Chapter 4

# Applying the Concepts: Excel Instructions

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

1. Open IMDb.csv in Excel.
2. Verify the Meta\_score variable is in Column H.
3. Type the following formulas into empty cells.
   1. Mean:

=ROUND(AVERAGE(H:H),1)

* 1. 90% Trimmed Average:

=ROUND(TRIMMEAN(H:H, 0.1),1)

* 1. Median:

=ROUND(MEDIAN(H:H),1)

* 1. Mode:

=ROUND(MODE(H:H),1)

## Calculate the Range, Variance, and Standard Deviation

1. Open IMDb.csv in Excel.
2. Verify the Meta\_score variable is in Column H.
3. Type the following formulas into empty cells.
   1. Range:

=MAX(H:H)-MIN(H:H)

* 1. Variance:

=ROUND(VAR.S(H:H),1)

* 1. Standard Deviation:

=ROUND(STDEV.S(H:H),1)

## Describe the Shape of a Histogram

1. Open the IMDb.csv file in Excel.
2. Select the column F containing IMDb\_rating.
3. Go to the Insert tab.
4. Choose Histogram from the Chart group.
5. Adjust the number of bins:
   1. Right-click the x-axis under the bins to open a navigation menu.
   2. Select Format Axis.
   3. Under Bins, set the ‘Number of bins’ to 16.
6. Customize the chart:
   1. Set the chart title to ‘Histogram of IMDb Ratings’.
   2. Label the x-axis as IMDb Rating’ and the y-axis as ‘Frequency’.

## Investigate Resistance

1. Open the IMDb.csv file in Excel and use the variable Gross in column O. Calculate the summary statistics:
   1. Verify the Gross variable is in Column O.
   2. Type the following formulas into empty cells.
      1. Mean:

=ROUND(AVERAGE(O:O),1)

* + 1. 90% Trimmed Average:

=ROUND(TRIMMEAN(O:O, 0.1),1)

* + 1. Median:

=ROUND(MEDIAN(O:O),1)

* + 1. Mode:

=ROUND(MODE(O:O),1)

* + 1. Range:

=MAX(O:O)-MIN(O:O)

* + 1. Variance:

=ROUND(VAR.S(O:O),1)

* + 1. Standard Deviation:

=ROUND(STDEV.S(O:O),1)

1. After completing these calculations, add a movie to the bottom of the Gross column O, making $10 billion in gross revenue. Gross revenue is listed in $Million, so a value of 10,000.0 should be entered into cell O833.
2. Follow the same steps as the previous activities to calculate the following summary statistics by including the new outliers:
   1. Mean:

=ROUND(AVERAGE(O:O),1)

* 1. 90% Trimmed Average:

=ROUND(TRIMMEAN(O:O, 0.1),1)

* 1. Median:

=ROUND(MEDIAN(O:O),1)

* 1. Mode:

=ROUND(MODE(O:O),1)

* 1. Range:

=MAX(O:O)-MIN(O:O)

* 1. Variance:

=ROUND(VAR.S(O:O),1)

* 1. Standard Deviation:

=ROUND(STDEV.S(O:O),1)

## Plot Correlations Between Two Quantitative Variables

1. Open IMDb.csv in Excel.
2. Select columns H (Meta\_score) and O (Gross) simultaneously.
3. Go to Insert→Recommended Charts→Scatter.
4. Rescale the y-axis.
   1. Right-click the y-axis.
   2. Select Format Axis…
   3. Under Bounds, change the Maximum to 1000.0
5. To add a trendline (regression line):
   1. Click on a data point in the scatterplot.
   2. Right-click→Add Trendline→choose Linear.
   3. Change the color to red.
6. Label axes:
   1. Click on the chart, then navigate to the ‘Chart Design’ tab.
   2. Click on Add Chart Element→Axis Titles
   3. Add appropriate titles (e.g., ‘Critic Rating’ for x-axis, ‘User Rating’ for y-axis).
7. For the correlation coefficient, in an empty cell, type the formula: =CORREL(H:H,O:O).
8. Repeat these steps for the other variable pair (Runtime and IMDb\_rating, setting the y-axis scale from 7 to 9.5 and the x-axis from 50 to 250 to see the effect more clearly).

## Match Box Plots with Histograms

Currently not available in Excel.

## Associate Two Categorical Variables

1. Open IMDb.csv in Excel.
2. Insert a pivot table:
   1. Highlight Columns A:O by selecting them simultaneously.
   2. Click Insert→PivotTable.
   3. Choose a new worksheet or existing location.
3. Configure the pivot table using the PivotTable Fields box to the right.
   1. Drag Genre to Rows.
   2. Drag Certificate to Columns.
4. Drag Series\_Title to Values. Keep the default of ‘Count of Series\_Title”.
5. Filter the Genre variable to include only ‘Drama’, ‘Action’, ‘Animation’, and ‘Comedy’.
   1. Click the dropdown arrow of "Row Labels" in cell A4.
   2. A list of all unique values for Genre will appear.
   3. Uncheck the 'Select All' option to clear all selections.
   4. Scroll through the list and find ‘Drama’, ‘Action’, ‘Animation’, and ‘Comedy’, and check the boxes next to each.
   5. Click 'OK' or press 'Enter'.
   6. After applying the filter, only the rows where "Genre" equals ‘Drama’, ‘Action’, ‘Animation’, and ‘Comedy’ will be displayed.
6. Filter the Certificate variable to include only ‘UA’, ‘A’, ‘U’, and ‘R’.
   1. Click the dropdown arrow of "Column Labels" in cell B3.
   2. A list of all unique values for the Certificate will appear.
   3. Uncheck the 'Select All' option to clear all selections.
   4. Scroll through the list, find ‘UA’, ‘A’, ‘U’, and ‘R’, and check the boxes next to each.
   5. Click 'OK' or press 'Enter'.
   6. After applying the filter, only the columns where the "Certificate" equals ‘UA’, ‘A’, ‘U’, and ‘R’ will be displayed.
7. Create chart:
   1. Click anywhere inside the pivot table.
   2. Click Insert→Chart→Column→Clustered Column.

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

1. Open IMDb.csv in Excel.
2. Verify that IMDB\_Rating is in Column F. Use this variable to calculate Mean and Standard Deviation.
   1. Find the mean
      1. Type into Cell R2 “Mean:”
      2. Type into Cell S2:

=AVERAGE(F:F)

* 1. Find the standard deviation
     1. Type into Cell R3 “Standard Deviation:”
     2. Type into Cell S3:

=STDEV.S(F:F)

1. Use Global\_sales to calculate quartiles and IQR:
   1. Find Q1 (25th Percentile)
      1. Type into Cell R4 “Q1 (25th Percentile)”
      2. Type into Cell S4:

=QUARTILE.INC(F:F, 1)

* 1. Find Q3 (75th Percentile)
     1. Type into Cell R5 “Q3 (75th Percentile)”
     2. Type into Cell S5:

=QUARTILE.INC(F:F, 3)

* 1. Find IQR
     1. Type into Cell R6 “IQR (Q3 - Q1)”
     2. Type into Cell S6:

=S5 - S4

1. Calculate Outlier Thresholds:
   1. Quartile Method:
      1. Mild Low: Q1 – 1.5 \* IQR
         1. Type into Cell R7 “Quartile Mild Low”
         2. Type into Cell S7:

=S4 - 1.5\*S6

* + 1. Mild High: Q3 + 1.5 \* IQR
       1. Type into Cell R8 “Quartile Mild High”
       2. Type into Cell S8:

=S5 + 1.5\*S6

* + 1. Regular Low: Q1 – 3 \* IQR
       1. Type into Cell R9 “Quartile Regular Low”
       2. Type into Cell S9:

=S4 - 3\*S6

* + 1. Regular High: Q3 + 3 \* IQR
       1. Type into Cell R10 “Quartile Regular High”
       2. Type into Cell S10:

=S5 + 3\*S6

* 1. Mean/SD Method for Outliers:
     1. Mild Low: Mean – 2 \* SD
        1. Type into Cell R11 “Mean/SD Mild Low”
        2. Type into Cell S11:

=S2 - 2\*S3

* + 1. Mild High: Mean + 2 \* SD
       1. Type into Cell R12 “Mean/SD Mild High”
       2. Type into Cell S12:

=S2 + 2\*S3

* + 1. Regular Low: Mean – 3 \* SD
       1. Type into Cell R13 “Mean/SD Regular Low”
       2. Type into Cell S13:

=S2 - 3\*S3

* + 1. Regular High: Mean + 3 \* SD
       1. Type into Cell R14 “Mean/SD Regular High”
       2. Type into Cell S14:

=S2 + 3\*S3

1. Count Number of High Outliers   
   Use the following formulas to count outliers above each high threshold.
   1. Calculate the Quartile Mild Outlier High
      1. Type into Cell R15   
         “Outlier Count for Quartile Mild High”
      2. Type into Cell S15:

=COUNTIF(F:F, ">"&S8)

* 1. Calculate the Quartile Regular Outlier High
     1. Type into Cell R16   
        “Outlier Count for Quartile Regular High”
     2. Type into Cell S16:

=COUNTIF(F:F, ">"&S10)

* 1. Calculate the Mean/SD Mild Outlier High
     1. Type into Cell R17 “Outlier Count for Mean/SD Mild High”
     2. Type into Cell S17:

=COUNTIF(F:F, ">"&S12)

* 1. Calculate the Mean/SD Regular Outlier High
     1. Type into Cell R18 “Outlier Count for Mean/SD Mild High”
     2. Type into Cell S18:

=COUNTIF(F:F, ">"&S14)

1. Count Number of Low Outliers   
   Use the following formulas to count outliers below each low threshold.
   1. Calculate the Quartile Mild Outlier High
      1. Type into Cell R19   
         “Outlier Count for Quartile Method Mild Low”
      2. Type into Cell S19:

=COUNTIF(F:F, "<"&S7)

* 1. Calculate the Quartile Regular Outlier High
     1. Type into Cell R20   
        “Outlier Count for Quartile Method Regular Low”
     2. Type into Cell S20:

=COUNTIF(F:F, "<"&S9)

* 1. Calculate the Mean/SD Mild Outlier High
     1. Type into Cell R21 “Outlier Count for Mean/SD Mild Low”
     2. Type into Cell S21:

=COUNTIF(F:F, "<"&S11)

* 1. Calculate the Mean/SD Regular Outlier High
     1. Type into Cell R22 “Outlier Count for Mean/SD Mild Low”
     2. Type into Cell S22:

=COUNTIF(F:F, "<"&S13)