**Statistical Methods for Data Analysis: Frequentist Inference**

The following 2 steps are examples of frequentists performed

TEST 1

Hypothesis Test of Difference Between Proportions using two-proportion z-test. We will be testing the difference in the proportion of yes's in telephone and cellular to determine statistical significance

Null Hypothesis: There is no difference between the proportion of yes's in contact methods telephone and cellular

Alternate Hypothesis: There is a difference between the proportion of yes's in contact methods telephone and cellular

Steps:

Calculate proportion of yes in telephone contact and cellular contact respectively

Execute formula to calculate: p = (p1 \* n1 + p2 \* n2) / (n1 + n2) where p1 and p2 are proportions o cellular and telephone respectively, and n1 and n2 are corresponding counts

Calculate standard error as: np.sqrt( p \* ( 1 - p ) \* ((1/n1) + (1/n2)))

Calculate z-score as : (p1 - p2) / standard\_error

Calculate p value as : 1 - norm.cdf(z\_score)

Z-score yielded 29.37739625383253 while P-Value is < .00001.

Conclusion: The difference is very significant, therefore we reject the null hypothesis

TEST 2

Use T-Test to test two numeric variables: nr.employed ad emp.var.rage, these are two highly positively correlated independent variables:S

Steps

Import stats model

Compute: t, p = ttest\_ind(x, y, equal\_var=False)

Where x = x= train\_df['nr.employed'] and Y = y= train\_df['emp.var.rate']

Z\_Score = 14508.012150383513 ; P-Value is < .00001