Business Statistics

Agenda - Hypothesis Testing - Week 2

1. Hypothesis Testing
   a. Introduction
   b. Hypothesis Formulation
2. Basic concepts of Hypothesis Testing
   a. Importance of null
   b. Importance of test statistic
   c. Type I and Type 2 errors
   d. Hypothesis testing template
3. Performing a Hypothesis Test
   a. Some key ideas
   b. Assumptions
   c. Critical point
   d. Rejection region approach
   e. p-value approach
4. One-Tailed and Two-Tailed Tests
5. Confidence Interval and Hypothesis Test
6. Some important tests
   a. Test for one mean
   b. Test for equality of means
   c. Test for equality of means - equal std dev
   d. Test for equality of means - unequal std dev
   e. Paired test for equality of means
   f. Test for one proportion
   g. Test for two proportions
   h. Test for one variance
   i. Test for equality of variances
   j. Test of independence
   k. ANOVA test

Hypothesis Testing

Real World Problem

Suppose you are a quality analyst at a bulb manufacturing company and analyze the reliability of bulbs. Historically, 70% of the bulbs pass the reliability test.

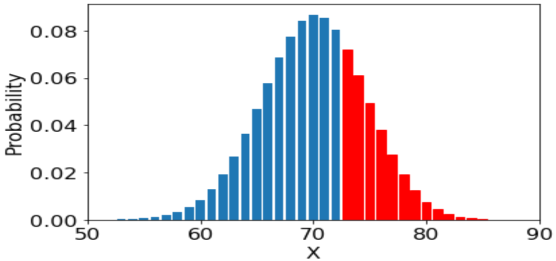
Now, a slightly altered manufacturing process(B) has been introduced to produce the bulbs.

Can you conclude whether the new process improves the reliability of the bulbs or not by checking the number of reliable bulbs in a sample?

Gathering evidence for statistical Inference

We selected a random sample of 100 bulbs out of which 73 are reliable. Does this provide strong evidence that the new manufacturing process is more reliable?

If the new manufacturing process was only as good as the current process - What is the probability of getting 73 or more reliable bulbs in a sample of 100 bulbs?

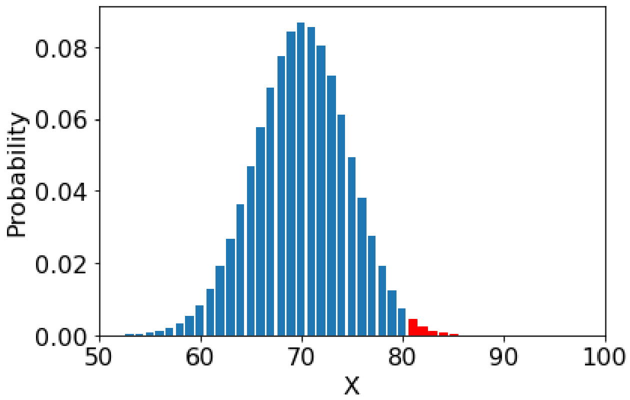


The probability of getting 73 or more reliable bulbs in a sample of 100 bulbs is ~0.30.

Thus, there is no strong evidence that the new process improves reliability

Gathering evidence for statistical Inference

A similar experiment was run with yet another manufacturing process (C). A sample of 100 bulbs produced using this process had 81 reliable bulbs.



The probability of getting 81 or more reliable bulbs in a sample of 100 bulbs is ~0.01.

Thus, there is strong evidence that the new process improves reliability

Why Hypothesis?

Estimation

The problem of estimation is considered, when there is no previous knowledge of the population parameter. The problem is simpler in that case. A random sample is taken, a sample statistic is computed and an appropriate point and interval estimate is suggested.

Hypothesis Testing

Often the interest is not in the numerical value of the point estimate of the parameter, but in knowing the plausibility of a hypothesis about the population parameter by using sample data. Estimation is not enough to arrive at a conclusion in such cases.

What is Hypothesis?

Often we are interested in population parameter(s)

A hypothesis is a conjecture about the population parameter(s)

For example, a bulb manufacturing company is interested in knowing whether the new manufacturing process improves reliability of the bulbs.

The objective of the Hypothesis Testing is to SET a value for the parameter(s) and perform a statistical TEST to see whether that value is tenable in the light of the evidence gathered from the sample.

Overview of Applications

Applications of Hypothesis Testing

- **Testing Research Hypotheses**
  - e.g. a new automobile system increases the mean mpg performance
- **Testing the validity of a claim**
  - e.g. a manufacturer claims that 1L soft drink bottles are filled with an average of at least 0.99L
- **Testing the business decisions**
  - e.g. new online ad has resulted in higher online conversion rates for an E-commerce website

Stating the Hypothesis

Null and Alternative Hypotheses - Two mutually exclusive statements about the population parameter(s)

Null Hypothesis (H_0)

The presumed current state of the matter or status quo.

E.g. The new process for manufacturing bulbs does not improve reliability.

Alternative Hypothesis (H_a)

The rival opinion or research hypothesis or an improvement target.

E.g. The new process for manufacturing bulbs improves reliability.

Null & Alternative Formulation : Example

Mean length of lumber is specified to be 8.5m for a certain building project. A construction engineer wants to make sure that the shipments she received adhere to that specification.

The population parameter about which the hypothesis will be formed is **population mean μ**.

The hypotheses are

$$H_0 : \mu = 8.5$$

$$H_a : \mu \neq 8.5$$

Null & Alternative Formulation : Example

There is a belief that 20% of men on business travel abroad brings a significant other with them. A chain hotel claims that number is too low.

The population parameter about which the hypothesis will be formed is **population proportion π.**

The hypotheses are

$$H_0 : \pi = 0.2$$

$$H_a : \pi > 0.2$$

Tips to formulate Null & Alternative

Am I testing a status quo that already exists?

Null Hypothesis

Negation of the research question

Always contains equality (=, >= , <=)

Am I testing an assumption or claim that is beyond what I know?

Alternate Hypothesis

Research question to be proven

Doesn’t contain equality (≠, >, <)

Basic Concepts of Hypothesis Testing

Importance of Null

Null hypothesis is assumed to be true unless reasonably strong evidence to the contrary is found.

Based on a random sample a decision is made whether there exists reasonably strong evidence against the null hypothesis.

Evidence is strong (satisfies the predetermined decision rule) → **Reject the null hypothesis** in favour of alternative hypothesis

Evidence is not strong (does not satisfy the predetermined decision rule) → **Fail to reject the null hypothesis** in favour of alternative hypothesis

Importance of Test Statistic

The test statistic is calculated from the sample data and tested against the predetermined Decision Rule.

The test statistic is a random variable that follows a standard distribution such as Normal, T, F, Chi-square etc. Sometimes the tests are named after the test statistic

Since hypothesis testing is done on the basis of sampling distribution, the decisions made are probabilistic.

Hence, it is very important to understand the errors associated with hypothesis testing.

Type I and Type II Error

Type I and Type II Errors

Level of significance

Power of the test

	H_0 is True	H_0 is False
Reject H_0	**Type I Error** **Prob = α**	**Correct decision** **Prob = 1 - β**
Fail to reject H_0	**Correct decision** **Prob = 1 - α**	**Type II Error** **Prob = β**

Type I and Type II Errors : Example

Null Hypothesis: The patient doesn't have cancer

Alternate Hypothesis: The patient has cancer

- **Type I error (false positive)**: "The patient doesn't have cancer but doctors says she does"
- **Type II error (false negative)**: "The patient does have cancer but report says she doesn't"

Template for Hypothesis Testing

Hypothesis Testing Template

1	Identify the key question	*What is the research question that you are trying to answer?*
2	Establish the hypotheses	*What is the metric of interest? Define the Null and Alternate Hypothesis.*
3	Understand and prepare data	*What data do you have? Do you understand what it means? Can it be used directly?*
4	Identify the right test	*Choose the method for testing based on the last three points*
5	Check the assumptions	*Ensure that data satisfies the assumption for the test.*
6	Perform the test	*Get to conclusion based on the results (p-value)*

Performing a hypothesis test

Some key ideas first

Level of Significance (α)

- Probability of rejecting the null hypothesis when it is true
- Fixed before the hypothesis test.

p-value

- Probability of observing test statistic or more extreme results than the computed test statistic, under the null hypothesis.
- Depends on the sample data. Alpha is pre-fixed but p-value depends on the value of the test statistic

Acceptance or Rejection Region

- The total area under the distribution curve of the test statistic is partitioned into acceptance and rejection region
- Reject the null hypothesis when the test statistic lies in the rejection region, Else we fail to reject it

Let’s start simple

Consider the following questions in hypothesis testing

What are the null and alternative hypotheses?	What is an appropriate test statistic?
What is preset level of significance?	How to check whether the data is giving significant evidence against the null hypothesis or not?

Let’s see an example and understand the significance of the above questions

For simplicity, we will assume that the population standard deviation is known and the sample size is more than 30.

Example

It is known from experience that for a certain E-commerce company the mean delivery time of the products is 5 days with a standard deviation of 1.3 days.

The new customer service manager of the company is afraid that the company is slipping and collects a random sample of 45 orders. The mean delivery time of these samples comes out to be 5.25 days.

Is there enough statistical evidence for the manager's apprehension that the mean delivery time of products is greater than 5 days.

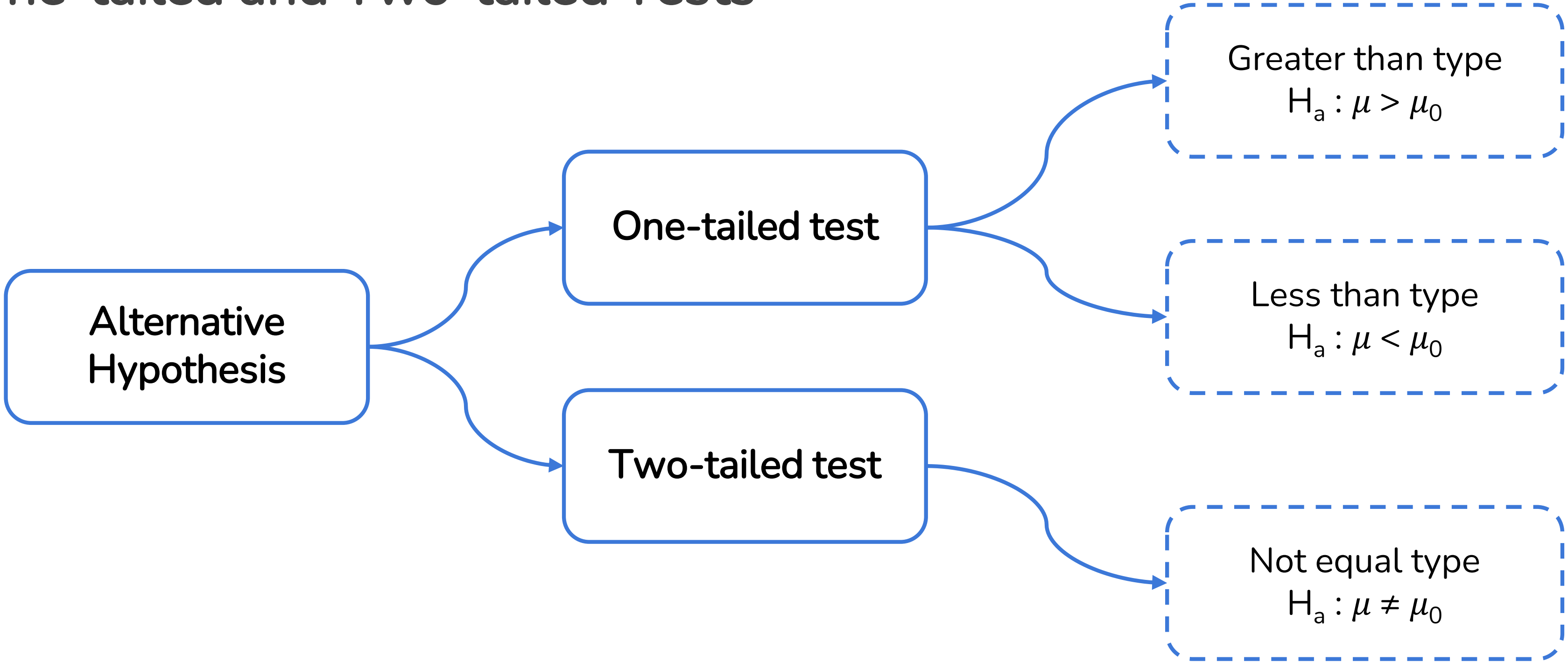
This is clearly a one-tailed test, concerning **population mean μ, the mean delivery time of products.**

First test - z-test for One Mean

Significance of the test	Assumptions	Test Statistic Distribution
Test for population mean $H_0 : \mu = \mu_0$	• Continuous data • Normally distributed population or sample size > 30 • Known population standard deviation σ • Random sampling from the population	Standard Normal distribution

One-tailed and Two-tailed Tests

One-tailed and Two-tailed Tests

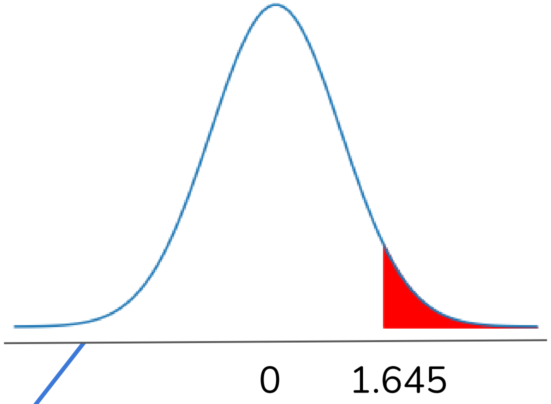


Choice of One tailed vs Two tailed depends on the nature of the problem, not on the sample data!

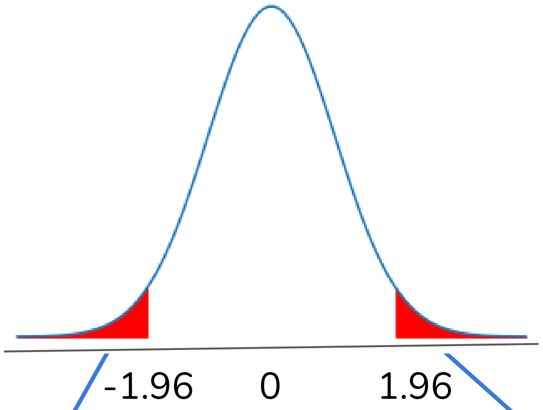
Difference between One-tailed and Two-tailed Tests

Test statistic value **does not change** for two-tailed or one-tailed test.

Only the critical value(s) / p-value associated with the test statistic changes



The difference is not tested on this side and the hypothesis test has greater power on the other side



The difference is tested on both the sides.

Connecting the dots with Confidence Intervals

Confidence Interval vs Hypothesis Testing

Suppose we calculate the (100 - 5)% confidence interval for the mean

We also conduct the Z-test for the mean with a 5% significance level.

The hypotheses of the Z-test are

$$H_0 : \mu = \mu_0 \text{ against } H_a : \mu \neq \mu_0$$

Is there any relationship between the estimated confidence interval and the hypothesis test?

The confidence interval contains all values of μ_0 for which the null hypothesis will not be rejected.

Some important Statistical Tests

Hypothesis Testing Frameworks

Choice of test depends on test statistic and data availability

Means

Compare the sample mean to the population mean when std dev is known
1-sample z-test

Compare the sample mean to the population mean when std dev is unknown
1-sample t-test

Compare the sample means from 2 independent populations when std devs are known
2-sample ind. z-test

Compare the sample means from 2 independent populations when std devs are unknown
2-sample ind. t-test

Compare the sample means from 2 related populations when std devs are unknown
Paired t-test

Compare the sample means from 2 or more independent populations
ANOVA Test

Proportions

Compare the sample proportion to the population proportion
1-sample z-test

Compare the sample proportions from two populations
2-sample z-test

Variances

Compare the sample variance to the population variance
Chi-Square test

Compare the sample variances from two populations
F-test

Frequencies

Check whether the categorical variables from a population are independent
Chi-Square Test of Independence

Test for one mean

Example

A certain food aggregator ZYX is facing stiff competition from its main rival SWG during Corona period. To retain business, ZYX is advertising that, within a radius of 5 km from the restaurant where the order is placed, it can still deliver in 40 minutes or less on the average (and changed condition has not made any impact on them).

The delivery times in minutes of 25 randomly selected deliveries are given in a CSV file.

Assuming the delivery distribution is approximately normal, is there enough evidence that ZYX's claim is false?

This is clearly a one-tailed hypothesis problem, concerning population mean μ, the average delivery time.

Test for One Mean - Unknown Std Dev

Significance of the test	**Assumptions**	**Test Statistic Distribution**
Test for population mean $H_0 : \mu = \mu_0$	● Continuous data ● Normally distributed population and sample size < 30 ● Unknown population standard deviation ● Random sampling from the population	t distribution (The test is also known as **One-sample t-test**)

Test for equality of means (Known std dev)

Example

To compare customer satisfaction levels of two competing media channels, 150 customers of Channel 1 and 300 customers of Channel 2 were randomly selected and were asked to rate their channels on a scale of 1-5, with 1 being least satisfied and 5 most satisfied. (The survey results are summarized in a CSV file)

Test at 0.05 level of significance whether the data provide sufficient evidence to conclude that channel 1 has a higher mean satisfaction rating than channel 2.

This is a two-sample problem where the channel 1 and channel 2 populations are independent. Further, this is a one-tailed hypothesis problem, concerning population means μ_1 and μ_2, the mean customer satisfaction for channel 1 and channel 2 respectively.

Test for Equality of Means - Known Std Devs

Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population means $H_0 : \mu_1 = \mu_2$	● Continuous data ● Normally distributed population or sample size > 30 ● Independent populations ● Known population standard deviations σ_1 and σ_2 ● Random sampling from the population	Standard Normal distribution (The test is also known as **Two independent sample z-test**)

Test for equality of means (Equal and unknown std dev)

Example

In the lockdown period, because of working from home and increased screen time, many opted for listening to FM Radio for entertainment rather than watching Cable TV. An advertisement agency randomly collected daily usage time data (in minutes) from both type of users and stored it in a CSV file.

Assuming daily Radio and TV usage time are normally distributed, do we have enough evidence to conclude that there is any difference between daily TV and Radio usage time at 0.05 significance level?

This is a two-sample problem where FM Radio and Cable TV users are assumed independent. Further, this is a two-tailed hypothesis problem, concerning population means μ_1 and μ_2, the daily mean usage time of Radio and TV respectively.

Test for Equality of Means : Equal Std Devs

Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population means $H_0 : \mu_1 = \mu_2$	● Continuous data ● Normally distributed populations ● Independent populations ● Equal population standard deviations ● Random sampling from the population	t distribution (The test is also known as **Two independent sample t-test**)

Test for equality of means (Unequal and unknown std dev)

Example

SAT verbal scores of two groups of students are given in a CSV file. The first group, **College**, contains scores of students whose parents have at least a bachelor's degree and the second group, **High School**, contains scores of students whose parents do not have any college degree.

The Education Department is interested to know whether the sample data support the theory that students show a higher population mean verbal score on SAT if their parents attain a higher level of education.

Assuming SAT verbal scores for two populations are normally distributed, do we have enough statistical evidence for this at 5% significance level?

This is a two-sample problem as the College and High School populations are different. Further, this is a one-tailed hypothesis problem, concerning population means μ_1 and μ_2, the mean verbal score on SAT for College and High School groups.

Test for Equality of Means : Unequal Std Devs

Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population means $H_0 : \mu_1 = \mu_2$	• Continuous data • Normally distributed populations • Independent populations • Unequal population standard deviations • Random sampling from the population	t distribution (The test is also known as **Two independent sample t-test**)

Paired Test for equality of means

Example

Typical prices of single-family homes in Florida are given for a sample of 15 metropolitan areas (in 1000 USD) for 2002 and 2003 in a CSV file.

Assuming the house prices are normally distributed, do we have enough statistical evidence to say that there is an increase in the house price in one year at 0.05 significance level?

This is a paired sample problem as the two observations (for 2002 and 2003) are taken on one sampled unit (a metropolitan area). Further, this is a one-tailed hypothesis problem, concerning population means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, the mean house price in 2002 and 2003 respectively.

Paired test for Equality of Means

Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population means $H_0: \mu_1 = \mu_2$	- Continuous data - Normally distributed populations - Independent observations - Random sampling from the population	t distribution (The test is also known as **Paired t-test**)

Test for One Proportion

Example

A researcher claims that Democratic party will win in the next United States Presidential election.

To test her belief the researcher randomly surveyed 90 people and 24 out of them said that they voted for Democratic party.

Is there enough evidence at $\boldsymbol{\alpha} = 0.05$ to support this claim?

This is clearly a one-tailed test, concerning population proportion p, the proportion of people voted from Democratic party.

Test for One Proportion

Significance of the test	Assumptions	Test Statistic Distribution
Test for population proportion $H_0 : p = p_0$	• Binomially distributed population • Random sampling from the population • When both mean (np) and n(1-p) are greater than or equal to 10, the binomial distribution can be approximated by a normal distribution	Standard Normal distribution (The test is also known as **One proportion z-test**)

Test for Two Proportions

Example

A car manufacturer aims to improve its products' quality by reducing the defects. So, the manufacturer randomly checks the efficiency of two assembly lines in the shop floor. In line 1, there are 20 defects out of 200 samples and In line 2, there are 25 defects out of 400 samples.

At 5% level of significance, do we have enough statistical evidence to conclude that the two assembly procedures are different?

This is clearly a one-tailed test, concerning two population proportion p_1 and p_2, the proportion of defects in assembly line 1 and assembly line 2 respectively.

Test for Two Proportions

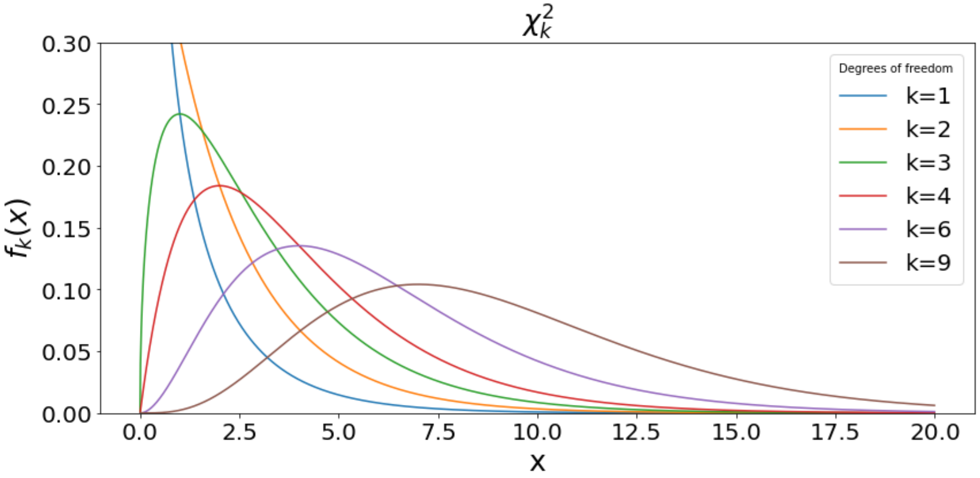
Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population proportions $H_0 : p_1 = p_2$	• Binomially distributed populations • Independent populations • Random sampling from the populations • When both mean (np) and n(1-p) are greater than or equal to 10, the binomial distribution can be approximated by a normal distribution	Standard Normal distribution (The test is also known as **Two proportions z-test**)

Test for One Variance

Test for Variance

Variance tests are used for a comparison of variability, often as a predecessor for other tests

Let us take many samples of the same size from a normal population and find the sample variances



They follow a **chi-square (χ^2) distribution**, which is dependent on the degrees of freedom

Example

It is conjectured that the standard deviation for the annual return of mid cap mutual funds is 22.4%, when all such funds are considered and over a long period of time. The sample standard deviation of a certain mid cap mutual fund based on a random sample of size 32 is observed to be 26.4%.

Do we have enough evidence to claim that the standard deviation of the chosen mutual fund is greater than the conjectured standard deviation for mid cap mutual funds at 0.05 level of significance?

This is clearly a one-tailed test, concerning population variance, the variance for mid cap mutual funds.

Test for One Variance

Significance of the test	Assumptions	Test Statistic Distribution
Test for population variance $H_0 : \sigma^2 = \sigma_0^2$	• Continuous data • Normally distributed population • Random sampling from the population	Chi Square distribution (The test is also known as **Chi-square test for variance**)

Test for Equality of Variances

Example

The variance of a process is an important quality of the process. A large variance implies that the process needs better control and there is opportunity to improve.

The data (Bags.csv) includes weights for two different sets of bags manufactured from two different machines. It is assumed that the weights for two sets of bags follow normal distribution.

Do we have enough statistical evidence at 5% significance level to conclude that there is a significant difference between the variances of the bag weights for the two machines.

This is clearly a two-tailed test, concerning two population variances, the variance for bag 1 weights and the variance for bag 2 weights.

Test for Equality of Variances

Significance of the test	Assumptions	Test Statistic Distribution
Test for equality of two population variances $H_0 : \sigma_1^2 = \sigma_2^2$	• Normally distributed populations • Independent populations • Larger variance should be placed in the numerator	F distribution (The test is also known as **F-test for variances**)

Test of Independence

Chi Square Test for Independence

2x2 contingency table that describes two variables (smoking and gender) at two levels each and stores the number of observations at each cell

	Male	Female	Total
Smoker	120	100	220
Non-smoker	60	140	200
Total	180	240	420

We are interested to know whether the **two variables are independent**

H_0: Smoking and gender are independent.

H_a: Smoking and gender are not independent.

Example

The following table summarizes beverage preference across different age-groups.

	Beverage Preference		
Age	Tea/Coffee	Soft Drink	Others
21 - 34	25	90	20
35 - 55	40	35	25
> 55	24	15	30

Does beverage preference depend on age?

This is a problem of Chi-Square test of independence, concerning the two independent categorical variables, Age and Beverage Preference.

Chi-Square Test for Independence

Significance of the test	Assumptions	Test Statistic Distribution
In a contingency table **H_0 : The row and column variables are independent**	• Categorical variables • Expected value of the number of sample observations in each level of the variable is at least 5 • Random sampling from the population	Chi Square distribution (The test is also known as **Chi-square test of independence**)

Analysis of Variance (ANOVA)

ANOVA Test : Introduction

Analysis **o**f **Va**riance (ANOVA) is used to determine whether the means of more than two independent populations are significantly different.

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$
H_a: at least one of these means if not the same

Why do we call it ANOVA? - The mathematical tools to calculate the p-value rely heavily on using the variances of the populations.

ANOVA is used in various problems such as comparing the yields of the crop from several varieties of seeds, comparing the gasoline mileage of various types of automobiles, etc.

ANOVA Test : Some important terms

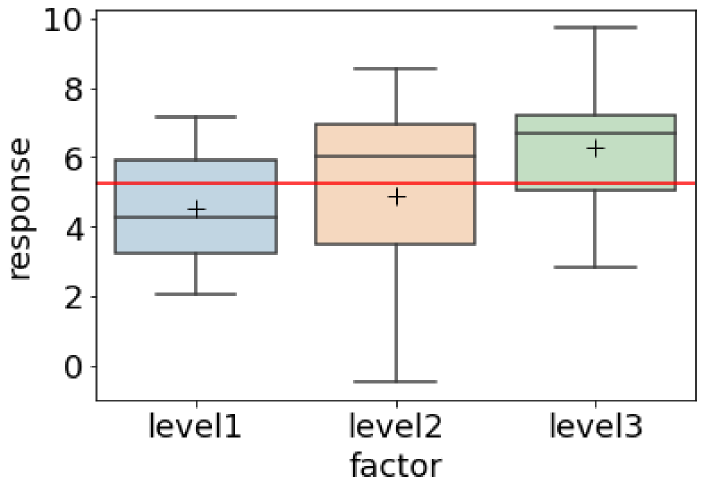
Response: Dependent variable which is continuous and assumed to follow a normal distribution

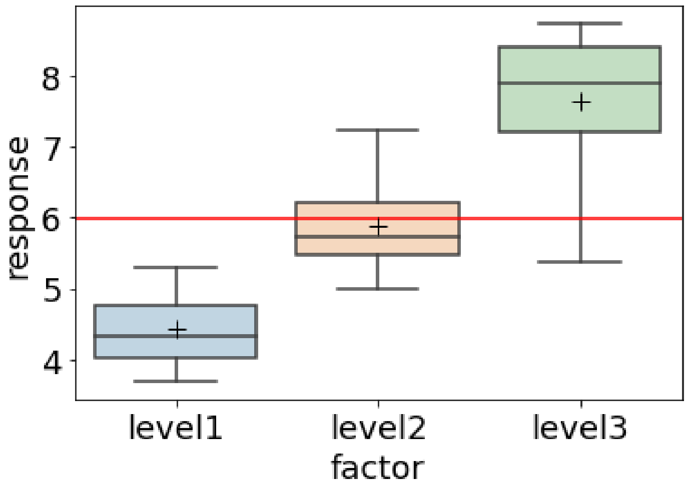
Consider, an example where interest lies in comparing the weekly volume of sales by different teams of sales executives.

Factor: Independent explanatory variable with several levels

ANOVA Test : One-way ANOVA

One-way ANOVA is used when the response variable depends on a single factor.





Between group variation is lower

Between group variation is higher

ANOVA Test : How it works

F-Statistic is the ratio of the between group variations to within group variations.

$$F - statistic = \frac{Between\ group\ variations}{Within\ group\ variations}$$

A large value of F-Statistic indicates that there is more variation between groups than within groups.

Thus, it will provide evidence against the null hypothesis.

Example

Traffic management inspector in a certain city wants to understand whether carbon emissions from different cars are different. The inspector has reasons to believe that Fuel type may be one important factor responsible for differences in carbon emission.

For this purpose, the inspector has taken random samples from all registered cars on the road in that city and would like to test if the amount of carbon emission release depends on fuel type at 5% significance level.

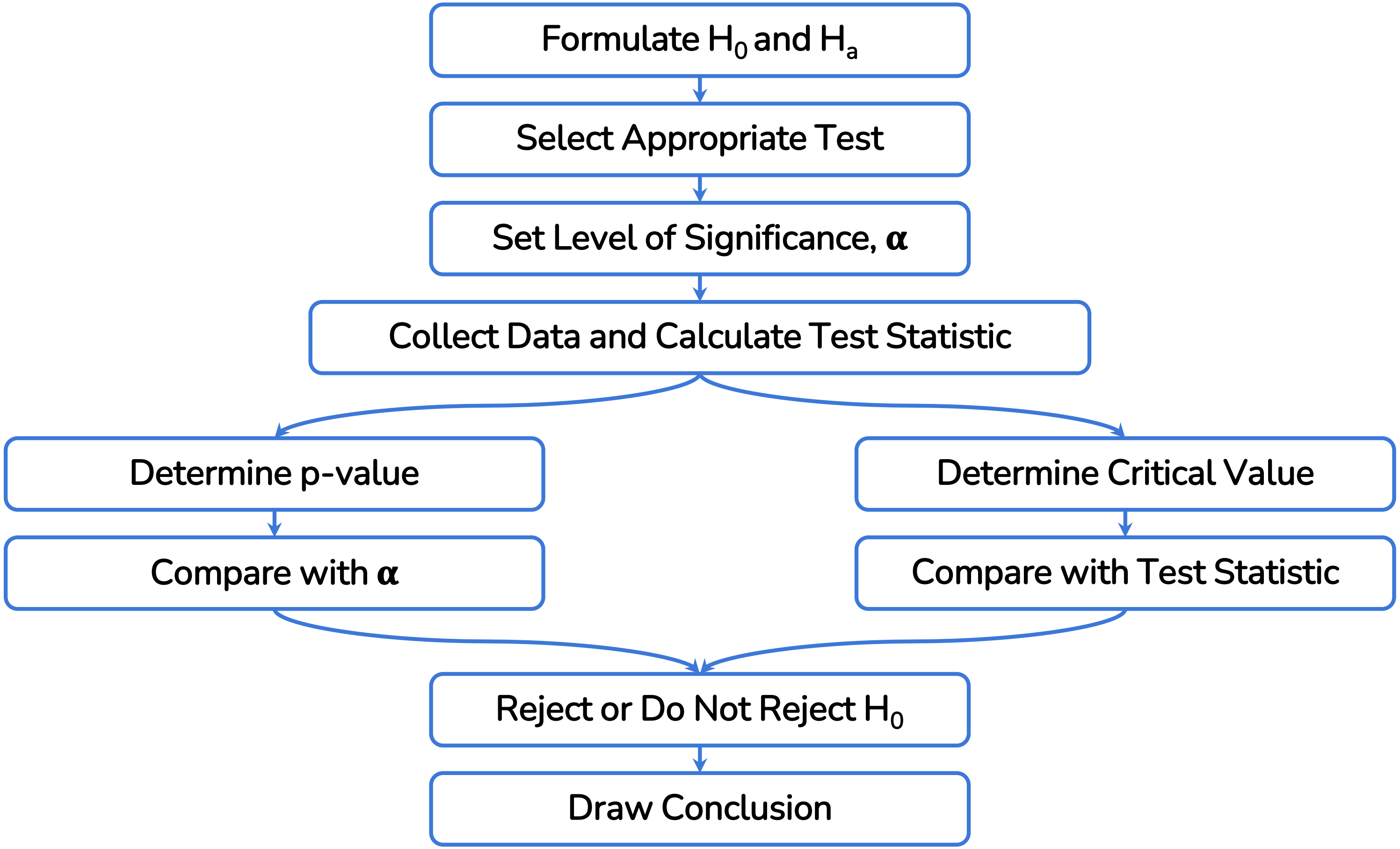
Here, we will compare the means of emission for the three different fuel types.

ANOVA Test : One-way ANOVA

Significance of the test	**Assumptions**	**Test Statistic Distribution**
Test for means for more than two populations **H_0 : All population means are equal**	• The populations are normally distributed • Samples are independent simple random samples • Population variances are equal	F distribution (The test is also known as **One-way ANOVA F-test**)

Now let’s summarize

Hypothesis Testing Steps



A flowchart to help you choose