Chapter 4

# Applying the Concepts: StatCrunch Instructions

## Calculate Mean, Mode, Median, 90% Trimmed Average

1. Open the IMDb.csv data in StatCrunch
2. Stat → Summary Stats → Columns
3. Columns → Meta\_score
4. Statistics → Mean, Median, Mode
5. Click Compute!
6. Calculate the 90% trimmed average.
   1. Select Stat → Summary Stats → Columns
   2. Columns → Meta\_score
   3. In the “Where:” box, enter:

*Meta\_score >percentile(Meta\_score,.05) and Meta\_score < percentile(Meta\_score,.95)*

* 1. Statistics → Mean
  2. Click Compute!

## Calculate the Range, Variance, and Standard Deviation

1. Open the IMDb.csv data in StatCrunch
2. Stat → Summary Stats → Columns
3. Select columns → Meta\_score
4. Statistics → Variance, Std. dev., and Range
5. Click Compute!

## Describe the Shape of a Histogram

1. Open the IMDb.csv data in StatCrunch
2. Graph → Histogram.
3. Select columns → IMDB\_rating
4. Under bins:
   1. Start at: 7.5
   2. Width: 0.19
5. Click Compute!

## Investigate Resistance

1. Open the IMDb.csv data in StatCrunch
   1. Stat → Summary Stats → Columns
   2. Columns → Gross
   3. Statistics → Mean, Variance, Std. dev., Median, Range, Mode
   4. Click Compute!
2. Calculate the 90% trimmed average
   1. Select Stat → Summary Stats → Columns
      1. Under Columns, select Gross
      2. In the “Where:” Box, enter:

*Gross >=percentile(Gross,.05) and Gross<= percentile(Gross,.95)*

* + 1. Under “Statistics” select Mean
    2. Click Compute!

1. Add a movie making $10 billion in gross revenue. Gross revenue is listed in millions, so a value of 10,000.0 should be entered.
   1. Locate the cell under the last observation in the Gross column.
   2. Add a new entry of 10000 for Gross; everything else can be left blank.
2. Calculate summary statistics by including the new outliers:
   1. Stat → Summary Stats → Columns
   2. Columns → Gross.
   3. Statistics → Mean, Median, Mode, Range, Std. dev., Mode
   4. Click Compute!
3. Calculate the 90% trimmed average
   1. Stat → Summary Stats → Columns
   2. Columns → Gross
   3. In the “Where:” box, type:

*Gross >percentile(Gross,.05) and Gross< percentile(Gross,.95)*

* 1. Statistics → Mean
  2. Click Compute!

## Plot Correlations Between Two Quantitative Variables

1. Open the IMDb.csv data in StatCrunch
2. Graph→Scatter Plot
3. X variable → Meta\_score
4. Y variable → Gross
5. Overlay polynomial order:
   1. Select ‘1’
6. Click Compute!
7. To calculate the correlation coefficient:
   1. Stat → Summary Stats → Correlation
   2. Columns → Meta\_score and Gross
   3. Click Compute!
8. Repeat steps 2-7 using Runtime as the “X Variable” and IMDb\_rating for the “Y Variable”

## Match Box Plots with Histograms

StatCrunch functionality projected for 2025.

## Associate Two Categorical Variables

1. Open the IMDb.csv data in StatCrunch.
2. Graph → Bar Plot → With Data
3. Column → Genre
4. In the “Where:” box enter:   
   *(Genre = “Drama” or Genre = “Action” or Genre = “Animation” or Genre = “Comedy”) AND (Certificate = “UA” or Certificate = “A” or Certificate = “U” or Certificate = “R”)*
5. Group by → Certificate
6. Click Compute!

## Determine Outliers Using the Quartile and Mean/Standard Deviation Methods

1. Open the IMDb.csv file in StatCrunch.
2. Calculate the Quartile Method thresholds using the IMDB\_Rating variable.
   1. Select Data → Compute → Multiple Expressions
   2. In the first row’s Name box, type “Q Mild Lo” (note: we use Q for Quartile Method)
   3. In the first row’s Expression box, enter: *Q1(IMDB\_Rating)-1.5\*iqr(IMDB\_Rating)*
   4. In the second row’s Name box, type “Q Mild Hi”
   5. In the second row’s Expression box, enter: *Q3(IMDB\_Rating)+1.5\*iqr(IMDB\_Rating*)
   6. In the third row’s Name box, type “Q Reg Lo”
   7. In the third row’s Expression box, enter: *Q1(IMDB\_Rating)-3\*iqr(IMDB\_Rating)*
   8. Click the + button at the end of the third row to add a fourth row.
   9. In the fourth row’s Name box, type “Q Reg Hi”
   10. In the fourth row’s Expression box, enter: *Q3(IMDB\_Rating)+3\*iqr(IMDB\_Rating)*
   11. All four rows should keep the default check box under Save
   12. Click Compute!
   13. Look for the answers in the first row of new columns in the StatCrunch sheet.
3. Calculate the Mean/SD Method thresholds using the IMDB\_Rating variable.
   1. Data→Compute→Multiple Expressions
   2. In the first row’s Name box, type “M Mild Lo” (note: we use M for Mean/SD Method)
   3. In the first row’s Expression box, enter: *mean(IMDB\_Rating)-2\*std(IMDB\_Rating)*
   4. In the second row’s Name box, type “M Mild Hi”
   5. In the second row’s Expression box, enter: *mean(IMDB\_Rating)+2\*std(IMDB\_Rating)*
   6. In the third row’s Name box, type “M Reg Lo”
   7. In the third row’s Expression box, enter: *mean(IMDB\_Rating)-3\*std(IMDB\_Rating)*
   8. Click the + button at the end of the third row to add a fourth row.
   9. In the fourth row’s Name box, type “M Reg Hi”
   10. In the fourth row’s Expression box, enter: *mean(IMDB\_Rating)+3\*std(IMDB\_Rating)*
   11. All four rows should keep the default check box under
   12. Click Compute!
   13. Look for the answers in the first row of new columns in the StatCrunch sheet.
4. Count the observations for High Outliers
   1. Select Stat → Summary Stats → Columns
   2. Under "Select column(s)" select IMDB\_Rating.
   3. Count Outliers using the Quartile Method
      1. For Mild High outliers, enter into the “Where:” box:   
         *IMDB\_Rating > Q3(IMDB\_Rating) + (1.5 \* iqr(IMDB\_Rating))*
   4. Under “Statistics” select n.
   5. Click Compute!
   6. Repeat for the remaining outlier calculations, replacing 4(c)(i) with the following “Where:” field inputs:
      1. Quartile Method for Regular High outliers:

*IMDB\_Rating > Q3(IMDB\_Rating) + (3 \* iqr(IMDB\_Rating))*

* + 1. Mean/SD Method for Mild High outliers:  
       *IMDB\_Rating > mean(IMDB\_Rating) + (2\*std(IMDB\_Rating))*
    2. Mean/SD Method for Regular High outliers  
       *IMDB\_Rating > mean(IMDB\_Rating) + (3\*std(IMDB\_Rating))*

1. Count the observations for Low Outliers
   1. Stat → Summary Stats → Columns
   2. Select columns → IMDB\_Rating
   3. Count Outliers using the Quartile Method
      1. For Mild Low outliers, enter into the “Where:” box:   
         *IMDB\_Rating < Q1(IMDB\_Rating) - (1.5 \* iqr(IMDB\_Rating))*
   4. Statistics → n
   5. Click Compute!
   6. Repeat for the remaining outlier calculations, replacing 4(c)(i) with the following “Where:” field inputs:
      1. Quartile Method for Regular Low outliers:

*IMDB\_Rating < Q1(IMDB\_Rating) - (3 \* iqr(IMDB\_Rating))*

* + 1. Mean/SD Method for Mild Low outliers:  
       *IMDB\_Rating < mean(IMDB\_Rating) - (2\*std(IMDB\_Rating)*)
    2. Mean/SD Method for Regular Low outliers:  
       I*MDB\_Rating < mean(IMDB\_Rating) - (3\*std(IMDB\_Rating))*