STATISTICAL TECHNIQUES FOR THE BIOMEDICAL SCIENCES
BMR 617
SPRING 2023

Assignment 3. Comparing data sets to the normal distribution.

In this assignment, you'll examine a couple of data sets and compare their distribution to the normal distribution. We will not do this in a particularly rigorous way at present. Later in the course we'll discuss some standard tests for this.

1. Download the mouse metabolic data from Dr. Kim's lab into R with the tidyverse function `read_csv`. The data can be found at `https://denvirlab.marshall.edu/BMR617-2021/data/TH-B6-metabolic.csv`
2. Looking at only the cholesterol data for all 29 mice regardless of strain or diet, find the mean, median, and standard deviation of the cholesterol level.
3. If the cholesterol levels were normally distributed, how many of the 29 mice would have cholesterol levels within one standard deviation of the mean? How many would have cholesterol levels within two standard deviations of the mean?
4. Using the actual cholesterol data, how many of the actual values in the data set are within one standard deviation of the mean? (Hint: one way is to filter the data set for `Cholesterol > m-s & Cholesterol < m+s`, where m is the mean and s the standard deviation: then look at the number of rows in the resulting data set. Another way is to `pull` the data into variable (`chol`, for example), and then do `sum(chol > m-s & chol < m+s)`.)
   How many of the actual cholesterol values are within two standard deviations of the mean?
5. Do you think the cholesterol values are sampled from a normal distribution? Justify your answer based on the previous steps.

Repeat steps 1-5 using the Insulin data values.

As usual, submit a Word document which answers the questions, and an R script file which generates the results you used to answer them.