Hypothesis Testing

April 12, 2021

1 Hypothesis Testing

- A/B Testing is to statistically validate if new experiment (new hypothesis) brings improvement, or the improvement could be because of the chance.
- A statistic is chosen, it could be mean, conversion rate etc for the sample of population for both control group or test group.
- Shape of the rejection region is determined, i.e. single tailed rejection, or double tailed rejection etc.
- p-value, the false rejection probability of null hypothesis is computed for the statistic of test region.
- Based on the test we choose, it follows a distribution, and we can derive the p-value. Lets take example of Z-test and Chi-Square Test.
- **Z-Test**
  - Are the Means of Two Population Equal? If two populations comes from same parameters?
  - X_i and Y_i are samples drawn from Bernoulli Random Variables $X \sim \mathit{Bernoulli}\,(\theta_X)$ and $Y \sim \mathit{Bernoulli}\,(\theta_Y)$. Variance of Bernoulli random variable is computed from parameter p as $p(1-p)$, which is respectively represented as σ_X^2 and σ_Y^2.
  - $H_0 : \theta_X = \theta_Y$, and $H_1 : \theta_X! = \theta_Y$. This would require double tailed rejection region.
  - Estimator for means would be $\hat{\theta}_X = \frac{1}{n_1}\Sigma_{i=1}^{n1} X_i$, and $\hat{\theta}_Y = \frac{1}{n_2}\Sigma_{i=1}^{n2} Y_i$
  - $\hat{\theta}_X$ and $\hat{\theta}_Y$ are random variables drawn from $N\left(\theta_X, \sigma_X^2/n1\right)$ and from $N\left(\theta_Y, \sigma_Y^2/n2\right)$, where n_X and n_Y are population sizes. This is proven by reduction of variable under central limit theorem.
  - Random Variable of difference between mean of X and mean of Y would be parameterised by mean $\theta_X - \theta_Y$ and variance of $\sigma_X^2/n1 + \sigma_Y^2/n2$. This is proven by distribution of sum of normally distributed random numbers.
  - Under H_0, both X and Y come from same population, hence difference of RV mean of X and mean of Y would be
$$\hat{\theta}_X - \hat{\theta}_Y \sim N\left(0, \sigma_X^2/n1 + \sigma_Y^2/n2\right)$$
  - Which could be written as,
$$\frac{\hat{\theta}_X - \hat{\theta}_Y}{\sigma_X^2/n1 + \sigma_Y^2/n2} \sim N\,(0,1)$$
  - We can estimate the variance $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ from the samples. Hence,
$$Z = \frac{\hat{\theta}_X - \hat{\theta}_Y}{\hat{\sigma}_X^2/n1 + \hat{\sigma}_Y^2/n2} \sim N\,(0,1)$$
  - lets take an example, for given A and B buckets, of session length, we can take mean of each bucket, compute variance of each bucket, and we can compute Z, and its probability under H_0. If the p-value is less than significance level $\alpha = 0.05$, then we would reject H_0 with false rejection probability of p-value.
  - **Student's T-test** is done when numbers of samples are small, i.e. less than 30. As in that case it would not be well approximated by normal distribution.
- **Chi-Square Test**
  - If proportions of populations in different classes are from same parameters?

- Suppose that n observations in a random sample from a population are classified into k mutually exclusive classes with respective observed numbers x_i (for i = 1,2,...,k).
- Under null hypothesis, H_0, we have a probability p_i of a sample falling into i-th class. Such that

$$\sum_i p_i = 1$$

- We know observed instances in each class i, which would be x_i and also know expected instances in each class i using null hypothesis, which would be $m_i = n * p_i$
- A variable X^2 would follow χ^2 distribution, which is

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i}$$

- once we compute observed X^2, and get the p-value from χ^2 distribution table, we would get the probability of false rejection. If p-value is less than $\alpha = 0.05$, we would reject the null hypothesis.