Statistics Cheat Sheet

Inferential Statistics – Hypothesis Testing

Hypothesis Tests: Used to test business hypothesis using data. Most often business data is based on samples, and therefore we need to assess degree of confidence around the observed sample outcome and determine to what degree the outcome could be driven by random chance variation.

Hypothesis Testing Framework:
Ho: Null Hypothesis - Always the Hypothesis that there is no change/difference/impact (comparing an observed sample mean to an expected population mean, or comparing different sample means)
H1: Alternate Hypothesis – The negation of the Null, implying there is a change/difference/impact between observed sample outcome and expected population outcome, or between different sample outcomes.

The Alternate Hypothesis can be set up as:

A two tail test: Alternate Hypothesis is framed as there is a difference between sample and population or between samples. Compare p-value to $\alpha/2$ to determine outcome of test

A one tail test: Alternate Hypothesis is framed as" sample mean < pop. mean", or "sample mean > pop. mean"

Test Distribution: The appropriate distribution to be used to calculate the probability of observed outcome. Very often is the Normal Distribution if sample size > 30 because of the Central Limit Theorem, but can be any of multiple distributions based on behaviour of outcome variable

Significance Level (α): The maximum allowable degree of variation due to random chance. Usually set to 5%
Confidence = 1- α

P-value : The calculated probability of seeing the observed sample outcome simply because of random chance variation. In most business hypothesis, we would like the p-value to be as low as possible (less than Significance level to reject the Null)

Inferential Statistics – Sample Hypothesis Test

A company is assessing if implementing an on demand work from home policy leads to reduction in productive hours spent on the job. A randomly chosen group of employees from multiple teams are allowed to decide when they want to work from home independently, and track productive hours using the employee productivity tracking CRM system, for one month (24 days). The average hours recorded on the system from the group of employees allowed to work from home was 7.1 hours, with std deviation of 0.6 hours, compared to all other employees in the same time that worked only from the office which 7.42 hours, with a standard deviation of 1.1 hours.

Should the company conclude that providing a work from home option leads to decreased productivity?

Step 1: Set up Hypotheses:

Null Hypothesis (H0): There is no reduction of employee productivity in the work from home option

Alternate Hypothesis (H1): There is a reduction of employee productivity in the work from home option

Note- One tail test

Step 2: Decide on appropriate test

In this case, we are comparing a sample to a population outcome. Since total # of observations are less than 30, we have to use the T-test.

Step 3: Decide on a level of confidence required

The company would prefer to be at least 95% confident

Step 4: Perform the test

T-Value = (7.1 – 7.42)/(1.1/(24^0.5)) = -1.42. T.Dist(-1.42, 23, True) = 0.08

Note- We need probability of seeing 7.1 hours or less, so we don't need to use 1-

Step 5: Conclusion

Compare p-value (0.08) to significance level (0.05)

Based on the calculated p-value, the company CANNOT reject the Null Hypothesis, and therefore concluded that there is NO reduction in productivity

Choosing the right Hypothesis Test

Choosing the right Hypothesis Test - Basic version		
Single Sample	Test	Notes
n > 30, Pop Std Deviation known	Z test (Using a normal distribution approx via	Remember to divide Pop Std Deviation by square root of sample size (n)
n > 30, Pop Std Deviation NOT known	T test	Can use Sample std deviation, but it needs to be divided by square root of sample size
n < 30	T Test	Divide the sample std deviation by square root of sample size to calculate T distance
Two Samples	Test	Notes
Two samples, different units/participants in each sample	T-test, Independent Sample	Samples need not be of the same size. Check variance of both samples to choose either Equal Variance or Unequal variance versions of the Independent samples T Test
Two samples, same units/participants in	T- test, Paired Sample	Samples need to be of the same size
More than Two Samples	Test	Notes
Continous outcome variable, Discrete	ANOVA	
Count data	Chi Square test	Use formula to calculated expected values