

Homework 4, Due Tuesday November 16, 2021, 11:00 PM, 10 Points

Directions: Do all the problems listed below. Submit 1 single pdf file through Canvas.

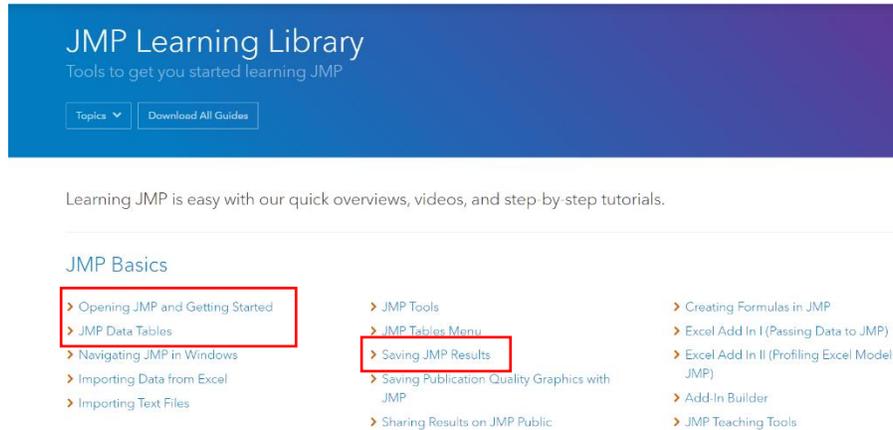
Note:

- You do not turn in anything for problem 1. In Problem 1 you'll watch and practice a few JMP tutorials.
- We'll mostly check for completion for problem 2. This is due to the highly subjective nature of the problem.
- **We may put out a few people's presentations as "solutions".** We'll ask for your permission before doing this.

Problem 1:

JMP provides a lot of useful tutorials at https://www.jmp.com/en_us/learning-library.html. You will watch a few of the videos and reproduce the results on your own but these reproduced results are not collected. The total length of the videos you'll watch is about 20 minutes.

1) Go to: https://www.jmp.com/en_us/learning-library.html, and watch the short videos on "Opening JMP and Getting Started", "JMP Data Tables", and "Saving JMP Results". See below. (You don't need to learn how to work with or save scripts but you do need to know how copy and paste your JMP results.) Reproduce the results on your own. I also recommend that you go over the beginner's tutorial when you open JMP for the first time.



2) Go to https://www.jmp.com/en_us/learning-library/graphical-displays-and-summaries.html, and watch the following short videos. Reproduce the results on your own.

- "Bar Charts and Frequency Distributions"
- "Mosaic Plot and Contingency Table"
We did not talk about mosaic plots but you are welcomed to use them. You can ignore the chi-squared discussion.
- "Histogram, Descriptive Stats and Stem and Leaf"
We did not talk about stem and leaf plots; they are not very common. Ignore stem and leaf plots.
- "Box Plots"
The boxplots I have shown you have been "outlier" box plots. We'll use "outlier" boxplots in this class.
- "Scatter Plots"
You can ignore the non-parametric density and the 95% ellipses discussions.
- This one is optional but I do recommend it: "Graph Builder"

Graphical Displays and Summaries

Learn how to use JMP to explore, describe, and understand data through visualizations and descriptive statistics. Create interactive graphs designed to display relationships, patterns, changes over time, and make comparisons.

> Bar Charts and Frequency Distributions

Visualize and numerically summarize the distribution of categorical variables.

> Pareto Plots and Pie Charts

Visualize and numerically summarize the distribution of categorical variables.

> Mosaic Plot and Contingency Table (Cross-tabs)

Explore and display a relationship between two categorical variables.

> Histograms, Descriptive Stats and Stem and Leaf

Visualize and numerically summarize the distribution of numerical variables.

> Dot Plots

Visualize the distribution of continuous variables.

> Box Plots

Visualize and numerically summarize the distribution of continuous variables.

> Scatter Plots

Explore and visualize the relationships between continuous variables.

> Run Charts (Line Graphs)

Display changes in a continuous variable over time.

> Graph Builder

Interactively create visualizations to explore and describe data. (Examples: dotplots, line plots, box plots, bar charts, histograms, heat maps, smoothers, contour plots, time series plots, interactive geographic maps, mosaic plots)

> Bubble Plots

Explore the relationship between two continuous variables and changes over time.

> Tabulating Data

Construct tables of descriptive statistics.

> Mapping in Graph Builder

Visualize data that correspond with geographical maps.

Problem 2:

Your goal is to conduct an exploratory data analysis (EDA) of the Ames Housing dataset. Keep the following questions in mind:

- Which predictors may be useful in predicting house prices?
- Could you gain any insight by looking at the relationship between the predictors?

You'll also get to know the data by doing EDA. This data will also be used for the competition (HW 6) and for your project.

The data comes from <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>. (You are welcomed to explore the site for ideas.) Use fairly cleaned-up and modified version of the data on Canvas: train_modified.csv. You should take 5-10 minutes to go over the data description file before you do anything.

There are a few categorical variables that have been coded as numeric. You'll have to tell JMP to treat them as categorical. The very last column, "SalePrice", is the response variable. The "Id" column is not a predictor.

After completing your EDA, summarize your results using PowerPoint or a similar presentation software (but convert to pdf format):

- You can assume that the audience is familiar with statistical graphical tools and tables.
- You should have some numerical or graphical summaries showing the relationship between your chosen predictors and the response. You will not be able to choose every single predictor. Just show the ones that look most promising to you. My recommendation is that you look at the relationship between the response and at least 7 predictors, and then choose the 3 most promising ones.
- You should have a few numerical or graphical summaries that show the relationship between the predictors. Reporting about 3 pairs of predictors is reasonable.
- At least one slide in a bullet point format should contain a summary of the most important things you learned.
- The first slide should be a title slide with your name on it.

Comment: If you were a paid consultant working with this data, then you should look at the response versus many more predictors, and many predictor-predictor summaries as well.