Lean Six Sigma Green Belt Certification Course

Hypothesis Testing

Learning Objectives

By the end of this lesson, you will be able to:

- Explain the basics of hypothesis testing
- Perform tests for Means, Variances, and Proportions
- Conduct hypothesis testing with non-normal data

Scenario

Java Coffee House implemented an improved process at facility B, which would help reduce defects. The desire is to compare the impact of the improved process at facility B to the standard process being performed at facility A.

How could you confidently say that the new process is significantly better than the standard process?

Yield A	Yield B
80	85
80	81
80	95
93	95
93	89
87	87
82	92
81	81
93	82

simplilearn

DIGITAL OPERATIONS

Basics

Hypothesis Testing

Hypothesis test is used to prove or disprove a theory or claim by comparing two or more samples or comparing a single sample to a defined value.

A Six Sigma project produced significant improvements in average performance

Product quality is independent of supplier

The average cycle time for processing similar products is the same between two different facilities

Hypothesis Testing

Handle uncertainty in data

Minimize subjectivity

Question assumptions

Manage decision error risk

Hypothesis Testing

- Determine what to compare and assume that there is no difference
- Decide how much risk of being wrong is acceptable
- Use data to calculate a test statistic
- Compare the test statistic to a critical test statistic
- Make a decision by comparing p-value to accepted risk value

Null Hypothesis vs. Alternative Hypothesis

Null Hypothesis

- Represented as H_0
- Statement of no change or difference
- Cannot be proved, only rejected
- Example: Movie is worth watching.

Alternative Hypothesis

- Represented as H_a
- Statement of change or difference
- Challenges the null hypothesis
- Example: Movie is not worth watching.

$H_0 : \mu_a = \mu_b$

$H_a : \mu_a \neq \mu_b$

Null is stating there is no difference in means

Alterative is stating there is a difference in means

Hypothesis Testing

Type I Error

- Rejecting a null hypothesis when it is true
- False positive
- Also known as Producer's Risk
- Significance level or 'α' is the chance of committing a Type 1 error
- The value of 'α' is 0.05 or 5%

Example: When a movie is worth watching, it is reviewed as 'not worth watching.'

Type II Error

- Accepting a null hypothesis when it is false
- False negative
- Also known as Consumer's Risk
- 'β' is the chance of committing a Type II Error
- The value of 'β' is 0.2 or 20%
- Any experiment should have as less β value as possible

Example: When a movie is not worth watching, it is reviewed as 'worth watching.'

Type I and Type II Errors

		Court's Decision	
		Accept Null: Defendant is NOT Guilty	Reject Null: Defendant is Guilty
The Truth	Null is True: Actually Innocent	Correct	Type I Error α-Risk
	Null is False: Actually Guilty	Type II Error β-Risk	Correct

Type I and Type II Errors

Probability of making one type of error can be reduced

A false null hypothesis may be accepted (type II error)

'A' is set at 0.05, which means the risk of committing a type I error will be 1 out of 20 experiments

It is important to decide what type of error should be less and set 'α' and 'β' accordingly

Type I and Type II Errors

Power of a Test

- The probability of correctly rejecting the null hypothesis when it is false
- The complement to Type II error and represented as $1-\beta$
- The probability of not committing a type II error
- Helps in improving the advantage of hypothesis testing
- With highest value should be preferred when given a choice of tests

Confidence Level

- The probability of correctly failing to reject the null hypothesis when it is true
- The complement to Type I error and represented as $1-\alpha$
- The probability of not committing a Type I error

Sample Size

- How much variation is present in the population? (σ)
- At what interval or tolerance does the true population mean need to be estimated? (±Δ)
- How much representation error is allowed in the sample? (α)

The sample size for **continuous data** can be determined by the formula:

$$n = \left[\frac{Z_{1-\alpha/2} * \sigma}{\Delta} \right]^2$$

The sample size for **discrete data** can be determined by the formula:

$$n = \left[\frac{Z_{1-\alpha/2}}{\Delta} \right]^2 p(1-p)$$

$\sigma = \sqrt{p(1-p)}$ p=percent of non-defective

Standard Sample Size Formula

To calculate the standard sample size for continuous data, the value of α is taken as 5%.

According to Z table, the $Z_{97.5} = 1.96$.

The standardized sample size formula

$$n = \left[\frac{1.96 * \sigma}{\Delta} \right]^2$$

Standard Sample Size Formula

Q

The population standard deviation for the time, to resolve customer problems, is 30 hours. What should be the size of a sample that can estimate the average problem resolution time within ± 5 hours tolerance with 99% confidence?

A

$\Delta = 5$, $\sigma = 30$, $\alpha = 0.01$, and $Z_{99.5} = 2.58$

Sample size = $[(2.575*30)/5]^2 = 238.70 = 239$

Standard Sample Size Formula

To calculate the standard sample size for discrete data, the average population proportion of non-defective is ‘p’ and value of α is taken as 5%.

$$n = \left[\frac{1.96}{\Delta}\right]^2 p(1-p)$$

Where Δ = Tolerance allowed on either side of the population proportion average in percentage.

Standard Sample Size Formula

Q

The non-defective population proportion for pen manufacturing is 80%. What should be the sample size to draw a sample that can estimate the proportion of compliant pens within ± 5% with an alpha of 5%?

A

$\Delta = 0.05$, σ 2= 0.8 (1-0.8), $\alpha = 0.05$, and $Z_{97.5} = 1.96$

Sample size = $(1.96/0.05)^2 *0.8*0.2 = 245.86 = 246$

simplilearn

P-Values

The P-value is the probability that any differences observed are due to random chance or common cause variation.

A small p-value indicates a small probability that the observed results could happen under the assumption that the null hypothesis is true.

Low p-values lead us to reject the null hypothesis.

If lower than alpha risk, the null is rejected in favor of the alternative.

High p-values indicate that we haven’t gathered sufficient data or evidence to not reject the null hypothesis.

Test Statistics

A test statistic is used to determine the validity of a hypothesis test.

The test statistic is calculated based on the observations.

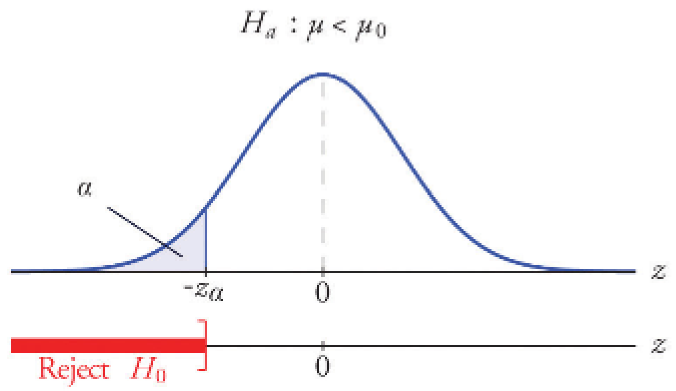
If the null hypothesis is true the test statistic will be a random variable with a known distribution.

If we can conclude that the test statistic came from the population, then the null is not rejected.

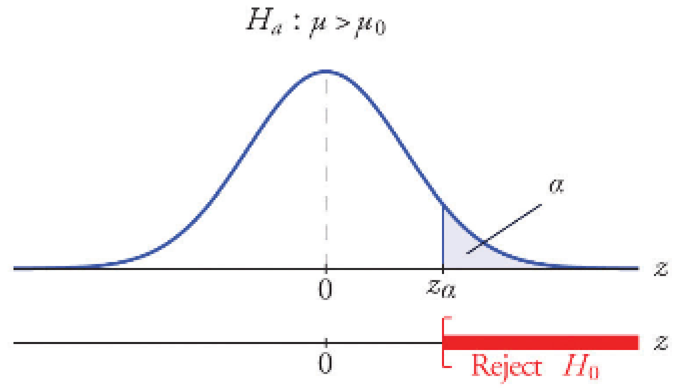
If it is not likely for the test statistic to come from population, then the null is rejected.

This is determined by comparing the test statistic to a critical value.

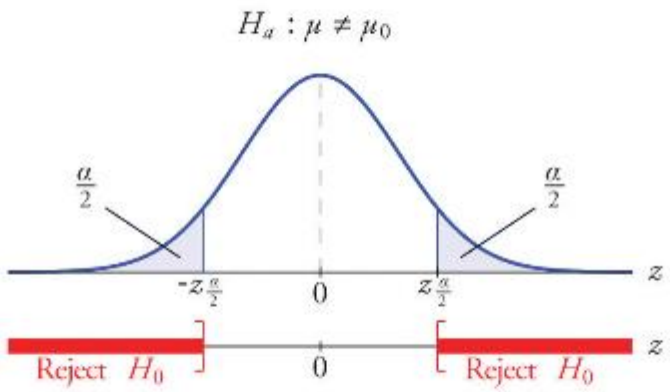
The calculation of the test statistic and critical value will depend on the particular hypothesis test.



Directional, Left-tail Test



Directional, Right-tail Test



Non-Directional, Two-tail Test

2 Tailed Probability and 1 Tailed Probability

Use of 2-tailed probability and 1-tailed probability depends on the direction of the alternative hypothesis.

2-Tailed Probability	1-Tailed Probability
If the alternate hypothesis tests more than one direction, either less or more, use a 2-tailed probability value from the test. **Example**: If Mean of A is not equal to Mean of B, then it is 2-tailed probability.	If the alternate hypothesis tests one direction, use a 1-tailed probability value from the test. **Example**: If Mean of A is greater than Mean of B, then it is 1-tailed probability.

Hypothesis Test Conclusions

Compare a calculated test statistic to a critical value	• If test statistic > critical value, Reject H_0 • If test statistic < critical value, Fail to reject H_0
Compare the p-value to the alpha risk	• If p-value < alpha risk, Reject H_0 • If p-value > alpha risk, Fail to reject H_0

Statistical and Practical Significance of Hypothesis Test

Significant but not Practical

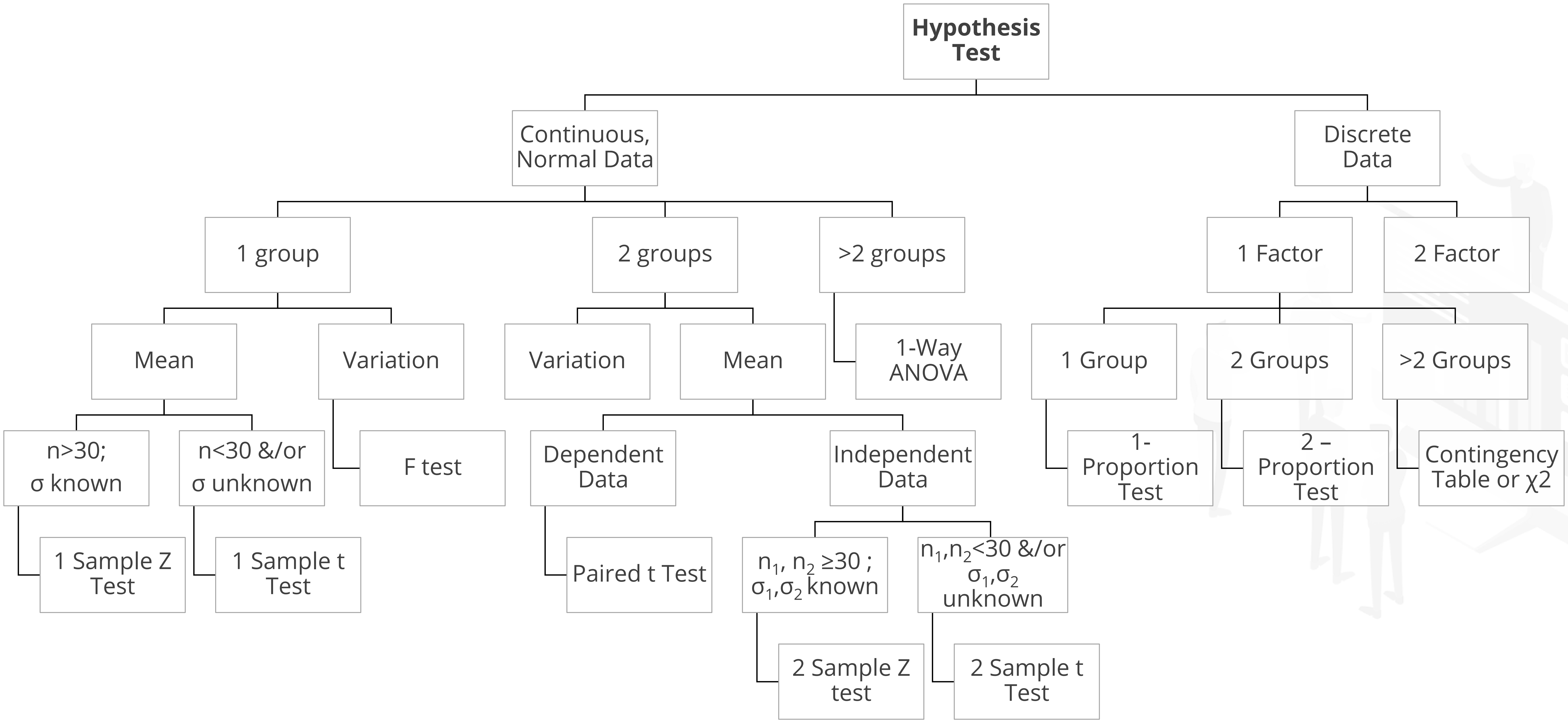
The Java Coffee House distribution center is comparing two methods for fulfilling orders in which the new method is significantly faster than the standard method by 30 minutes; however to justify the cost of changing to the new method, a reduction of an hour is required.

Practical but not Significant

A call center is piloting new technology to process payments which has increased transactions by 500 per day. Although the new technology has a higher average that will have a practical impact to the business, there is no statistically significant difference since $p \geq 0.05$

Tests for Means, Variances, and Proportions

Hypothesis Testing Roadmap



Hypothesis Testing Formulas

Test	Test For		Test Statistic	Distribution	Conditions
1 Sample Z test	Population mean (μ)	$\mu = \mu_0$	$\frac{(\bar{x} - \mu_o)}{\sigma / \sqrt{n}}$	Z	Normal distribution or $n > 30$; σ known
1 Sample t test	Population mean (μ)	$\mu = \mu_0$	$\frac{(\bar{x} - \mu_o)}{s / \sqrt{n}}$	t_{n-1}	$n < 30$, and/or σ unknown
1 Proportion test	Population proportion (p)	$p = p_0$	$\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1 - p_o)}{n}}}$	Z	$n\hat{p},\ n(1 - \hat{p}) \geq 10$
2 Sample Z test	Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Z	Both normal distributions, or $n_1, n_2 \geq 30$; σ_1, σ_2 known
2 Sample t test	Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	t distribution with df = the smaller of $n_1 - 1$ and $n_2 - 1$	$n_1, n_2 < 30$; and/or σ_1, σ_2 unknown
Paired t Test	Mean difference μ_d (paired data)	$\mu_d = 0$	$\frac{(\bar{d} - \mu_d)}{s_d / \sqrt{n}}$	t_{n-1}	$n < 30$ pairs of data and/or σ_d unknown
2 Proportion test	Difference of two proportions ($p_1 - p_2$)	$p_1 - p_2 = 0$	$\frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	Z	$n\hat{p},\ n(1 - \hat{p}) \geq 10$ for each group

CALCULATE TEST STATISTICS

Contingency Table or χ^2

H_0: Factor A is independent of Factor B

H_a: Factor A is dependent of Factor B

Test Statistic:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_o = Observed frequency

f_e = Expected frequency

1 Sample Test

H_0: Average height of North American men is165 cm

H_a: Average height of North American men is not 165 cm

Z-Test

Sample Size (n) = 117

Sample Average = 164.5 cm

Z Test Statistic = 1.04

t-test

Sample Size (n) = 25

Sample Average = 164.5 cm

t Test Statistic = 0.5

1 Sample Test

In hypothesis test for variance, Chi-square test is used.

H_0: Proportion of wins in Australia or abroad is independent of the country played against

H_a: Proportion of wins in Australia or abroad is dependent on the country played against

$\chi^2_{\text{Critical}} = 6.251$

$\chi^2_{\text{Calculated}} = 1.36$

Result:

Since calculated value is less than the critical value, the proportion of wins of Australia hockey team is independent of the country played or place.

1 Sample Test

H_0: Proportion of smokers among men in a place named R is 0.10 (p_0)

H_a: Proportion of smokers among men is R is different that 0.10

H_0: $p = p_0$ against Ha: $p \neq p_0$

Among n = 150 adult men interviewed, 23 were found smokers.

Sample proportion p = 23/150 = 0.153

Compute test statistic: $Z_{calc} = \frac{p - p_0}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.153 - 0.10}{\sqrt{\frac{0.153(0.847)}{150}}} = \frac{0.053}{0.029} = 1.80$

Reject H0 at level of significance α if $z_{calc} > z_{1-a/2}$

Since, $z_{.975} = 1.96$, the null hypothesis is not rejected at 5% level of significance.

Result: It can be concluded based on the sample that the proportion of smokers in R is 0.10.

Means of Two Groups

Understand the significant difference in the outcome of the two processes

Understand whether a new process is better than an old process

Understand whether the two samples belong to the same population or a different population

Benchmark the existing process with another process

2 Sample Test

H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$

Two samples of sizes n_1 = 125 and n_2 = 110 are taken from the two populations

$\bar{X}_1$ = 167.3, $\bar{X}_2$ = 165.8, s_1 = 4.2, s_2 = 5.0 are the sample means and SDs respectively

Compute test statistic

$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}} = \frac{167.3 - 165.8}{\sqrt{\frac{4.2^2}{125} + \frac{5.0^2}{110}}} = 2.47$$

Reject H_0 at level of significance **α** if I Computed t I > $t_{DF,a/2}$; DF = (n_1 – 1) + (n_2 – 1) = 124 + 109 = 233

Since $t_{233,0.025}$ = $1.97_{[\text{“=T.INV.2T(.05,233)”}]}$, the null hypothesis is rejected at 5% level of significance

2 Sample Test

Q

Obtained earnings data for Company A = 31 years
Obtained earnings data for Company B = 41 years
Sample standard deviation of Company A's earnings = $4.40
Sample standard deviation of Company B's earnings = $3.90.

Determine whether the earnings of Company A have a greater standard deviation than those of Company B at 5% level of significance.

A

$H_0 : \sigma_A^2 = \sigma_B^2$; the variance of Company A's earnings is equal to the variance of Company B's earnings.
$H_a : \sigma_A^2 \neq \sigma_B^2$; the variance of Company A's earnings is different.
σ_A^2= variance of Company A's earnings.
σ_B^2= variance of Company B's earnings.

F-Test Example

df_A (degrees of freedom of A) = 31 – 1 = 30

df_B (degrees of freedom of B) = 41 – 1 = 40

Critical value from F-table = 1.74

[“=F.INV.RT(0.50,30,40)”].

Calculation of F-test statistic:

$F = (S_A^2/S_B^2) = 4.40^2/3.90^2 = 1.273$

$\sigma_A > \sigma_B$. In calculating the F-test statistic, always put the greater variance in the numerator.

Results

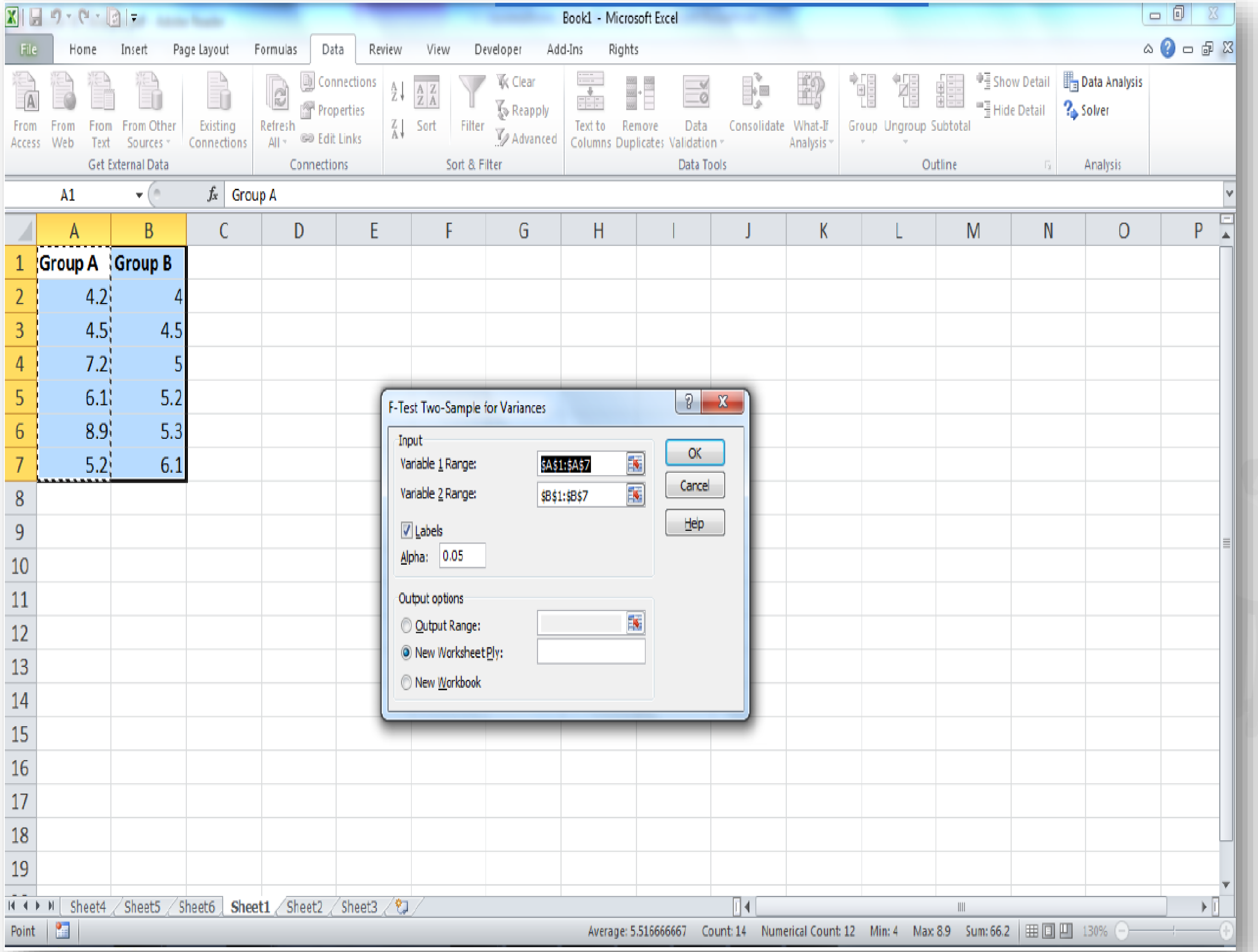
The F-test statistic (1.273) is not greater than the critical value (1.74). Therefore, at 5% significance level, the null hypothesis cannot be rejected.

F-Test Example

A restaurant wants to explore the overuse of avocados. It suspects that there is a difference in the way the two chefs use avocados and the number of avocados used, in pounds.

Group A (Chef 1)	Group B (Chef 2)
4.2	4
4.5	4.5
7.2	5
6.1	5.2
8.9	5.3
5.2	6.1

Conducting F-Test In MS Excel



simplilearn

F-Test Example

Null Hypothesis

- There is no significant statistical difference between the variances of the two groups
- This is Common Cause of Variation

Alternate Hypothesis

- There is a significant statistical difference between the variances of the two groups
- This is Special Cause of Variation

F-Test Example

- The p-value is 0.03.
- If p-value is < 0.05, null must be rejected.
- Null hypothesis with 97% confidence is rejected.
- The fact that variation could only occur due to Common Cause of Variation is rejected.
- There could be Assignable Causes of Variation or Special Causes of Variation.

F-Test Two-Sample for Variances		
	Variable 1	Variable 2
Mean	6.016666667	5.016666667
Variance	3.197666667	0.517666667
Observations	6	6
df	5	5
F	6.177076626	
P(F<=f) one-tail	0.033652302	
F Critical one-tail	5.050329058	

t-Test

Open MS Excel, click **Data**, and click **Data Analysis**.

Select **2-Sample Independent t-test** assuming unequal variances.

In Variable 1 range, select the data set for Group A.

In Variable 2 range, select the data set for Group B.

Keep the "Hypothesized Mean Difference" as 0.

Click **OK**.

t-Test

Group A (Chef 1)	Group B (Chef 2)
4.2	4
4.5	4.5
7.2	5
6.1	5.2
8.9	5.3
5.2	6.1

t-Test

Null Hypothesis

- There is no significant statistical difference between the means of the two groups
- This is Common Cause of Variation

Alternate Hypothesis

- There is a significant statistical difference between the means of the two groups
- This is Special Cause of Variation

H_0 : Mean of Group A = Mean of Group B

H_a : Mean of Group A ≠ Mean of Group B

t-Test

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	**Variable 2**
Mean	6.016666667	5.016666667
Variance	3.197666667	0.517666667
Observations	6	6
Hypothesized Mean	0	
df	7	
T Stat	1.270798616	
P(T<=t) one-tail	0.122200546	
T Critical one-tail	1.894578605	
P(T<=t) two-tail	0.244401092	
T Critical two-tail	2.364624252	

- The p-value of 2-tailed probability testing is 0.24
- This value is greater than 0.05
- The null hypothesis is not rejected
- Both the groups are statistically same

t-Test

The paired t-test is conducted before and after the process to measure:

Customer satisfaction before and after improvements

Employee performance before and after training

Jan	Feb
360	365
324	325
377	359
336	352
383	397
361	351
369	367
349	397
301	335
354	338
344	349
329	393
337	370
387	400
378	411

ANOVA

Used to compare the means of more than two samples

Stands for analysis of variance

Helps in understanding that all sample means are not equal

Shortlisted samples can further be tested

Generalizes the t-test to include more than two samples.

ANOVA

Q

The table shows the takeaway food delivery time of three different outlets.
Is there any evidence that the averages for the three outlets are not the same?

A

The null hypothesis will assume that the three means are equal. If the null hypothesis is rejected, it would mean that there are at least two outlets that are different in their average delivery time.

Outlet 1	Outlet 2	Outlet 3
48	50	49
49	48	48
48	36	39
53	50	49
58	50	34
50	62	33
46	45	57
50	47	48
49	51	47
47	44	39

ANOVA

Enter the data in an Excel spreadsheet → Select the ANOVA – single factor test from the Data Analysis "Toolpack" → Select all the cells for analysis

Outlet 1	Outlet 2	Outlet 3
48	50	49
49	48	48
48	36	39
53	50	49
58	50	34
50	62	33
46	45	57
50	47	48
49	51	47
47	44	39

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	10	498	49.8	11.95556
Column 2	10	483	48.3	42.9
Column 3	10	443	44.3	58.9

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	161.6667	2	80.83333	2.131764	0.138165	3.354131
Within Groups	1023.8	27	37.91852			
Total	1185.467	29				

Chi-Square Distribution

Chi-square distribution (χ^2-distribution) or Chi-squared

- Is a widely used probability distribution in inferential statistics;
- Needs one sample for the test to be conducted; and
- Has k-1 degrees of freedom and is the distribution of a sum of the squares of k independent standard normal random variables.

$$\chi^2_{\text{Calculated}} = \sum \frac{(f_0 - f_e)^2}{f_e}$$

Where,

- $\chi^2_{\text{Calculated}}$ = chi-square index
- F_o = An observed frequency
- F_e = An expected frequency

Chi-Square Distribution

	No. of wins in Australia and abroad against following countries				
	S. Africa	Pakistan	England	N Korea	Total
Australia	3	6	5	7	21
Abroad	2	2	4	2	10
Total	5	8	9	9	

Estimated Population Parameters

Sample Statistics

		No. of wins in Australia and abroad against following countries				
		S. Africa	Pakistan	England	N Korea	Total
Australia	Observed	3	6	5	7	21
	Expected	3.39	5.42	6.10	6.10	21
	Observed	2	2	4	2	10
	Expected	1.61	2.58	2.90	2.90	10
	Total	5	8	9	9	

Chi-Square Distribution

f_0	f_e	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
3	3.39	0.15	0.04
6	5.42	0.34	0.06
5	6.10	1.20	0.20
7	6.10	0.82	0.13
2	1.61	0.15	0.09
2	2.58	0.34	0.13
4	2.90	1.20	0.41
2	2.90	0.82	0.28
			1.36

$$\mathcal{X}^2_{\text{Calculated}} = \Sigma \frac{(f_0 - f_e)^2}{f_e} = 1.36$$

Chi-Square Distribution

There is a different Chi-square distribution for each of the different numbers of degrees of freedom. For Chi-square distribution, degrees of freedom are calculated according to the number of rows and columns in the contingency table.

Degrees of freedom = (2 - 1)*(4-1) = 3

Assuming α = 10%, $\chi^2_{Critical} = 6.251$ [="CHISQ.INV.RT(0.1,3)']

$\chi^2_{Calculated} = 1.36$

DIGITAL OPERATIONS

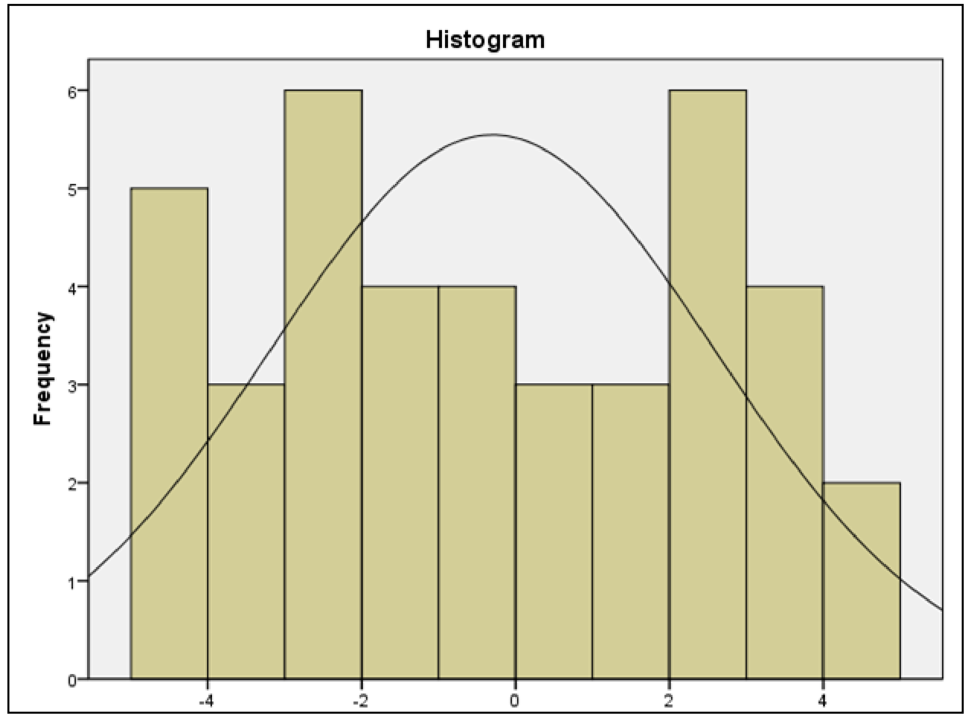
Hypothesis Testing with Non-normal Data

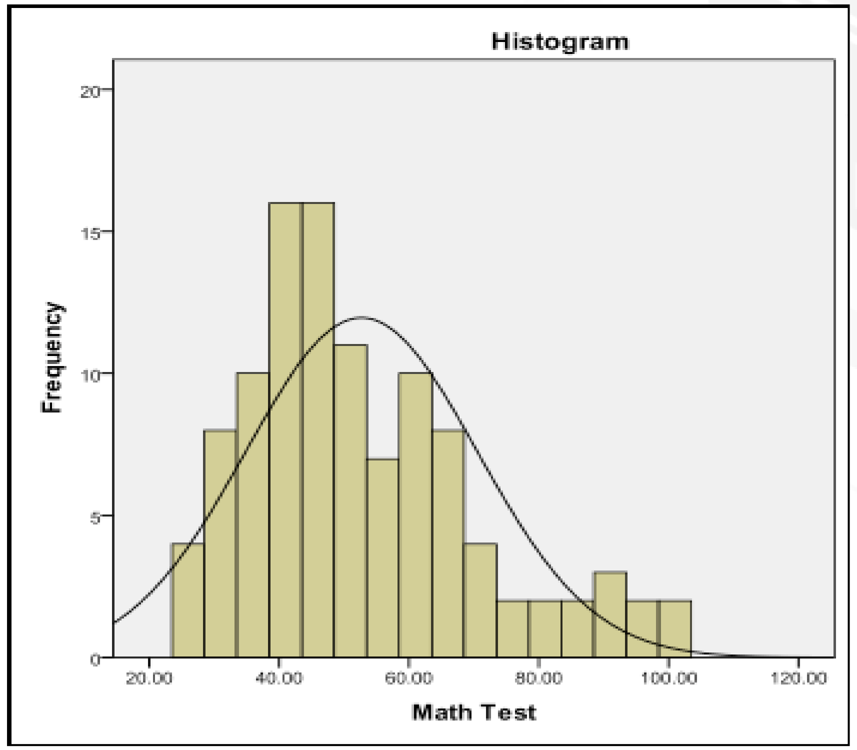
simplilearn

Non-Parametric Test

Non-parametric tests do not make any assumptions about a distribution model where the data could fit.

Non-parametric tests compare groups of medians using the relative ranks of the data within the groups.





Non-Parametric Test

Non-parametric Test	Main Characteristics	Corresponding Parametric Tests
1 sample Sign test	Test on the median, for non symmetric distribution	1 sample t or Z test
1 sample Wilcoxon test	Test on the median for symmetric distribution (pairwise averages)	1 sample t or Z test
Mann and Whitney test	Test on ranks to compare center of 2 groups	2 samples t or Z test
Kruskal-Wallis test	Test on ranks to compare center of 2 or more groups; based on the Chi Square Distribution More powerful than Mood's median test but less robust to outliers	One Way ANOVA
Mood's median test	Test on the overall median; based on the Chi Square test More robust to outliers than the Kruskal-Wallis test, but less powerful	One Way ANOVA
Freidman Test	Test on ranks, based on the Chi Squared distribution	Two-way randomized ANOVA

Same rules apply: If p < α reject the Null Hypothesis

Non-Parametric Test

Mann-Whitney test is a non-parametric test used to compare the center between two unpaired groups

Group	Data
G1	14
	2
	5
	16
	9
G2	4
	2
	18
	14
	8

Sorted Data	Group	Rank A
2	G1	1
2	G2	2
4	G2	3
5	G1	4
8	G2	5
9	G1	6
14	G1	7
14	G2	8
16	G1	9
18	G2	10

Avg. = 1.5

Avg. = 7.5

Final Rank
1.5
1.5
3
4
5
6
7.5
7.5
9
10

G1 Rank (R1)
1.5
4
6
7.5
9
Total = 28
n1 = 5

G2 Rank (R2)
1.5
3
5
7.5
10
Total = 27
n2 = 5

Non-Parametric Test

$$U1 = R_1 - \frac{n_1(n_1+1)}{2} = 28 - \frac{5\,(5+1)}{2} = 13$$

$$U2 = R_2 - \frac{n_2(n_2+1)}{2} = 27 - \frac{5\,(5+1)}{2} = 12$$

$$U_{calc} = MIN\,(U1, U2) = 12$$

n_2	α	n_1: 3	4	5	6	7	8	9	10	11	12	13	14	15	16
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18

The obtained U has to be equal to or less than this critical value.

The calculated value is not equal to or less than 2. Therefore, there is no statistical difference between the means of the two groups.

Non-Parametric Test

The Kruskal-Wallis test is used for testing the source of origin of the samples.

Only way to analyze the variance by ranks

Medians of two or more samples are compared to find the source of origin

Does not assume the normal distribution of the residuals

Testing the ratings of a product from three different groups to see if the ratings are the same or different

Non-Parametric Test

The Mood's Median is a non-parametric test used to test the equality of medians from two or more different populations. The test works if:

- The output (Y) variable is continuous, discrete-ordinal, or discrete count.
- The input (X) variable is discrete with two or more attributes.

Find median of the combined data set → Find the number of values in each sample that are greater than the median →

Form a contingency table → Find expected value for each cell →

Find Chi-square value → Compare Chi-square value to critical Chi-Square value

Example:
To determine whether temperature changes in the ocean water near a nuclear power plant will have a significant effect on the animal life in the region, an environmental group places groups of fish in four bowls that are identical in except for water temperature. Six months later, they measure the weights of the fish.

Non-Parametric Test

The Friedman Test is a form of non-parametric test that does not make any assumptions on the origin of the sample.

A marketing company wants to compare the relative effectiveness of three different modes of advertising:

Direct Mail   Newspaper   Magazine Advertisements.

The company conducts a randomized block design experiment. For 14 customers, the marketing company used all three modes during a 1-year period and recorded the percentage response to each type of advertising.

Non-Parametric Test

1 Sample Sign test is the simplest of all non-parametric tests that can be used instead of one sample t test

Count the number of positive values → Count the number of negative values → Test the values

HR of a large company analyzes its payroll to determine whether the company's median salary differs from the industry average.

Non-Parametric Test

1 Sample Wilcoxon Test is a non-parametric test.

It is equivalent to 1 Sample t-Test and is more powerful than 1 Sample sign test.

It is used to estimate the population median and compare it to a target or reference value.

It assumes the existing sample is randomly taken from a population.

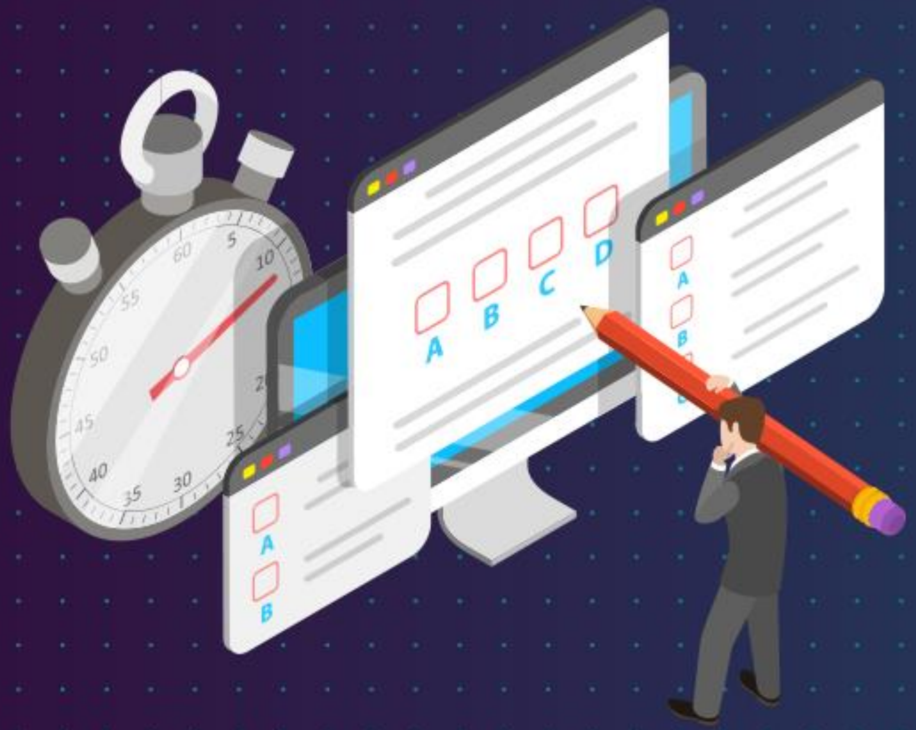
The median customer satisfaction score of an organization has always been 3.7. Management wants to see if it has changed. They conducted a survey and got the results grouped by the customer type

Conclusion:

- If median = 3.7 then Fail to Reject Null
- If median ≠ 3.7 then Reject Null

Key Takeaways

- Hypothesis test is used to prove or disprove a theory or claim by testing two or more samples or comparing a sample to a defined value.
- In deciding to reject or not reject the null hypothesis, we can make two possible decision errors—Type I and Type II errors.
- To calculate the standard sample size for discrete data, the average population proportion non-defective is ‘p’ and value of α is taken as 5%.
- The P-value is the probability that any differences observed are due to random chance or common cause variation.
- A test statistic is used to determine the validity of a hypothesis test.
- Non-parametric tests include Mann-Whitney Test, Kruskal-Wallis Test, Mood’s Median Test , Freidman Test, 1 Sample Sign Test, and 1 Sample Wilcoxon Test.



Knowledge Check

Knowledge Check

1

If p value is 0.05 and the confidence level is 90%, what would be the hypothesis test conclusion?

A. Fail to reject the null hypothesis

B. Reject the null hypothesis

C. Reject the alternative hypothesis

D. Not enough information

Knowledge Check 1	**If p value is 0.05 and the confidence level is 90%, what would be the hypothesis test conclusion?**

A. Fail to reject the null hypothesis

B. Reject the null hypothesis

C. Reject the alternative hypothesis

D. Not enough information

The correct answer is **B**

Alpha value is 10%. Since confidence level is 90% and p is smaller than the alpha value, the null hypothesis is rejected

Knowledge Check 2	**An Assembly team desired to see if there was any performance improvement after completing a Six Sigma project. What hypothesis could be used?**

A. F-test

B. Two Sample t-test

C. One Sample t-test

D. Paired t-test

Knowledge Check 2

An Assembly team desired to see if there was any performance improvement after completing a Six Sigma project. What hypothesis could be used?

A. F-test

B. Two Sample t-test

C. One Sample t-test

D. Paired t-test

The correct answer is **D**

To see if a process has improved, a paired t-test should be used to compare the before and after improvement state.

Knowledge Check 3

Which non-parametric test is similar to a single factor ANOVA?

A. Sample Sign

B. Wilcoxon Sign test

C. Mood’s Median

D. Freidman’s Test

Knowledge Check 3

Which non-parametric test is similar to a single factor ANOVA?

A. Sample Sign

B. Wilcoxon Sign test

C. Mood's Median

D. Freidman's Test

The correct answer is **C**

Mood's median is similar to a single factor ANOVA . It is able to test for the difference in medians for more than 2 groups.

Knowledge Check 4	**A team wants to test if a new drug reduced pain in the patients. What would be the Type II error?**

A. The new drug really works and team concludes it works

B. The new drug does not work and team concludes it works

C. The new drug really works and team concludes it does not work

D. The new drug does not work and the team concludes it does not work

Knowledge Check

4

A team wants to test if a new drug reduced pain in the patients. What would be the Type II error?

A. The new drug really works and team concludes it works

B. The new drug does not work and team concludes it works

C. The new drug really works and team concludes it does not work

D. The new drug does not work and the team concludes it does not work

The correct answer is **C**

Type II error fails to reject the null hypothesis when it is false. Therefore, if the null hypothesis is the drug, it does not cause a difference in pain levels.

Knowledge Check 5	**Which hypothesis test is used to compare the variance for two groups with normal data?**

A. Z Test

B. F Test

C. t-Test

D. χ2 Test

Knowledge Check

5

Which hypothesis test is used to compare the variance for two groups with normal data?

A. Z Test

B. F Test

C. t-Test

D. $\chi 2$ Test

The correct answer is **B**

The F test is used to compare variance for two or more groups with normal data.

Knowledge Check

6

The population standard deviation for the time, to resolve customer problems, is 20 hours. What should be the size of a sample that can estimate the average problem resolution time within ± 2 hours tolerance with 95% confidence?

A. 385

B. 384

C. 386

D. 400

Knowledge Check

6

The population standard deviation for the time, to resolve customer problems, is 20 hours. What should be the size of a sample that can estimate the average problem resolution time within ± 2 hours tolerance with 95% confidence?

A. 385

B. 384

C. 386

D. 400

The correct answer is **A**

Explanation: Since the confidence level is 95% we can use the standard sample size formula $n = \left[\frac{1.96^{*}\,\sigma}{\Delta}\right]^2 = \left[\frac{1.96^{*}20}{2}\right]^2 = 384.16$. We round up to nearest integer so answer is A.

simplilearn

Knowledge Check 7

After conducting a hypothesis test at 5% significance level the test statistic was 30, the critical value was 25, and the p-value was .10. What can we conclude?

A. Reject the null hypothesis

B. Fail to reject the null hypothesis

C. Not enough information to decide

D. Something went wrong with the test

Knowledge Check 7	**After conducting a hypothesis test at 5% significance level the test statistic was 30, the critical value was 25, and the p-value was .10. What can we conclude?**

A. Reject the null hypothesis

B. Fail to reject the null hypothesis

C. Not enough information to decide

D. Something went wrong with the test

The correct answer is **D**

The p-value indicates failure to reject the null and the test statistic indicates rejection of the null. Therefore, there is a discrepancy because both methods should always lead to the same conclusion.

Knowledge Check

8

Which non-parametric test is similar to a 1 sample t test?

A. Freidman

B. Kruskal-Wallis

C. Wilcoxon Signed Rank

D. Mood’s Median

simplilearn

Knowledge Check
8

Which non-parametric test is similar to a 1 sample t test?

A. Freidman

B. Kruskal-Wallis

C. Wilcoxon Signed Rank

D. Mood’s Median

The correct answer is **C**

Wilcoxon Signed Rank test is is similar to a 1 sample t test. It is also known as the 1 Sample Wilcoxon Test.