

Probability & Statistics (MS-301) Software Engineering

Lecture # 1 (Statistics)

Prepared by Dr. Asma Zaffar

سُبْحَانَكَ يَا أَرْحَمَ الرَّاحِمِينَ

Sir Syed University of Engineering & Technology

Department of Software Engineering

COURSE INFORMATION SHEET

Course Code	:	MS-301
Course Title	:	Probability and Statistics
Credit Hours	:	3+0
Prerequisites	:	MS-103: Calculus & Analytical Geometry
Instructor Name	:	Dr. Asma Zaffar
Email & Contact Info.	:	aszafar@ssuet.edu.pk ; Ext: 222

COURSE OUTLINE :

Descriptive Statistics: Basic definitions, measures of central tendency and variation, z-scores, frequency distribution, graphical representation of data, symmetry and skewness, quintiles (percentiles, deciles and quartiles).

Regression and correlation: Linear regression models and correlation coefficient.

Probability theory: Basic definitions and rules of probability, conditional probability, counting techniques.

Random variable: Concept of random variable, discrete and continuous random variable.

Discrete and continuous probability distributions: Uniform, binomial, multinomial, negative binomial, geometric, Poisson, Normal; normal approximation and exponential distributions and their applications.

Statistical inference: Estimation of parameters such as mean and variance.

Sampling theory: Sampling distribution of mean, t-distribution and sampling procedures.

Hypothesis testing: z-test, t-test and goodness of fit test.

COURSE LEARNING OUTCOMES (CLOs):

Upon completion of this course, students will be able to:

CLO #	CLO Statement	Bloom's Taxonomy (Maximum level)
CLO-1	Understanding the concept of group and ungroup descriptive statistics, correlation and regression. Analyze and apply these concepts on real life problems	C4 (Analyzing)
CLO-2	Understanding the concept of probability, random variables, probability distribution functions (discrete and continuous) and probability distributions (discrete and continuous). Evaluating and applying these concepts to real life problems	C5 (Evaluating)
CLO-3	Compute hypothesis, sampling and testing for understanding real life and engineering problems	C3 (Applying)

COURSE PLAN



Week #	Date	Course plan	Recommended Reading	Assessment Tools
Week 1		Introduction to Statistics: Basic definitions, Classification of data, frequency distribution.		
		Construction of frequency distribution table	Notes will be provided	
		Graphical Representation: Histogram, Bar chart, Multiple bar chart, Polygon, Ogive, Pie chart, Stem and Leaf, and dot plot	*1.2 Page 24	
Week 2		Measures of Central Tendency (Group and Ungroup Data): Mean, Mode, Median (ungroup data)	*1.3 Page 34	
		Mean, Mode, Median (group data)	1.20-1.22 Page 31	
		Measures of Variation: Variance, Standard Deviation	1.13-1.17 Page 30-31	
Week 3		Deviation about mean	Notes will be provided	
		Deviation about mean Median	Notes will be provided	Assignment 1

		Quintiles (Quartiles, Deciles and Percentiles)	continued	
Week 4		Regression and Correlation: Linear Regression and Scatter diagram, Regression line.	11.1-11.14 Page 398-399	Quiz 1
			continued	
		Correlation coefficient.	11.43,11.44 Page 435	
Week 5		Counting Techniques:	2.21-2.47 Page 51-52	
		Multiplication rule, factorial, arrangements with and without replacements	continued	
		Permutation and Combination.	continued	
Week 6		Probability Theory: Basic definition, axioms, Sample space, Events, types of events, relationship between events.	2.1-2.10 Page 42-43	
		Rules of Probability: Additive rule, Conditional Probability, Multiplication rule, Baye's rule	2.49-2.69 Page 59-61	
		Continued...	2.73-2.94 Page 70-71	
Week 7		Continued...	continued	Quiz 2
		Random Variable: Concept of random variable, discrete and continuous random variable	3.1-3.4 Page 91	

		and continuous random variable		
		Discrete Probability Distributions functions	3.5-3.18 Page 91-93	
Week 8	MID TERM EXAMINATION			
Week 9		Continuous Probability Distributions functions	continued	
		Cumulative probability distribution (discrete and continuous)	3.19-3.36 Page 93-94	
		Joint Probability Distributions (discrete)	3.37-3.60 Page 104-105	
Week 10		Joint Probability Distributions (continuous)	continued	Assignment 2
		Marginal and Conditional Probability Distributions	continued	
		Statistical Independence	continued	
Week 11		Discrete Probability Distributions: Uniform Distribution	5.1 Page 150	
		Binomial Distribution	5.2-5.18 Page 150-151	
		Negative Binomial Distribution	5.49-5.55 Page 164-165	

Week 12		Geometric Distributions.	continued	
		Multinomial Distribution	5.19-5.23 Page 151-152	Quiz 3
		Poisson Distribution	5.56-5.68 Page 165	
Week 13		Continuous Probability Distributions: Normal Distribution	6.4-6.17 Page 186-187	
		Continued...	continued	
		Continued...	continued	
Week 14		Normal approximation to Binomial and its applications	6.24-6.38 Page 870 & 878	
		Continued...	continued	Assignment 3
		Continued...	continued	
Week 15		Sampling Theory: Sampling Distribution, Sampling Distribution of Mean, t- Distribution	8.17-8.30 Page 241-242	
		Sampling Distribution of Mean	continued	
		t- Distribution	continued	
Week 16		Hypothesis Testing: z-test, t-test and goodness of fit	Notes will be provided	
		Continued...	12.5 Page 887-888	
		Review		
FINAL EXAMINATION				

WHAT IS STATISTICS?

1.1 ORIGIN AND INTRODUCTION OF STATISTICS

The word Statistics has been derived from the Latin word "Status" or an Italian word "Statistia" both of these words mean a political state. In old days kings and queen employed a person known as a statistician to maintain the record of their wealth and ornaments. With the passage of time this word changes its shape and is now being used as "Statistics".

1.2 WHAT STATISTICS IS?

In statistics first of all we collect the data regarding the field of our interest. Then to derive some conclusion from the collected data certain statistical methods are applied.

1.2.1 MEANING OF STATISTICS

Statistics is a social science and different statisticians use it in different meanings. Nowadays the word "Statistics" is used to convey three different meanings:-

1.3 DEFINITION OF STATISTICS

Since statistics is a social science, therefore, it does not have a unique definition. Every one have defined it in his own way.

Some says statistics is a science of counting. An other is of the views that statistics deals with estimates and probabilities. Yet an other statistician says that in statistics we play with figures. All those definitions have their own importance but they donot cover the entire scope of statistics.

After going through the above definition a complete definition given by Horace, Secrist:

"Statistics are the aggregate of facts, affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according a reasonable standard of accuracy, collected in systematic manner for a pre-determined purpose and place in comparison".

DATA

Ungroup
data/
Raw data

2, 4, 8, 9, 10,
11, 13, ...

Group data
data consist of in two
columns

	0-4	5-8	9-12	13-16	17-20
frequency $\leftarrow f$	5	6	7	9	

	C-I	2-4	5-7	8-10	11-13
class interval	f	8	10	11	

PRESENTATION OF DATA

In the last chapter, we started to discuss the stages of statistical enquiries. We have already discussed the first stage of statistical enquiry i.e. collection of data either from primary or secondary source in previous chapter. The next stage of a statistical enquiry is presentation of collected data (raw data) in a way that it will:

- (i) Point out the most significant features of the data at a glance;
- (ii) Condense the mass of data in such a manner that similarities and dissimilarities can easily be cleared;
- (iii) Facilitate comparison; and
- (iv) Provide basis for further statistical analysis.

To achieve the above mentioned objectives, we usually use the following two methods for presentation of data:

1. Classification and tabulation,
2. Graphical or diagrammatic presentation of data.

CLASSIFICATION AND TABULATION

The two separate heading "Classification" and "Tabulation" should not lead the readers to assume that these are two distinct processes. Infact, they go together, classification is the first step in tabulation. Before the data are put in tabular form they have to be classified in different classes or group having common characteristics. After this step the data are displayed under different columns and rows so that their relationship can be easily understood. For example data given in the statement of example 1 is classification of data whereas the procedure of placing given facts in form of table like that given in the solution, is tabulation.

Discrete Frequency Distribution

Class Number	No. of Servants (Variable)	No. of Families (Frequency)
1	0	39
2	1	20
3	2	6
4	3	4
5	4	1
Total	—	70

Continuous Frequency Distribution

Class Number	Height (inches) (Variable)	No. of Students (Frequency)
1	60 – 63	10
2	63 – 66	16
3	66 – 69	14
4	69 – 72	7
5	72 – 75	3
Total	—	50

Questions?

