

1. Open the carseats.csv data in StatCrunch.
2. Stat → Summary Stats → Correlation
3. Select columns → Sales, Advertising
4. Click Compute!

Fit a Regression Line:

1. Open the carseats.csv data in StatCrunch.
2. Stat → Regression → Simple Linear
3. X variable → Advertising
4. Y variable → Sales
5. Click Compute!

## Extrapolate Weight

1. Enter the age and weight data from the text into two columns in a new table.
2. Stat → Regression → Simple Linear
3. X variable → age
4. Y variable → weight
5. Beneath Prediction of Y, type in the X value(s) box: *40*
6. Click Compute!

## Exploring R-Squared in Multiple Linear Regression

Fit Regression Model:

1. Open the carseats.csv data in StatCrunch.
2. Stat → Regression → Multiple Linear
3. Y variable → Sales
4. X variables → (*the explanatory variable(s) in your model)*
   1. If your model includes categorical variables then you will need to create Indicator variables of them **first** (Data → Indicator → Select your categorical variable); then include all but one of them in your model. For example, if a categorical variable has three levels/values then you will only include two of the corresponding indicator variables in your model.
   2. If your model includes US, then include the “*US = Yes”* variable
   3. If your model includes Shelf\_Location, then include both the “*Shelf\_Location = Good”* and the “*Shelf\_Location = Medium”* variables
5. Click Compute!
6. Record R-squared value in a new column with the name “R-Squared” at the top, and record the number of variables in your model in a new column with the name “Number of Variables” at the top (e.g., for the first row “Number of Variables” = 1 because there is one variable in the regression, and “R-Squared” = .004)
7. Repeat steps 2-6 for each model.

Create Summary Plot of Models:

1. Graph → Scatter Plot
2. X variable → *your column with the number of variables in each model*
3. Y variable → *your column with the R-squared value for each model*
4. Click Compute!

## Fit, Evaluate, and Interpret Decision Trees

StatCrunch functionality projected for 2025.

## Fit, Evaluate, and Interpret Random Forests

StatCrunch functionality projected for 2025.

## Fit and Evaluate k-Nearest Neighbors

StatCrunch functionality projected for 2025.

## Compare Regression Models

StatCrunch functionality projected for 2025.

## Fit, Evaluate, and Interpret Logistic Regression

StatCrunch functionality projected for 2025.

## Compare Classification Models

StatCrunch functionality projected for 2025.