STAT 2600 Recitation
02/02/2021

Recitation Week 3 Problems

1) Load the 'tidyverse' and 'datasets' packages (note if you are not using google collab you will have to install these as well as load them). Explore the documentation for 'datasets' and pick out an interesting dataset. Save it to a new variable as a tibble.

2) Explore the dataset - print the whole dataset; access and print one column; access and print one row. Change one value in the dataset. *Optional* rename a column.

3) For the rest of the questions we will create a new dataset and use that. Create a tibble with three columns: date, product, price. Add 730 rows - one for each date between 1/1/2018 and 12/31/2019. Fill in the price column as follows: (1) for 1/1/2018 - 12/31/2018 list a random number generated from a normal distribution with mean 100 and standard deviation 10; (2) for 1/1/2019 - 12/31/2019 list a random number generated from a uniform distribution on 100 to 200. (Hint: explore the functions 'rnorm()' and 'runif()').

4) If these were real price fluctuations, how would you analyze them to determine if there was a significant change in pricing behavior between 2018 and 2020?

5) Calculate the average price in 2018 and the average price in 2019 and compare them - is this evidence of price increases? What test would you run to make sure of this result?

6) Create a new column that lists the "mean price difference" for each date. Does this increase or decrease from 2018 to 2019? You could also look at the variance in price - how does this change over time?