Statistics

# Hypothesis Testing

1. Process
   1. Specify the null hypothesis
   2. Calculate the statistics:
      1. Pooled t test
      2. Paired t test
      3. Welch’s test
      4. ANOVA
   3. Determine the alpha & p-value/critical value
      1. Alpha is the type 1 error: even though the null hypothesis is true, we rejected it

(False Positive)

* + 1. P-Value based: The probability of observing the statistics assuming the null hypothesis is true.
       1. We **calculate only the p-value based on the statistic**
    2. Critical region based: We **calculate the t statistic as well as the critical value**
       1. Compare the test statistic with the critical value.
       2. If greater than critical.
  1. Check for the hypothesis
     1. For p-value: We are checking the probability that the observed statistic exists when we the null hypothesis is true. So, in case it is less than our alpha and we can reject the null hypothesis.
     2. Critical Values: This is denoted by t(alpha,dof) for right tailed. If the test statistic t\* is greater than the critical value, we reject the null hypothesis.

1. Examples
   1. Data 1 – R/python
   2. Data 2- R/python
   3. Data 3 - R/python

# Parameter Estimation

Parameter estimation for a probability distribution from the dataset.

1. Maximum likelihood : mean, standard deviation
2. Rank regression
3. Probability plotting
4. Bayesian

Sampling Methods

1. Probability Sampling
   1. Simple Random
   2. Stratified
   3. Systematic
   4. Cluster
   5. Multi Stage
2. Non Probability Sampling
   1. Convenience sampling
   2. Purposive sampling
   3. Quota Sampling
   4. Referral

Resampling Methods: Cross Validation and Bootstrap

A/B testing

