STT 3850 Midterm Study Guide

Andrew Thorp

October 9, 2017

Statistics

Characterizing a set of data (3 S's)

- Shape: how the data is distributed
- Low outliers make a dataset skewed to the Right
- High outliers make a dataset skewed to the Left
- Normal distributions have fairly even outliers on either side

Center: Where the data is centered around

- Normal: If the dataset has a normal distribution (shape) this can be calculated using the `mean($data)` function.
- Skewed (left or right): The mean will be misrepresent the center. Calculate a skewed center using `median($data)`.

Spread: How far the data differs from the center

- Normal: If the dataset has a normal distribution, then the standard deviation applies to both sides of the data and so it represents the spread.
- Skewed (left or right): If the dataset is skewed on either side, the deviation above and below the center will not be the same, so you must calculate it using `IQR($data)` for the interquartile range.

Hypothesis testing (5 step procedure)

1. Specify the Null and ALternative hypothesis
   - Null hypothesis notated as $H_0 : M = value$ or $\bar{M_1} - \bar{M_2} = 0$
   - Alternative hypothesis notated as $H_A : M \neq 0$

The Null hypothesis is assumed to be true, and you can either find evidence to support the alternative hypothesis or failt to find evidence to suppor the null hypothesis. 2.

Markdown Dplyr Ggplot2