**A MINI PROJECT REPORT**

**ON**

**“Probability Distribution and Sampling Theory”**

Submitted in the partial fulfillment of the requirements for

The degree of

**BACHELOR of Engineering IN Computer Engineering**

**By**

1) 55\_adnan Shaikh

2) 38\_binitdev Pandey

3) 31\_KANCHAN MENGUNE

4)03\_ZEESHAN ANSARI

**UNDER THE GUIDANCE OF**

**Prof. Vasudeo Nazrikar**



Department of Computer Engineering  
Saraswati College of Engineering, Kharghar, Navi Mumbai  
University of Mumbai  
2020-21

**Saraswati College of Engineering, Kharghar**

**Vision:**

To be universally accepted as autonomous center of learning in Engineering Education and Research.

**Mission:**

* To educate students to become responsible and quality technocrats to fulfil society and industry needs.
* To nurture student’s creativity and skills for taking up challenges in all facets of life.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Department of Computer Engineering**

**Vision:**

To be among renowned institution in Computer Engineering Education and Research by developing globally competent graduates.

**Mission:**

* To produce quality Engineering graduates by imparting quality training, hands on experience and value education.
* To pursue research and new technologies in Computer Engineering and across interdisciplinary areas that extends the scope of Computer Engineering and benefit humanity.
* To provide stimulating learning ambience to enhance innovative ideas, problem solving ability, leadership qualities, team-spirit and ethical responsibilities.

****

**DEPARTMENT OF COMPUTER ENGINEERING**

**PROGRAM EDUCATIONAL OBJECTIVE’S**

1. To embed a strong foundation of Computer Engineering fundamentals to identify, solve, analyze and design real time engineering problems as a professional or entrepreneur for the benefit of society.
2. To motivate and prepare students for lifelong learning & research to manifest global competitiveness.
3. To equip students with communication, teamwork and leadership skills to accept challenges in all the facets of life ethically.

****

**DEPARTMENT OF COMPUTER ENGINEERING**

**PROGRAM OUTCOMES**

1. Apply the knowledge of Mathematics, Science and Engineering Fundamentals to solve complex Computer Engineering Problems.
2. Identify, formulate and analyze Computer Engineering Problems and derive conclusion using First Principle of Mathematics, Engineering Science and Computer Science.
3. Investigate Complex Computer Engineering problems to find appropriate solution leading to valid conclusion.
4. Design a software System, components, Process to meet specified needs with appropriate attention to health and Safety Standards, Environmental and Societal Considerations.
5. Create, select and apply appropriate techniques, resources and advance Engineering software to analyze tools and design for Computer Engineering Problems.
6. Understand the Impact of Computer Engineering solution on society and environment for Sustainable development.
7. Understand Societal, health, Safety, cultural, Legal issues and Responsibilities relevant to Engineering Profession.
8. Apply Professional ethics, accountability and equity in Engineering Profession.
9. Work Effectively as a member and leader in multidisciplinary team for a common goal.
10. Communicate effectively within a Profession and Society at large.
11. Appropriately incorporate principles of Management and Finance in one’s own Work.
12. Identify educational needs and engage in lifelong learning in a Changing World of Technology.

****

**DEPARTMENT OF COMPUTER ENGINEERING**

**PROGRAMME SPECIFIC OUTCOME**

1. Formulate and analyze complex engineering problems in computer engineering (Networking/Big data/ Intelligent Systems/Cloud Computing/Real time systems).
2. Plan and develop efficient, reliable, secure and customized application software using cost effective emerging software tools ethically.



**(Approved by AICTE, recg. By Maharashtra Govt. DTE ,Affiliated to Mumbai University)**

**PLOT NO. 46/46A, SECTOR NO 5, BEHIND MSEB SUBSTATION, KHARGHAR, NAVI MUMBAI-410210**

**Tel. : 022-27743706 to 11 \* Fax : 022-27743712 \* Website: www.sce.edu.in**

**CERTIFICATE**

*This is to certify that the requirements for the mini project report entitled ”****Probability Distribution And Sampling Theory****” have been successfully completed by the following students:*

1) 55\_adnan Shaikh

2) 38\_binitdev Pandey

3) 31\_KANCHAN MENGUNE

4)03\_ZEESHAN ANSARI

In partial fulfillment of Sem –IV , **Bachelor of Engineering of Mumbai University in Computer Engineering** of Saraswati college of Engineering , Kharghar during the academic year 2020-21.

**Internal Guide**  **External Examiner**

Prof. Vasudeo Nazirkar

**Mini Project Co-ordinator**  **Headof Department**

Prof. Monali DeshmukhProf. Sujata Bhairnallykar

**DECLARATION**

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included. I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

1) 55\_adnan Shaikh

2) 38\_binitdev Pandey

3) 31\_KANCHAN MENGUNE

4)03\_ZEESHAN ANSARI

Date:

**ACKNOWLEDGEMENT**

After the completion of this work, words are not enough to express feelings about all those who helped us to reach goal.

It’s a great pleasure and moment of immense satisfaction for us to express my profound gratitude to Prof. Vasudeo Nazirkar , whose constant encouragement enabled us to work enthusiastically. His perpetual motivation, patience and excellent expertise in discussion during progress of the project work have benefited us to an extent, which is beyond expression.

We would also like to give our sincere thanks to **Prof.Sujata Bhairnallykar, Head of Department**, and **Prof.Monali Deshmukh ,Mini Project co-coordinator** from Department of Computer Engineering, Saraswati college of Engineering, Kharghar, Navi Mumbai, for their guidance, encouragement and support during a project.

I am thankful to **Dr. ManjushaDeshmukh, Principal,**Saraswati College of Engineering, Kharghar, Navi Mumbai for providing an outstanding academic environment, also for providing the adequate facilities.

Last but not the least we would also like to thank all the staffs of Saraswati college of Engineering (Computer Engineering Department) for their valuable guidance with their interest and valuable suggestions brightened

1) 55\_adnan Shaikh

2) 38\_binitdev Pandey

3) 31\_KANCHAN MENGUNE

4)03\_ZEESHAN ANSARI

# 

# List of Figures

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Name** | **Page No.** |
| 1 | Introduction | 11 |
| 2 | Simple sampling of attributes | 12 |
| 3 | Test of significance of large samples | 13-14 |
| 4 | Comparison of large samples | 15-16 |
| 5 | Test of significance for means of two large samples | 17-20 |
| 6 | Applications of large sampling | 21 |
| 7 | References | 22 |

**Introduction:**

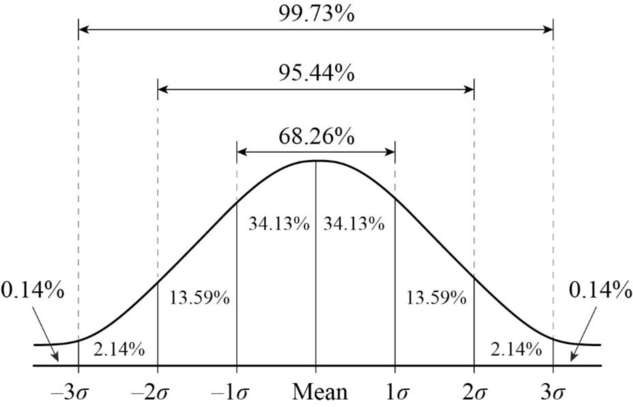
* A sample in a statistics is a small section selected from a large population.
* It is essential that a sample must be a random selection so that each member of the population has the same chance of being included in the sample. Thus the fundamental assumption underlying theory of sampling is Random sampling.
* A special case of random sampling in which each event has the same probability of success and the chance of success of different events are independent whether previous trials have been made or not, is known as simple sampling.
* If a sample quantity is greater thirty then it is consider to be a large sample.
* A large sample is assumed to be normal.
* By testing our Hypothesis for a large sample we can predict many things about the original population

**SIMPLE SAMPLING OF ATTRIBUTES**

* The sampling of attributes may be regarded as the selection of samples from a population whose members possess the attribute K or not K. The presence of K may be called a success and its absence a failure. Suppose we draw a simple sample of n items. Clearly it is same as a series of n independent trials with the same probability p of success. The probabilities of 0,1, 2,..., n successes are the terms in the binomial expansion of (q + p)n where q = 1 - p.
* Mean of distribution is np and Standard deviation is i.e., the expected value of success in a sample of size n is np and the standard error is .
* If we consider the proportion of success, then
* (*i*) mean proportion of successes = np/n = p
* (*ii*) standard error of the proportion of success =
* (*iii*) precision of the proportion of success =

**TEST OF SIGNIFICANCE OF LARGE SAMPLES**

* For a normal distribution, only 5% of the members lie outside while only 1% of the members lie outside
* If x be the observed number of success in the sample and z is the standard normal variate then .
* Thus we have the following test of significance:
* (i) If |z| < 1.96, then difference between the observed and expected number of success is not significant.
* (ii) If |z| > 1.96, then difference is significant at 5% level of significance.
* (iii) If |z| > 2.58, then difference is significant at 1% level of significance.

****

**Example:**

Q.) A die was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing, do the data indicate an unbiased die?

Soln: Let us assume the die is unbiased

Then the probability of throwing 5 or 6 with one die =

+

probability of failure =

Expected number of success = --------(n = 9000, Given)

Observed number of success = -------------(Given)

Standard deviation =

Hence =

As z > 2.58, the hypothesis has to be rejected at 1% level of significance and we conclude that the dice is biased.

**COMPARISON OF LARGE SAMPLES**

Two large sample of sizes n1 and n2 are taken from two populations giving proportions of attributes A’s as p1,p2 respectively.

1. On the hypothesis that the populations are similar as regards the attribute A, we combine the two samples to find an estimate of the common value of proportion of A’s in the populations which is given by

If e1,e2 be the standard errors in the two samples then

e12 = and e22 =

If e be the standard error of the difference between p1 and p2, then

e2  = e12 + e22  =

Therefore,

If z>3, the difference between p1 and p2 is real one.

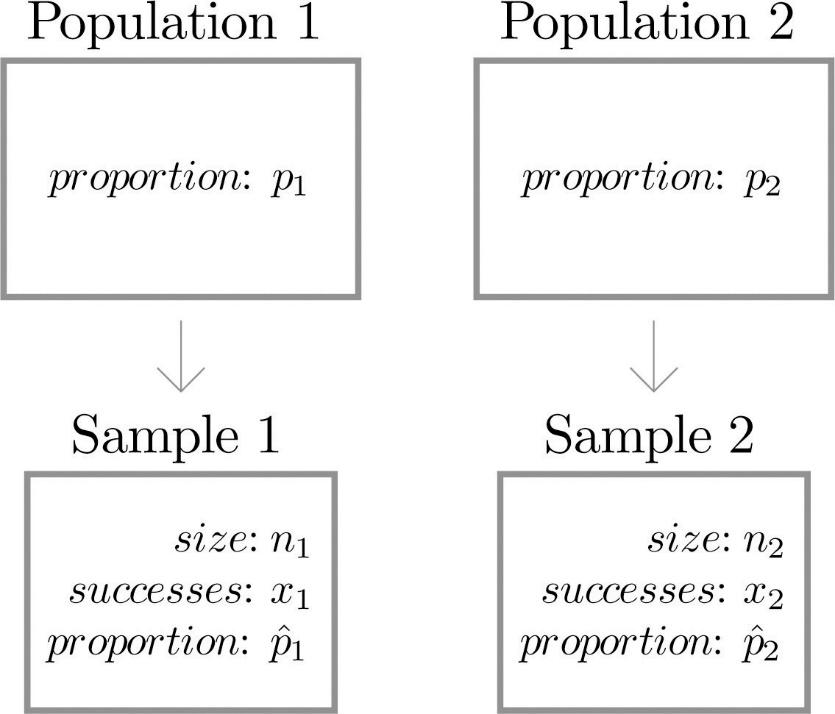
If z<2, the difference may be due to fluctuations of simple sampling.

But if z lies between 2 and 3, then the difference is significant at 5% level of significance.

(b) If the proportions of A’s are not the same in the two populations from which the samples are drawn, but p1 and p2 are true values of proportions then S.E. e of the difference p1-p2 is given by

e2 =

If < 3, the difference could have arisen due to fluctuations of simple sampling.



**TEST OF SIGNIFICANCE FOR MEANS OF TWO LARGE SAMPLES**

1. Suppose two random samples of sizes n1 and n2 have been drawn from the same population with S.D. . We wish to test whether the difference between the sample means x1 and x2 is significant or is merely due to fluctuations of sampling.

If the samples are independent, then the standard error e of the difference of their means is given by

e2  = e12 + e22

Where e1 = , e2 = are the S.E.s of the means of the two samples.

Therefore, e = . Hence

is normally distributed with mean zero and S.D. 1.

Test of significance (n1, n2 being large):

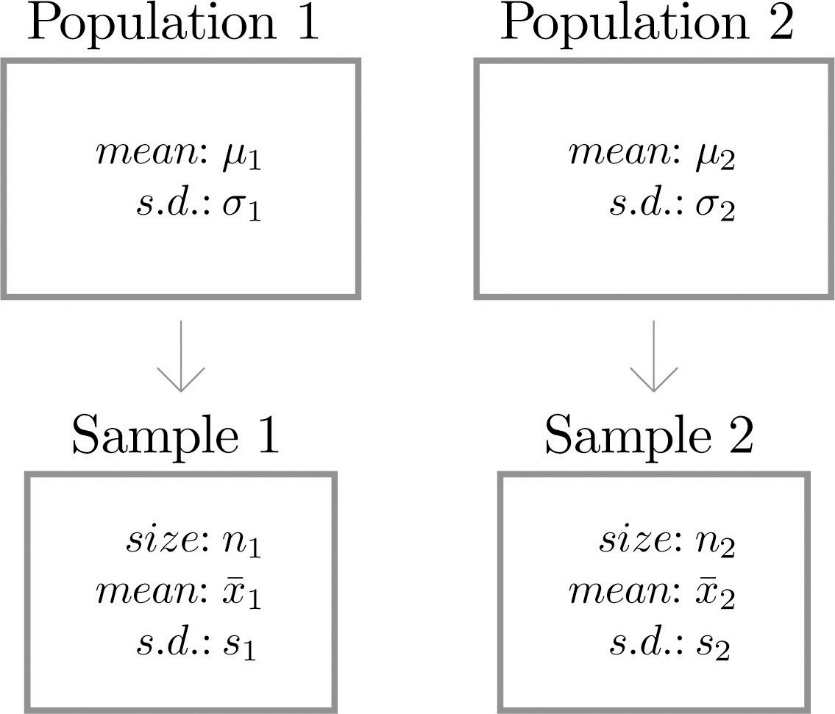
If z > 1.96, then the difference is significant at 5% level of significance.

If z > 3, it is highly probable that either the samples have not been drawn from the same population or the sampling is not

(b) If the samples are known to be drawn from different populations with means , 2 and standard deviations and 2. Then the standard error e of their means is given by

e =

Assuming that the two populations have the same mean (i.e., = ), the difference of the means of the samples will be normally distributed with mean zero and S.D. e. Now the same procedure of test of significance is applied.



Examples:

Q1) In a city A 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of a random sample of1600 school boys had the same defect Is the difference between the proportions significant ?

Soln: We have n1 = 900 and n2 = 1600

and p1 = , p2 =

Therefore, = = 0.19

and q = 1 – p = 0.81

Thus e2  = = 0.19 0.81 () = 0.0017

giving e = 0.04 nearly

Also p1 – p2 = 0.015. Therefore, = = 0.37

As z < 1, the difference between the proportions is not significant.

Q2) The means of simple samples of sizes 1000 and 2000 are 67.5 and 68.0 cm respectively. Can the samples be regarded as drawn from the same population of S.D. 2.5 cm.

Soln: We have x1 = 67.5, x2 = 68.0

n1 = 1000, n2 = 2000

On the hypothesis, that the samples are drawn from the same population of S.D. = 2.5, we get

=

Hence the difference between the sample means i.e., 5.1 is very much greater than 1.96 and is therefore significant. Thus, the samples cannot be regarded as drawn from the same population.

**Applications Of Large Sampling**

* Sampling mostly use in research by selecting section of data (sample) which must be a representative of population and analyzing this sample and applying the result to original population, this reduces the inspection cost.
* Sampling have many applications in Data Science.
* Designing of Web Crawlers which optimizes search engine of web browser.
* In Marketing Research to check sale of product in different areas and relation of sale in two different areas.
* In Machine learning to analyze large data.
* Weather forecasting uses sampling theory to predict the weather.
* Predicting rate of some disease in a Large population.

References:

1. Higher Engineering Mathematics by

**Dr. B.S. Grewal** (for Questions and Introduction)

2.<https://www.quora.com/>

(for Applications)