# Hypothesis Testing

## Lesson 8

# Hypothesis – Meaning

- A statistical hypothesis is an assumption about a population which may or may not be true. Hypothesis testing is a set of formal procedures used by statisticians to either accept or reject statistical hypotheses. Statistical hypotheses are of two types:
  - √ Null hypothesis, H0 - represents a hypothesis of chance basis.
  - √ Alternative hypothesis, Ha-represents a hypothesis of observations which are influenced by some non-random cause.

# Hypothesis - Example

- Suppose we wanted to check whether a coin was fair and balanced. A null hypothesis might say, that half flips will be of head and half will of tails whereas alternative hypothesis might say that flips of head and tail may be very different.
- H0:P=0.5
- Ha:P≠0.5
- For example if we flipped the coin 50 times, in which 40 Heads and 10 Tails results. Using result, we need to reject the null hypothesis and would conclude, based on the evidence, that the coin was probably not fair and balanced.

# Formal Process for Hypothesis Testing

- State the hypotheses - This step involves stating both null and alternative hypotheses. The hypotheses should be stated in such a way that they are mutually exclusive. If one is true then other must be false.
- Formulate an analysis plan - The analysis plan is to describe how to use the sample data to evaluate the null hypothesis. The evaluation process focuses around a single test statistic.
- Analyze sample data - Find the value of the test statistic (using properties like mean score, proportion, t statistic, z-score, etc.) stated in the analysis plan.
- Interpret results - Apply the decisions stated in the analysis plan. If the value of the test statistic is very unlikely based on the null hypothesis, then reject the null hypothesis.

# Margin of Error & Types of Errors

- The margin of error *m* of interval estimation is defined to be the value added or subtracted from the sample mean which determines the length of the interval: $Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- A Type I Error is rejecting the null hypothesis in favor of a false alternative hypothesis
- A Type II Error is failing to reject a false null hypothesis in favor of a true alternative hypothesis
- The probability of a Type I error is typically known as Alpha, while the probability of a Type II error is typically known as Beta.

# Power

- Bullard describes multiple ways to interpret power correctly:
  - √ Power is the probability of rejecting the null hypothesis when, in fact, it is false.
  - √ Power is the probability of making a correct decision (to reject the null hypothesis) when the null hypothesis is false.
  - √ Power is the probability that a test of significance will pick up on an effect that is present.
  - √ Power is the probability that a test of significance will detect a deviation from the null hypothesis, should such a deviation exist.
  - √ Power is the probability of avoiding a Type II error.

Simply put, power is the probability of not making a Type II error

- Bullard also states there are the following four primary factors affecting power:
  - √ Significance level (or alpha)
  - √ Sample size
  - √ Variability, or variance, in the measured response variable
  - √ Magnitude of the effect of the variable

# Parametric and Non-parametric Tests

In statistics, parametric and nonparametric methodologies refer to those in which a set of data has a normal vs. a non-normal distribution, respectively. Parametric tests make certain assumptions about a data set; namely, that the data are drawn from a population with a specific (normal) distribution. Non-parametric tests make fewer assumptions about the data set. The majority of elementary statistical methods are parametric, and parametric tests generally have higher statistical power. If the necessary assumptions cannot be made about a data set, non-parametric tests can be used. Nonparametric tests don't require that your data follow the normal distribution. They're also known as distribution-free tests and can provide benefits in certain situations.

| Parametric Tests | Non-Parametric Tests |
|---|---|
| Paired T Test | Wilcoxon Rank Sum Test |
| Unpaired T Test | Mann-Whitney U Test |
| Pearson Correlation | Spearman Correlation |
| One way ANOVA | Kruskal Wallis Test |

# Reasons to Use Parametric and Non Parametric Tests

- Parametric Tests:
  - √ Parametric tests can perform well with skewed and non-normal distributions
  - √ Parametric tests can perform well when the spread of each group is different.
  - √ Parametric tests usually have more statistical power than nonparametric tests. Thus, you are more likely to detect a significant effect when one truly exists.
- Non-Parametric Tests:
  - √ Your area of study is better represented by the median
  - √ You have a very small sample size
  - √ You have ordinal data, ranked data, or outliers that you can't remove

# Parametric Test for Testing of Means - Single Sample T-test

- A single sample t-test (or one sample t-test) is used to compare the mean of a single sample of scores to a known or hypothetical population mean. So, for example, it could be used to determine whether the mean diastolic blood pressure of a particular group differs from 85, a value determined by a previous study.
- *Requirements*
  - √ The data is normally distributed
  - √ Scale of measurement should be interval or ratio
  - √ A randomized sample from a defined population
- *Null Hypothesis*
  - √ $H_0$: $M - \mu = 0$, where M is the sample mean and μ is the population or hypothesized mean.
- As above, the null hypothesis is that there is no difference between the sample mean and the known or hypothesized population mean.
- Equation: $$t = \frac{M - \mu}{\sqrt{\dfrac{\sum X^2 - ((\sum X)^2/N)}{(N-1)(N)}}}$$

# Parametric Test for Testing of Means – Independent T-Test for Two Samples

- The independent t-test, also called the two sample t-test, independent-samples t-test or student's t-test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups.
- The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal:
  - √ $H_0: u_1 = u_2$
- In most cases, we are looking to see if we can show that we can reject the null hypothesis and accept the alternative hypothesis, which is that the population means are not equal:
  - √ $H_A: u_1 \neq u_2$
- To do this, we need to set a significance level (also called alpha) that allows us to either reject or accept the alternative hypothesis. Most commonly, this value is set at 0.05.
- In order to run an independent t-test, you need the following:
  - √ One independent, categorical variable that has two levels/groups.
  - √ One continuous dependent variable.

- Unrelated groups
  - √ Unrelated groups, also called unpaired groups or independent groups, are groups in which the cases (e.g., participants) in each group are different. Often we are investigating differences in individuals, which means that when comparing two groups, an individual in one group cannot also be a member of the other group and vice versa. An example would be gender - an individual would have to be classified as either male or female – not both.

# Assumptions for Independent T-Test for Two Samples

- The independent t-test requires that the dependent variable is approximately normally distributed within each group.
- The variances of the two groups you are measuring are equal in the population.
- Independent t-test formula
  - √ Let A and B represent the two groups to compare.
  - √ Let mA and mB represent the means of groups A and B, respectively.
  - √ Let nA and nB represent the sizes of group A and B, respectively.
  - √ The t test statistic value to test whether the means are different can be calculated as follow :

$$t = \frac{m_A - m_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}}$$

- Once t-test statistic value is determined, you have to read in t-test table the critical value of Student's t distribution corresponding to the significance level alpha of your choice (5%). The degrees of freedom (df) used in this test are :
  - √ Df=Na+Nb-2
- If the absolute value of the t-test statistics (|t|) is greater than the critical value, then the difference is significant. Otherwise it isn't. The level of significance or (p-value) corresponds to the risk indicated by the t-test table for the calculated |t| value.

# Parametric Test for Testing of Means – Paired Sample T-Test

The paired sample *t*-test, sometimes called the dependent sample *t*-test, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample *t*-test, each subject or entity is measured twice, resulting in *pairs* of observations. Common applications of the paired sample *t*-test include case-control studies or repeated-measures designs.

Hypothesis

- √ The null hypothesis (H0) assumes that the true mean difference (μd) is equal to zero.
- √ The two-tailed alternative hypothesis (H1) assumes that μd is not equal to zero.
- √ The upper-tailed alternative hypothesis (H1) assumes that μd is greater than zero.
- √ The lower-tailed alternative hypothesis (H1) assumes that μd is less than zero.

The mathematical representations of the null and alternative hypotheses are defined below:

- √ $H0: \mu d = 0$
- √ $H1: \mu d \neq 0$ (two-tailed)
- √ $H1: \mu d > 0$ (upper-tailed)
- √ $H1: \mu d < 0$ (lower-tailed)

**Assumptions:**

- √ The dependent variable must be continuous (interval/ratio).
- √ The observations are independent of one another.
- √ The dependent variable should be approximately normally distributed.
- √ The dependent variable should not contain any outliers.

**Calculations:**

- √ Test Statistic : $t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$
- √ Note that the standard error of $\bar{d}$ is $\frac{s_d}{\sqrt{n}}$ where sd is the standard deviation of the differences.
- √ As before, we compare the t-statistic to the critical value of t (which can be found in the table using degrees of freedom and the pre-selected level of significance, α). If the absolute value of the calculated t-statistic is larger than the critical value of t, we reject the null hypothesis.

# Limitations of Hypothesis Testing

- The tests should not be used in a mechanical fashion. It should be kept in view that testing is not decision-making itself; the tests are only useful aids for decision-making. Hence "proper interpretation of statistical evidence is important to intelligent decisions."
- Test do not explain the reasons as to why does the difference exist, say between the means of the two samples. They simply indicate whether the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell us as to which is/are the other reason(s) causing the difference.
- Results of significance tests are based on probabilities and as such cannot be expressed with full certainty. When a test shows that a difference is statistically significant, then it simply suggests that the difference is probably not due to chance.
- Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypothesis. This is specially so in case of small samples where the probability of drawing erring inferences happens to be generally higher. For greater reliability, the size of samples be sufficiently enlarged.

All these limitations suggest that in problems of statistical significance, the inference techniques (or the tests) must be combined with adequate knowledge of the subject-matter along with the ability of good judgement.

# End