manipalglobal
Academy of Data Science

Statistical Techniques for Data Science

Testing of Hypothesis

Introduction to Hypothesis Testing

Dr. Subhabaha Pal
Manipal Global Academy of Data Science

Objective

manipalglobal
Academy of Data Science

After attending this session, you will be able to –

- **Define Hypothesis testing**
- **Define Null Hypothesis, Alternative Hypothesis and Test Statistic**
- **Describe Type-I and Type-II errors**
- **Define α and β values**
- **Describe confidence level, Power of a test and P-value**
- **Describe Parametric and non-parametric test**

Hypothesis Testing

manipalglobal
Academy of Data Science

- A Statistical Hypothesis is an assumption about a population parameter – The assumption may or may not be true
- Hypothesis testing refers to the formal procedure used in statistics to accept or reject statistical hypotheses
- Generally the best way to determine whether a statistical hypothesis is true would be to examine the entire population
- Since examining the whole population is not often practical, researchers typically examine a random sample from the population to infer about some parameters about the population
- If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected

Hypothesis

manipalglobal
Academy of Data Science

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

Non-Statistical Hypothesis Testing - Example

- 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 H_0: The defendant is innocent
  - The alternative hypothesis H_1: The defendant is guilty
- The jury does not know which hypothesis is true as he was not present at the site of crime and he is supposed to make a decision on the basis of evidence presented by the prosecution (Police)
- In the judicial system, at the beginning, the jury first believes that the defendant is innocent
- The prosecutor needs to show substantial evidence against the defendant so that the latter can be proved guilty
- The main logic behind this system is for reducing the chance of innocent person getting penalized

Statistical Hypothesis Testing - Example

manipalglobal Academy of Data Science

- **A rented car company is trying to optimize cost and for that they are thinking of the option of opting for radial tyres in their fleet instead of the regular belted tyres in order to increase fuel efficiency**
- **Here, the following hypotheses need to be considered –**

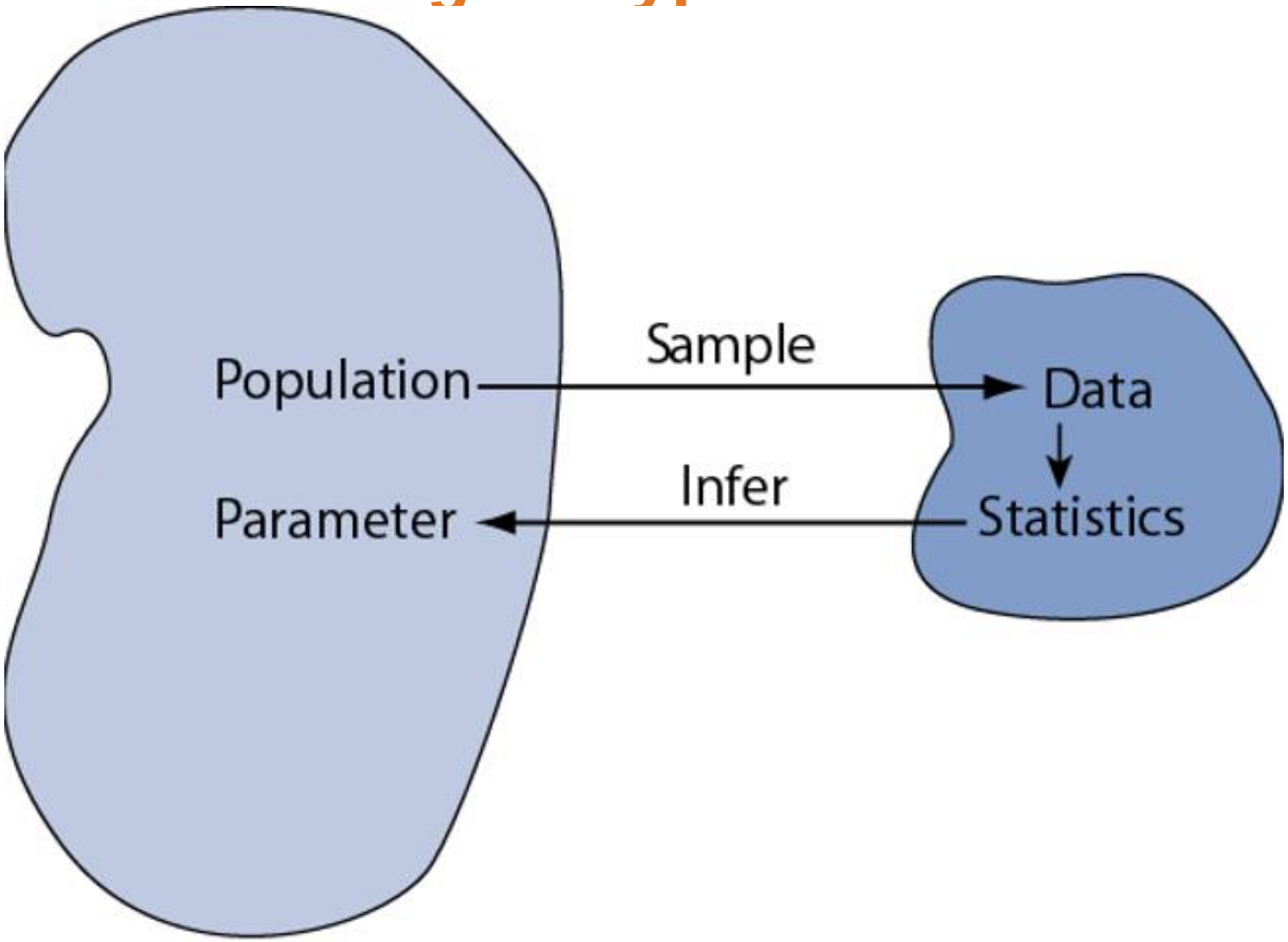
  Null Hypothesis - H_0 : The mean fuel consumption in cars fitted with radial tyres and regular belted tyres are the same, i.e., $H_0 : \mu_1 = \mu_2$
- **Alternative hypotheses can be multiple –**

  H_1 : The mean fuel consumption in cars fitted with radial tyres may be inferior to regular belted tyres, i.e., $H_1 : \mu_1 < \mu_2$

  H_1 : The mean fuel consumption in cars fitted with radial tyres may be better than regular belted tyres, i.e., $H_1 : \mu_1 > \mu_2$

  H_1 : The mean fuel consumption in cars fitted with radial tyres and regular belted tyres may be different, i.e., $H_1 : \mu_1 \neq \mu_2$

How Testing of Hypothesis works?



Statistical Hypothesis Testing - Example

manipalglobal
Academy of Data Science

- In the previous problem, the company can not measure the mean fuel consumption in all cars with both type of tyres
- It needs to select sample of cars having both type of tyres and then based on the results after calculating mean mileage in each case, the company needs to infer about the whole population of cars having radial tyres as well as regular belted tyres
- The decision that the company needs to take will be based on the sample only and the process they will be using in order to infer about the population from the sample data is statistical hypothesis testing
- Statistical hypothesis testing is a process involving rules which decide whether to accept the null hypothesis or not?
- The decision is always made based on the sample and not on the population and it leads to possibility of error between the decision made and reality
- We generally make decision in Statistical Hypothesis Testing based on test statistic value

Test Statistic

manipalglobal
Academy of Data Science

- A test statistic is a standardized value that is calculated from the sample data during a hypothesis testing process
- Test statistic is used to determine whether to reject a null hypothesis
- The test statistic compares the data with what is expected under the null hypothesis
- The test statistic is used to calculate the p-value
- A test statistic measures the degree of agreement between a sample data and the null hypothesis
- The observed value of test statistic changes randomly from one random sample to different sample
- A test statistic contains information about the data that is relevant for deciding whether to reject the null hypothesis
- The sampling distribution of the test statistic under the null hypothesis is called the null distribution
- When the data show strong evidence against the assumptions in the null hypothesis, the magnitude of the test statistic becomes too large or too small depending on the alternative hypothesis simultaneously causing test's p-value to become small enough to reject null hypothesis

Test Statistic Examples

manipalglobal
Academy of Data Science

- **Different test statistics are calculated in case of different tests**

Hypothesis test	Test statistic
Z-test	Z-statistic
t-tests	t-statistic
ANOVA	F-statistic
Chi-square tests	Chi-square statistic

- **These test statistics will be discussed later**

Errors in Testing of Hypothesis

manipalglobal
Academy of Data Science

- Since the whole population can not be examined and decision on the population parameter in the testing process is taken based on a small sample drawn from the whole population, there remains chance that population properties are not transferred to the sample and the decision taken based on the sample does not reflect true population property
- It is similar to the fact that in judicial system, judge or jury makes his judgement based on the evidence he has as he is not present at the time of actual crime
- Hence, if the evidences don't reflect the true scenario of crime, the judge is bound to make erroneous decision
- The next slide discusses in details the different types of errors possible in hypothesis testing process

Errors in Testing of Hypothesis

manipalglobal
Academy of Data Science

Jury example Non-Statistical example			Mileage 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

Jury example Non-Statistical example			Mileage 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

- **Type I Error – Incorrect rejection of the true hypothesis**
- **Type II Error – Incorrect acceptance of the false hypothesis**

Procedure of Hypothesis Testing

manipalglobal Academy of Data Science

- **Set up a Hypothesis – Set up null and alternative hypothesis**
- **Set up a suitable significance level – Need to select suitable level of significance – the confidence with which an experimenter rejects or retains null hypothesis depends on the significance level adopted – the level of significance, usually specified by 'α' , is generally specified before any samples are drawn – In practice, 5 percent level of significance is generally adopted which means there are about 5 chances out of 100 that null hypothesis will be rejected when the null hypothesis is true or it can be said that the experimenter is 95% confident that he has made the right decision – In general, one uses a level of significance of α =0.05, indicating that one is willing to accept a 5 percent chance of being wrong to reject H_0**
- **Determination of a Suitable Test Statistic – The 3rd step is the determination of a suitable test statistic and its distribution – the general form of test statistic is as follows –**

$$Test\ Statistic = \frac{Sample\ Statistic - Hypothesised\ Population\ Parameter}{Standard\ Error\ of\ the\ Sample\ Statistic}$$

- **Determination of the Critical Region – It is important to specify, before the sample is taken, that which values of the test statistic will lead to a rejection of H_0 and which will lead to acceptance of H_0 – The region whose values will lead to rejection is called the critical region – When the level of significance is 'α', the optimal critical region for a two-sided test consists of that 'α/2' percent of the area in the right-hand tail of the distribution plus that 'α/2' percent of the area in the left-hand tail – Establishing a critical region is similar to determining a 100(1- α)% confidence interval**

Procedure of Hypothesis Testing

- **Performing Computation – The fifth step in testing of hypothesis is the performance of various computations from a random sample of size n, necessary for the test statistic determined before – after that it is needed to see whether the sample results falls in the critical region or in the acceptance region**
- **Making Decision – A statistical decision or conclusion comprises of either accepting the null hypothesis or rejecting the null hypothesis – the decision will depend on whether the computed value of the test criterion falls in the region of rejection or the region of acceptance – if the hypothesis is tested at 5 percent level of significance and the observed set of results has a probability less than 5 percent, the null hypothesis is rejected and the difference between the sample statistic and the hypothetical population parameter is considered to be significant – if the test statistic falls in the region of non-rejection, the null hypothesis is accepted and the difference between the sample statistic and the hypothetical population parameter is not regarded as significant, i.e., it can be explained by the chance variations**

Hypothesis Testing

manipalglobal Academy of Data Science

- Probability of each decision outcome is a conditional probability and elements in the same column sum up to 1.0, since the events to which they are associated are complementary events
- $\alpha \; and \; \beta$ are not independent of each other, nor they are independent of sample size n
- When n is fixed, if α is lowered, then β normally rises and vice versa
- If n is increased, it is possible for both $\alpha \; and \; \beta$ to decrease

Jury example			Mileage example		
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

- Since, increasing the sample size involves money and time, therefore, it is important to decide how much additional money and time, someone is willing to spare on increasing the sample size in order to reduce the size of $\alpha \; and \; \beta$
- Power of a test is equal to (1 – Probability of Type II error) and is the probability of rejecting a hypothesis when the hypothesis is false
- It is not possible to decrease probabilities of Type I and Type II errors simultaneously and we fix probability of Type I error at a certain level in order to perform testing experiment
- Only way to simultaneously decrease type I and type II error is by increasing sample size n

P-Value

- P-value is the probability associated with the test statistic
- P-value or 'Calculated probability' is the probability of finding the observed, or more extreme, results when the null hypothesis (H_0) of a study is true
- If the test statistic is available, the p-value is the probability of getting observations equal to the test statistic as well as more extreme than the test statistic given the distribution of observations
- If p-value is less than 0.05 as obtained from the test statistic, it means that there is less than 5 percent chance of getting the value of the test statistic value as computed from the sample if the null hypothesis is true
- In case of p-value is less than 0.05 (or less than the chosen significance level), the null hypothesis is rejected and if it is more, then the null hypothesis is accepted

Parametric and Non-parametric Tests

manipalglobal
Academy of Data Science

- **Parametric statistical test is the one that makes assumptions about the parameters of the population from which the sample is drawn and the test is done to understand the correctness of the assumption**
- **Non-parametric test does not require any assumption about the parameter of the population**
- **The following tests are parametric tests - Student's T-test, F-test, Z-test and ANOVA**
- **The following tests are non-parametric tests – Chi-square test, Fisher's exact probabilities, Mann-Whitney U-test, Wilcoxon Signed Rank Test, Kolmogorov-Smirnov test, Kruskal-Wallis Test**
- **Non-parametric tests are simple and easy to understand**
- **Non-parametric test does not involve sampling theory** and no assumption is to be made regarding the parent population
- **Non-parametric test is generally applied for nominal scale data**

manipalglobal
Academy of Data Science