HW3: Chapter 5 of MD: Due Tuesday the 25th

Your Name Here

Mar 17, 2021

Note: Be sure to turn in this hw as a well organized html/pdf (printed) document.

Following is the yaml I used. Please replace the default yaml with this so it would generate a pdf.

```
---
title: "HW3: Chapter 5 of MD: Due Tuesday the 25th"
author: "Your Name Here"
date: 'Mar 17, 2021'
output:
  pdf_document:
    latex_engine: xelatex
---
```

Question 01

The quality of the orange juice produced by a manufacturer (e.g., Minute Maid, Tropicana) is constantly monitored. There are numerous sensory and chemical components that combine to make the best tasting orange juice. For example, one manufacturer has developed a quantitative index of the "sweetness" of orange juice. (The higher the index, the sweeter the juice.)

Is there a relationship between the sweetness index and a chemical measure such as the amount of water-soluble pectin (parts per million) in the orange juice?

Data collected on those two variables for 24 production runs at a juice manufacturing plant are listed in **OJUICE**.

How to import data:

1. Download the data from my website under this week (**do not try to open in the computer**) and upload the data file in to your working directory (the folder where your homework is).
2. Click on the uploaded data file. Read the message and click "Yes".
3. Remove "#" and run the following code, if you see the "head" of the dataframe, the data is in!

```
#load("OJUICE.Rdata")
#head(OJUICE)
#skim(OJUICE)
```

1. Suppose a manufacturer wants to use **simple linear regression** to predict the sweetness (y) from the amount of pectin (x)

   i. **Conduct Exploratory data analysis (Note: This consists of three parts)**

      a) Looking at the raw values:

      b) Computing and describing the summary statistics likes means, medians, and standard deviations. Specially the correlation coefficient:

      c) Creating and describing data visualizations:

ii. **Fit the model and write the equation.**

iii. **Give a practical interpretation of the value of b_0, if possible.**

iv. **Give a practical interpretation of the value of b_1, if possible.**

v. **Predict the sweetness index if amount of pectin in the orange juice is 300 ppm.**

vi. **Find out a function to do part v) in `R`. (We did not discuss a code for this in class)**

vii. **Conduct diagnostics and comment on how "good" the model is.**

Question 02

`Credit` dataset from the `ISLR` library consist of infomation like income, credit limit, and education level for $n = 400$ credit card holders. Note that this dataset is not based on actual individuals, it is a simulated dataset used for educational purposes.

1. **Suppose we want to use linear regression to predict the `Balance` (y) from the `Ethnicity` (x)**

   i. *Conduct Exploratory data analysis (Note: This constists of three parts)*

      a) Looking at the raw values:

      b) Computing and describing the summary statistics likes means, medians, and standard deviations. Specially the correlation coefficient:

      c) Creating and describing data visualizations, individual and comparisons:

   ii. **Fit the model and write the equation.**

   iii. **Give a practical interpretations of ALL coefficients and Fitted values.**
