manipal PROlearn

Statistical Technique for Data Science

Testing of Hypothesis

Prof.Gangaboraiah PhD (Stats)

Learning objectives of unit 4

- At the end of the session on unit 4, the student should be able to
  - Define hypothesis, null and alternative hypothesis, Type-I and Type-II errors, α-errors, β-errors, confidence level, power of the test and P-value
  - Definition of Statistical test, Difference between large sample and small sample tests, Different types of tests
  - Understand and apply different parametric tests like Z-test, Student's t-tests, F-test, ANOVA, Repeated measures of ANOVA.
  - Understand and apply different non-parametric tests like Chi-square tests, Fisher's exact probabilities, Mann-Whitney U-test, Wilcoxon signed rand test, Kolmogorov-Smirnov tests, Kruskal-Wallis test
  - Understand Bayesian analysis and apply to a situational analysis

Hypothesis: A statement on the parameter(s) which is yet to be proved or established

Null Hypothesis: Hypothesis of no difference or neutral or may be due to Sampling variation

Alternative/ Research Hypothesis: Hypothesis of difference which is yet to be proved/ established

Example on
Non-statistical Hypothesis Testing

A criminal trial is an example of hypothesis testing without the Statistics.

In a trial a Jury must decide between two hypotheses. The null hypothesis is

H_0: The defendant is innocent

H_1: The defendant is guilty

The jury does not know which hypothesis is true. He should make a decision on the basis of evidence presented.

Example

A taxi company manager is trying to decide whether the use of radial tires or regular belted tires improves fuel economy.

The variable measured is quantitative, therefore

H_0 : The mean fuel consumption in cars fitted with radial tires and regular beleted tires will be same

$$\text{ie., } H_0 : \mu_1 = \mu_2$$

H_1 : The mean fuel consumption in cars fitted with radial tires may be inferior to regular belted tires

$$\text{ie., } H_1 : \mu_1 < \mu_2$$

H_1 : The mean fuel consumption in cars fitted with radial tires may be better than regular belted tires

$$\text{ie., } H_1 : \mu_1 > \mu_2$$

H_1 : The mean fuel consumption in cars fitted with radial tires and regular belted tires may be different

$$\text{ie., } H_1 : \mu_1 \neq \mu_2$$

Two judges have to judge independently whether the defendant is innocent or guilty on the basis of evidence. Lack of sufficient evidence may lead to erroneous decisions like false positive or false negative. Suppose based on evidences, if we are interested in finding proportion of false positivity in the judgement, then the hypothesis to be tested will be

If variable measured is qualitative, then

H_0 : The proportion of false positive decisions between two judges will be same

ie., $H_0 : P_1 = P_2$

H_1 : The proportion of false positive decisions of Judge 1 may be less than Judge 2

ie., $H_1 : P_1 < P_2$

$$H_1: \text{The proportion of false positive decisions of Judge 1 may be more than Judge 2}$$

$$\text{ie., } H_1: P_1 > P_2$$

$$H_1: \text{The proportion of false positive decisions may be different between two Judges}$$

$$\text{ie., } H_1: P_1 \neq P_2$$

Test (Statistical test)

It is a statistical rule which decides whether to accept the null hypothesis or not?

Warning

Decision is made based on the sample not on the population

↓

This leads to possibility of error between the decision made and the reality

Parametric Test

Large sample test $(n > 30)$

Standard Normal test (Z - test)

Small sample test $(n \leq 30)$

Student's unpaired t - test

Student's paired t - test

Analysis of Variance (ANOVA)

Repeated measures of ANOVA

Non-Parametric Test

- Chi – square test
- Fisher's Exact Probabilities
- McNemar Chi - square test
- Mann - Whitney U - test
- Wilcoxon Signed Rank test
- Kruskal Wallis test
- Friedmann test

Computation of test statistic

Based on the sample data the test-statistic should be computed using

$$t = \frac{\text{Mean 1 - Mean 2}}{\text{SE (Diff. b'n mean)}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t_{(\alpha;\, n_1 + n_2 - 2)}$$

Errors in Decision Making

Non-Statistical example			Statistical example		
Verdict	Actual Situation (H_0)		Decision	Null Hypothesis (H_0)	
	Innocent	Guilty		True	False
Innocent	Correct decision	Wrong decision	Do not Reject (Accept)	Correct decision	Wrong decision
Guilty	Wrong decision	Correct decision	Reject	Wrong decision	Correct decision

Non-Statistical example			Statistical example		
Verdict	Actual situation (H_0)		Decision	Null Hypothesis (H_0)	
	Innocent	Guilty		True	False
Innocent	Correct decision	Type-II error	Do not Reject (Accept)	Correct decision	Type-II error
Guilty	Type I-error	Correct decision	Reject	Type I-error	Correct decision

Non-Statistical example			Statistical example		
Verdict	Actual situation (H_0)		Decision	Null Hypothesis (H_0)	
	Innocent	Guilty		True	False
Innocent	Confidence level (1-α)	β-error	Do not Reject (Accept)	Confidence level (1-α)	β-error
Guilty	α-error	Power 1-β	Reject	α-error	Power 1-β

Steps involved in testing of hypothesis

1. State null and alternative hypotheses
2. Specify the level of significance α
3. Define the probability distribution the data follows
4. Compute the test statistic based defined population
5. Define the rejection criteria/ critical regional
6. Conclusion

Parametric tests

Parametric Test

Large sample test $(n > 30)$

Standard Normal Test (Z - test)

Small sample test $(n \leq 30)$

Student's t - test

Unpaired t - test

Paired t - test

Analysis of Variance (ANOVA)

Repeated measures of ANOVA

Large sample test

- A statistical test is called large sample if the sample size n > 30.

Z-test

- Z-test is called standard normal variate test.
- This test can be used for testing the
  - Mean of a single population (μ)
  - Difference between means of two populations ($\mu_1 - \mu_2$)
  - Proportion of a single population (P)
  - Difference between proportions of two populations ($P_1 - P_2$)

Testing the mean of a single population

- Assume that the samples are drawn from normal distribution
- The population variances should be known

Steps involved in testing mean of a population

1. Stating null and alternative hypothesis

$$H_0: \mu = \mu_0$$

versus

(i) $H_1: \mu < \mu_0$

or

(ii) $H_1: \mu > \mu_0$

or

(iii) $H_1: \mu \neq \mu_0$

Steps involved in testing mean of a population

2. The level of significance α = 0.05

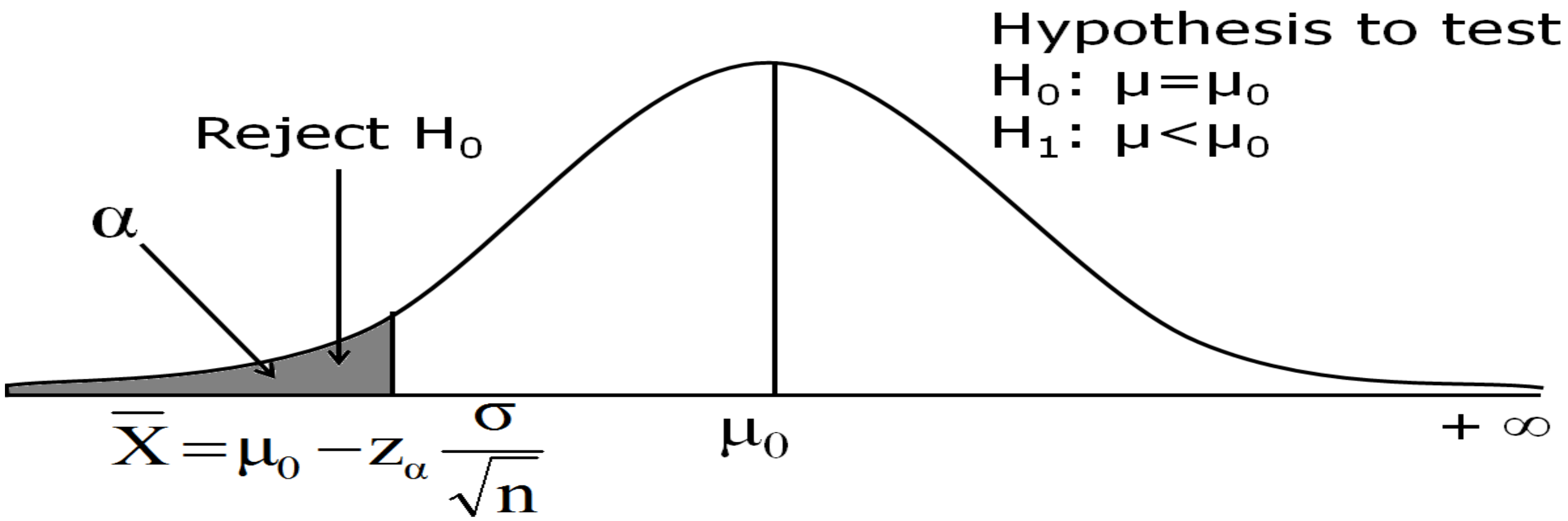
3. Normal distribution

4. The test statistic is

$$Z = \frac{\bar{x} - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} \approx N(0, 1)$$

5. Define the rejection criteria/ critical regional

(i) Reject H_0 if computed value of Z is less than the critical value, ie., $P(Z < -z_\alpha)$, otherwise do not reject H_0

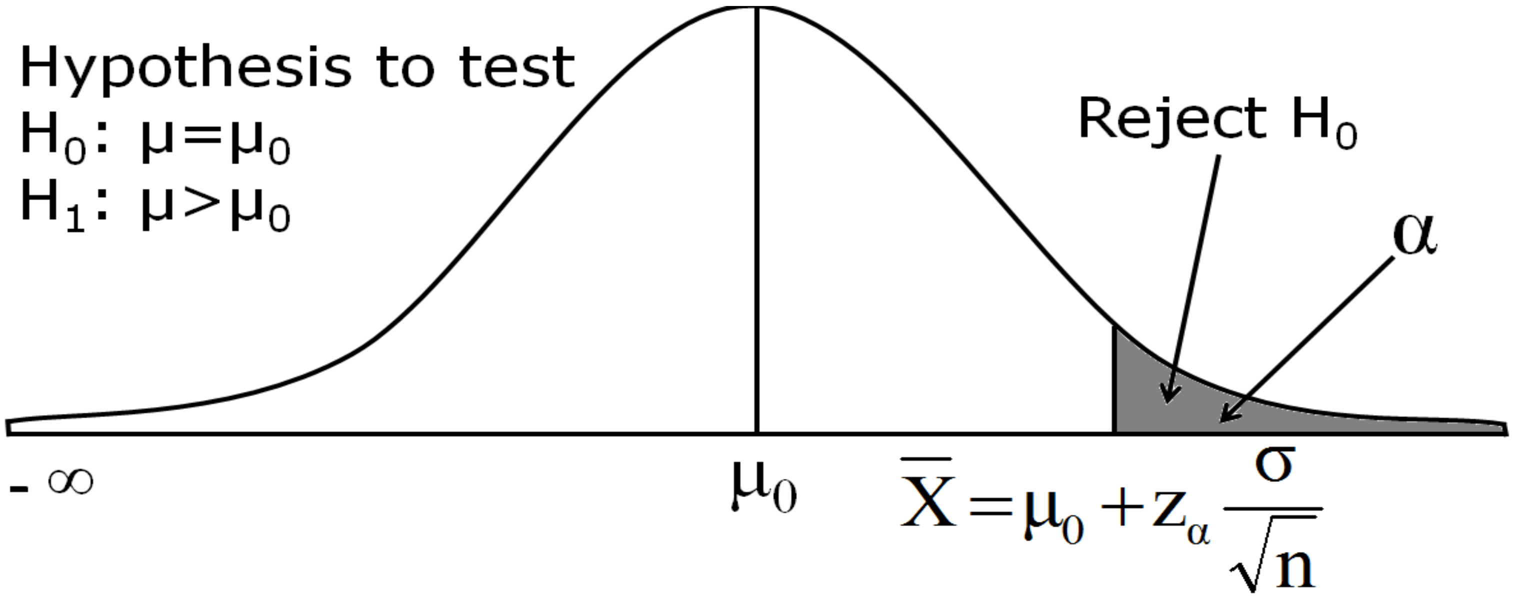


6. Conclusion

Steps involved in testing mean of a population

5. Define the rejection criteria/ critical regional

(ii) Reject H_0 if computed value of Z is greater than the critical value, ie., $P(Z > z_\alpha)$, otherwise do not reject H_0

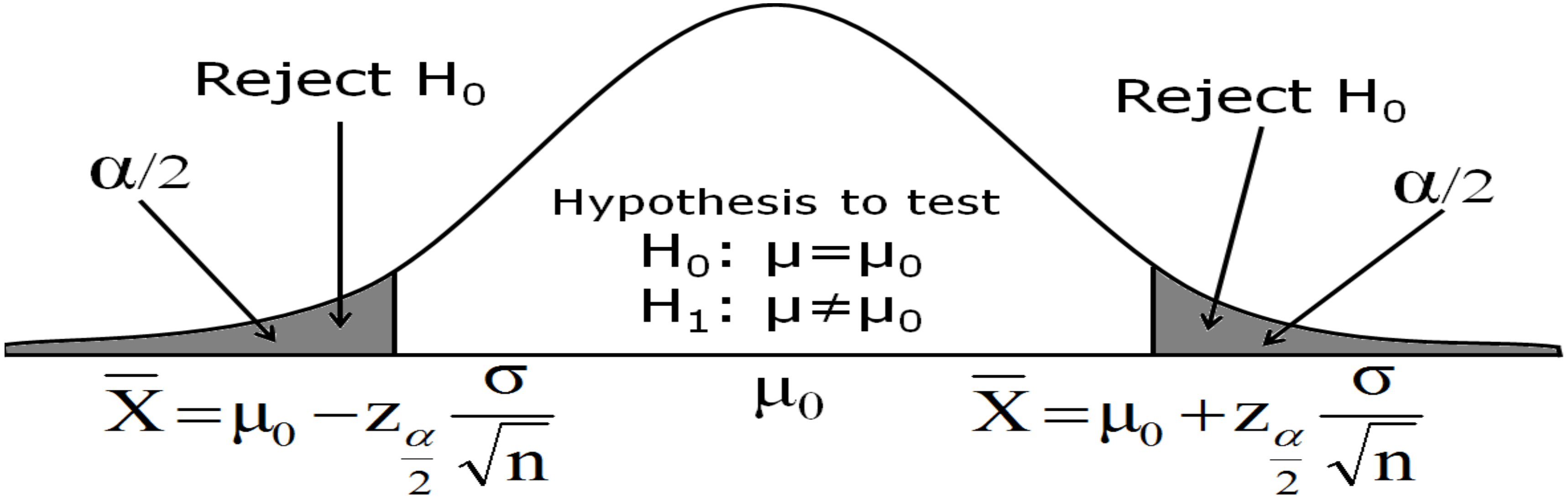


6. Conclusion

Steps involved in testing mean of a population

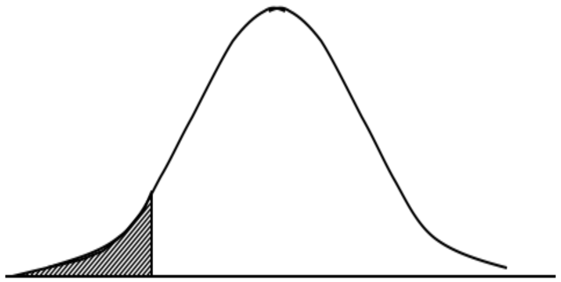
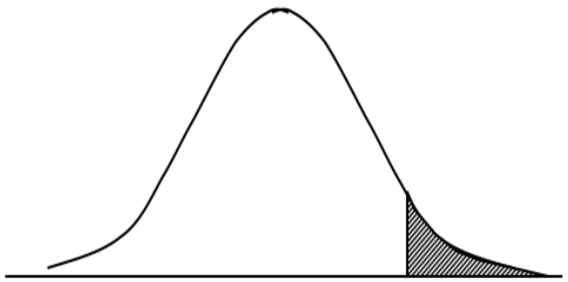
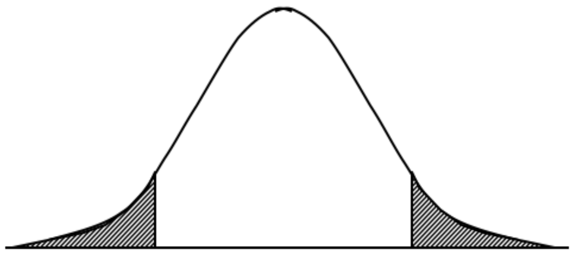
5. Define the rejection criteria/ critical regional

(iii) Reject H_0 if computed value of Z is less than or greater than the critical value, ie., $P(Z < -z_{\alpha/2})$ or $P(Z > z_{\alpha/2})$, otherwise do not reject H_0



6. Conclusion

Summary of One- and Two-Tail Tests…

One-Tail Test (left tail)	One-Tail Test (right tail)	Two-Tail Test
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$
		

Testing the mean difference of a two population

Assumptions on Z-test

- ♠ Assume that the samples are drawn from normal distribution
- ♠ The population variances should be known
- ♠ The two groups should be independent
- ♠ The data should be allocated randomly to both groups
- ♠ The sample size should be more than 30 (i.e., n > 30)

Steps involved in testing mean difference of two population

1. Stating null and alternative hypothesis

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 < \mu_2$

Or

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 > \mu_2$

Or

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

Steps involved in testing mean of a population

2. The level of significance α = 0.05

3. Normal distribution

4. The test statistic is

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1)$$

5. Define the rejection criteria/ critical regional

a. Reject H_0 if computed value of Z is less than the critical value, ie., $P(Z < -z_{\alpha})$, otherwise do not reject H_0

b. Reject H_0 if computed value of Z is greater than the critical value, ie., $P(Z > z_{\alpha})$, otherwise do not reject H_0

c. Reject H_0 if computed value of Z is less than or greater than the critical value, ie., $P(Z < - z_{\alpha/2})$ or $P(Z > z_{\alpha/2})$, otherwise do not reject H_0

d. 6. Conclusion

Testing the proportion of a single population

- **Assume that the samples are drawn from normal distribution**

Steps involved in testing proportion of a population

1. Stating null and alternative hypothesis

$$H_0: P = P_0$$

versus

(i) $H_1: P < P_0$

or

(ii) $H_1: P > P_0$

or

(iii) $H_1: P \neq P_0$

Steps involved in testing mean of a population

2. The level of significance $\alpha = 0.05$

3. Normal distribution

4. The test statistic is

$$Z = \frac{p - P_0}{\left(\sqrt{\frac{pq}{n}}\right)} \approx N(0, 1)$$

The rejection criteria for H_0 is same as described in testing for the mean

Testing the difference between proportion of two population

- **Assume that the samples are drawn from normal distribution**
- **The population variances should be known**

1. Stating the null and alternative hypothesis

- $H_0: P_1 = P_2$ vs $H_1: P_1 < P_2$

Or

- $H_0: P_1 = P_2$ vs $H_1: P_1 > P_2$

Or

- $H_0: P_1 = P_2$ vs $H_1: P_1 \neq P_2$

Z-test for difference between two proportions

Based on the sample data the test-statistic should be computed using

$$Z = \frac{\text{Proportion 1 - Proportion 2}}{\text{SE (Diff. b'n proportions)}}$$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx N(0, 1)$$

where

$$p = \frac{x_1 + x_2}{n_1 + n_2}, q = 1 - p$$

Small sample test

Student's t-test

- Independent sample t-test
- Unpaired t-test
- Paired t-test

Small sample test

Independent sample t-test

- Testing mean of population
- Testing difference between means of two populations

Assumptions on t-test

- ♠ Assume that the samples are drawn from normal distribution
- ♠ The population variances (σ^2) are not known
- ♠ The two groups should be independent

Assumptions on t-test

- Generally, the sample size should be less than 30 ($n \leq 30$). However, even if the sample size is more than 30 ($n>30$), but the Population variances are unknown, the Student's unpaired t-test will continued to be applied.

- This is because due to Central limit theorem (CLT) for $n>30$, the data converges to Standard Normal Distribution (SND)

Assumptions on t-test

- Samples drawn from two different populations independently (Not necessarily of equal sample size).

Testing the mean of a single population

- Assume that the samples are drawn from normal distribution
- The population variances should be known

Steps involved in testing mean of a population

1. Stating null and alternative hypothesis

$$H_0: \mu = \mu_0$$

versus

(i) $H_1: \mu < \mu_0$

or

(ii) $H_1: \mu > \mu_0$

or

(iii) $H_1: \mu \neq \mu_0$

Steps involved in testing mean of a population

2. The level of significance α = 0.05

3. Student's t-distribution

4. The test statistic is

$$t = \frac{\bar{x} - \mu_0}{\left(\frac{S}{\sqrt{n}}\right)} \approx t_{(\alpha;n-1)}$$

5. Define the rejection criteria/ critical regional

(i) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_\alpha)$, otherwise do not reject H_0 for one-tailed test

(ii) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_{\alpha/2})$, otherwise do not reject H_0 for two-tailed test

6. Conclusion

Steps involved in testing mean difference of two population

1. Stating null and alternative hypothesis

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 < \mu_2$

Or

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 > \mu_2$

Or

- $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

Steps involved in testing mean of a population

2. The level of significance α = 0.05

3. Student's t-distribution

4. The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx t_{(\alpha;\, n_1+n_2-2)}$$

$$\text{where } S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

5. Define the rejection criteria/ critical regional

(i) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_{\alpha})$, otherwise do not reject H_0 for one-tailed test

(ii) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_{\alpha/2})$, otherwise do not reject H_0 for two-tailed test

6. Conclusion

Paired sample t-test

Observations recorded on the same set of subjects at two different time intervals

Eg. Before and after treatment or Pre-test and Post-test

Assumptions on Paired t-test

- ♠ The difference between before and after observation should be normally distributed.

Steps involved in testing mean of paired samples

1. Stating null and alternative hypothesis

$$H_0: \mu = \mu_d = 0$$

versus

(i) $H_1: \mu < \mu_d$

or

(ii) $H_1: \mu > \mu_d$

or

(iii) $H_1: \mu \neq \mu_d$

Steps involved in testing mean of a population

2. The level of significance α = 0.05

3. Student's t-distribution

4. The test statistic is $t = \dfrac{\text{Difference b'n before and after observations}}{\text{SE (Diff. b'n before and after observations)}}$

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} \approx t_{(\alpha,\, n\text{-}1)}$$

5. Define the rejection criteria/ critical regional

(i) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_{\alpha})$, otherwise do not reject H_0 for one-tailed test

(ii) Reject H_0 if computed value of $|t|$ is greater than the critical value, ie., $P(|t| > t_{\alpha/2})$, otherwise do not reject H_0 for two-tailed test

6. Conclusion

Analysis of Variance (ANOVA)

- ►Samples should have been drawn from populations which are normally distributed
- ►Used to test equality of more than two population means

ANALYSIS OF VARIANCE

When there are more than two groups to be compared, it is not correct to compare the groups in pairs, as this type of comparison will not take the within variability into consideration

The Analysis procedure used in such comparisons is known as ANALYSIS OF VARIANCE

In this analysis the variability of observations between and within groups are taken into consideration

The total variability is split in these components and test is applied

The test used to compare these variability is F test

Assumptions on ANOVA

- Assume that the samples are drawn from normal distribution
- The population variances should be equal
- The groups should be independent

Assumptions on ANOVA

- The data should be allocated randomly to both groups

- The sample size should be more than 30

Non-Statistical example			Statistical example		
Verdict	Actual situation (H_0)		Decision	Null Hypothesis (H_0)	
	Innocent	Guilty		True	False
Innocent	Confidence level (1-α)	β-error	Do not Reject (Accept)	Confidence level (1-α)	β-error
Guilty	α-error	Power 1-β	Reject	α-error	Power 1-β

Definition of P-value

The P-value tell us the strength of the evidence against the null hypothesis that the true difference in the population is zero

Definition of P-value

- With any research study, there is a possibility that the observed differences were a chance event
- The only way to know that a difference is really present with certainty, the entire population would need to be studied
- The research community and statisticians had to pick a level of uncertainty at which they could live

Definition of P-value

- This level of uncertainty is called *Type 1* error or a false-positive rate (α)
- More commonly called a *p*-value
- In general, $p \leq 0.05$ is the agreed upon level
- In other words, the probability that the difference that we observed in our sample occurred by chance is less than 5%
  - Therefore we can reject the H_o

Definition of P-value

Stating the Conclusions of our Results

- **When the P-value is small, we reject the null hypothesis or, equivalently, we accept the alternative hypothesis.**
  - **"Small" is defined as a P-value $\leq \alpha$, where α = acceptable false (+) rate (usually 0.05).**

Definition of *P*-value

Stating the Conclusions of our Results

- **When the P-value is not small, we conclude that we cannot reject the null hypothesis or, equivalently, there is not enough evidence to reject the null hypothesis.**
  - **“Not small” is defined as a P-value > α, where α = acceptable false (+) rate (usually 0.05).**

Interpreting the P-value

If the P-value is less than 1% (< 0.01), there is *overwhelming evidence* that supports the alternative hypothesis.

If the P-value is between 1% and 5%, there is a *strong evidence* that supports the alternative hypothesis.

If the P-value is between 5% and 10% there is a *weak evidence* that supports the alternative hypothesis.

If the P-value exceeds 10%, there is *no evidence* that supports the alternative hypothesis.

Steps involved in testing of hypothesis

1. State null and alternative hypotheses
2. Specify the level of significance α
3. Define the probability distribution the data follows
4. Compute the test statistic based defined population
5. Define the rejection criteria/ critical regional
6. Conclusion

THANK YOU

Manipal ProLearn
#7, Service Road, Pragathi Nagar, Electronic City,
Bengaluru 560100
contact@manipalprolearn.com | manipalprolearn.com

Highlights of the program

- **First-of-its-kind program in the Data Science space**, equipping learners with competencies and skills boosting their visibility and credibility for future employment prospects.
- **Industry-relevant course curriculum**, with applications in multiple domains, where such talent is in demand
- **Highly experienced Subject Matter Experts (SMEs)** from academia, IT and Data Science industry
- **Enhanced learning experience** through the digital LMS - **EduNxt**
- State-of-the-art infrastructure, latest technology and a well-equipped, 77,000 square feet residential campus.

Highlights of the program

- Delivery Models: Online, blended, face to face
- Domain Expertise in Banking, Retail, Healthcare
- Industry Partnerships with Genpact, IBM-BDU, Coursera
- Centre Of Excellence with Deakin University
- Academic Partnership with Manipal University