**Module 2 HW**

**DUE: Monday September 9th by 11:59 PM**

Professor Nick Williams Fall 2024

Economics 4010, University of Cincinnati

## **What are you being asked to do:**

* Starting from a template R program, demonstrate that you can produce descriptive statistics and visualizations using R.
* Demonstrate that you can produce descriptive statistics and visualizations when you are not provided with template R code
* When requested, provide an interpretation of your descriptive statistics and visualizations.

## Why is this important?

This term we use a variety of data sets as we learn econometrics, and use econometrics tools to investigate causal relationships. To do this you need to know:

* How to use R
* How to use an R script so that your efforts are reproducible
* How to ensure the variables in your data set are measured and reported correctly (i.e. “good”)
* How to produce and interpret descriptive statistics and visualizations.

This homework provides you with an opportunity to demonstrate that you have learned each of these tasks.

## Directions

* **Turn in your answers in a Word document through Canvas. You should also turn in a copy of your R script**. Not a copy of your Console, but a copy of your R Script that I can run.
* **Make sure you look at my example R scripts from the lectures and learning exercise.** I am NOT asking you to use any R code that we have not already used in at least one of those R scripts.
* In some circumstance, copying results from RStudio into Word is acceptable, but note that in many instances I ask you to interpret or explain. Below I make it clear when I want to write an answer.
* **IMPORTANT: Make sure you read and follow my directions in the companion Word document “Practicing Professionalism”**
* As stated in the syllabus, homework will be accepted one day late with a 25% penalty. Later homework will not be accepted.
* As stated in the syllabus: Feel free to work with others. However, I recommend that you spend time working on the problems by yourself before working with others. **Every student must turn in their own, *unique* answers to the homework.**
* **Please read each question carefully, and multiple times.**
* See my R Hints on the last page of this homework.
* The homework will be carefully graded out of 100 total points.

# Questions

1. I have put a template R file that you can use to get started, and the required R dataset on Canvas. This data contains a random sample of 1000 customer's orders from a company that uses catalogs to sell its products.

# -- List of Variables ---

# Salary = Total household salary, in $

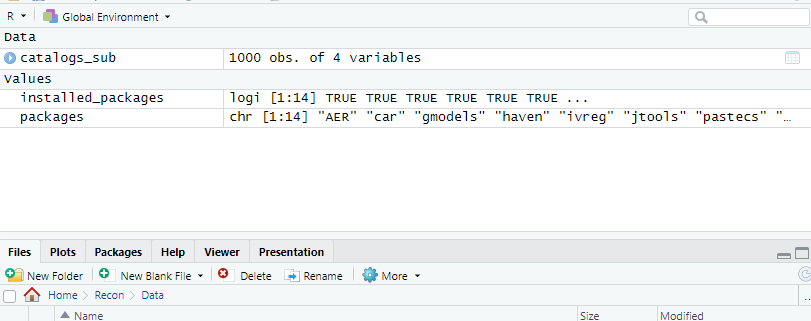
# Children = Number of children in the household

# Catalogs = Number of catalogs sent to the household in a year

# AmountSpent= Total amount spent by household, in $

**BE CAREFUL. Capitalization matters in R. So “Catalogs” is different than “catalogs”, etc.**

* 1. Load the template R script into RStudio.
     1. Insert the last date you modified this code.
     2. Insert your name in place of Brilliant Student.
  2. Load the required packages, and read in the data. Once you have done this correctly take a screen shot of your RStudio page that includes your Environment window, and insert this image into the Word document where you will answer the questions below.



* 1. Answer each of the below in words, without copying results or tables directly from R
     1. What percentage of the sample had exactly 2 children in the household?

14.6%

* + 1. What percentage of the sample had 2 or fewer children in the household?

87.5%

* + 1. What percentage of the sample had 18 or more catalogs sent to their household?

46.6%

* 1. Answer each of the below in words, without copying results or tables directly from R
     1. How many households in the sample with no children received 12 or more catalogs? How many with 1 or more children received 12 or more catalogs?

462 – 96 = 366 households with no children received 12 or more catalogs

267+146+125-73-36-47 = 382 households with 1 or more children received 12 or more catalogs

* + 1. Given that the household received 6 catalogs in a year, what percentage of these households had 1 or more children?

252 – 96 = 156

156/252 = 61.9%

* 1. Create the R code to look at the descriptive statistics for the household’s salary.
     1. Copy this table into your Word document.

Salary

----------------- -----------

Mean 56103.90

Std.Dev 30616.31

Min 10100.00

Q1 29850.00

Median 53700.00

Q3 77050.00

Max 168800.00

MAD 34841.10

IQR 47050.00

CV 0.55

Skewness 0.42

SE.Skewness 0.08

Kurtosis -0.57

N.Valid 1000.00

Pct.Valid 100.00

* + 1. Looking at these descriptive statistics, briefly describe the distribution of Salary.

The mean is 56k, median 53k,

The lower bound is 30k and upper is 77k with a max of 168k

* 1. Create a subset of the data that only includes only households that received 18 or more catalogs. Using this subset, create a visualization using boxplots of AmountSpent with the group being “Catalogs”.
     1. Copy these boxplots into your Word document.

A graph with a number of lines

Description automatically generated with medium confidence

* + 1. Looking at these boxplots, briefly compare the distribution of the amount spent between the households receiving 18 catalogs and those receiving 24.

The medians are very similar and so are the Q1s but Q3 and max spending differ significantly. Which makes sense when you consider how wealth is distributed

* 1. On the entire dataset (not the subset in part f), create a histogram of amount spent. Make sure you choose a bin width that is most informative.
     1. Copy this histogram into your Word document.

A graph of a graph

Description automatically generated

Mean is blue and median is red

* + 1. Looking at this histogram, briefly describe the distribution of amount spent

The median looks to be about 9k or so, just below 10k. Considering almost half the households have no children and they are the ones who spent the most on catalogs the other half or so spent significantly less, which aligns with the histogram.

## R Hints

# In the template program, there are several places I have inserted ????. You need to insert the correct R code for these ??????

# By far the easiest way to get the correct R code is to look at the example programs I used in the lectures and/or learning exercises.

# In particular, for part e) I have provided an example in the 02\_LE01.R file for how to write this code.

# In particular, for part f), I have provided an example of how to create a subset in the file 02\_Visualizations and Desc Stats.R used in the lectures.

# And in part g), I have provided an example in the 02\_LE01.R file for how to write this code.