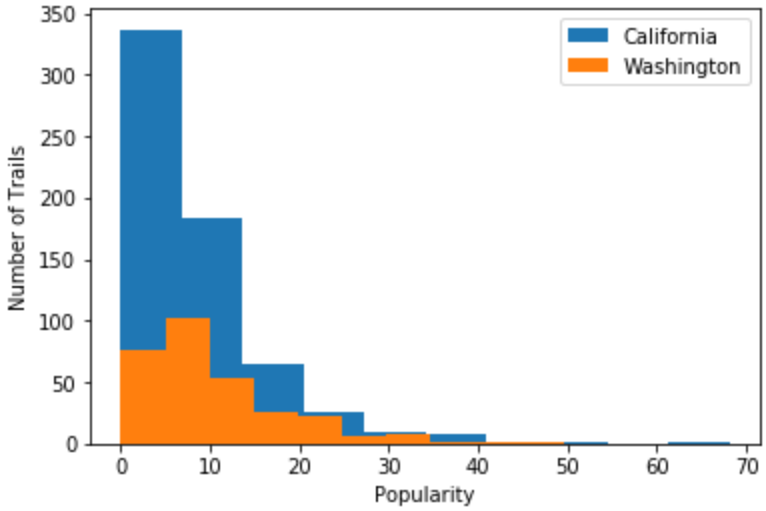
Google link (comments enabled):
https://docs.google.com/document/d/1wBOYmz8iU8gWlRALZJnwPlodVvlevlatII1u82iOopY/edit?usp=sharing

Capstone 1 Inferential Statistical Analysis

Evan Hintz

Within my National Parks data I'm interested in finding out how different activities and features of trails affect the trail popularity, rating, and usage (all data tracked by AllTrails.com). Through some exploratory data analysis I found the differences in specific states' data curious. I decided to use an inferential statistical method to analyze how the dependent variables (popularity, rating, usage) changed between states.

First I wrote out the hypothesis to test:
Null hypothesis: The average rating for California's trails are equal to those of Washington's
Alternative hypothesis: The average ratings are not equal
Using an alpha of 5%, this two-tailed test will have a z-score of 1.96
After the necessary calculations my Z-test value was -11.29; which is far to the left of the distribution and strongly recommends the null hypothesis to be rejected. This concludes that the average rating for California's trails are not equal to those of Washington's. Similar results were found when comparing popularity and usage.



Above is a histogram of CA and WA's trail popularity. It is obvious here that CA has many more trails, however it's worth noting that both state's distribution of trail popularity are very similar. (Histograms of the other variables look very similar as well)

While CA has more than twice as many trails as WA I decided to run the same tests with Wyoming and Utah, states that only have a difference of 8 trails. While the number of trails may not necessarily change the outcome, CA also has a much higher population as well as tourist traffic than the other two states. After running through my program with new inputs the end conclusion is similar (Z-test = 05.33), the rating along with the popularity of the two states' trails are not equal. Oddly enough, the visitor usage does fall within the 1.96 Z-score distribution at .098 for these two states, so we can state with significance that visitor usage between Wyoming and Utah are equal.

While these tests haven't answered my initial questions of how the trail activities and features change the three dependent variables they have provided some good insights on both the differences in some of the states as well as the workings of the trail population, average rating, and visitor usage.