Group 9, Project 1

Data Analysis Project Proposal: MMR Immunization Trends

Data Source: <https://www.kaggle.com/jessemostipak/measles-immunization-rates-in-us-schools>

Kaggle: Measles Immunization Rates in US Schools

Questions we are asking:

* **Priority Question**: Are public schools more likely to have a higher rate of vaccinated children than charter or private institutions?
  + We did ultimately find that there was a significant difference between public and private schools, as well as between all other types of schools save between kindergarten and private schools after removing outliers discovered in the process of analyzing the data.
  + While we had intended to focus on just private and public schools, we expanded our dataset to include kindergarten and other to allow us to look at a greater number of schools and states with usable data. This posed interesting challenges in assessing overall relationships. Through many ANOVA and two sample t-tests, we were able to look at the relationships between all four of the represented school types and found that school type does, generally, have an impact on the immunization rates.
* *Is there a difference in the percentage of students vaccinated by state?* 
  + *Compare two vastly “different” states or regions*
  + We were able to compare California and Ohio data, as they shared the same school type profiles of both public and private schools. This allowed us to see that California had higher rates of vaccination, by mean and median, in every category except private school medians, which were identical. Had our data contained more usable data from more states, we would have liked to compare additional state or region pairs.

Testable Hypotheses:

* **Priority Hypothesis:** Public institutions will show higher rates of MMR immunizations than either charter or private institutions.
  + Null: There is no relationship between the type of school (public or private) and the proportion of school-aged children vaccinated. (ANOVA)
    - p =1.53 e-248 was the value received for testing all school types against one another with ANOVA when we include outliers. Without outliers, our p = 0.0. These both indicate that there is a significant difference of at least one school type.
    - When we performed ANOVA tests by removing each school type individually, we found both with and without outliers in our data set that removing public schools gave us the largest p values, but they were both still smaller than 0.05.
    - We performed paired t-tests in order to see if there were non-significant relationships between two school types, where we found that when we excluded outliers that private and kindergarten school types were not significantly different (p = 0.229). This was the only hypothesis test run where we failed to reject the null that there is no difference between immunization rates of different school types.

Overall, our hypothesis tests showed that there is a significant difference between immunization rates among all school types with public schools being the least closely related to all of the other types.