Parsh Gandhi

7th July 2022

ATDP secondary division

Data Science Lab with R & Python, Section 1

Reflection #3

Treatment- Something(drug, price, web headline) to which a subject is exposed

Treatment group- A group of subjects exposed to a specific treatment

Control group- a group of subjects exposed to no(or standard) treatment

Randomization- The process of randomly assigning subjects to treatments

Subjects- The items(web visitors, patients, etc.) that are exposed to treatments

Test statistic- The metric used to measure the effect of the treatment

Null hypothesis- The hypothesis that chance is to blame

Alternative hypothesis- counterpoint to the null(what you hope to prove)

One-way test- Hypothesis test that counts chance results only in one direction

Two-way test- Hypothesis test that counts chance results in two directions

Permutation test- The procedure of combining two or more samples together and randomly (or exhaustively) reallocating the observations to resamples

Resampling- Drawing additional samples("resamples") from an observed data set

With or without replacement- In sampling, whether or not an item is returned to the sample before the next draw

P-value- Given a chance model that embodies the null hypothesis, the p value is the probability of obtaining results as a unusual or extreme as the observed results

Alpha- The probability threshold of "unusualness" that chance results must surpass for actual outcomes to be deemed statistically significant

Type 1 error- Mistakenly concluding an effect is real(when it is due to chance)

Type 2 error- Mistakenly concluding an effect is due to chance(when it is real)

Test statistic- A metric for the difference or effect of interest

T-statistic- A standardized version of common test statistics such as means

T-distribution- A reference distribution(in this case derived from the null hypothesis), to which the observed t-statistic can be compared

An A/B test is an experiment regarding two groups to find out which of two different headlines, treatments, or similar things are superior. A/B testing is a test mainly to compare things and is used commonly for online web pages and marketing but can be used for much more. To conduct an A/B test you will first need subjects. These subjects will be assigned to different groups called the control and test groups, usually randomly to keep out bias. Finally you will need a product to test and a test statistic, which is the thing to be measured in the experiment. Sometimes however you might need more than just A/B if there are additional treatments, more prices, or more than two values are being compared and are included. Hypothesis tests are tests made on assumptions made regarding a population parameter. A hypothesis test is a further analysis on an A/B test which means it can give an explanation behind the observed difference of groups A and B. T tests are tests that can help approximate the distribution of a single sample mean. I have a question about the t tests used in the real world, could you provide an example, and when are t tests used, in what circumstances. Resampling is

useful in data science as a way to sample values over and over again from an observed data set by assessing the random variability of the data set.