Cadets: _____
Section: _____

Using Excel to Build Better Models

File: `TL3_UsedCarData.xlsx`

Save As: `TechLab3_YYYYMMDD_LastnameFirstName.xlsx`

Purpose:

This lab is designed to increase your proficiency with Excel for data analysis. You will:

- Construct scatter plots.
- Plot trendlines.
- Use Excel to solve a system of non-linear equations.
- Calculate predictions using Excel.
- Add the predicted model to the scatter plot of your data.
- Calculate SSE for the predictive model you found.
- Minimize SSE to find an optimal solution.

---

1. **BEFORE YOU OPEN EXCEL**: You're given a 1990 sedan as your first car and are trying to predict what you could sell it for next year to help you buy the next vehicle. Every year since 1990 the car depreciates by 17% of its value. You've also looked up information and found you can sell it for parts for $2000. If the car was originally purchased for $18,501.33 in 1990, what will it be worth in 2026? Define your variables, identify assumptions that are reasonable and necessary, and construct a first-principles model.

2. Plot the data in Excel.
   (a) Highlight the data in columns `age_years` and `resale_value`.
   (b) Then follow `Insert` → `Charts` → `Scatter Plot` () → `Markers Only` ()
   (c) Change the title and axis labels to match the context of our problem.
   (d) Click on the plus sign () and add an exponential trendline to your plot.
   (e) Display the equation and R^2 value on the chart by clicking on the trendline and scrolling to the bottom of `Format Trendline` and checking the boxes next to `Display Equation on chart` and `Display R-squared value on chart`. Write the equation and the R^2 below.

3. The equation doesn't look like it fits very well; let's see if we can do better by hand.

   (a) Locate the following table titled Solver Parameters in `K2`:

a	
b	
d	

   (b) Type the parameters from your first-principles model into the chart.

   (c) Locate the cell `D2` labeled `y_hat`, under the header Using Solver. Enter the following equation into cell `D3`: `= $L$3*$L$4^B3+$L$5`

   (d) Auto-populate the equation down the column to calculate the predicted value of the vehicle over time.

4. Add this new data to your chart.

   (a) Right click on your chart and click on `Select Data`.

   (b) In the new window, click on `Add` (Add).

   (c) In the box for series name, type the following: `By Hand Calculation`.

   (d) Then click in the box that says `Series X values:` and highlight the data in the `age_years` column.

   (e) Finally, delete everything in the `Series Y values:` and select the data in the `y_hat` column (column `I`).

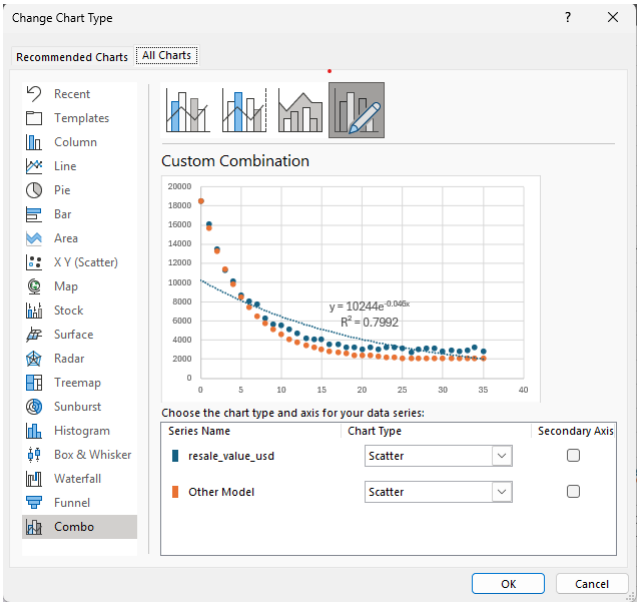
   (f) Select `Okay`, then `OK` to exit out of the windows.

5. The first-principles model we developed is a continuous function, so let's change the style from dots to a solid line.

   (a) Click on the dots representing the model you just added.

   (b) Select the `Chart Design` tab in the ribbon and then select the `Change Chart Type` button ( ).

   (c) A new dialog box should open as shown below:



(d) Find the series name of the model you are adding to the scatter plot and change the Chart Type using the drop down menu to: `Scatter with Smooth Lines`.

(e) Then, select `OK`.

6. While this model already looks better than the Excel trendline, let's calculate the SSE for each to compare.

(a) In cell `G3` type in the equation for the Excel trendline: `=10244*EXP(-0.046*B3)`.

(b) Auto-populate this equation down column `G`.

(c) In cell `H3`, calculate the error between the Excel trendline and the raw data using the equation `=C3-G3`

(d) Then calculate the squared error in cell `I3`, by squaring the value of cell `H2`, `=H3^2`.

(e) Highlight cells `H3` and `I3` and auto-populate the equations down the column.

(f) Locate cell `M8`, and type `=SUM(I3:I38)`. This is the SSE for the model Excel used for a trendline. Write this number here:

(g) Now follow steps (d) - (g), but in columns `E` and `F` to calculate the SSE for the model you developed by hand ensuring you adjust the cell references to use `y_hat`. Calculate the SSE in cell `L8`. Write the SSE for the by hand calculation here:

7. You should note that the by hand SSE is smaller than the SSE for the Excel model. But can we make it better? Let's enable Excel Solver to find out.

(a) Install Solver

i. Go to `File` → `Options` and select the `Add-ins` option.

ii. At the very bottom of the window, you'll see the word "Manage:" with `Excel Add-ins` in the cell next to it. Select the `Go...` button next to this.

iii. In the window that pops up, check the box next to `Solver Add-in`, then select `OK`.

iv. Go to the `Data` tab on the main ribbon, and in the ribbon to the very right should be a section called `Analysis` and you should see the option to select `Solver` (

).

(b) Use Solver

i. Click on the `Solver` button and it should open a new dialog box.

ii. In the `Set Objective` cell, delete the current option, and then click on cell `L8`.

iii. We want to minimize SSE, so click on `Min` underneath the `Set Objective:` cell.

iv. Click on the cell under the words `By Changing Variable Cells:` and highlight cells `L3`, `L4`, and `L5`.

v. Finally, un-check the box next to the words `Make Unconstrained Variables Non-Negative`.

vi. At the very bottom of the dialog box, click `Solve`.

vii. This should open a new dialog box titled `Solver Results` that provides you with the result. The text at the top should say: "Solver has converged to the current solution. All Constraints are satisfied."

(c) Go back to cells `L3` to `L5`, did these values change? Write the new equation for the exponential function that Solver found.

(d) Now look at cell `L8`, is this value lower than the original SSE you calculated for the by hand function? Write this value here:

(e) Finally, look at your graph. How well does the continuous function plot against your data?

(f) Using the model with the lowest SSE, predict the value of the car in 2026?

(g) Look at step 7.b.v. By un-checking this box, what assumption are we making? Does this assumption affect your results? What would it mean in the context of this problem if any of these parameters were negative.

(h) Refer back to your initial assumptions, what are some limitations of this model?