STA 124: INTRODUCTION TO PROBABILITY DISTRIBUTION

University of Ilorin
Centre for Open & Distance Learning
CODL

Published by the Centre for Open and Distance Learning,
University of Ilorin, Nigeria

E-mail: codl@unilorin.edu.ng
Website: https://codl.unilorin.edu.ng

Course Development Team

Subject Matter Expert

Dr. N. A. IKOBA
Department of Statistics
University of Ilorin, Nigeria

Instructional Designers

Olawale Koledafe
Center for Open and Distance (CODL)
University of Ilorin, Nigeria

Mathew Kayode Owolabi
Department of Educational Technology,
University of Ilorin, Nigeria

Samuel Adeleke Adekanye
Department of Educational Technology,
University of Ilorin, Nigeria

Aloba Samuel Olasupo

Language Editors

Bankole Ogechi Ijeoma
Center for Open and Distance (CODL)
University of Ilorin, Nigeria

Abdulwahab Mahmud

From the Vice Chancellor

Courseware development for instructional use by the Centre for Open and Distance Learning (CODL) has been achieved through the dedication of authors and the team involved in quality assurance based on the core values of the University of Ilorin. The availability, relevance and use of the courseware cannot be timelier than now that the whole world has to bring online education to the front burner. A necessary equipping for addressing some of the weaknesses of regular classroom teaching and learning has thus been achieved in this effort.

This basic course material is available in different electronic modes to ease access and use for the students. They are available on the University's website for download to students and others who have interest in learning from the contents. This is UNILORIN CODL's way of extending knowledge and promoting skills acquisition as open source to those who are interested. As expected, graduates of the University of Ilorin are equipped with requisite skills and competencies for excellence in life. That same expectation applies to all users of these learning materials.

Needless to say, that availability and delivery of the courseware to achieve expected CODL goals are of essence. Ultimate attention is paid to quality and excellence in these complementary processes of teaching and learning. Students are confident that they have the best available to them in every sense.

It is hoped that students will make the best use of these valuable course materials.

Professor S. A. Abdulkareem
Vice Chancellor

Foreword

Courseware is the livewire of Open and Distance Learning. Whereas some institutions and tutors depend entirely on Open Educational Resources (OER), CODL at the University of Ilorin considered it necessary to develop its materials. Rich as OERs are and widely as they are deployed for supporting online education, adding to them in content and quality by individuals and institutions guarantees progress.

Pursuing this goal has brought the best out of the Course Development Team across Faculties in the University. Despite giving attention to competing assignments within their work setting, the team has created time and eventually delivered. The development of the courseware is similar in many ways to the experience of a pregnant mother eagerly looking forward to the delivery date.

As with the eagerness for a coming baby, great expectation pervaded the air from the University Administration, CODL, Faculty and the writers themselves. Careful attention to quality standards, ODL compliance and UNILORIN CODL House Style brought the best out from the course development team. Response to quality assurance with respect to writing, subject matter content, language and instructional design by the authors, reviewers, editors and designers, though painstaking, has yielded the course materials now made available primarily to CODL students as open resources.

Aiming at parity of standards and esteem with regular university programmes is usually an expectation from students on open and distance education programmes. The reason being that stakeholders hold the view that graduates of face-to-face teaching and learning are superior to those exposed to online education. CODL has the dual mode mandate. This implies a combination of face-to-face with open and distance education.

With this in mind, the Centre has developed its courseware to combine the strength of both modes to bring out the best from the students. CODL students and other categories of students of the University of Ilorin and similar institutions will find the courseware to be their most dependable companion for the acquisition of knowledge, skills and competences in the respective courses and programmes.

Activities, assessments, assignments, exercises, reports, discussions and projects at various points in the courseware are targeted at achieving the Outcomes of teaching ang learning. The courseware is interactive and directly points the attention of students and users to key issues helpful to their particular learning. The student's understanding has been viewed as a necessary ingredient at every point. Each course has also been broken into modules and their component units in an ordered sequence. In it all, developers look forward to successful completion by CODL students.

Courseware for the Bachelor of Science in Computer Science housed primarily in the Faculty of Communication and Information Science provide the foundational model for Open and Distance Learning in the Centre for Open and Distance Learning at the University of Ilorin.

Henry O. Owolabi
Director, CODL

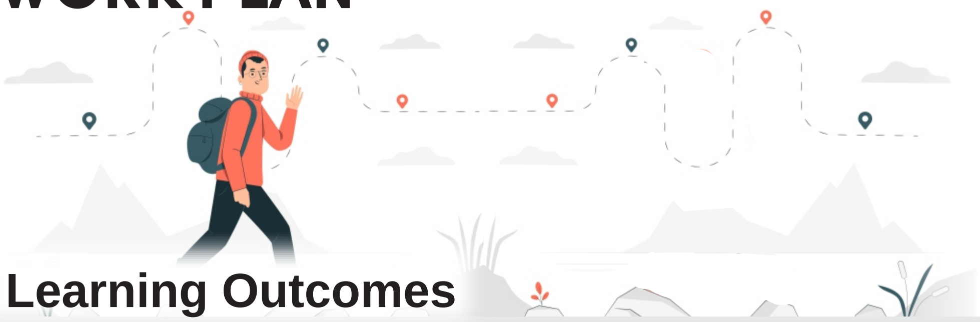
INTRODUCTION

The basic concept of a random variable shall be considered in detail, as well as the types of random variables. The course introduces the student to the concepts of probability: experiment, outcome, trial, chance, event and sample space, Expectations, variance, distribution functions; bivariate distribution: marginal distribution, conditional distribution. Some common probability distributions: Bernoulli, Binomial, Poisson, Geometric, Uniform and Normal distributions are also considered. Simple linear regression is also introduced, together with regression concepts: correlation (Pearson correlation and Rank correlation); and association of attributes.

Course Goal

The course is designed to introduce students in the Mathematical Sciences, Physical Sciences, Computer Science and Engineering to the application of statistics in their disciplines. The course is to intimate the students with the usefulness of statistics in their various fields of study and enlighten the student on the importance of statistics in carrying out meaningful researches in their various areas of study. It is also to develop in students, the ability to apply their knowledge and skills to the solution of theoretical and practical problems in Statistics.

WORK PLAN



Learning Outcomes

At the end of this course, you should be able to:

i. Identify different random variables, and how to generate them;
ii. Explain the concepts of probability;
iii. Analyse data using the concepts of bivariate distribution, marginal and conditional distributions;

Week 01 / Week 02

iv. Identify different probability distributions;
v. Fit a simple linear regression model;
vi. Analyse an event using the concept of correlation; and
vii. Explain the concept of association of attribute.

Week 03

Course Guide

Module 1

Introductory concepts

Unit 1 - Sample space, events and random variable

Unit 2 - Functions of a Random variable: probability mass function, probability density function, cumulative distribution function Technology

Module 2

Moments of a Random Variable

Unit 1 - Mathematical expectations of a random variable

Unit 2 - Mathematical expectations of a random variable Module

Module 3

Discrete Probability Distributions

Unit 1 - Bernoulli distribution, Binomial distribution

Unit 2 - Geometric distribution

Unit 3 - Poisson distribution

Course Infomation

This is a compulsory course for students in the Departments of Chemistry, Industrial Chemistry, Geology, Mathematics, Computer Science. You are expected to participate in all the course activities and have minimum of 75% attendance to be able to write the final examination

Pre-requisite | NIL

II. explain the internet architecture and its protocols;

Week 03

Module 4

Continuous Probability Distributions

Unit 1 - Uniform distribution

Unit 2 - Normal distribution

Module 5

Concepts of Bivariate Distributions

Unit 1 - Bivariate probability distribution

Unit 2 - Marginal and conditional distribution of bivariate distribution

Module 6

Mathematical Relationships Between Bivariate Distributions

Unit 1 - Linear regression

Unit 2 - Pearson correlation

Unit 3 - Rank correlation

Unit 4 - Association of attributes

Course Requirements

Requirements for success

The CODL Programme is designed for learners who are absent from the lecturer in time and space. Therefore, you should refer to your Student Handbook, available on the website and in hard copy form, to get information on the procedure of distance/e-learning. You can contact the CODL helpdesk which is available 24/7 for every of your enquiry.

Visit CODL virtual classroom on http://codllms.unilorin.edu.ng. Then, log in with your credentials and click on CSC 111. Download and read through the unit of instruction for each week before the scheduled time of interaction with the course tutor/facilitator. You should also download and watch the relevant video and listen to the podcast so that you will understand and follow the course facilitator.

At the scheduled time, you are expected to log in to the classroom for interaction.

Self-assessment component of the courseware is available as exercises to help you learn and master the content you have gone through.

You are to answer the Tutor Marked Assignment (TMA) for each unit and submit for assessment.

Embedded Support Devices

Support menus for guide and references

Throughout your interaction with this course material, you will notice some set of icons used for easier navigation of this course materials. We advise that you familiarize yourself with each of these icons as they will help you in no small ways in achieving success and easy completion of this course. Find in the table below, the complete icon set and their meaning.

Introduction	Learning Outcomes	Main Content

Summary	Tutor Marked Assignment	Self Assessment
Web Resources	Downloadable Resources	Discuss with Colleagues
References	Futher Reading	Self Exploration

Grading and Assessment

TMA	CA	Exam	Total
20%	20%	60%	100%

Picture: Statistical Analysis

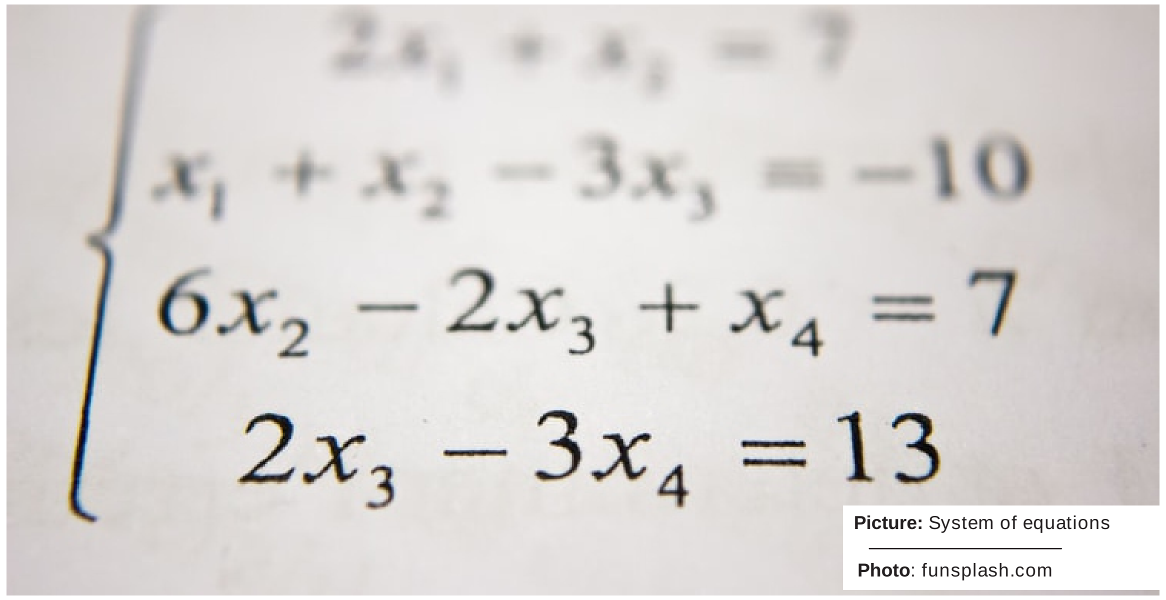
Photo: pexels.com

Introductory Concepts

Units

Unit 1 - Sample space, events and random variable

Unit 2 - Functions of a Random variable: probability mass function, probability density function, cumulative distribution function Technology



Picture: System of equations

Photo: funsplash.com

UNIT 1

Sample Space, Events and Random Variable

Introduction

I welcome you to the first unit of the course, STA 124 – Introduction to Probability Distribution. In this unit, you will explore the basic concepts of sample space, event, probability experiments and random variable. Throughout your journey in this course, you will be guided by examples of some simple experiments that demonstrate the concepts and enable you to identify whether a random variable is discrete or continuous, depending on the nature of the sample space.

Learning Outcomes

At the end of this unit, you should be able to:

1. Define an outcome of a probability experiment;
2. Define an event;
3. Obtain the sample space for some simple experiments like tossing a die, tossing a coin, tossing two or more coins, and tossing two dice.
4. Distinguish between a discrete random variable and a continuous random variable;
5. Give some everyday examples of random variables.

Main Content

Do you know that the theory of probability grew out of the study of various games of chance using coins, dice, and card? Since these devices lend themselves well to the application of concepts of probability, they will be used extensively in describing the basic concepts.

Basic Concepts

10 mins

A probability experiment is a chance process that leads to well defined results called outcomes.

The process whereby you flip a coin, roll a die, or draw a card from a deck are called probability experiments.

An outcome is the result of a single trial of a probability experiment. A trial means flipping a coin once, rolling one die once, or the like. When you toss a coin, there are two possible outcomes: head or tail. In the roll of a single die, there are six possible outcomes: 1, 2, 3, 4, 5, or 6. In any experiment, the set of all possible outcomes is called the sample space.

In many problems, one must find the probability of two or more outcomes. For this reason, it is necessary to distinguish between an outcome and an event.

An event consists of a set of outcomes of a probability experiment. An event can be one outcome or more than one outcome. For example, if a die is rolled and 6 shows, this result is called an outcome, since it is a result of a single trial.

An event with one outcome is called a simple event. The event of getting an odd number when a die is rolled is called a compound event since it consists of three outcomes or three simple events. In general, a compound event consists of two or more outcomes or simple events.

A sample space, (S) is the set of all possible outcomes of a probability experiment. The individual outcomes in the sample space are also called sample points.

The sample spaces for some probability experiments are shown below

Experiment	Sample space
Toss one coin	Head, Tail
Roll a die	1, 2, 3, 4, 5, 6
Answer a true/false question	True, False
Toss two coins	Head-Head, Tail-Tail, Head-Tail, Tail-Head
Answer a multiple-choice question having four options	Options A, B, C and D

It is important to realize that when two coins are tossed, there are four possible outcomes, as shown in the fourth experiment above.

SAQ 2-5

Random Variable

Let me tell you that **random variable, random quantity, aleatory variable, or stochastic variable** are described informally as a variable whose values depend on outcomes of a random phenomenon or physical experiment.

Now, a random variable's possible values might represent the possible outcomes of a yet-to-be-performed experiment, or the possible outcomes of a past experiment whose already-existing value is uncertain (for example, because of imprecise measurements).

The mathematical function, X(w), is a single-valued real function that assigns a real number, called the value of X(w), to each sample point $w \in S$ That is, it is a mapping of the sample space onto the real line. The sample space S is called the domain of the random variable X. Also, the collection of all numbers that are values of X is called the range of the random variable X.

So, as a function, a random variable is required to be measurable, which allows for probabilities to be assigned to sets of its possible values. It is common that the outcomes depend on some physical variables that are not predictable. For example, when you toss a fair coin, the final outcome of heads or tails depends on the uncertain physical conditions. Which outcome will be observed is not certain. The coin could get caught in a crack in the floor, but such a possibility is conventionally excluded from consideration,

A random variable takes real values. Physical examples of random variables include noise voltage at a given time and place, temperature at a given time and place, height of the next person to enter your room, the weight of students in your class, the number of children born in your family, the size of my chicken, and so on. The color of a randomly picked apple is not a random variable since its value is not a real number. However, if the experiment of picking the apple is categorized as either red or not red, then the color of a randomly picked apple could be viewed as a random variable (the sample space is red, not red).

However, for every random variable, there is a corresponding probability distribution, which specifies the probability of its values. Random variables can be discrete, that is, taking any of a specified finite or countable list of values, endowed with a probability mass function characteristic of the random variable's probability distribution; or continuous, taking any numerical value in an interval or collection of intervals, via a corresponding probability density function that is characteristic of the random variable's probability distribution; or a mixture of both types.

Do you know that two random variables with the same probability distribution can still differ in terms of their associations with, or independence from, other random variables? The realizations of a random variable, that is, the results of randomly choosing values according to the variable's probability distribution function, are called random variaties.

Activity 1	Throw a die for 50times and record the sequence of outcomes for the 50 throws. Use your mobile phone to capture your readings and share with yourcollegues on the Learning Management System's Forum.

Example 1: When you throw a pair of dice, let (x,w) represent the number shown. The sample space S(c,d) for all the possible outcome is given below:

$$S(c,d) = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,4) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

Example 2: When you throw a pair of dice, let X(w) represent the sum of the first and second die. The possible values of the random variable are: X(w)= 2, 3, 4, ...,11,12. The corresponding sample space is given below:

$$S = \left\{ \begin{array}{cccccc} 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 8 & 9 & 10 & 11 \\ 7 & 8 & 9 & 10 & 11 & 12 \end{array} \right\}$$

Example 3: Let X = the number of heads obtained when you toss three fair coins. The sample space for the toss of three fair coins is {TTT; THH; HTH; HHT; HTT; THT; TTH; HHH}. So the values of X are x = 0, 1, 2, 3. Notice that for this example, the x values are countable outcomes. Because you can count the possible values that X can take on and the outcomes are random (the x values 0, 1, 2, 3), X is a discrete random variable.

Summary

Thus far, I had explained the basic definitions governing the theory of probability. Some simple probability experiments were also given. The definition of a random variable and how it has wide applicability in everyday life was also given in this unit. We also discussed the two types of random variables: the discrete and continuous random variables, with several examples provided.

Self-Assessment Questions

1. Differentiate between an outcome and an event.
2. Define a random variable and give four examples of random variables.
3. What is the difference between a discrete random variable and a continuous random variable?
4. Classify the following random variables as either discrete or continuous.

(b) The weight of a box of cereal labeled "18 ounces."

(c) The duration of the next outgoing telephone call from a business office.

(d) The number of kernels of popcorn in a 1-pound container.

(e) The number of applicants for a job.

5. Classify each random variable as either discrete or continuous.

a. The time between customers entering a checkout lane at a retail store.

b. The weight of refuse on a truck arriving at a landfill.

c. The number of passengers in a passenger vehicle on a highway at rush hour.

d. The number of clerical errors on a medical chart.

e. The number of accident-free days in one month at a factory.

6. Classify each random variable as either discrete or continuous.

a. The number of boys in a randomly selected three-child family.

b. The temperature of a cup of coffee served at a restaurant.

c. The number of no-shows for every 100 reservations made with a commercial airline.

d. The number of vehicles owned by a randomly selected household.

e The average amount spent on electricity each July by a randomly selected household in a certain state.

7. Obtain the sample space for each of the following experiments:

a. The number of heads in throwing two coins together?

b. The number of tails in throwing three coins together?

c. The outcomes in throwing a die once?

Tutor Marked Assignment

1. A random variable which takes on a non-countable infinite number of values is said to be

2. Obtain the sample space for tossing a coin two times

3. Obtain the sample space for tossing three coins together.

4. Given that the sample space is the set of the first 20 whole numbers, and let the event A= the even numbers, then A=.

5. Classify each random variable as either discrete or continuous.

a. The number of patrons arriving at a restaurant between 5:00 p.m. and 6:00p.m

b. The number of new cases of influenza in a particular county in the coming month.

c. The air pressure of a tyre on a car.

d. The amount of rain recorded at an airport one day.

e. The number of students who actually register for classes at a university in a semester.

Further Reading

- https://en.wikipedia.org/wiki/Random_variable

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Statistical Analysis

Photo: pexels.com

UNIT 2

Functions of a Random Variable: Probability Mass Function, Probability Density Function, Cumulative Distribution Function

Introduction

In this study unit, I will introduce you to the fundamental functions of a random variable which are: probability mass function, probability density function and cumulative distribution function for discrete and continuous random variables. I will also explain these functions with examples.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Differentiate between the probability mass function and probability density function;
2. Enumerate the properties of the CDF.
3. Explain the relationship between the CDF and PDF;
4. Obtain the PMF of some random variables emanating from simple probability experiments like tossing a coin or throwing a die.

Main Content

SAQ 1,5

Probability Mass Function

15 mins

Definition: Let X denote a discrete random variable which can assume possible values given by $x_1, x_2, x_3, \cdots, x_n$ arranged in increasing order of magnitude. If these values take on the probabilities given by
$P(X = x_k) = p(x_k) = p_k, k = 1,2,3 \cdots,$ then the **probability mass function (PMF)** can be written as

$$p(x_k) = P(X = x_k)$$

In other words, if X is a random variable that can assume values $x_1, x_2, x_3, \cdots, x_n$ with associated probabilities $p_1, p_2, p_3, \cdots, p_n$ then the set of ordered pairs (x_k, p_k), k = 1, 2, ..., n is called the probability function or the probbility mass function (PMF) of random variable X.

So, the Probability Mass Function is thus the probability or possibility that a particular outcome will occur in a single trial of a probability experiment.

Example 1: When you throw a pair of fair dice, the sample space S = {(1,1), (1,2),(1,3),…, (6,6)} where n(s) = 36, and n(s) is the **number of sample points in S**; let the random variable X represent the sum of the number shown in the two dice, that is,

$$X(c, d) = c + d$$

where (c,d) is a possible outcome in the sample space S.

The probability mass function, p(x) is given in the table below.

X	2	3	4	5	6	7	8	9	10	11	12
p(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Activity 2

Throw a die for 50times and tabulate the outcomes for the 50 throws as shown in the table below. Use your mobile phone to capture your readings and share with your collegues on the Learning Management System's Forum.

Outcome	1	2	3	4	5	6
Number of Occurences						

Example 2: If an unbiased coin is thrown twice. Let X(S) represent the number of heads that appear. Here, S = {TT, TH, HT, HH} and X(S) = {0, 1, 2}. Let E_0, E_1, E_2 represent the events that no head appears, one head appears and two heads appear respectively.

$\therefore E_0 = (\{TT\}, \; E_1 = \{HT, TH\}, \; E_2 = \{HH\}$ and $P(E_0) = \frac{1}{4}, P(E_1) = \frac{2}{4}, P(E_2) = 1/4.$

Therefore, the probability mass function p(x) and the corresponding values of the random variable are given in the table below, called the **probability distribution** of X.

X	0	1	2
p(x)	1/4	2/4	¼

SAQ 3

Properties of the probability mass function

1. $0 \leq p(x) \leq 1 \; for \; every \; x$
2. $\sum_{x_i} p(x_i) = 1$
3. $P(E) = \sum_{x_i \in E} p(x_i)$

Explanation

Property 1 essentially states that all probability mass functions have values lying between 0 and 1.

Property 2 specifies that the sum of all the probabilities must be equal to 1.

Property 3 states that given an event E, which is a set of possible outcomes in the sample space, the probability of that event E is equal to the sum of the probabilities of the individual outcomes x_i. Indeed, property 3 specifies what is called the cumulative distribution function (CDF).

Probability Density Function of a Continuous Random Variable

SAQ 2,5

May I let you know at this point that **discrete random variables** have a set of possible values that are either **finite** or **countably infinite**. However, there exists another group of random variables that can assume an uncountable set of possible values. Random variables that are defined on an interval in the real line e.g $(-\infty, \infty)$, (0,5), (2,3), (0,1), etc, are called **continuous random variables**. Thus, a continuous random variable is defined as follows: a random variable X is said to be continuous if there exists a non-negative function fx), defined for all real numbers x, having the property that for any set A of real numbers,

$$P(X \in A) = \int_A f(x)dx$$

The function f(x) is called the Probability Density Function (PDF) of the random variable X and is defined by

$$f(x) = Pr(X = x) = \frac{d}{dx}F(x) = F'(x)$$

where F(x) is called the cumulative distribution function (CDF)

Properties of the Probability Density Function

SAQ 4

The following are the properties of the PDF f(x):

1. $f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$ which implies that $P(X{=}a) = \int_a^a f(x)dx = 0$
4. $P(X < a) = P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx$

Unit 2 | Module 1 | STA 124
Functions of a Random variable: Probability mass function, Probability density function, Cumulative distribution function.

Cumulative Distribution Function of a Random Variable

SAQ 6

Definition:

Note that for any random variable X, the cumulative distribution function (CDF), F(x) is defined as

$$F(x) = P(X \leq x)$$

The CDF of a discrete random variable X is given as

$$F(x) = Pr(X \leq x) = \sum_{k \leq x} p(k)$$

The CDF of a discrete random variable is a step function. That is, if X takes on values $x_1, x_2, \cdots$, where $x_1 < x_2 < \cdots$ then the value of F(x) is constant in the interval between x_{i-1} and x_i and then takes a jump of size $p(x_i)$ at x_i, i=1,2,.... Thus, F(x) represents the sum of all the probability masses from $-\infty$ to x.

The CDF of a continuous random variable X is given as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(k)dk$$

That is, F(x) denotes the probability that the random variable X takes on a value that is less than or equal to x.

You should note that while the PMF and PDF specify the probability at a point, x, (hence the name probability mass or probability density), the CDF on the other hand, specifies the probability below or above a point x or in an interval (hence the name cumulative distribution). The CDF is therefore an accumulation of points representing the PMF or PDF.

Properties of the Cumulative Distribution Function

1. F(x) is a non-decreasing function, which means that if $x_1 < x_2$ then $F(x_1) \leq F(x_2)$
   Thus, F(x) can increase or stay level, but it cannot decrease.

2. $0 \leq F(x) \leq 1$.
3. $F(+\infty) = 1$.
4. $F(-\infty) = 0$
5. $P(a < X \leq b) = F(b) - F(a)$.
6. $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$

Relationship Between the Probability Density Function (F(x)) and the Cumulative Distribution Function (F(x))

i. If $F(x) = Pr(X \leq x)$ is the cumulative distribution function (CDF) of a continuous random variable X, whose proability density function is f(x), then

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)du$$

This means that the CDF of a continuous random variable is obtained by integrating the PDF in the range $(-\infty, x)$, which are the lower and upper limits, respectively.

ii. If f(x) is a probability density function and F(x) is the corresponding cumulative distribution function of a continuous random variable X, then

$$f(x) = \frac{d}{dx}\mathrm{F(x)} = \mathrm{F'(x)}$$

Thus, the first derivative of the CDF gives the corresponding probability density function of the continuous random variable X.

From the two relationships stated above, we can interpret that the CDF can be obtained via integration of the corresponding PDF, and conversely the PDF can be obtained via differentiation of the CDF. Therefore, the knowledge of one of these distributions is sufficient in describing any probability distribution.

Example 3: Given that the probability density function of a continuous random variable is given as

$$f(x) = tx^3, \qquad 0 < x < 4$$

where t is a constant. Find (a) the value of t; (b) the cumulative distribution function.

Solution

(a) Since f(x) is a PDF, then

$$\int_0^4 f(x)dx = 1$$

That is,

$$\int_0^4 tx^3dx = \frac{tx^4}{4}\Big]_0^4 = \frac{t4^4}{4} - 0 = 1$$

Thus, rearranging the equation and making t the subject of the formula, yields t=1/64. Hence the value of t for which f(x) is a PDF is 1/64.

(b) By definition of the Cumulative Distribution Function,

$$F(x) = P(X \leq x) = \int_0^x \frac{k^3}{64}dk = \frac{1}{64} * \frac{x^4}{4} = \frac{x^4}{256}$$

Example 4: Use the CDF given below to find $P(1 \leq X \leq 2)$.

$$F(x) = P(X \leq x) = \frac{x^4}{256}$$

Solution

We can compute the required probability by substituting the values 2 and 1 into the CDF and subtracting the smaller from the larger one. That is,

$$P(1 \leq X \leq 2) = F(2) - F(1) \quad = \frac{2^4}{256} - \frac{1^4}{256} = \frac{15}{256}$$

It could be seen therefore that to obtain the probability of an event occurring in a specified interval, the CDF is the appropriate function to use. This applies to both discrete and continuous probability distributions

Summary

Thus far, I had explained the probability mass function and the probability density function are used to obtain the probability at a sample point in the sample space, for discrete and continuous random variables, respectively. The PMF and PDF are both probabilities, and as such, must sum to 1, must not be negative and cannot be greater than 1.

Furthermore, the Cumulative Distribution Function is the probability over a set

of sample points (in the case of a discrete random variable) or an interval (in the case of a continuous random variable) and is an aggregate of the individual PMF and PDF of the points in the the set.

It may interest you that the knowledge of the PDF f(x) of a continuous random variable implies knowledge of the CDF F(x), as they are both related. We can obtain CDF from the PDF via integration, while we can obtain PDF from the CDF via differentiation.

When we are using the PMF and PDF, it is possible we can compute any specified probability for probability experiments and random occurrences.

Self-Assessment Questions

1. Define the probability mass function (PMF) of a discrete random variable.
2. Define the probability density function (PDF) of a continuous random variable.
3. List the properties of the probability mass function.
4. State the properties of the probability density function.
5. What is the difference between the probability mass function and probability density function?
6. Define the cumulative distribution function (CDF) of a random variable.
7. Let X be a discrete random variable. What conditions must be satisfied to make p(x) the probability mass function of X?
8. Let X be a discrete random variable with probability mass function given by

X	1	2	3	4	5	6	7	8	9	10
f(x)	0	1/10	0	0	16/40	6/36	5/36	4/36	3/36	2/36

Compute (a) F(5) (b) F(9) (c) F(10) (d) F(2).

9. Given that $f(x) = 2x, 0 < x < 1,$ obtain the expression for F(x) and use it to determine (a) F(0.5) (b) F(0.7) (c) F(0.2) (d) F(1)

Tutor Marked Assignment

1. The distribution function F(x) is a monotonically...... function.
2. Given that a random variable can take the values -4, 3, and 2 with probabilities 1/4, 1/2 and 1/4, respectively. Compute (a) $P(2 \leq X \leq 3$ (b) F(2) © F(3).
3. Given that f(x)=2x/25, 0<x<5, find Pr(1<X<3).
4. Given that f(x)=2x/25, 0<x<5. Find Pr(-2<X<4)
5. Suppose $f(x) = kx^3$, $0 < x < 4$ find the constant for which f(x) is the probability density function.
6. Suppose $f(x) = kx^3$, $0 < x < 4$ find P(1<X<2)
7. If f(x)=1/2, 0< x < 2, then F(1.5)=
8. Given that $p(x) = \frac{x^2}{55}, x = 1,2,3,4,5$

Further Reading

- https://en.wikipedia.org/wiki/Probability_mass_function
- https://en.wikipedia.org/wiki/Probability_density_function
- https://en.wikipedia.org/wiki/Cumulative_distribution_function

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: National Cinema Mesum, Torino, Italy

source: unsplash.com

Module 2

Moments of a Random Variable

Units

Unit 1 - Mathematical Expectations of a Random Variable
Unit 2 - Variance and standard deviation of a random variable

Picture: calculator-scientific

Photo: freepik.com

UNIT 1

Mathematical Expectations of a Random Variable

Introduction

In this study unit, I will be explaining the concept of expectation and applies the principle to determine a basic measure of any probability distribution – the mean. Other measures like the variance and standard deviation are also obtained, together with their mathematical properties.

Learning Outcomes

After you have studied this unit, you should be able to:

1. Define the expectation of a random variable;
2. Identify some basic mathematical properties of expectation;
3. Compute the expectation of some discrete and continuous random variables emanating from simple probability experiments; and
4. Compute the expectation of some simple functions of random variables.

Main Content

SAQ 1,5

Mathematical Expectation

I want you to know that the **expected value** or **mean** or **first moment** of a random variable may be viewed as the "long-term" average of the variable. That is, if the probability experiment is repeated over and over, then the expected value or mean would be the average of the numerical values obtained at each replication of the experiment.

For example, suppose that the experiment consists of tossing a fair coin and recording the outcome. Although the probability of getting heads is 1/2, this does not mean that in multiple trials that exactly half the tosses would land as heads. You might toss a fair coin ten times and record nine heads. Probability does not describe the short-term results of an experiment. Instead, it gives information about what can be expected in the long term.

SAQ 1,2,3

Expectation of a Discrete Random Variable

Let X be a random variable, and $x_1, x_2, \cdots, x_n$ the list of possible outcomes for X. Then the *mean* of the distribution and the *expected value of X* are the same quantity, given by:

$$\mu = E(X) = \Sigma x_i \, p(x_i)$$

That is, to find the expected value or long term average, μ, simply multiply each value of the discrete random variable by its probability and add the products. Thus, the expectation is a weighted average of the possible values of the random variable X.

Example 1: A fair coin is tossed three times, let X denote the number of heads which appears. The sample space contains 8 elements, S = {TTT, HTT, THT, TTH, HHT, HTH, THH, HHH}. Since 0, 1, 2 or 3 heads can be obtained, then X(s) = (0, 1, 2, 3). From the sample space, the probability of any of the occurrences (0, 1, 2, or 3 heads), can be obtained by counting the number of outcomes and dividing by the total number of elements in the sample space. This is presented in the probability distribution below

X	0	1	2	3
p(x)	1/8	3/8	3/8	1/8

Therefore $E(X) = \sum xp(x) = 3/2 = 1.5$

The expected value of X is gotten by summing up the product of each outcome (x) with the corresponding PMF p(x).

Example 2: A part-time football team trains zero, one, or two days a week. The probability that they do not train any day in the week is 0.2, the probability that they train once in a week is 0.3, and the probability that they train two days in a week is 0.5. Find the long-term average or expected value, μ, of the number of days per week the team trains.

Solution

Let the random variable X=the number of days the team trains in a week. The distribution, detailing the values of X (0, 1, 2) and the corresponding probabilities (0.2, 0.3, 0.5) will be used to compute the expected number of days per week the team trains.

That is, $E(X) = \sum xp(x) = 0(0.2) + 1(0.3) + 2(0.5) = 1.3$

This implies that, on the average, the team trains or do practice sessions 1.3 times a week.

Example 3: It is known that a certain student studies 0, 1,2, 3, 4, or 5 times a week during the semester. Let X be the number of times the student studies in a week. The probability distribution is given in the table below.

X	0	1	2	3	4	5
p(x)	0.1	0.2	0.1	0.1	0.2	0.3

Obtain the expected number of times that the student studies in a week.

Solution

$$E(X) = \sum xp(x) = 0(0.1) + 1(0.2) + 2(0.1) + 3(0.1) + 4(0.2) + 5(0.3) = 3$$

Thus, the expected number of times that the student studies in a week is 3 times.

Activity 3	From your table in Activity 2, obtain the probability of each outcome of the experiment.

Properties of Expectation

i. If t is a constant and X is a random variable, then $E(tX) = tE(X) = t\mu$

ii. $E(X + t) = E(X) + t = \mu + t$ where t is a constant.

iii. If X and Y are two independent random variables, then $E(X + Y) = E(X) + E(Y)$

iv. If X and Y are two independent random variables, then $E(XY) = E(X)E(Y)$

Expectation of a Continuous Random Variable

Let X be a continuous random variable defined on a specified interval (a,b), with PDF f(x). The expectation E(X) is given as

$$E(X) = \int_a^b xf(x)dx$$

Thus, the expectation or first moment of the random variable is the integral over the range of X, of the product of x and the PDF, f(x).

Activity 4:	Given the PDF $f(x) = \frac{x}{2}, 0 < x < 2$ Obtain E(X)

Solution

$$E(X) = \int_0^2 xf(x)dx = \int_0^2 \frac{x^2}{2}dx = \frac{1}{2}\frac{x^3}{3}\Big]_0^2 = \frac{(2^3 - 0^3)}{2(3)} = \frac{8}{6} = \frac{4}{3}$$

Hence, the expected value of the random variable X is 4/3.

Activity 4:	Given the PDF $f(x) = \frac{x}{2}, 0 < x < 2$ Obtain E(X)

Solution

$$E(X) = \int_0^2 xf(x)dx = \int_0^2 \frac{x^2}{2}dx = \frac{1}{2}\frac{x^3}{3}\Big]_0^2 = \frac{(2^3 - 0^3)}{2(3)} = \frac{8}{6} = \frac{4}{3}$$

Hence, the expected value of the random variable X is 4/3.

Activity 5:	If $f(x) = \frac{1}{2}, 0 < x < 2$ find E(X).

Solution

This is an example of the uniform continuous distribution.

$$E(X) = \int_0^2 xf(x)dx = \int_0^2 \frac{x}{2}dx = \frac{1}{2}\frac{x^2}{2}\Big]_0^2 = \frac{4}{4} = 1$$

Thus the mean or expected value of the distribution is 1.

SAQ 4,5

Expectation of Function of a Random Variable

Definition: May I let you know that a function of a random variable is a rule that transforms a point from a sample space into a real number line. If X and Y are random variables on the sample space S, then Y is said to be a function of X provided Y can be represented by Y(s) =g (X(s)) for every $s \in S$.

Some examples of functions of random variables are Y=2X+t, $Y = X^2 + 3X$ and Y=tX.

A function of a random variable is also a random variable.

Definition: Let X be a random variable and Y=g(X) a function of X. Then the expectation of Y is given as

$$E(Y) = E(g(X))$$

Activity 6:	An unbiased green octahedral die is tossed. If Y denote twice the number that appears, find (i) E(Y); (ii) E(Z), if Z= 3 + Y ; and E(Z) if Z = 1/3Y + 1.

Solution

An octahedral die has 8 faces with the following numbers: 1, 2, 3, 4, 5, 6, 7, 8. The sample space for Y=2X, where X is the value of the face that shows up in any toss of the die, is given as S=(2, 4, 6, 8, 10, 12, 14, 16). Since the die is unbiased, then each outcome is equally likely, hence have the same probability 1/8..

i. $$E(Y) = \sum yp(y) = \frac{1}{8(2+4+6+8+10+12+14+16)} = \frac{72}{8} = 9$$

ii. E(Z)=E(3+Y)=3+E(Y)=3+9=12.

iii. E(Z)=E(1/3Y+1)=1/3E(Y)+1=1/3(9)+1=3+1=4.

Explanation

In order to compute the expectation of a function of a random variable Y=g(X), the expectation of the underlying random variable, X should first be computed. Thereafter, the expected value of Y could be easily obtained using the properties of expectation.

Summary

In this unit, you have learnt that the concept of mathematical expectation of a random variable is used to capture the long-run or average behaviour of the random variable. It is one of the important measures in probability theory and can be estimated for both discrete and continuous probability distributions. Expectation has some mathematical properties which determine its behaviour, as specified in this unit. Several examples were given to illustrate how expectation is computed for both discrete and continuous probability distributions.

Self-Assessment Questions

1. A fair coin is tossed twice and the number of heads observed. (a) determine the sample space for the experiment (b) obtain the probability distribution of X, the number of heads that show up (c) compute E(X).

2. A fair coin is tossed twice. If X denotes the number of heads that appears, find the expected value of Y if (a) $Y = 2X - 1$; (b) $Y = 2 + 3X$; (C) $Y = \frac{1}{2}X + 1$
3. Given that a random variable can take the values -4, 3, and 2 with probabilities 1/4, 1/2 and 1/4, respectively. Compute E(X).
4. An unbiased coin is tossed three times and the number of heads observed. Compute the expected number of heads, E(X).
5. If X is a random variable and b is a constant, then evaluate the following: (a) E(X+b), (b) E(bX+b)

Tutor Marked Assignment

1. Given that f(x)=2x/25, 0<x<5. Compute E(X)
2. Given that f(x)=2x, 0< x < 1, then E(X)=
3. If f(x)=1, 0< x < 1, then E(X)=
4. If f(x)=1/2, 0< x < 2, then E(X)=
5. Given that $p(x) = \frac{x^2}{55}, x = 1,2,3,4,5$, compute E(X).
6. A fair coin is tossed four times. Let random variable X denote the number of heads obtained. Determine (i) the distribution of X; (ii) E(X); (iii) if Y=2X+1, obtain E(Y); (iv) if Y=2+X, find E(Y-2).

Further Reading

- https://en.wikipedia.org/wiki/Expected_value
- https://www.youtube.com/watch?v=qDOmZrn6qJQ

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.

- ·Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- ·Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: growth analytic concept illustration

Photo: freepik.com

UNIT 2

Variance and Standard Deviation of a Random Variable

Introduction

In the previous unit, we explainedthe concept of expectation and applies the principle to determine a basic measure of any probability distribution. In this unit, I will be explaining and calculating the concept of mathematical expectation that is extended to the variance and standard deviation which are important measures of dispersion or spread of a probability distribution and the general procedure for obtaining these measures is presented. The properties of the variance are also given.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the second moment about zero of a random variable;
2. Explain the meaning of the variance of a random variable;
3. Obtain the variance of a random variable;
4. Determine the variance of a function of a random variable; and
5. Obtain the variance of specified random variables from simple probability experiments and random phenomena.

Main Content

SAQ 1

Random Variable

10 mins

Definition: Lets say we are given a random variable X, with probability mass function p(x) (or probability density function f(x) for the case of a continuous random variable), the **nth moment about 0** or the **nth raw moment** is given as

$$E(X^n) = \begin{cases} \sum_{x \in X} x^n p(x), & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} x^n f(x)dx, & \text{if } X \text{ is a continuous random variabe} \end{cases}$$

May I let yo know that the first moment about zero is the expected value of the random variable, while the second moment about zero $E(X^2)$ is used importantly to obtain the variance of the probability distribution.

Also, higher moments are also important in computing some properties of a probability distribution like *skewness* and *kurtosis.*

Definition: So if X is a random variable with probability mass function p(x) or probability density function f(x) and expectation $E(X) = \mu$, then the **variance** of X, var(X) is given as

$$Var(X) = E\big(X - E(X)\big)^2 = E(X^2) - \big(E(X)\big)^2$$

Now, the variance of a random variable is also called the **second moment about the mean** or **second central moment**. The variance of a distribution gives a quantitative value to the amount of dispersion or spread of the sample points around the mean.

Given a probability distribution (discrete or continuous) with expected value or first moment about zero, $E(X) = \mu$, then the variance of the distribution can be expressed as

$$Var(X) = E(X-\mu)^2 = \begin{cases} \sum_{x \in X} (x-\mu)^2 p(x), & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, & \text{if } X \text{ is a continuous random variabe} \end{cases}$$

For a probability distribution, the mean of the random variable describes the measure of the long-run or theoretical average, but it does not tell anything about the spread of the distribution. In order to measure this spread or variability, the variance and standard deviation are used.

However, to find the variance for a discrete random variable, subtract the expected value of the random variable from each outcome and square the difference. Then multiply each difference by its corresponding probability and add the products. The formula is

$$Var(X) = \sigma^2 = \sum(x-\mu)^2 p(x) = \sum x^2 p(x) - \mu^2 = E(X^2) - \mu^2$$

The variance of a continuous random variable is obtained via integration.

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = E(X^2) - \mu^2$$

The **standard deviation** σ is the square root of the variance, that is

$$\sigma = \sqrt{Var(X)} = \sqrt{E\big(X - E(X)\big)^2}$$

The standard deviation specifies the deviation from the mean and is of the same unit as the mean of the distribution.

Properties of Variance

Given a random variable X and a constant t, then the following are true about the variance of the random variable.

1. $Var(X+t) = Var(X)$.
2. $Var(tX) = t^2 Var(X)$
3. The variance is always a positive number, hence it cannot be negative.

Explanation

SAQ 2,3, 4,5

Property 1 implies that the variance of the sum of a random variable and a constant is equal to the variance of the random variable itself. It also means that the variance of a constant is equal to zero. Indeed, there is no variation within a constant or fixed variable, hence its variance is zero.

Property 2 means that the variance of the product of a random variable and a constant is equal to the square of the constant multiplied by the variance of the random variable itself.

Property 2 also implies that Var(-X)=Var(X). That is, if you have a negative constant multiplying the random variable X, then the variance will be the square of that negative constant (this square will always be positive) multiplied by the variance of the random variable itself.

Example 1: You tossed a fair die once. Let the random variable X represent the face that shows up. Obtain (a) the distribution of X (b) E(X) (c) $E(X^2)$ (d) Var(X) (e) the standard deviation of X.

Solution

(a) p(x)=1/6, for x=1, 2, 3, 4, 5, 6. is the required probability distribution of X.

(b) $E(X) = \sum xp(x) = \frac{21}{6} = 7/3$

(c) $E(X^2) = \sum x^2 p(x) = 91/6$

(d) $Var(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \left(\frac{7}{3}\right)^2 = 9.722$

(e) $sd(X) = \sqrt{Var(X)} = \sqrt{9.722} = 3.12$

Example2: Find the mean, variance and standard deviation of the random variable X having the distribution below

x_i	A	b
$p(x_i)$	R	t

Solution

$Mean = E(X) = ar + bt$

$E(X^2) = a^2r + b^2t$

$Var(X) = E(X^2) - (E(X))^2 = a^2r + b^2t - (ar + bt)^2 = rt(a - b)^2$

Therefore,

$SD(X) = \sqrt{Var(X)} = (a - b)\sqrt{rt}$

Variance of a Function of a Random Variable

Do you know that the principle used in computing the expected value of a function of a random variable (discrete or continuous) can also be applied to compute the variance of a function of a random variable? Given a random variable X with mean μ and variance σ^2. Define another random variable Y=h(X). Then the variance of random variable Y is

$$Var(Y) = E(Y - E(Y))^2 = E(h(X) - E(h(X)))^2 = E[h(X)^2] - [E(h(X))]^2$$

Thus the variance of any function of a random variable is obtained by obtaining both the expectation $E(h(X))$ and the second moment $E[h(X)^2]$. The next step is to subtract the square of $E(h(X))$ from $E[h(X)^2]$.

Example3: Given a random variable X, with $E(X) = 2$ and $E(X^2) = 8$, determine the variance and standard deviation of the following functions: (a) Y=2+X (b) Y=3X-1 (c) Y=2-X (d) Y=-X.

Solution

Given E(X)=2, $E(X^2) = 8$.

(a) $Var(Y) = Var(2 + X) = Var(X) = E(X^2) - (E(X))^2 = 8 - 4 = 2$

$sd(Y) = \sqrt{Var(Y)} = \sqrt{2} = 1.4142$

Thus the variance of Y, in this case, is equal to the variance of X, which is a direct consequence of the first property of variance earlier presented. And the standard deviation is just the square root of the variance.

(b) $Var(Y) = Var(3X - 1) = 3^2Var(X) - 0 = 9(2) = 18$

$sd(Y) = \sqrt{Var(Y)} = \sqrt{18} = 4.2426$

$Var(Y) = Var(2 - X) = 0 + Var(X) = 2$

$sd(Y) = \sqrt{Var(Y)} = \sqrt{2} = 1.4142$

Summary

In this study unit, we have discussed the higher moments of a probability distribution. The second moment about zero was introduced, which provides an alternative way of computing the variance of the distribution. The variance and the standard deviationare important measures of the dispersion or spread of a probability distribution. The variance is also called the second moment about the mean. The second raw moment, variance and standard deviation were obtained for several discrete and continuous distributions. Finally, the variance and standard deviation of some simple functions of random variables were obtained.

Self-Assessment Questions

1. A fair coin is tossed twice. If X denotes the number of heads appearing, and given that Y=h(X) find E(Y) if (a) h(X)=$X^2 - 1$, (b) $h(X) = X^2 + 4$, (c) $h(X) = X^2 + 2X + 3$.
2. Given that $p(x) = \frac{x^2}{55}, x = 1,2,3,4,5,$ compute var(X).
3. Given that $p(x) = \frac{x^2}{91}, x = 1,2,3,4,5,6$ compute $E(X^2)$
4. Given that $p(x) = \frac{x^2}{91}, x = 1,2,3,4,5,6$ compute $var(X)$
5. Suppose f(x)=1/6, 0 < x < 6, then var(X)=

Tutor Marked Assignment

1. Given that a random variable can take the values -4, 3, and 2 with probabilities 1/4, 1/2 and 1/4, respectively. Compute $E(X^2)$
2. Given that a random variable can take the values -4, 3, and 2 with probabilities 1/4, 1/2 and 1/4, respectively. Compute Var(X).
3. An unbiased coin is tossed three times. If a discrete random variable X denotes the number of tails that appear, find $E(X^2)$
4. An unbiased coin is tossed three times. If a discrete random variable X denotes the number of tails that appear, find Var(X).

5. A discrete random variable X can take only two values - a and b, while the probability mass function is f(a)=r, and f(b)=t, respectively. Then Var(X)=

Further Reading

- https://www.statsdirect.com/help/basic_descriptive_statistics/standard_deviation.htm

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: statistics analysis

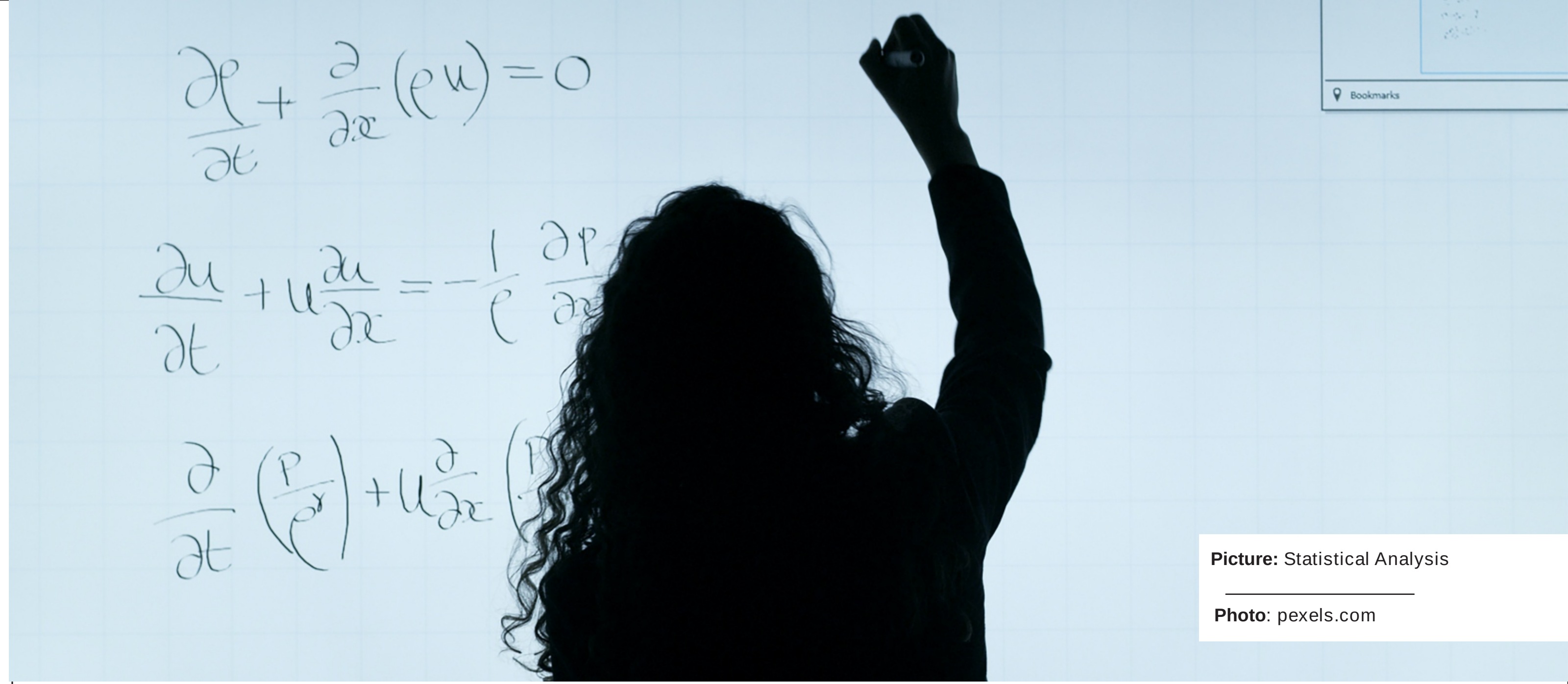
source: pexels.com

Module 3

Discrete Probability Distributions

Units

Unit 1 - Bernoulli Distribution, Binomial Distribution
Unit 2 - Geometric Distribution
Unit 3 - Poisson Distribution



Picture: Statistical Analysis

Photo: pexels.com

UNIT 1

Bernoulli Distribution, Binomial Distribution

Introduction

In this unit, I will be unfolding the theory of sample spaces, events and probability mass function are extended to some common discrete probability distributions, which emanate from simple experiments like tossing a coin or rolling dice and observing the outcomes. The Bernoulli distribution is first introduced and its basic properties are given. The Bernoulli distribution is then extended to the binomial distribution and its properties are also presented.ements.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Identify simple real-life experiments and processes that can be modelled by the Bernoulli distribution;
2. Define a Bernoulli random variable, its probability mass function and at least three properties of the distribution;
3. Define a binomial random variable, its probability mass function and state at least three properties of the distribution;
4. Identify the relationship between the Bernoulli random variable and the binomial random variable; and
5. Provide some examples of probability experiments that could be modelled by the binomial distribution;

Main Content

Bernoulli Distribution

SAQ 1,6,7

15 mins

May I interest you to know that a **Bernoulli trial** (named after Jacob Bernoulli) is any experiment with only two possible outcomes: success or failure? It was popularized by Jacob Bernoulli, who did much research in the area towards the end of the 17th century. Many examples of Bernoulli trials exist, like tossing a coin, whether a student will gain admission into the university or not, whether a woman will become pregnant or not, whether a pregnant woman will give birth to a boy or not, whether a football player will score a goal or not, whether a student will have an excellent grade in an exam or not, and so many others.

Given an experiment whose outcomes can be classified into one of two classes, for example, pass or fail, on or off, head or tail. Let X = 1 for success with probability p and X = 0for a failure, with probability q=1-p, then the probability mass function is given by

$$p(x) = P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

Alternatively, the PMF can be represented as

$$p(x) = p^x(1-p)^{1-x} = p^x q^{1-x}, \quad x = 0, 1$$

A random variable with the above probability mass function is said to be a Bernoulli random variable and the probability distribution is called a Bernoulli distribution.

Conditions for Bernoulli Experiment

1. There are only two possible outcomes, called "success" and "failure," for each trial.

2. The trials are independent and are repeated using identical conditions. Because the trials areindependent, the outcome of one trial does not help in predicting the outcome of another trial.

3. The letter p denotes the probability of a success on one trial, and q denotes the probability of afailure on one trial, so $p + q = 1$. Since the trials are independent, p remains the same for each trial.

The graph of the probability mass function of the Bernoulli distribution is given below

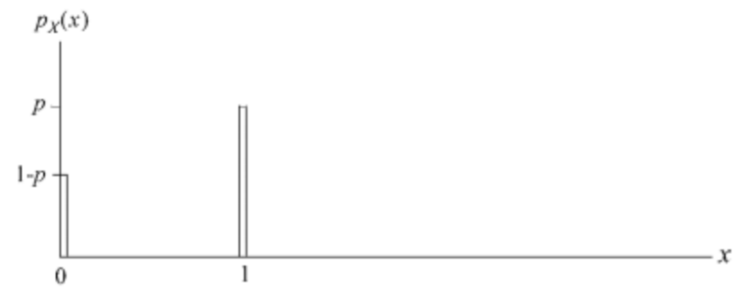


Figure 1: Probability mass function of the Bernoulli distribution with probability of success p.

Properties of the Bernoulli Distribution

SAQ 9

1. E(X) = p

2. $E(X^2) = p$

3. Var(X) = pq.

4. $F(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$

Example 1: The probability that a student passes an exam is 0.3. (a) What is the mean of the underlying Bernoulli distribution? (b) Compute the variance.

Solution

(a) p=0.3, q=1-p=0.7 $E(X) = p = 0.3$, (b) Var(X)=pq=0.3(0.7)=0.21

The utility of the Bernoulli distribution is that it is the building block for the binomial distribution, as the binomial experiment is in fact n independent Bernoulli trials with the same probability of success p.

SAQ 4,5,6,8

Binomial Distribution

Be notified that any experiment that involves two or more independent trials which can yield only one of two possible outcomes (success or failure) and has a fixed probability of success p can be described by the **binomial distribution.**

Define the random variable X as the number of successes in n, independent trials each having the same probability of success p. Then X has a binomial distribution with parameters n and p, represented as B(n,p), that is $X \sim B(n,p)$, where n is the number of trials and p is the probability of success at each trial.

The probability density function of the binomial distribution is given as

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \cdots, n$$

where

$$\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$$

Conditions for Binomal Experiment

You should note that the binomial experiment takes place when the number of successes is counted in one or more Bernoulli trials Binomial distribution is a special discrete probability distribution. There are **four conditions** that the experiment has to meet to be considered a **binomial experiment:**

1. There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter *n* denotes the number of trials.
2. There are only two possible outcomes, called "success" and "failure," for each trial.
3. The *n* trials are independent and are repeated using identical conditions. Because the *n* trials are independent, the outcome of one trial does not help in predicting the outcome of another trial.

The letter *p* denotes the probability of a success on one trial, and *q* denotes the probability of a failure on one trial, so $p + q = 1$. Since the trials are independent, *p* remains the same for each trial.

SAQ 10

Properties of the Binomial Distribution

i. It has n independent trials.

ii. It has a constant probability of success p and the probability of failure q = 1-p from trial to trial.

iii. There is an assigned probability to non-occurrence of events.

iv. The mean E(X) = np and the variance Var(X)= npq.

v. Each trial can result in one of only two possible outcomes called success or failure.

Example 2: If X has a binomial distribution with n = 5 and p = 1/2. Find (i) P(X=1) (ii) P(X>1) (iii) E(X) (iv) Var(X) and (v)

Solution

X~B(5,0.5). x=0,1,2,3,4,5. The probability mass function is given as

$$p(x) = \binom{5}{x} (0.5)^x (0.5)^{5-x}, \qquad x = 0, 1, 2, \cdots, 5$$

Since the probability of success is equal to the probability of failure, the PMF can be rearranged as

$$p(x) = \binom{5}{x} (0.5)^{x+5-x} = \binom{5}{x} (0.5)^5$$

i. $P(X=1) = p(1) = \binom{5}{1}(0.5)^5 = \frac{5}{2^5} = \frac{5}{32} = 0.15625$

ii. $P(X>1) = 1 - P(X \le 1) = 1 - \big(p(0) + p(1)\big) = 1 - \left(\frac{1}{32} + \frac{5}{32}\right) = \frac{26}{32} = \frac{11}{16} = 0.6875$

iii. $E(X) = np = 5\left(\frac{1}{2}\right) = 2.5$

iv. $Var(X) = npq = 5\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{5}{4} = 1.25$

v. Since $Var(X) = E(X^2) - (E(X))^2$ then rearranging and making $E(X^2)$ the subject of the formula yields

$$E(X^2) = Var(X) + (E(X))^2 = 1.25 + (2.5)^2 = 7.5$$

Thus, given the parameters n and p of the binomial distribution, the expected value, variance, cumulative probabilities, as well as the PMF of the distribution can be easily obtained.

Activity 4	A set of eight pregnant women live in a community. If ultrasound scan reveals that each of them is carrying a child and we are interested in determining the probability of either a woman give birth to a male or female child, which probability distribution will be appropriate?

Summary

In this unit, you have learnt that Bernoulli distribution was presented and its principle extended to define the binomial distribution. Some common examples emanating from these simple distributions were provided. A Bernoulli trial consists of any experiment or occurrence with only two possible outcomes (success or failure) having a fixed probability of success. The binomial experiment specifies two or more replication of a Bernoulli trial, with the random variable being the number of successes observed in n replications of the Bernoulli experiment. The properties of these distributions were presented and some illustrations of the application of these distributions were also provided.

Self-Assessment Questions

1. Suppose a Bernoulli experiment has the success probability p=0.7, compute the mean and variance of the distribution.

2. A fair coin is tossed 15 times. What is the mean and variance of obtaining heads?
3. A fair coin is tossed three times and the sequence of outcomes observed. What is the probability that exactly two heads occur?
4. What probability distribution has a standard deviation given as $\sqrt{npq}$?
5. An experiment consists of tossing a fair six-sided die 6 times and observing the number of odd numbers that turn up. What is the probability of having at least 2 odd numbers in the six tosses of the die?
6. What are the relationships between a Bernoulli distribution and a Binomial distribution?
7. Give five everyday examples of Bernoulli trials.
8. Provide five everyday examples of binomial experiments.
9. Mention 3 properties of the Bernoulli distribution.
10 Mention 4 properties of the binomial distribution.

Tutor Marked Assignment

1. An unbiased coin is tossed 3 times. Find the mean and variance of the number of heads that turn up.
2. In a family of 6 children, what is the probability of having at least two boys if each outcome (boy or girl) is equally likely?
3. X is binomially distributed with n = 10 and p = 0.5, (X ~ B(10,0.5)), Calculate Pr(x<3)
4. X is binomially distributed with n = 10 and p = 0.5, (X ~ b(10, 0.5)), Calculate Pr(x>3)
5. X is binomially distributed with n = 10 and p = 0.5, (X ~ b(10, 0.5)), Calculate Pr(1 ≤ x < 4)

6. If the probability is p=0.40 that a divorcee will remarry within three years, find the probability that out of ten (10) divorcees at most three of them will remarry within three years?

7. If the probability is p=0.40 that a divorcee will remarry within three years, find the probability that out of ten (10) divorcees at least three of them will remarry within three years?

8. If the probability is p=0.60 that a candidate with a UTME aggregate score greater than 240 will get admitted to read Engineering, find the probability that three out of nine candidates with UTME aggregate scores greater than 240 will get admitted to read Engineering?

9. If the probability is p=0.60 that a candidate with a UTME aggregate score greater than 240 will get admitted to read Engineering, find the probability that at least one out of nine candidates with UTME aggregate scores greater than 240 will get admitted to read Engineering?

10. Given that X is binomially distributed with n = 6 and p = 0.3, calculate $Pr(1 < x < 4)$?

Further Reading

- https://en.wikipedia.org/wiki/Bernoulli_distribution
- https://www.youtube.com/watch?v=bT1p5tJwn_0
- https://en.wikipedia.org/wiki/Binomial_distribution
- https://www.youtube.com/watch?v=qIzC1-9PwQo

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Pencil on system of equations

Photo: unsplash.com

UNIT 2

Geometric Distribution

Introduction

The geometric distribution and its statistical properties are presented in this study unit. Examples of some random processes that follow the geometric distribution are also presented.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define two variants of the geometric random variable; Enumerate the properties of the CDF.
2. Highlight some of the properties of the geometric distribution;
3. Identify scenarios that lead to geometric distribution;
4. Provide some examples of the application of the geometric distribution; and
5. Compute the mean and variance of the geometric distribution.

Main Content

Geometric Distribution

15 mins

Note that the geometric distribution is either of two probability distributions:

i. The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set $\{1, 2, 3, \ldots\}$

ii. The probability distribution of the number $Y = X - 1$ of failures before the first success, supported on the set $\{0, 1, 2, 3, \ldots\}$

Any of the two specifications above refers to the geometric distribution.

Now these two different geometric distributions should not be confused with each other. Often, the name **shifted geometric distribution** is adopted for thefirst one (distribution of the number X); however, to avoid ambiguity, it is considered wise to indicate which is intended, by mentioning the domain explicitly.

Let me say that the geometric distribution gives the probability that the first occurrence of success requires k independent trials, each with success probability p.

Suppose that independent trials each having a probability, p of success is performed until a success occurs. If we let X be the number of trials until the first success, then the probability mass function of the geometric distribution with parameter p is given as

$$p(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \cdots$$

Now for the geometric distribution, the random variable X is the number of trials until the first success, and this number is not known beforehand, hence it is not fixed, unlike the binomial distribution which has a fixed number of trials. This is a major difference between the binomial distribution and the geometric distribution.

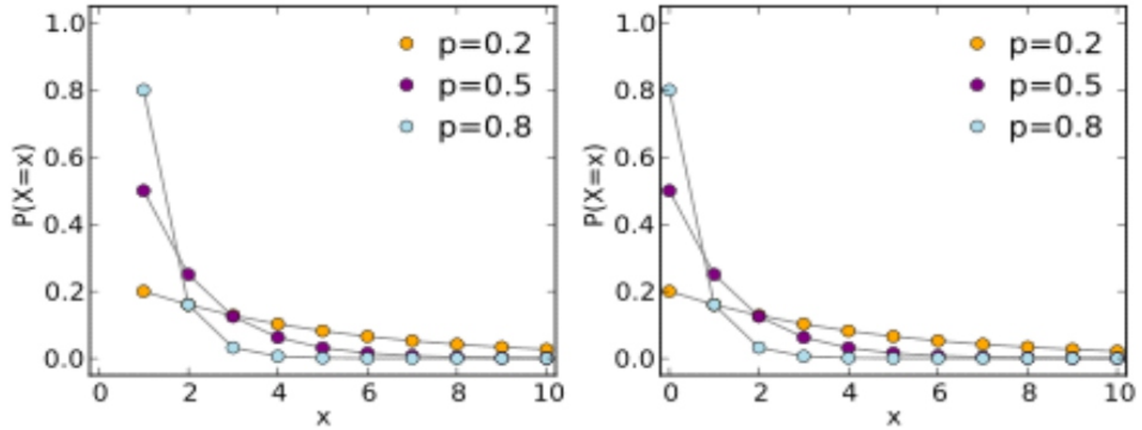


Figure 1: Probability mass function of the two variants of the geometric distribution.

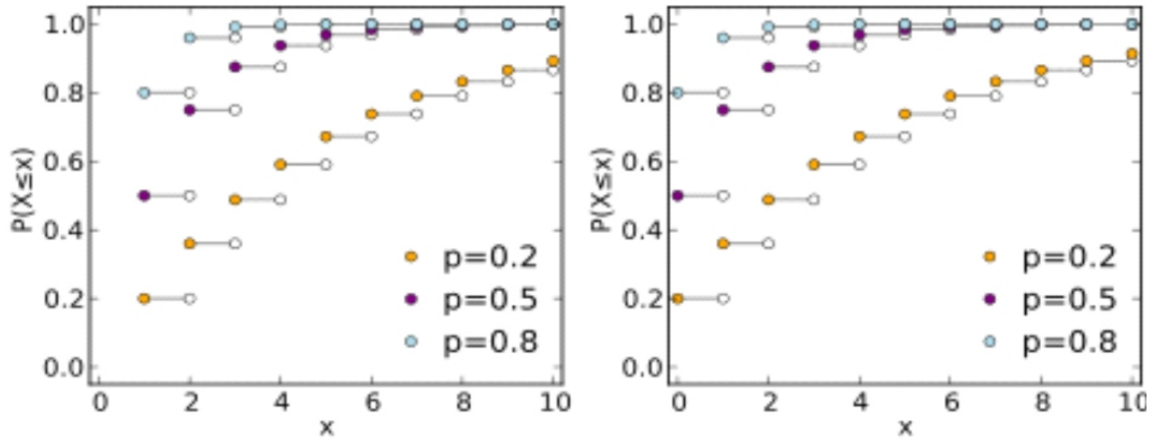


Figure 2: Cumulative distribution function of the two variants of the geometric distribution.

Properties of the Geometric Distribution

1. There is a sequence of independent trials (the outcome of one trial does not depend on another).
2. Only two outcomes e.g. success or failure are possible at each trial.
3. There is a constant probability of success at each trial.
4. X is the number of trials for the first success to appear (different from that of binomial distribution).

5. $E(X) = 1/p$, and $Var(X) = q/p^2$.

6. $P(X \leq x) = F(x) = 1 - (1-p)^x$

7. The mode of the geometric distribution is 1.

Some Applications of the Geometric Distribution

SAQ 1-6

i. A newlywed couple plans to have children and will continue until the first girl. Questions of interest could be: what is the probability that there are no boys before the first girl, one boy before the first girl, two boys before the first girl, and so on?

ii. A doctor is seeking an anti-depressant for a newly diagnosed patient. Suppose that, of the available anti-depressant drugs, the probability that any particular drug will be effective for a particular patient is p. What is the probability that the first drug found to be effective for this patient is the first drug tried, the second drug tried, and so on? What is the expected number of drugs that will be tried to find one that is effective?

iii. A patient is waiting for a suitable matching kidney donor for a transplant. If the probability that a randomly selected donor is a suitable match is p, what is the expected number of donors who will be tested before a matching donor is found?

iv. A resilient student has vowed to continue registering for and writing an entrance exam until he passes it. Suppose that the probability that he passes the exam is p, how many exams will the student expectedly write before he gains admission into the university of his choice?

Activity 1	Let us consider that a container has 6 white balls and 4 black balls. Balls are randomly selected one at a time until a black ball is picked. Assume that each ball selected is replaced before the next one selected. What is the probability that (i) Exactly 3 draws are needed; (ii) Not more than two draws are needed; (iii) the mean number of draws until success (iv) the variance of the number of draws until a success.

Unit 2 | Module 3 | STA 124 Geometric Distribution

Solution

$$p(x) = p(1-p)^{x-1}, x = 1, 2, \cdots$$

The random variable X is the number of draws needed to pick the first black ball.

The probability of picking a black ball is 4/10=0.4.

Hence $p(x) = 0.4(0.6)^{x-1}$

i. $p(3) = (0.4)(0.6)^2 = 0.144$

ii. P(not more than two draws needed) $P(X \leq 2) = p(1) + p(2) = 0.4 + 0.4(0.6) = 0.64$

iii. $E(X) = \frac{1}{p} = \frac{1}{0.4} = 2.5$

iv. $Var(X) = \frac{q}{p^2} = \frac{0.6}{0.4^2} = 3.75$

Activity 2	A businessman has decided to keep applying to different banks for a loan until he gets the loan. Suppose that the probability of the man getting a loan from any bank is 0.1 (a) what is the expected number of applications he will make? (b) what is the probability that he will make not more than 3 applications? (c) compute the variance of the number of applications the businessman will make before success.

Solution

a. The probability of success, p=0.1 and the expected number of applications is

$$E(X) = \frac{1}{p} = \frac{1}{0.1} = 10$$

b. The probability mass function of the geometric distribution is

$$p(x) = p(1-p)^{x-1}, x = 1, 2, \cdots$$

$$P(X \leq 3) = p(1) + p(2) + p(3) = p + p(1-p) + p(1-p)^2$$
$$= 0.1 + 0.1(0.9) + 0.1(0.9)^2 = 0.271$$

C. $Var(X) = \frac{q}{p^2} = \frac{0.9}{(0.1)^2} = 90$

Summary

The two categorizations of the geometric distribution were presented in this unit. Greater attention was put at the categorization detailing the number of trials until the first success of independent Bernoulli trials with probability of success, p. The geometric distribution is different from the binomial distribution in the sense that, the number of trials is the random variable of interest, unlike the binomial distribution whose corresponding random variable is the number of successes in a fixed number of Bernoulli trials.

Lastly, the properties of the geometric distribution were also specified, together with real-life applications of the distribution.

Self-Assessment Questions

1. A woman decided to stop giving birth only when she has had a baby boy. What is the probability that she had the baby boy at the fourth delivery if she has an equal likelihood of giving birth to a boy or a girl?

2. A woman decided to stop giving birth only when she has had a baby boy. What is the probability that she had the baby boy between the third and the sixth deliveries? (Pr(boy)=Pr(girl)=0.5).

3. The lifetime risk of developing pancreatic cancer is about one in78. Let X = the number ofpeople you ask until one says he or she has pancreatic cancer. (a). What is the probability that you ask ten people before one says he or she has pancreaticcancer? (b). What is the probability that you must ask 20 people? (c). Find the mean and standard deviation of X.

4. Balls are tossed at random into 50 boxes. Find the expected number of tosses required to get the first ball in the fourth box.

5. The probability that a missile hits a target is p. If missiles are fired independently at a target, until it is hit, what is the probability that it takes more than three missiles to hit the target?

Tutor Marked Assignment

1. A football enthusiast in Ilorin randomly selects people from a randomly chosen area of the town until he finds a person who attended the last home football game of the local league side, Kwara United. Let p, the probability that he succeeds in finding such a person, equal 0.20. Let X denote the number of people he selects until he finds his first success. What is the probability that the football enthusiast must select 4 people before he finds a person who attended the last home football game?

2. A football enthusiast in Ilorin randomly selects people from a randomly chosen area of the town until he finds a person who attended the last home football game of the local league side, Kwara United. Let p, the probability that he succeeds in finding such a person, equal 0.20. Let X denote the number of people he selects until he finds his first success. What is the probability that he must select more than 6 people before he finds a person who attended the last home football game?

3. You play a game of chance that you can either win or lose until you lose. Your probability of losingis p = 0.57. What is the probability that it takes five games until you lose?

4. A safety engineer feels that 25% of all industrial accidents in her plant are caused by the failure of employees to follow instructions. She decides to look at the accident reports (selected randomly and replaced in the pile after reading) until she finds one that shows an accident caused by the failure of employees to follow instructions. (a) On average, how many reports would the safety engineer expect to look at until she finds a report showing an accident caused by employee failure to follow instructions? (b) What is the probability that the safety engineer will have to examine at least three reports until she finds a report showing an accident caused by employee failure to follow instructions?

5. Assume that the probability of a defective computer component is 0.02. Components are randomly selected and tested until one component fails.

   a. Find the probability that the first defect is caused by the seventh component tested.

b. How many components do you expect to test until one is found to be defective?

Further Reading

- https://en.wikipedia.org/wiki/Geometric_distribution
- https://www.youtube.com/watch?v=d5iAWPnrH6w

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Statistics

Photo: freepik.com

UNIT 3

Poisson Distribution

Introduction

The Poisson distribution is named after the famous French mathematician, Simeon-Denis Poisson (1781-1840), who published his results about the distribution in 1838. One of the earliest applications of the Poisson distribution was to the number of Prussian soldiers annually kicked to death by their horses. The fit to a Poisson distribution is remarkably good, although the distribution is appropriate as long as the probabilities of success in the binomial experiment do not vary too much.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Identify the conditions that give rise to Poisson experiments; Enumerate the properties of the CDF.
2. Define the probability mass function of the Poisson distribution;
3. Identify some basic properties of the distribution; and
4. Provide examples of real-life occurrences that follow the Poisson distribution.

Main Content

SAQ 2

Poisson Distribution

7 mins

I want you to know that the Poisson distribution is a limiting form of a binomial distribution B(n,p). For the binomial distribution, the goal was to look for the probability of a specific value of success in *n* trials. For the Poisson distribution, interest is now on looking for the specific number of occurrences in a specific time interval (e.g. 1 hour, 1 day, 1 week, 2 weeks, 1 year, etc).

The Poisson distribution is a discrete probability distribution that is useful when *n*, the number of trials is large and *p*, the probability of success is small and when the random phenomena occur over a specified period.

In addition to being used for the stated conditions (i.e., *n* is large, *p* is small, and the variables occur over some time), the Poisson distribution can be used when a density of items is distributed over a given area or volume, such as the number of plants growing per acre or the number of defects in a given length of videotape.

SAQ 1

Conditions of a Poisson Experiment

1. The experiment consists of counting the number of events occurring in a fixed interval of time or space if these events happen with a known average rate and independently of the time since the last event.
2. The probability of the event remains constant for each interval of equal length.
3. The number of occurrences in one fixed interval is 0independent of the number of occurrences in other fixed intervals.
4. The random variable *X* denotes the number of occurrences in the interval of interest.

For example, a book editor might be interested in the number of words spelt incorrectly in a particular book. It might be that, on the average, there are five words spelt incorrectly in 100 pages. The appropriate interval is therefore 100 pages.

Let the random variable X represent the number of occurrences of an event in specified time or space interval with mean λ occurrences per unit time (or space), then X is said to have a Poisson distribution with parameter λ, that is $X{\sim}P(\lambda)$.

The probability mass function of the Poisson distribution is

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0,1,2,\cdots$$

The letter *e* is a constant approximately equal to 2.7183.

The Poisson distribution provides a realistic model for many random phenomena. Since the values of a Poisson random variable are the non-negative integers, any random phenomenon for which a count of some sort is of interest is a candidate for modeling by assuming a Poisson distribution. Such a count might be the number of fatal traffic accidents per week in a given state, the number of radioactive particle emissions per unit of time, the number of telephone calls per hour coming into the switchboard of a large business, the number of defects per unit of some material and so on.

SAQ 3

Properties of the Poisson Distribution

If a discrete random variable X follows the Poisson distribution, then the following are true for the distribution:

1. X assumes integer values 0, 1, 2, ...
2. The average number of occurrences per unit of time (or any other specified measure) is known.
3. There is a non-zero probability of non-occurrence of the event in any specified interval.

4. E(X)= λ and *Var(X)*= λ. Thus uniquely, the mean and the variance of the Poisson distribution is equal to the Poisson parameter, λ

5. The Poisson distribution can also be used to approximate the binomial distribution when the expected value np is less than 5. In such instances $\lambda = np$

SAQ 4-10

Examples

Example1: If there are 200 typographical errors randomly distributed in a 500-page manuscript, let us see how we can calculate the probability that a given page contains exactly 3 errors.

Solution

Since there are 200 errors in 500 pages, then the mean number of errors per page is $\lambda = \frac{200}{500} = 0.4$ errors per page. Hence X~P(0.4).

$$P(X = 3) = p(3) = \frac{(0.4)^3}{3!}e^{-0.4} = 0.0072$$

There is less than a 1% chance that any given page in the manuscript will contain exactly 3 errors.

Example2: Suppose that the average number of telephone calls arriving at the switchboard of Dangote PLC is 30 calls per hour.(I) What is the probability that no calls will arrive in a 3-minute period?(ii)What is the probability that more than five calls will arrive in a 5–minute period?

Solution

In order to obtain the rate λ, we must determine the relevant interval for each question.

(i) If 30 calls arrive on the average every hour, then the average number of calls that will arrive in a 3-minute period would be $(30/60) \times 3 = 3/2 = 1.5$ calls per 3-minute interval.

Hence $X \sim P(1.5)$ and $P(X = 0) = p(0) = \frac{(1.5)^0}{0!}e^{-1.5} = 0.2231$

There is a 22.3% possibility that no call will arrive at the switchboard in a 3-minute period.

(ii) Following the same argument in (i) above, $\lambda = (30/60) \times 5 = 2.5$ calls per 5-minute period.

$$P(X > 5) = 1 - P(X \le 5) = 1 - \left(p(0) + p(1) + p(2) + p(3) + p(4) + p(5)\right) = 0.42$$

Therefore, the probability of having more than 5 calls in a 5-minute period is 0.42.

Example 3: If approximately 1% of the people in a room of 250 people are divorcees, (a) Let us find the probability that exactly 8 people will be divorcees? (b) What is the probability that there will be no divorcee? (c) What is the probability that there will be at least one divorcee?

Solution

(a) This is ordinarily a binomial problem with n=250 and p=0.01. However, we can easily approximate it with the Poisson distribution, since

which is less than 5, hence we can approximate it with a Poisson distribution with $\lambda = np = 2.5$.

Therefore

$$P(X = 8) = p(8) = \frac{(2.5)^8}{8!}e^{-2.5} = 0.0031$$

(b) $P(X = 0) = p(0) = e^{-2.5} = 0.0821$

(c) $P(X \ge 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = 1 - 0.0821 = 0.9179$

There is thus a 92% chance that one or more divorcees will be found in the room.

Activity 5

Go to any busy road around you and make a count of the vehicles that passby within a space of 10minutes. Tabulate your observations as follows.

Vehicle type	Cars	Tricycles	Buses	Others
No. of Observation				

Calculate the poison variable from your table

Summary

You have learned that the Poisson distribution was named after the French mathematician, Simeon-Denis Poisson and can be used to model any counting process. It is also a limiting form of the binomial distribution when n is large and p is small so that np is less than 5. Such events with a very small probability of occurring are called rare events. The Poisson distribution has its mean equal to the variance.

Self-Assessment Questions

1. List the conditions that give rise to Poisson experiments.
2. Define the probability mass function of the Poisson distribution.
3. State the properties of the Poisson distribution.
4. Give 6 examples of Poisson random variables.
5. If X follows a Poisson distribution with parameter λ=2.93, find the probability that (a) $P(X \leq 1)$; (b) $P(x < 3)$; (c) $P(1 < x < 4)$; (d) $P(x > 2)$.
6. If X follows a Poisson distribution with parameter λ=1.83, calculate (a) Pr(X=4); (b) Pr(X≤1).
7. Suppose that particles are emitted from a radioactive source and that the number of particles (x) emitted during a one-hour period has a Poisson distribution with parameter Φ=3. . Evaluate P(x≤1).
8. Suppose that flaws in plywood occur at random with an average of one flaw per 50 square feet. What is the probability that a 4 foot by 7-foot sheet will have no flaws?
9. Suppose that flaws in a particular plywood occur at random with an average of 4 flaws per 50 square feet. What is the probability that a 5 foot by 5-foot sheet will have no flaws?
10. Suppose X~P(0.5), compute

Tutor Marked Assignment

1. Suppose that particles are emitted from a radioactive source and that the number of particles (x) emitted during a one-hour period has a Poisson distribution with parameter =4. Evaluate the P(X=0)?
2. Suppose that flaws in plywood occur at random with an average of one flaw per 50 square feet. What is the probability that a 32 square feet sheet will have (a) no flaws (b) at most one flaw?
3. A sales firm receives, on average, 3 calls per hour on its toll-free number. For any given hour, find the probability that it will receive (a) At most 3 calls (b) At least 3 calls (c) 5 or more calls?
4. If approximately 2% of the people in a room of 200 people are left-handed, find the probability that exactly 5 people there are left-handed.
5. Suppose that in a town of about 1000 inhabitants, the probability that a person lives to 100 years is 0.0025, what is the probability that 5 people will live to 100 years in that town?

Further Reading

- https://en.wikipedia.org/wiki/Poisson_distribution
- https://brilliant.org/wiki/poisson-distribution/

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: mathematical

source: pexels.com

Module 4
Continuous Probability Distributions

Units

Unit 1 - Uniform Distribution
Unit 2 - Normal Distribution

Picture: Statistics

Photo: freepik.com

UNIT 1

Uniform Distribution

Introduction

Welcome to the Unit 1 of the fourth module of this course. In this unit, you will learn about the Uniform distribution. The uniform distribution is a distribution whose values are evenly distributed over its range. For instance, when you throw a fair die, you have equal probability of 1/6 on your outcome. Such outcomes are said to be uniformly distributed. Also in this unit, we will attempt to explain the two types of uniform distribution.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Identify situations that may be modeled by the uniform distribution;
2. Differentiate between the discrete and continuous uniform distribution;
3. Obtain the mean, variance and cumulative distribution of various uniform random variables;
4. Compute various probabilities for both the discrete and continuous uniform distributions; and
5. Provide some applications of the uniform distribution.

Main Content

Uniform Distribution

15 mins

May I let you know that the discrete uniform distribution is a symmetric probability distribution whereby a finite number of values are equally likely to be observed; every one of n values has equal probability 1/n.

SAQ 1,3

Discrete Uniform Distribution

Definition: A discrete random variable X taking values 1,2, ..., n such that the probability mass function is given as

$$P(X = x) = p(x) = \frac{1}{n}, \quad x = 1,2,\dots,n$$

is said to have a **discrete uniform distribution.**

Note that an alternative and more general categorization of the discrete uniform distribution is that the random variable X takes values between a and b such that the PMF is given as

$$p(x) = \frac{1}{b-a+1}, x = a, a+1, a+2, \dots, b$$

The uniform distribution is appropriate for any situation in which each of the possible outcomes of the random process or experiment has an equal probability of occurring. That is, the uniform distribution is appropriate for *equiprobable* discrete distributions.

The cumulative distribution function of the discrete uniform distribution form $x \in (a, b)$ is

$$P(X \leq x) = F(x) = \frac{x-a+1}{b-a+1}$$

The range of X may be between 1 and n, or any discrete range e.g. between 3 and 10, between -4 and +4, and so on. What should be noted is that each possible outcome in the sample space has the same probability, which is equal to the reciprocal of the number of elements in the sample space.

SAQ 1,3

We have many examples of the discrete uniform distribution. A simple example is throwing a fair die. The possible values are 1, 2, 3, 4, 5, 6, and each time the die is thrown the probability of a given score is 1/6. If two dice are thrown and their values added, the resulting distribution is no longer uniform since not all sums have equal probability. Other simple examples of the discrete uniform distribution are tossing a fair coin, any Bernoulli trial with p=1/2, randomly picking n students from a population, and so on.

The discrete uniform distribution itself is inherently non-parametric. It is convenient, however, to represent its values generally by all integers in an interval [a,b], so that a and b become the main parameters of the distribution (often one simply considers the interval [1,n] with the single parameter n).

SAQ 1,3

Continuous Uniform Distribution

Definition: A continuous random variable X, defined on the interval (a,b) on the set of real numbers, has a **continuous uniform distribution** if

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

I want you to note that the continuous uniform distribution or rectangular distribution is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's range are equiprobable. The range is defined by the two parameters, a and b, which are its **minimum** and **maximum** values. The distribution is often abbreviated U(a,b).

One of the example of the continuous uniform distribution is the CDF of any continuous random variable.

The cumulative distribution function of the continuous uniform distribution for $x \in (a, b)$ is given

$$P(X \le x) = F(x) = \frac{x-a}{b-a}$$

The probability that a uniformly distributed random variable falls within any interval of fixed length is independent of the location of the interval itself (but it is dependent on the interval size), so long as the interval is contained in the distribution's range.

Restricting a = 0 and b = 1, the resulting distribution U(0,1) is called a standard uniform distribution. An important property of the standard uniform distribution is that if *X* has a standard uniform distribution, then so does 1 - *X*. This property can be used for generating antithetic variates, among other things.

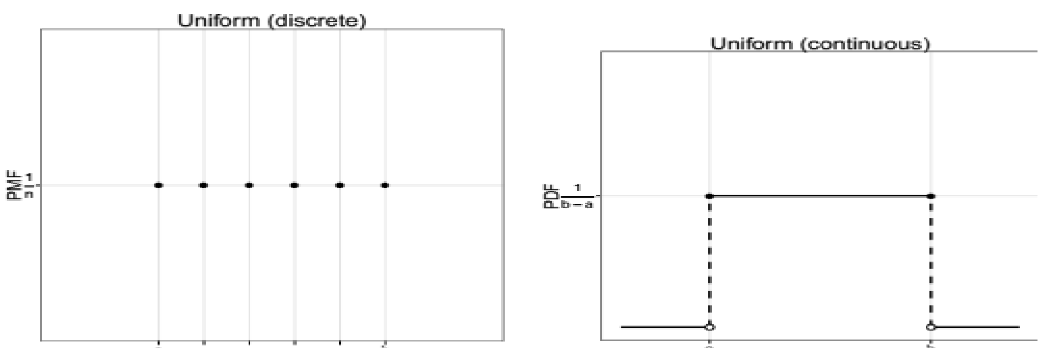


Figure 1: Graph of the probability mass function of the discrete uniform distribution and the probability density function of the continuous uniform distribution.

SAQ 4-10

Applications of the Uniform Distribution

1. In statistics, when a p-value is used as a test statistic for a simple null hypothesis, and the distribution of the test statistic is continuous, then the p-value is uniformly distributed between 0 and 1 if the null hypothesis is true.

2. There are many applications in which it is useful to run simulation experiments. Many programming languages come with implementations to generate pseudo-random numbers which are effectively distributed according to the standard uniform distribution.

3. The uniform distribution is useful for sampling from arbitrary distributions. A general method is the inverse transform sampling method, which uses the cumulative distribution function (CDF) of the target random variable. This method is very useful in theoretical work.

Table 1 below provides some properties of the uniform distribution.

Table 1: Summary of some properties of the discrete and continuous uniform distribution.

	Range	PMF/ PDF	E(X)	Var(X)
Discrete	$x = a, a+1, \ldots, b$	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Continuous	$x \in (a, b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

SAQ 5-10

Application 1: A fair octahedral die is tossed once.(a) What is the probability mass function of this experiment? (b) compute the mean; (c) compute the variance; (d) compute $E(X^2)$; (e) Find the probability of having a score not greater than 3?

Solution

This is an example of the discrete uniform distribution. The random variable X takes the values 1,2,3,4,5,6,7,8. The minimum a=1 and the maximum b=8.

(a) The PMF is given as $p(x) = \frac{1}{8}, \quad x = 1,2,\ldots,8$

(b) $E(X) = \frac{a+b}{2} = \frac{1+8}{2} = \frac{9}{2} = 4.5$

(c) $Var(X) = \frac{(b-a+1)^2 - 1}{12} = \frac{63}{12} = 5.25$

(d) $E(X^2) = Var(X) + \left(E(X)\right)^2 = 5.25 + (4.5)^2 = 25.5$

(e) $P(X \leq 3) = F(3) = \frac{x-a+1}{b-a+1} = \frac{3}{8} = 0.375$

It can be seen that applying the principles already specified in this unit, we could compute any of the above entities and probabilities, given any discrete uniform distribution.

Application 2: Given a continuous uniform distribution

$$f(x) = \frac{1}{8}, \qquad 0 \leq x \leq 8$$

(a) compute the mean; (b) compute the variance; (c) compute $E(X^2)$; (d) find the probability of having a score not greater than 3? (e) obtain $\Pr(1 \leq x \leq 6)$

Solution

(a) $E(X) = \frac{a+b}{2} = \frac{0+8}{2} = \frac{8}{2} = 4$

(b) $Var(X) = \frac{(b-a)^2}{12} = \frac{64}{12} = 5.33$

(c) $E(X^2) = Var(X) + \left(E(X)\right)^2 = 5.33 + (4)^2 = 21.33$

(d) $P(X \leq 3) = F(3) = \frac{x-a}{b-a} = \frac{3}{8} = 0.375$

(e) $\Pr(1 \leq x \leq 6) = F(6) - F(1) = \frac{6}{8} - \frac{1}{8} = \frac{5}{8} = 0.625$

It can be deduced from both examples that the discrete and continuous uniform distributions differ in some ways. While the PMF and PDF of the two examples are the same (1/8), the corresponding variances and cumulative probabilities differ. The continuous variant of the distribution had a slightly higher value for the variance.

Indeed, the CDF of the discrete uniform distribution is a step function, the CDF of the continuous uniform distribution is a smooth function on the interval (0,1).

Summary

You have learned that the uniform distribution is a unique probability distribution that has both discrete and continuous variants. If X~U(a,b), (whether discrete or continuous), then the expected value and the variance of the distribution are of the same form, as presented in table 1. There are many applications of uniform distribution in practice. A simple example of the discrete uniform distribution is the tossing of a fair die, while an example of the continuous uniform distribution is the cumulative distribution function of any continuous random variable.

Self-Assessment Questions

1. Define both the discrete and continuous uniform distributions.
2. Give some common examples of discrete and continuous uniform distributions.
3. Different between the discrete and continuous uniform distributions.
4. Give 3 applications of the uniform distribution in statistics.
5. If X has a Uniform distribution over the interval [3, 10]. Obtain the mean and variance of X.
6. A twelve-faced die with the numbers 1,2,...,12 is tossed once. (a) Obtain the probability mass function; (b) Determine the probability of obtaining a number greater than 10? (c) obtain the mean score that turns up (d) Compute $P(1 \leq x \leq 5)$.
7. If X follows a Uniform distribution over the interval (0, 12). Find the Pr(3< X < 7).
8. Suppose X follows a Uniform distribution over the interval (0, 12), obtain the mean and variance of the distribution.
9. Obtain the mean and variance of the discrete random variable X which follows the uniform distribution with range [0, 12].

10. Obtain the mean and standard deviation of the discrete random variable X which follows the uniform distribution with range [2, 9].

Tutor Marked Assignment

1. If f(x)=1/2, 0< x < 2, then obtain F(1.5)
2. If f(x)=1/2, 0< x < 2, then obtain E(X) and Var(X).
3. If X follows a Uniform distribution over the interval (2, 10). Find the Pr(3< X < 7).
4. If X follows a Uniform distribution over the interval (2, 10). Find the Pr(X < 7).
5. If X follows a Uniform distribution over the interval (2, 10). Find the Pr(X>3).
6. If X follows a Uniform distribution over the interval (2, 15). Find the variance of x.
7. If X follows a Uniform distribution over the interval (0, 10). Find the standard deviation of X.
8. X~U(0,1). Compute $P(0.1 \le x \le 0.6)$.
9. X~U(0,1). Compute E(X) and Var(X)
10. A fair die is tossed once and the outcome recorded. What is the probability of having a number less than 4?

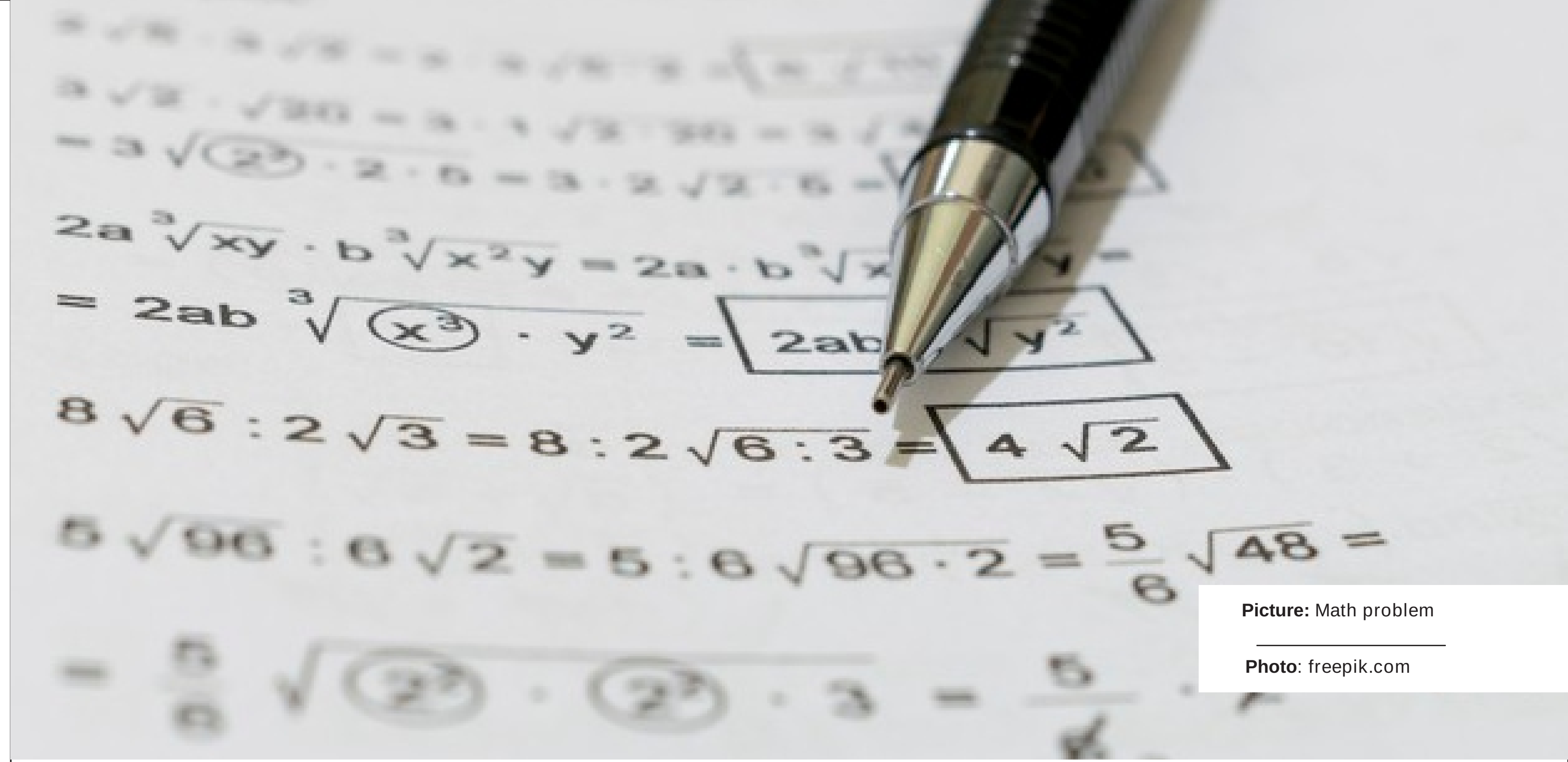
Further Reading

- https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)
- https://en.wikipedia.org/wiki/Discrete_uniform_distribution
- https://www.youtube.com/results?search_query=Uniform+distribution

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.

- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA



Picture: Math problem

Photo: freepik.com

UNIT 2

Normal Distribution

Introduction

In the previous unit, you have learnt about Uniform distribution and its types. Here, we will look at the Normal Distribution. A normal distribution is used to model a wide range of continuous random variables like Age, Height, Weight, Thickness, etc. We usually use the Mean and Variance as the parameters of the normal distribution

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the normal distribution;
2. State the important properties of the normal distribution;
3. Use the cumulative distribution of the normal distribution to obtain probabilities in an interval (i.e. Pr(a<X<b)).
4. Use the standard normal table to calculate cumulative probabilities of the normal distribution; and
5. Mention some applications of the normal distribution.

Main Content

SAQ 1

The Normal Distribution

10 mins

Do you know that the normal distribution is the most widely used distribution in statistics? Continuous data such as mass, length, etc, can often be modelled using a normal distribution.

Definition: A continuous random variable X is defined to be a normal random variable with parameters μ and σ^2 if its probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$$

The normal distribution has two parameters – the mean (μ) and variance (σ^2) If a random variable follows the normal distribution with mean μ and variance, σ^2, it can be written in shorthand form as $X \sim N(\mu, \sigma^2)$.

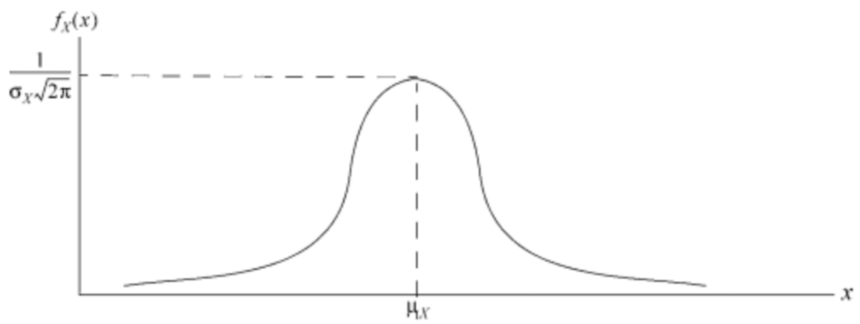


Figure 1: Probability density function of the normal distribution.

SAQ 2

Properties of the Normal Distribution

1. The curve is symmetrical about the vertical axis through the mean.
2. It has a bell shape.
3. The mean, mode and median coincide at the point μ
4. The variance σ^2 the shape of the curve: it tends to be flat when σ^2 is large,

and peaked when σ^2 is small.

5. The total area under the curve and above the horizontal axis is equal to 1.
6. $E(X) = \mu$ and $Var(X) = \sigma^2$, are the two parameters of the distribution.
7. P(-σ < X - μ < σ) = 0.68; P(-2σ < X - μ < 2σ) = 0.95; P(-3σ < X - μ < 3σ) = 0.997. That is, the probability that a value centered on the mean lies within 1, 2 and 3 times the standard deviation σ of the distribution is 0.68, 0.95 and 0.997, respectively.

SAQ 3

Cumulative distribution function of the Normal distribution

Note that the cumulative distribution function of the normal distribution is given as

$$F(x) - P(X \le x) - \int_{-\infty}^{x} f(k)dk$$

Due to the complicated nature of the normal distribution PDF, the above integral may not be evaluated explicitly. However, the value of the integral has been extensively tabulated for the normal distribution with zero mean and unit variance, that is $\mu = 0$ and $\sigma^2 = 1$.

If a random variable X follows the normal distribution N(μ, σ^2), then the transformation

$$Z = \frac{X-\mu}{\sigma}$$

is normally distributed with zero mean and unit variance, N(0,1), called th **standard normal distribution.**

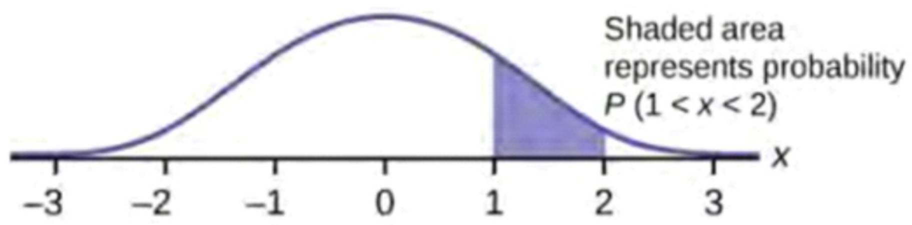


Figure 2: Graph of the cumulative distribution function of the standard normal distribution.

The corresponding CDF of the standard normal distribution is thus

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\left(\frac{y^2}{2}\right)}$$

where z is the standard normal transformation of the random variable Z.

The values of $\phi(z)$ have been tabulated and are available as the standard normal statistical table (Z-table).

If $X \sim N(\mu, \sigma^2)$, then

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \phi(b^*) - \phi(a^*).$$

where a^* and b^* are the Z-transformation of a and b, respectively.

It, therefore, suffices to transform the limits of the probability into Z-scores and look up the corresponding values in the standarized normal distribution table to obtain the cumulative probability.

Explanation

SAQ 4-7

Given a normal distribution with parameters μ and σ^2, in order to compute the probability of any specified interval (a,b), just obtain the corresponding z-transformation of the limits a and b and look up the transformed values in the statistical table. The final step is to subtract the table value for the lower limit from that of the higher limit to obtain the required probability.

Activity 1: Let us find the probability that P(1.5 < Z < 2.3)

Solution: Since the probability is already given in z-score, we only need to look up in the statistical tables ϕ(2.3) and ϕ 1.5 and then compute ϕ (2.3) - ϕ (1.5)= 0.9893 – 0.9332= 0.0561

Activity 2: The scores in an online test was found to be normally distributed with mean 50 and standard deviation 10.

(a) What is the probability of obtaining a score less than 5?

(b)What is the probability of obtaining a score between 45 and 55?

Solution:

(a) The first thing we have to do is to obtain the z-score of a=5, that is, subtract the mean from a and divide by the standard deviation (which is the square root of the variance). Let the z-score of a be represented by a^*. Hence a^*=0.5.

$P(X < 5) = P(Z < 0.5) = 0.6915$. (b) the limits are a=45 and b=55, while the corresponding z-scores are a^* = -0.5 and b^* = 0.5, Hence

$P(45 < x < 55) = P(-0.5 < Z < 0.5) = \phi(0.5) - \phi(-0.5) = 0.6915 - 0.3085$ = 0,3830.

Activity 3: The weight of boys in our class is normally distributed with mean 55kg and standard deviation 2kg. Find the probability that a randomly chosen boy will have a weight (i) between 50kg and 60kg (ii) more than 58kg (iii) less than 52kg.

Solution

(I) The first step is to convert the raw scores to standardized z-scores and read off the corresponding values from the statistical table.

$$P(50 < X < 60) = P(-2.5 < Z < 2.5) = \phi(2.5) - \phi(-2.5) = 0.9938 - 0.0062 = 0.9876$$

(ii) P(X > 58) = 1- P(X<58)=1-P(Z<1.5)= 1 – 0.9332= 0.0668.

Summary

We have discussed the normal distribution which is one of the most important probability distribution in statistics in this unit. The probability density function, cumulative distribution function and the Z-transformation, which transforms any Normal random variable to the standard normal distribution, were also discussed. Several other properties of the distribution were provided and its application to tests of hypotheses was given.

Self-Assessment Questions

1. Define a random variable that follows a Normal distribution, state its probability density function.
2. List 6 properties of the normal distribution.
3. Write out the cumulative distribution function of the normal distribution. Is this CDF easily computed?
4. The average ages of Nigerian Universities' Vice-Chancellors are normally distributed with mean 55 and variance 81. If a Vice-Chancellor is chosen at random, what is the probability that his age is (i) more than 52 years; (ii) less than 48 years; (iii) between 40 and 64 years?
5. The scores in a test are normally distributed with mean 60 and standard deviation 15, what is the probability of (i) obtaining a score less than 70; (ii) obtaining a score between 40 and 60?
6. In a sample of 97 Vice Chancellors with a mean age of 60 years and a standard deviation of 4 years, how many will be expected to be more than 63 years?
7. The mean and standard deviation in an exam was 60 and 15 respectively. Find the score in standard units of students receiving (i) 85; (ii) 40; (iii) 55 marks.

Tutor Marked Assignment

1. The age of Nigerian University's Vice Chancellors is normally distributed with mean 57 and variance 225. If a Vice-Chancellor is chosen at random what is the probability that his/her age is (i).more than 52 years (ii). less than 48 years (iii) between 40 and 64 years.
2. If $X \sim N(50,100)$ and $z=(x-\mu)/\sigma$, find $Pr(1.5<: Z<2.3)$.
3. The score in a test is distributed as normal with mean 50 and variance 100. What is the probability of obtaining a score less than 55?
4. If $X \sim N(55,4)$, find $Pr(X>58)$.

5. If $X \sim N(55,16)$, obtain the values of X corresponding to z=-1.96 and z=1.96.
6. If $X \sim N(70,100)$, find the values corresponding to the standard scores 0 and 1.70.
7. The lifetime of a type of storage battery is known to be distributed normally with mean 5 and variance 0.25. Find $Pr(X<6.4)$.
8. The weight of fresh students of the University of Ilorin is normally distributed with mean 45kg and variance 121kg. If a student is picked at random, what is the probability that his/her weight is less than 40kg?
9. The weight of fresh students of the University of Ilorin is normally distributed with mean 45kg and variance 121kg. If a student is picked at random, what is the probability that his/her weight is 45kg?
10. Two students in a class gave their weights as 54kg and 75kg respectively which are normally distributed, and when expressed in standard units (z) had values 0.6 and 1.8 respectively. Find the variance of the class weight?

Further Reading

- https://en.wikipedia.org/wiki/Normal_distribution
- https://www.youtube.com/watch?v=VjjGcRP0zU4
- https://www.youtube.com/watch?v=2MgYDrGcn6c
- https://www.mathsisfun.com/data/standard-normal-distribution.html

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Office supplies

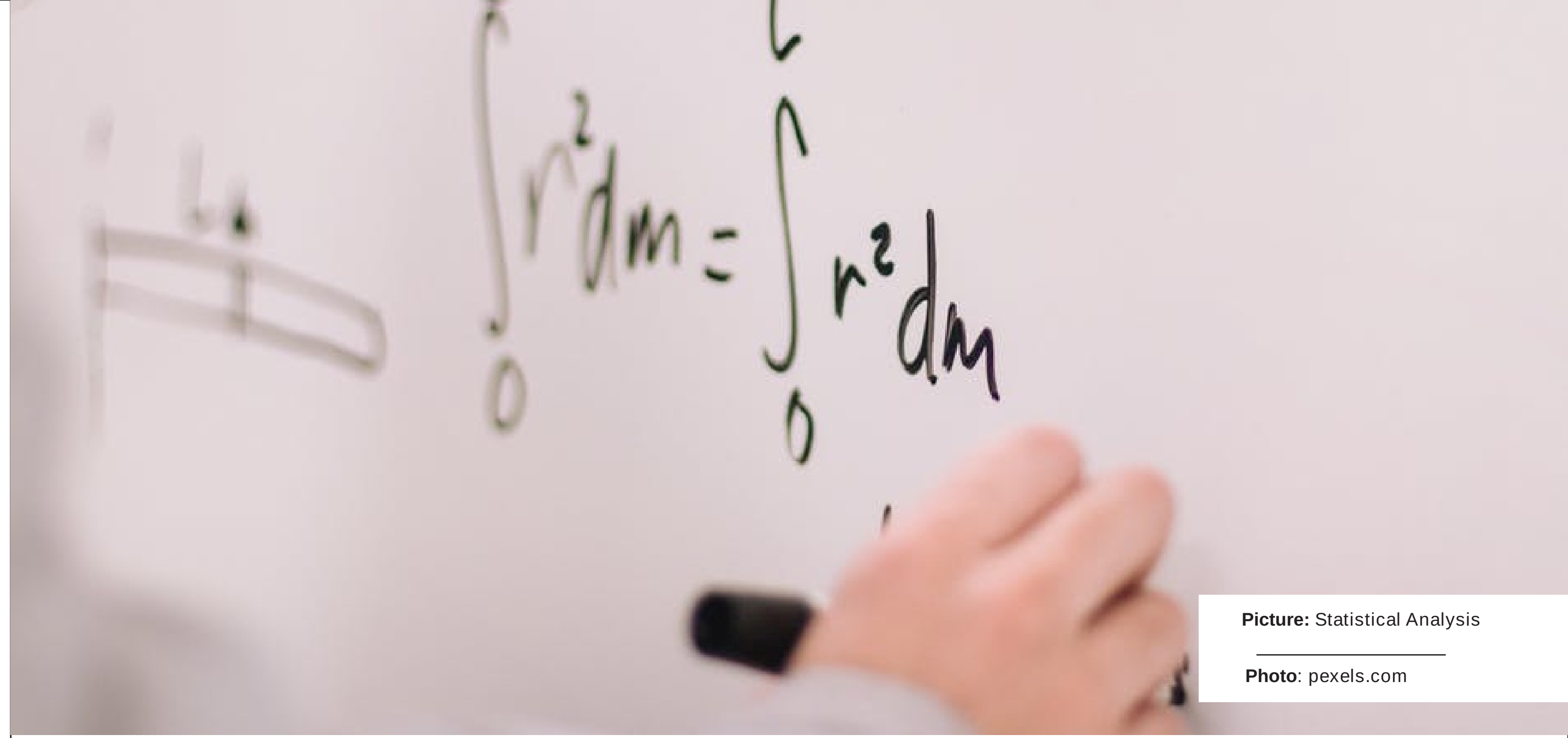
source: freepik.com

Module 5

Concepts of Bivariate Distributions

Units

Unit 1 - Bivariate probability distribution
Unit 2 - Conditional Distribution



Picture: Statistical Analysis

Photo: pexels.com

UNIT 1

Bivariate Probability Distribution

Introduction

In this unit, we will be discussing the theory of univariate probability distributions and how it is extended to bivariate distributions. Some basic properties of discrete and continuous bivariate distributions are provided. The important concept of statistical independence between two random variables is defined, as well as the marginal distributions. The covariance of a bivariate distribution is also presented.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the joint probability density function of a bivariate random variable (for both discrete and continuous cases);
2. Obtain the joint cumulative distribution function of a bivariate random variable;
3. Determine whether two random variables are independent;
4. Obtain the mean and variance of the marginal distributions;
5. Obtain the mean and variance of the marginal distributions;

Main Content

SAQ 1

Bivariate Probability Distribution

20 mins

A **bivariate distribution** is a probability that a certain event will occur when there are two **independent** random variables. For example, in the experiment of tossing two dice, the outcome of the first die is independent of the outcome of the second die. Thus the two outcomes give us two independent random variables. Since you are tossing the dice at the same time, the outcome of the first and second die gives a bivariate distribution.

Discrete Bivariate Distribution

Definition: Assuming *X* and *Y* are two discrete random variables, and let *S* denote the two-dimensional support (or the range) of *X* and *Y*. Then, the function *f(x,y) = P(X = x, Y = y)* is a **joint probability mass function** if it satisfies the following three conditions:

1. $0 \leq f(x,y) \leq 1$
2. $\sum\sum_{(x,y)\in S} f(x\ y) = 1$
3. $P[(X,Y) \in A] = \sum\sum_{x,y \in A} f(x,y)$ where A is a subset of the support *S*.

Explanation

SAQ 2

Condition 1 says that all joint PMF of a discrete bivariate random variable must have values between 0 and 1, as they are indeed probabilities. Condition 2 is another fundamental condition of univariate probability distribution extended to the bivariate case. It says that the sum of all the joint probabilities over the range of the two random variables (X,Y) is equal to 1. Condition 3 specifies the **joint cumulative distribution function,** which is the probability of event A and is equal to the joint sum of the PMF over subset A.

SAQ 3

Definition: Given a joint PMF f(x,y), where the random variables have support $X \in S_1$ and $Y \in S_2$ respectively. Then the **marginal probability mass function of X** $(f_1(x))$ and the marginal probability mass function of Y $(f_2(y))$, respectively are

$$f_1(x) = P(X = x) = \sum_y f(x,y), \qquad x \in S_1$$

$$f_2(y) = P(Y = y) = \sum_x f(x,y), \qquad y \in S_2$$

The marginal probability mass function of X, $f_1(x)$ is obtained by summing the joint PMF over the set of values of the random variable Y. It is the probability that a single outcome $x \in S_1$ occurs, given the joint PMF f(x,y).

Similarly, the marginal probability mass function of Y, $f_2(y)$ is obtained by summing the joint PMF over the set of values of X.

Definition: Two random variables X and Y are said to be independent if and only if

$$P(X = x, Y = y) = f(x,y) = f_1(x)f_2(y) = P(X = x)P(Y = y)$$

That is, random variables X and Y are independent if their joint PMF is equal to the product of their marginal PMF. One of the ways of establishing whether two random variables are independent is to obtain their marginal PMF $f_1(x)$ and $f_2(y)$ from their joint PMF $f(x,y)$. If $f_1(x)f_2(y) = f(x,y)$, then X and Y are said to be independent. Otherwise, the two variables X and Y are said to be **dependent.**

The above definition also holds for continuous random variables. In that case, the PMF is replaced by the PDF. Thus, for either discrete or continuous random variables X and Y, independence holds true if and only if the joint PMF (or PDF) is equal to the product of the marginal PMF (or PDF).

SAQ 4

Definition: Given a bivariate probability distribution with joint PMF f(x,y), the **joint expectation of X and Y, E(XY)** is given as

$$E(XY) = \sum_{Y}\sum_{X} xyf(x,y)$$

The joint expectation of X and Y, E(XY) is also an important way of testing the independence of random variables X and Y. This is because if X and Y are independent, then

$$E(XY) = \sum_{X} xf_1(x)\sum_{Y} yf_2(y) = E(X)E(Y)$$

where E(X) and E(Y) are the expected value (or mean) of X and Y, respectively.

Note: If this equality does not hold, then X and Y are not independent.

SAQ 5

The joint expectation of X and Y is also useful in computing the **covariance,** cov(X,Y) of the bivariate distribution:

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

You should note that the covariance is a measure of the joint dispersion of random variables X and Y around their individual means, and also lends support in establishing the strength of the linear relationship between the two random variables.

Activity

SAQ 6–8

Activity 1: An experiment consists of tossing two fair dice and recording the outcomes of the first and second die. Let random variable X represent the outcome of the first die and Y, the outcome of the second die. Obtain (a) the joint probability distribution; (b) the marginal probability distributions; (c) E(X), E(Y) and E(XY); (d) Cov(X,Y); (e) Determine whether X and Y are independent; (f) $P(X \le 2, Y \le 3)$

Solution

Each die has six faces numbered 1, 2, ..., 6, hence $x \in \{1,2,3,4,5,6\}$ and $y \in \{1,2,3,4,5,6\}$ possible events in the sample space and since the dice are fair, then all the events are is equiprobable and equal to 1/36. The table of the events is presented below:

		Y						
	$f(x,y)$	1	2	3	4	5	6	$f_1(x)$
X	1	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	2	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	3	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	4	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	5	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	6	1/36	1/36	1/36	1/36	1/36	1/36	6/36
	$f_2(y)$	6/36	6/36	6/36	6/36	6/36	6/36	1

(a) $f(x,y) = P(X = x, Y = y) = \frac{1}{36},\ x = 1,2,\dots,6, y = 1,2,\dots,6$

(b)
$$f_1(x) = \sum_{Y} f(x,y) = \frac{6}{36} = \frac{1}{6}, x = 1,2,\dots,6$$

$$f_2(y) = \sum_{X} f(x,y) = \frac{6}{36} = \frac{1}{6}, x = 1,2,\dots,6$$

(c)
$$E(X) = \sum_{x=1}^{6} xf_1(x) = (1+2+3+4+5+6)\left(\frac{1}{6}\right) = \frac{21}{6} = \frac{7}{3}$$

$$E(Y) = \sum_{x=1}^{6} yf_2(y) = (1+2+3+4+5+6)\left(\frac{1}{6}\right) = \frac{21}{6} = \frac{7}{3}$$

$$E(XY) = \sum_{y=1}^{6}\sum_{x=1}^{6} xyf(x,y) = \sum_{y=1}^{6} y(1+2+3+4+5+6)\left(\frac{1}{36}\right) =$$

$$\sum_{y=1}^{6} y(21)\left(\frac{1}{36}\right) = 21\left(\frac{21}{36}\right) = \frac{49}{9}$$

(d) $$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{49}{9} - \left(\frac{7}{3}\right)\left(\frac{7}{3}\right) = 0$$

(e) We can see that $f_1(x)f_2(x) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = f(x,y)$, hence, X and Y are independent.

We can also see that $E(XY) = E(X)E(Y)$, which then imply that $Cov(X,Y) = 0$.
Therefore, if X and Y are independent, then $E(XY) = E(X)E(Y)$, and $Cov(X,Y) = 0$
The covariance of two independent random variables is thus equal to zero.

(f) $$P(X \le 2, Y \le 3) = \sum_{y=1}^{3}\sum_{x=1}^{2} f(x,y) = \left(\frac{1}{36} + \frac{1}{36}\right)\left(\frac{1}{36} + \frac{1}{36} + \frac{1}{36}\right) = \frac{2}{36} \times \frac{3}{36} = \frac{1}{216}$$

Activity 2: Suppose $f(x,y) = \frac{x+y}{21}, x = 1,2,3; y = 1,2$. Obtain (a) the marginal PMFs; (b) E(X) and Var(X), (c) E(Y) and Var(Y), (d) Cov(X,Y).

Solution

(a) $$f_1(x) = \sum_{y=1}^{2} f(x,y) = \sum_{y=1}^{2} \frac{x+y}{21} = \frac{x+1}{21} + \frac{x+2}{21} = \frac{2x+3}{21}, \quad x = 1,2,3$$

$$f_2(y) = \sum_{x=1}^{3} f(x,y) = \sum_{x=1}^{3} \frac{x+y}{21} = \frac{1+y}{21} + \frac{2+y}{21} + \frac{3+y}{21} = \frac{6+3y}{21}, y = 1,2$$

(b) $$E(X) = \sum_{x=1}^{3} x f_1(x) = \sum_{x=1}^{3} \frac{2x^2+3x}{21} = \frac{2+3}{21} + \frac{2(4)+6}{21} + \frac{2(9)+9}{21} = \frac{23}{7} = 2.19$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \sum_{x=1}^{3} x^2 f_1(x) = \sum_{x=1}^{3} \frac{2x^3+3x^2}{21} = \frac{114}{21} = \frac{38}{7} = 5.43$$

$$Var(X) = \left(\frac{38}{7}\right) - \left(\frac{46}{21}\right)^2 = 0.6304$$

(c) $$E(Y) = \sum_{y=1}^{2} y f_2(y) = \sum_{y=1}^{2} \frac{6y+3y^2}{21} = \frac{11}{7}$$

$$E(Y^2) = \sum_{y=1}^{2} y^2 f_2(y) = \sum_{y=1}^{2} \frac{6y^2+3y^3}{21} = \frac{19}{7}$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{19}{7} - \left(\frac{11}{7}\right)^2 = \frac{12}{49} = 0.2449$$

(d) $$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$E(XY) = \sum_{y=1}^{2}\sum_{x=1}^{3} xyf(x,y) = \sum_{y=1}^{2}\sum_{x=1}^{3} \frac{xy(x+y)}{21} = \sum_{y=1}^{2}\sum_{x=1}^{3} \frac{x^2y+xy^2}{21} = \frac{72}{21} = \frac{24}{7} = 3.429$$

$$\therefore \quad Cov(X,Y) = \frac{24}{7} - \left(\frac{23}{7}\right)\left(\frac{11}{7}\right) = -\frac{85}{49} = -1.735$$

Thus, the covariance of X and Y is negative and is not equal to zero, implying that X and Y are not independent.

Definition: Let's assume that X and Y are two continuous random variables, defined on the range $x \in (a,b), y \in (c,d)$, where a, b, c and d are points on the real line $(-\infty, \infty)$ Then, the function f(x,y) is a joint probability density function if it satisfies the following conditions:

1. $f(x,y) \ge 0$
2. $\int_c^d \int_a^b f(x,y)dxdy = 1$
3. $P(X \le x, Y \le y) = \int_c^y \int_a^x f(x,y)dxdy$

The first condition means that the function must be nonnegative. The second condition implies that the volume defined by the support, the surface and the xy-plane must be 1. The third condition says that in order to determine the probability of an event A in a region $(X \in (a,x), Y \in (b,y)$, one must integrate the function f(x,y) over the space defined by the event A. That is, just as finding pobabilities associated with one continuous random variable involved finding areas under curves, finding probabilities associated with two continuous random variables involves finding volumes of solids that are defined by the event A in the xy-plane and the two-dimensional surface f(x,y).

We now proceed to define the marginal probability density function.

Definition: Given two continuous random variables X and Y, with joint PDF f(x,y) defined on the range $x \in (a,b), y \in (c,d)$. Then the marginal probability density function of X ($\boldsymbol{f_1(x)}$) and the marginal probability density function of Y ($\boldsymbol{f_2(y)}$), respectively are

$$f_1(x) = P(X = x) = \int_c^d f(x,y)dy$$

and

$$f_2(y) = P(Y = y) = \int_a^b f(x,y)dx$$

The marginal probability density function of X, $f_1(x)$ is obtained by integrating the joint PDF with respect to y over the range of the random variable Y. It is the probability that a single outcome $x \in (a,b)$ occurs, given the joint PDF f(x,y).

Similarly, the marginal probability density function of Y, $f_2(y)$ is obtained by integrating the joint PDF with respect to x over the range of X.

Definition: Given a continuous bivariate probability distribution with joint PDF f(x,y), the **joint expectation of X and Y, E(XY)** is given as

$$E(XY) = \int_c^d \int_a^b xyf(x,y)dxdy$$

As applied in the case of discrete bivariate distributions, the joint expectation can be used to establish the independence of continuous random variables. If X and Y are independent, then

$$E(XY) = \int_a^b xf_1(x) \int_c^d yf_2(y) = E(X)E(Y)$$

where E(X) and E(Y) are the expected value (or mean) of X and Y, respectively.

If this equality does not hold, then X and Y are not independent.

The covariance of X and Y is given as

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Activity 3: Given the joint probability density function

$f(x,y) = 4xy, 0 < x < 1, 0 < y < 1.$ Obtain (a) the marginal PDFs; (b) E(X) and Var(X), (c) E(Y) and Var(Y), (d) E(X,Y), (e) $P(X \leq 0.5, Y \leq 0.5)$, (f) are X and Y independent?

(a)
$$f_1(x) = \int_0^1 4xydy = \left.\frac{4xy^2}{2}\right|_0^1 = 2x(1^2 - 0) = 2x, \qquad 0 < x < 1$$

$$f_2(y) = \int_0^1 4xydx = \left.\frac{4x^2y}{2}\right|_0^1 = 2y(1^2 - 0) = 2y, \qquad 0 < y < 1$$

(b) Using the formulas given, then it can be shown that E(X)=2/3, $Var(X) = 1/18$. and $E(X^2) = 1/2$

(c) Similarly, E(Y)=2/3, $E(Y^2) = 1/2$ and $Var(Y) = 1/18$.

(d)
$$E(XY) = \int_0^1\int_0^1 xy(4xy)dxdy = \int_0^1 \left.\frac{4x^3y^2}{3}\right|_0^1 dy = \frac{4}{3}\left.\left(\frac{y^3}{3}\right)\right|_0^1 = \frac{4}{9}$$

(e)
$$P(X \leq 0.5, Y \leq 0.5) = \int_0^{0.5}\int_0^{0.5} 4xydxdy = \int_0^{0.5} \left.\frac{4x^2y}{2}\right|_0^{0.5} dy = \left.\frac{y^2}{4}\right|_0^{0.5} = \frac{1}{16}$$

(f) Since $f_1(x)f_2(y) = 2x(2y) = 4xy = f(x,y)$ random variables X and Y are independent.

Alternatively, it is seen that E(XY)=E(X)E(Y), also implying independence of X and Y.

Summary

Bivariate probability distributions have been introduced and discussed extensively in this unit. The basic conditions for a joint probability mass function (or joint probability density function) to exist were given. The concept of independence of two random variables was also presented. The joint expectation, marginal distribution and covariance were presented. Sufficient examples demonstrating the concepts were given to drive home the principles.

Self-Assessment Questions

1. For both discrete and continuous random variables, define the joint and marginal probability mass (or density) functions.
2. Give the necessary conditions for a function f(x,y) to be a valid PMF or PDF.
3. Give an example of a probability experiment that has a bivariate distribution.
4. Define the concept of independence of two random variables X and Y.
5. Define the covariance of two random variables X, and Y, Cov(X,Y)
6. Given that

$$f(x,y) = \begin{cases} \frac{1}{7}(x+3y), 0<x<1, & 0<y<2 \\ 0 \quad if \quad otherwise \end{cases}$$

Obtain (i) f1(x) (ii) f2(y) (iii) E(XY) (iv) Cov(X,Y) (v) $P(X \le 0.5, Y \le 1)$

7. Given a bivariate probability distribution

$$f(x,y) = \frac{xy^2}{15}, x = 1,2, \qquad y = 1,2$$

(a) Obtain the marginal distributions $f_1(x)$ and $f_2(y)$; compute (b) E(X) (c) E(Y) (d) Var(X), (e) Var(Y), (f) E(XY), (g) Cov(X,Y), (h) Are X and Y independent?

8. Calculate E(XY) if X and Y are continuous random variables having joint density function

$$f(x,y) = 6.75(x^2+3y^2); 0<x<1, \qquad 0<y<\frac{1}{3}$$

Tutor Marked Assignment

1. $f(x,y) = 0.3(2x^2y + y^2), 0 < x < 2; 0 < y < 1.$

(a) Obtain the marginal distributions $f_1(x)$ and $f_2(y)$; compute (b) E(X) (c) E(Y) (d) Var(X), (e) Var(Y), (f) E(XY), (g) Cov(X,Y).

2. Given that $f(x,y) = 1.2(x+y^2), 0<x<1, 0<y<1$, obtain the marginal PDF of X.
3. Given that $f(x,y) = x + y, 0<x<1, 0<y<1$. Determine whether X and Y are independent.
4. $f(x,y) = 2(x+y-2xy), 0<x<1, 0<y<1$, obtain E(XY)

Given that $f(x,y) = 1.2(x+y^2), 0<x<1, 0<y<1$, determine whether X and Y are independent.

Further Reading

- https://study.com/academy/lesson/bivariate-distributions-definition-examples.html
- https://newonlinecourses.science.psu.edu/stat414/node/104/
- https://newonlinecourses.science.psu.edu/stat414/node/107/

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Statistical Analysis

Photo: pexels.com

UNIT 2

Conditional Distribution

Introduction

In the previous unit, we have learned bivariate probability distribution. The conditional distribution of bivariate random variables will be presented in this unit. Your expedition through this unit will also avail you the opportunity to understand the steps through which the conditional distribution is obtained from the joint distribution. I will also explain the conditional expectation, conditional variance and conditional cumulative probability.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the conditional distribution of two random variables;
2. Obtain the conditional distribution for a joint probability distribution (discrete or continuous);
3. Define independence in terms of continuous distribution;
4. Obtain the mean and variance of the conditional distribution; and
5. Calculate cumulative probabilities of conditional distributions.

Main Content

Conditional Distribution

SAQ 1,3,6

15 mins

You should be aware that in probability theory, the **conditional probability** P(B|A) is the probability that an event B occurs after event A has already occurred.

Definition: We are given two jointly distributed random variables X and Y, the conditional probability distribution of X given Y (written as $f(x|y)$ is the probability distribution of X when Y is known to be a particular value. The conditional distribution contrasts with the marginal distribution of a random variable, which is its distribution without reference to the value of the other variable. Indeed, the conditional distribution can be defined in terms of the marginal distribution and the joint distribution. The conditional probability density function of X given Y, $f(x|y)$ is obtained by dividing the joint probability density function with the marginal probability density function of Y. That is

$$f(x|y) = \frac{f(x,y)}{f_2(y)}$$

Similarly, the conditional probability density function of Y given X is given as

$$f(y|x) = \frac{f(x,y)}{f_1(x)}$$

where $f_1(x)$ and $f_2(y)$ are the marginal probability density function of X and Y, respectively and $f(x,y)$ is the joint probability density function of X and Y.

So if the conditional distribution of Y given X is a continuous distribution, then its probability density function is known as the conditional density function. The properties of a conditional distribution, such as the moments, are often referred to by corresponding names such as the **conditional mean** and **conditional variance.**

More generally, we can refer to the conditional distribution of a subset of a set of more than two variables; this conditional distribution is contingent on the values of all the remaining variables, and if more than one variable is included in the subset then this conditional distribution is the conditional joint distribution of the included variables.

May I let you know that t he conditional probability density function (for continuous random variables X and Y) or conditional probability mass function (for discrete random variables) are true probability densities? hence they possess the characteristics of probability density functions. The value of the conditional distribution lies between 0 and 1, the sum (or integral) over the range of values is equal to 1, and the **cumulative distribution function** $P(X \le x|Y)$ and $P(Y \le y|X)$ are obtained in a similar manner to the marginal distribution or univariate probability distributions. These properties hold true for both discrete and continuous conditional distributions.

Moreso, the cumulative probabilities can also be obtained. For example, $P(X \le x|Y = y)$ and $P(Y \le y|X = x)$ can be obtained by substituting the specified value of the conditioned variable and then computing the corresponding cumulative probability.

Activity 1: Let us compute the conditional probability mass function $f(x|y)$ and $f(y|x)$ for the joint probability mass function given below.

$$f(x,y) = \frac{x+y}{21}, x = 1,2,3, \qquad y = 1,2$$

Solution

$$f(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{x+y}{21} / \frac{6+3y}{21} = \frac{x+y}{6+3y}, x = 1,2,3, \qquad y = 1,2$$

$$f(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{x+y}{21} / \frac{2x+3}{21} = \frac{x+y}{2x+3}, x = 1,2,3, \qquad y = 1,2$$

From the above activity, we can see that the conditional PMF is like a joint PMF. The difference is that the conditioned variable takes specific values.

Activity 2: Let us also compute $P(X \leq 2|Y=1)$ for the joint probability mass function given below.

$$f(x,y) = \frac{x+y}{21}, x = 1,2,3, \qquad y = 1,2$$

Solution

$$P(X \leq 2|Y=1) = \sum_{x=1}^{2} f(x|1) = \sum_{x=1}^{2} \frac{x+1}{6+3} = \frac{5}{9} = 0.55$$

Notice that the above argument also applies to the computation of the conditional cumulative probabilities for Y given X.

Now the interpretation of the conditional cumulative probability that was obtained in activity 2 above is that: the probability that X will have a value that is at least 2, given that Y takes a value of 1 is equal to 0.55. This probability is thus conditioned on an outcome that has already occurred (Y=1), hence the name conditional probability.

SAQ 4,7, 8,9,10

Sometimes it is of interest to compute the mean of a subset of a population that shares some property. For example, it may be of interest to compute the mean grade of those students who have passed an exam or the average age of professors who have doctoral degrees. The conditional expectation and the conditional variance are obtained the same way with the marginal distributions for both the discrete and continuous cases.

Activity 3: Let us compute the mean E(X|Y=1) and variance Var(X|Y=1) for the conditional distribution given below:

$$f(x|y) = \frac{x+y}{6++3y}, x = 1,2,3, \qquad y = 1,2$$

Solution

For us to compute the mean and variance of $f(x|1)$. the first step we are taking is to substitute the value of Y=1 into $f(x|y)$ and proceed to compute the mean and variance.

$$f(x|1) = \frac{x+1}{6+3(1)} = \frac{x+1}{9}, x = 1,2,3$$

The next step is to compute the mean for the above marginal PMF.

$$E(X|Y=1) = \sum_{x=1}^{3} xf(x|1) = \frac{\sum_{x=1}^{3}(x^2+x)}{9} = \frac{20}{9} = 2.22$$

$$E(X^2|Y=1) = \sum_{x=1}^{3} x^2 f(x|1) = \sum_{x=1}^{3} \frac{1}{9}(x^3+x^2) = \frac{2+12+36}{9} = \frac{50}{9} = 5.55$$

Therefore

$$Var(X) = E(X^2|Y=1) - \left(E(X|Y=1)\right)^2 = \frac{50}{9} - \left(\frac{20}{9}\right)^2 = \frac{50}{81} = 0.6173$$

SAQ 2,5

Independence and the Conditional Distribution

Let me tell notify you that when two random variables are independent, the joint probability density function of the two variables is equal to the product of the two marginal probability density functions. The concept of independence could also be defined from the viewpoint of the conditional distribution.

Definition: Let us say that X and Y are two random variables with joint probability density (or mass) function f(x,y), then X and Y are independent if

$$f(x|y) = \frac{f(x,y)}{f_2(y)} = f_1(x)$$

$$f(y|x) = \frac{f(x,y)}{f_2(y)} = f_2(y)$$

Thus, for two independent random variables X and Y with joint probability density (or mass) function f(x,y), the conditional distribution of X given Y is equal to the marginal distribution of X. Similarly, the conditional distribution of Y given X is equal to the marginal distribution of Y.

Note that the marginal distribution is the same as the conditional distribution for two independent random variables. Therefore, two random variables are said to be dependent if their conditional distributions differ from the marginal distributions.

Activity 4: we are given the joint probability density function

$f(x,y) = 4xy, 0 < x < 1, 0 < y < 1.$ Obtain (a) f(x|y) and f(y|x); (b) E(X|Y) and Var(X|Y); (c) Are X and Y independent?

Solution

a) for us to compute the conditional distributions f(x|y) and f(y|x), firstly, we have to obtain the marginal probability density functions:

$$f_1(x) = \int_0^1 4xydy = \frac{4xy^2}{2}\Big|_0^1 = 2x(1^2 - 0) = 2x, \qquad 0 < x < 1$$

$$f_2(y) = \int_0^1 4xydx = \frac{4x^2y}{2}\Big|_0^1 = 2y(1^2 - 0) = 2y, \qquad 0 < y < 1$$

Thereafter, the relationship between the joint PDF and the marginal PDF is then used to obtain the conditional probability density function.

$$f(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{4xy}{2y} = 2x$$
$$f(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{4xy}{2x} = 2y$$

We can see that the marginal probability density functions and the conditional distributions were the same.

(b) Using the formulas given, then E(X|Y)=2/3, $E(X^2|Y) = 1/2$ and $Var(X|Y) = 1/18$.

© Since $f(x|y) = f_1(x)$ and $f(y|x) = f_2(y)$, then X and Y are independent random variables.

Summary

In this unit, we have discussed the conditional probability distribution of joint random variables. It was shown that the conditional distribution of X given Y is just the ratio of the joint probability distribution and the marginal probability density (or mass) function of Y. The conditional expectation, conditional variance and conditional cumulative probabilities can be easily obtained from the conditional probability density function. The conditional distribution and the marginal distribution are the same when the two random variables are independent, otherwise, they differ in the case where the two random variables are not independent.

Self-Assessment Questions

1. Define the conditional probability mass function of X given Y.
2. Define independence in terms of the conditional probability distribution.
3. What is the formula for the expectations E(X|Y) and E(Y|X) for (a) discrete random variables; and (b) continuous random variables.
4. Under what conditions would the conditional variance Var(X|Y) be equal to the marginal variance Var(X)?
5. Explain how to compute the conditional cumulative probability $P(X \le x|Y = y)$.
6. Express f(x,y) in terms of the conditional distribution f(x|y) and the marginal distribution of Y.
7. Express f(x,y) in terms of the conditional distribution f(y|x) and the marginal distribution of X.
8. If X and Y are dependent random variables then f(x|y)=
9. Given $f(x,y) = x + y, 0 < x < 1,\ 0 < y < 1.$ Obtain f(x|y).

   Given $f(x,y) = x + y, 0 < x < 1,\ 0 < y < 1.$ Obtain (a) f(y|x); (b) E(Y|X); (c) $E(Y^2|X)$.

Tutor Marked Assignment

1. Given a bivariate probability distribution

   a. $f(x,y) = \frac{xy^2}{15}, x = 1,2,\ y = 1,2$. Obtain (a) f(y|x) (b) E(X|Y); (c) Var(X|Y); (d) Are X and Y independent?

2. $f(x,y) = 0.3(2x^2y + y^2), 0 < x < 2;\ 0 < y < 1$. obtain f(y|x).

3. Given that $f(x,y) = 1.2(x + y^2), 0 < x < 1, 0 < y < 1$, obtain f(x|y).

4. Given that $f(x,y) = x + y, 0 < x < 1, 0 < y < 1$. Determine whether X and Y are independent.

5. $f(x,y) = 2(x + y - 2xy), 0 < x < 1, 0 < y < 1$. (a) Obtan f(x|y); (b) Determine f(y|x); (c) Determine whether X and Y are independent.

Further Reading

- https://en.wikipedia.org/wiki/Conditional_probability_distribution
- https://www.statisticshowto.com/conditional-distribution
- https://www.khanacademy.org/math/ap-statistics/analyzing-categorical-ap/distributions-two-way-tables/v/marginal-distribution-and-conditional-distribution

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

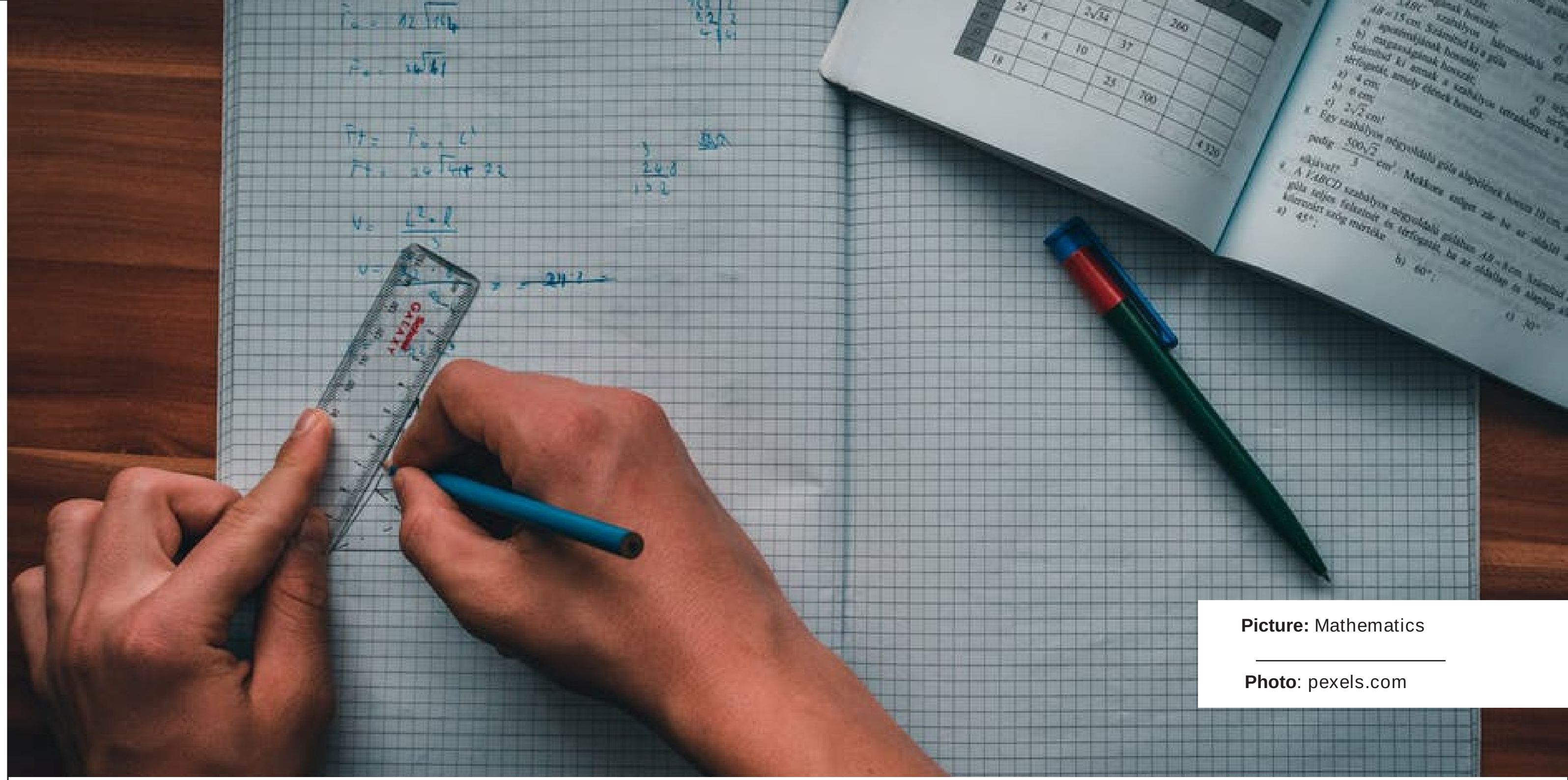
Picture: Statistical Analysis

source: pexels.com

Module 6

Mathematical Relationships Between Bivariate Distributions

Units



Picture: Mathematics

Photo: pexels.com

UNIT 1

Linear Regression

Introduction

At times, a distribution have a combination of two variables (e.g. X and Y), Welcome to Unit 1 of Module Six of this course. In this unit, I explain how you can determine the simple relationship between two variables. You will also be acquitted to the basic assumptions that governs simple linear regression.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the basic concepts of simple linear regression;
2. State the simple linear regression model and define its parameters;
3. List the assumptions of the simple linear regression model;
4. Estimate the parameters of the simple linear regression model; and
5. Predict the values of the dependent variable for fixed values of the independent variable.

Main Content

Linear Regression

SAQ 1,6–8

15 mins

May I interest you to know that an important area of inferential statistics involves determining whether a relationship exists between two or more numerical or quantitative variables.

Regression is a statistical method used to describe the nature of the relationship between variables, that is, positive or negative, linear or non-linear.

In simple regression, there are two variables – the **independent or explanatory variable**, and the **dependent or response variable**. The independent variable (X) is used to predict the dependent variable (Y).

In **multiple regression**, there are two or more independent variables that are used to predict the dependent or response variable.

A **simple linear regression model** is given as

$$Y = \alpha + \beta X + \varepsilon$$

where

α is called the intercept. It is the value of the dependent variable (Y) for which X=0.

β is the slope or gradient. It is the amount of change in Y for a unit change in X.

ε is the error variable, assumed to be normally distributed with mean zero and variance σ^2

The Scatter Plot

SAQ 5

Be aware that in simple regression, a scatter plot is used to portray the relationship between two variables; the relationship occurs in a sample of ordered (X,Y) pairs. It is constructed by plotting, on the rectangular coordinates axes, the X and Y measurements for each observation.

The scatter plot is used to illustrate diagrammatically, the relationship between the dependent and independent variables.

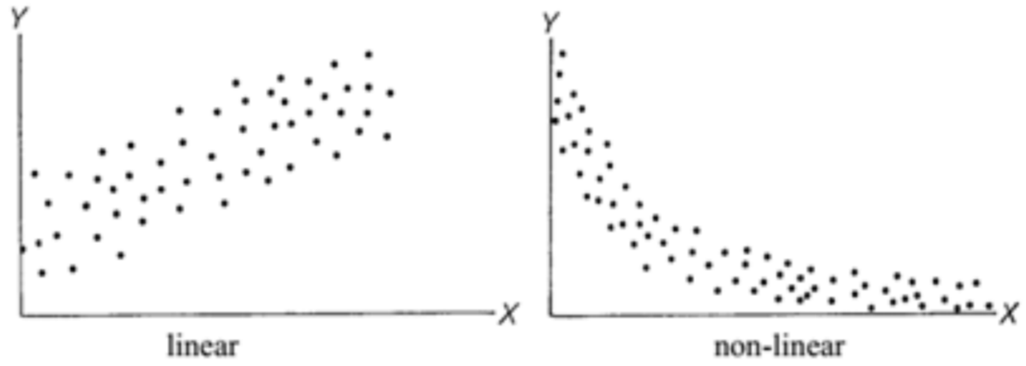


Figure 1: An example of a scatter diagram for linear and non-linear relationships.

Activity 1: Let us construct a scatter plot for the data obtained in a study on the number of absences and the final grades of seven randomly selected students from a statistics class, shown in the table below.

Student	Number of absences x	Final grade y (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78

Solution: The first step we are taking is to draw and label the x and y axes. Thereafter, the points (x,y) are plotted on the graph. The scatter plot is given below.

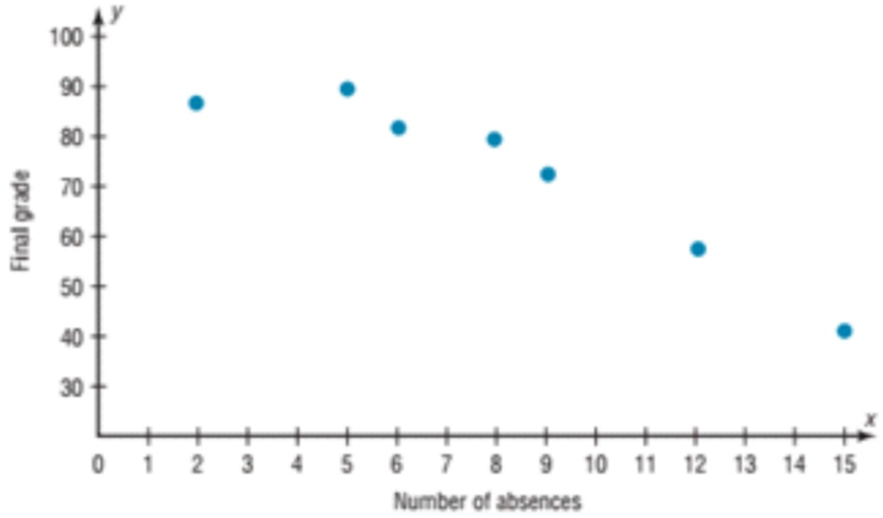


Figure 2: Scatter plot of the final grade of the student (y) against the number of absences (x) of seven randomly selected students of a statistics class.

SAQ 4

Let me bring to your notice that the scatter plot gives useful information on the possible nature of the linear relationship between the two variables, if indeed any relationship exists. Figure 2 suggests a linear relationship between the two variables. However, it is observed that the final grade (y) decreases as the number of absences (x) increases. This indicates a negative linear relationship between the independent variable and the dependent variable.

There could also be a positive linear relationship. In this case, y increases as x also increases.

SAQ 2

Assumptions of Simple Linear Regression

1. Simple regression implies that a linear relationship exists between Y and X such that $Y = \alpha + \beta x$ and is based on the following assumptions:
2. The variable X is called the independent variable while Y is the dependent variable.
3. X is measured without error term.

4. For each value of X, there is a family (or subgroup) of values
5. The variances of the subpopulation are equal
6. The means of the subpopulation of Y all lies on a straight line.

SAQ 3

Fitting the Regression Line

Do you know the methods of fitting the regression line? Methods of fitting the regression line include:

a. Freehand Method
b. Grand mean Method
c. Semi-average Method
d. Least squares

The Method of Least Squares

It is important to note that out of the four methods of fitting the regression line, Least squares is the most reliable of all the methods of fitting the regression line because it leads to a unique regression line and regression coefficient. In this method, the regression line is represented by the general equation of a straight line

$$Y = \alpha + \beta x$$

The values of α and β, and hence the equation above, are determined by minimizing

$$Q = \sum(y - \hat{y})^2 = \sum(y_i - \alpha - \beta x_i)^2 = \sum \varepsilon_i^{\,2}$$

which is the sum of squares of the distances between the points of the scatter diagram and the regression line.

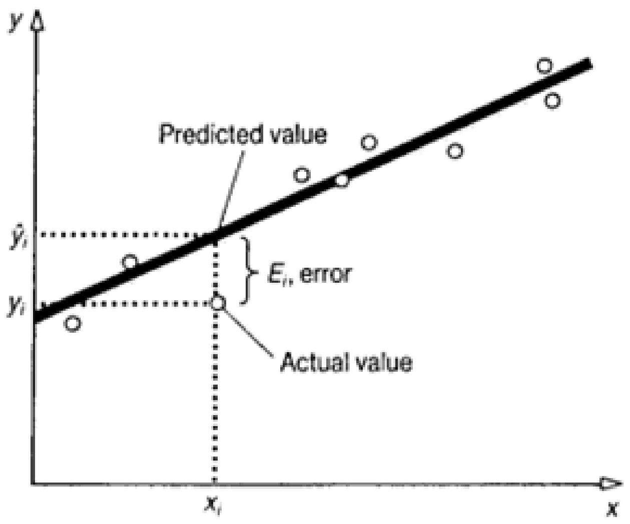


Figure 2: Sample observations and the estimated regression line.

The least squares estimate of the regression parameters is the values of α and β that minimizes Q. This is gotten through the partial derivatives of Q with respect to α and β, respectively, equated to zero, and then solving the equations.

$$\frac{\partial Q}{\partial \beta_0} = -2\sum(y_i - \alpha - \beta x_i) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum(y_i - \alpha - \beta x_i)\,x_i = 0$$

Rearranging the equations above, we have

$$\sum y_i = n\alpha + \beta \sum x_i \qquad (1)$$
$$\sum x_i y_i = \alpha \sum x_i + \beta \sum {x_i}^2 \qquad (2)$$

Equations (1) and (2) are called the Normal Equations. Solving these equations yield

$$\hat{\beta} = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum {x_i}^2 - (\sum x_i)^2/n} = \frac{n\sum x_i y_i - (\sum x_i \sum y_i)}{n\sum {x_i}^2 - (\sum x_i)^2}$$

And

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Thus, the estimated least squares regression equation is

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

Activity 2: Suppose the management of a hospital conducts a 5-month experiment to determine the effect of advertising on its revenue of the hospital. The table below represents the data (in N100,000):

Advertising Expenditure (X)	1	2	3	4	5
Revenue (Y)	10	10	20	20	40

Let us find the regression line, and predict the revenue if the sum of N450,000 is spent on advertisement in a month.

Solution:

x= Advertising expenditure (in N100,000), y=Revenue (in N100,000), n=5

In order to estimate the regression line, compute the following statistics:

$$\sum x_i, \sum y_i, \sum {x_i}^2 \text{ and } \sum x_i y_i.$$

x	y	x^2	xy	y^2
1	10	1	10	100
2	10	4	20	100
3	20	9	60	400
4	20	16	80	400
5	40	25	200	1600
15	**100**	**55**	**370**	**2600**

The required line is

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

and the estimates of the parameters are

$$\hat{\beta} = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum {x_i}^2 - (\sum x_i)^2/n} = \frac{370 - 15(100)/5}{55 - 15(15)/5} = \frac{70}{10} = 7$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{100}{5} = \frac{7(15)}{5} = 20 - 21 = -1$$

Therefore, the estimated regression line is

$$y = -1 + 7x$$

When x=4.5, then y=-1+7(4.5)= 30.5

Since the data are in units of N100,000, you must multiply the value of y obtained by N100,000. That is, the corresponding revenue for x=45,000 is N3,050,000.

Activity 3: The number of absences from school (x) for a group of 7 students and their final grades (y) in a statistics course are given below:

Student	Number of absences x	Final grade y (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78

Obtain the linear regression equation for the data.

Solution

In order to estimate the regression line, we must compute $\sum x_i, \sum y_i, \sum x_i^2$ and $\sum x_i y_i$. These are given as

$\sum x_i = 67, \sum y_i = 511, \sum x_i^2 = 719$ *and* $\sum x_i y_i = 4605$. The number of observations, n=7.

Using the formula for the estimates of the regression parameters (α, β) we have

$$\hat{\beta} = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum x_i^2 - (\sum x_i)^2/n} = \frac{4605 - 67(511)/7}{719 - 67(67)/7} = -3.6801$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{511}{7} - \frac{3.68017(67)}{7} = 108.224$$

Hence the required regression line is $y = 108.224 - 3.6801x$

Summary

The theory of linear regression, which is the linear relationship between two variables: the dependent variable and the independent variable, was presented in this unit. The scatter diagram which is the plot of the data values of the dependent variable (Y) against the independent variable (X) was introduced as a fair indicator of the type of linear relationship that may exist between the two variables. The linear regression model has the following underlying assumptions: assumptions of fixed-x, normality, equal spread, and independent errors.

Self-Assessment Questions

1. Define a simple linear regression?
2. What are the assumptions of a simple linear regression?
3. Mention 4 methods of fitting a linear regression line.
4. Distinguish between positive linear relationship and negative linear relationship;
5. What is the use of a scatter diagram in regression?
6. Given the bivariate data below

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

   Fit a simple linear regression model to the data.

7. Obtain the least squares regression line for the data below

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

8. Fit a simple linear regression line to the data below. What is the value of y when x=200?

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Tutor Marked Assignment

1. Obtain the simple linear regression model for the data below

w	67.6	87.1	117	115	118	113
x	12100	12600	12500	10900	7800	7000

2. A physician wishes to know whether there is a relationship between a father's weight (in pounds) and his newborn son's weight (in pounds). Obtain the simple linear regression model for this data

F`sgdq╪ v dhf gs, x	176	160	187	210	196	142	205	215
Snn's weight, y	6.6	8.2	9.2	7.1	8.8	9.3	7.4	8.6

3. It is desired to find the relationship between verbal aptitude (x) of students and their corresponding math scores (y). Fit a regression line on the data obtained below.

X	526	504	594	585	503	589
Y	530	522	606	588	517	589

4. A sample of 10 billionaires is selected, and the person's age (x) and net worth (y) are compared. The data are given below.

w	56	39	42	60	84	37	68	66	73	55
x	18	14	12	14	11	10	10	7	7	5

What is the relationship between the person's age and his net worth?

5. An architect wants to determine the relationship between the heights (in feet) (y)of abuilding and the number of stories in the building (x). The data for a sample of 10 buildings are shown. Explain the relationship.

w	64	54	40	31	45	38	42	41	37	40
x	841	725	635	616	615	582	535	520	511	485

Further Reading

- https://en.wikipedia.org/wiki/Linear_regression
- https://en.wikipedia.org/wiki/Simple_linear_regression
- https://www.bing.com/videos/search?q=linear+regression&docid=608012445362687617&mid=934781BA6BD71FAB1B0B934781BA6BD71FAB1B0B&view=detail&FORM=VIRE

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Variable symbol

Photo: freepik.com

UNIT 2

Pearson Correlation

Introduction

Welcome to unit 2 of module six. In the previous lesson, I have explained the concept of linear regression. Here we will look at how simple linear regression can be extended to determine the strength of the linear relationship using Pearson correlation coefficient. We will also consider the coefficient of determination and tests of hypothesis regarding the regression model and the correlation coefficient.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define the Pearson correlation coefficient;
2. Highlight the assumptions of the Pearson correlation coefficient;
3. Define the coefficient of determination;
4. Interpret estimates of the correlation coefficient obtained from a sample; and
5. Carry out a test of hypothesis on the Pearson correlation coefficient.

Main Content

SAQ 1

Pearson Correlation

15 mins

Correlation Coefficient: The correlation coefficient is used to determine the strength of the linear relationship between two variables. There are several types of correlation coefficients. The correlation coefficient considered in this unit is the **Pearson product moment correlation coefficient**, named after statistician Karl Pearson, who pioneered the research in this area.

Note that the correlation coefficient computed from the sample data measures the strength and direction of a linear relationship between two quantitative variables. The symbol for the sample correlation coefficient is r. The symbol for the population correlation coefficient is ρ (Greek letter rho).

The range of the correlation coefficient is from -1 to +1. If there is a strong positive linear relationship between the variables, the value of r will be close to +1. If there is a strong negative linear relationship between the variables, the value of r will be close to -1. When there is no linear relationship between the variables or only a weak relationship, the value of r will be close to 0.

Assumptions for the Correlation Coefficient

1. The sample is a random one.
2. The data pairs fall approximately on a straight line and are measured at the interval or ratio level.
3. The variables have a joint normal distribution.

The following facts about simple correlation should be noted.

i. Correlation between two variables x and y is **perfect** if all points of the scatter diagram lie on a straight line.

ii. Correlation between two variables x and y is **positive or direct** if y increases as x increases or if the regression line slopes upwards from left to right.

iii. Correlation between two variables x and y is **negative or inverse** if y decreases as x increases or vice versa, or if the regression line slopes downward from left to right.

iv. There is an absence of correlation **(zero correlation)** between two variables x and y if there is no definite pattern in the direction of the variables x and y.

SAQ 4

Limitations of Correlation

1. The sample correlation coefficient is a measure of linear relationship only. There may be an exact connection between the two variables but if it is not a straight line, r offers no help. It is very useful to study the scatter diagram carefully to see if a nonlinear relationship may exist.
2. Correlation does not imply causality. It simply measures the strength of the relationship between the two variables.
3. An unusual result or outliers may have a strong effect on the value of r.

SAQ 4

The correlation coefficient, **ρ** gives the strength of the linear relationship. The Pearson correlation coefficient for a sample is given as

$$r = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sqrt{(\sum x_i^2 - (\sum x_i)^2/n)(\sum y_i^2 - (\sum y_i)^2/n)}} = \frac{Cov(x,y)}{\sqrt{var(x)var(y)}}$$

The **coefficient of determination, r2** is a measure of the variation of the dependent variable that is explained by the regression line and the independent variable.

The **standard error of the estimate** is given as

$$s_{\hat{y}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum(y_i - \widehat{\beta_0} - \widehat{\beta_1}\sum x_i)^2}{n-2}}$$

SAQ 6-8

Activity 1: Suppose the management of a hospital conducts a 5-month experiment to determine the effect of advertising on revenue of the hospital. The table below represents the data (in N100,000):

Advertising Expenditure (x)	1	2	3	4	5
Revenue (y)	10	10	20	20	40

Compute the Pearson correlation coefficient and the coefficient of determination.

Solution:

x= Advertising expenditure (in N100,000), y=Revenue (in N100,000), n=5

The summary statistics are

$$\sum x_i = 15, \sum y_i = 100, \sum x_i^2 = 55, \sum x_i y_i = 370, \sum y_i^2 = 2600.$$

The Pearson correlation coefficient is given as

$$r = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} = \frac{70}{\sqrt{10(600)}} = \frac{70}{77.4596} = 0.904$$

$$r^2 = (0.904)^2 = 0.8172$$

This means that about 82% of the variability in y is explained by the regression equation.

Activity 2: The numbers of absences from school (x) for a group of 7 students and their final grades (y) in a statistics course are given below:

Student	Number of absences *x*	Final grade *y* (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78

Estimate the Pearson correlation coefficient and the coefficient of determination for the data.

Solution

The summary statistics are given below:

$$\sum x_i = 67, \sum y_i = 511, \sum x_i^2 = 719, \sum x_i y_i = 4605, \sum y_i^2 = 38993.$$

The number of observations, n=7.

Using the formula for the estimates of the regression parameters (α, β), we have

$$r = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} = -0.9442$$

$$r^2 = (-0.9442)^2 = 0.8915$$

The result implies that there is a strong negative linear relationship between the dependent and independent variables. Also, the estimate of the coefficient of determination shows that about 89% of the variation in the dependent variable is explained by the fitted linear regression line.

SAQ 1,3

Continuous Uniform Distribution

The sample correlation coefficient, r measures the strength of the linear relationship found in a sample. An important question is whether the size of r suggests the variables really are related (in the population). However, it is important to test whether the value obtained from data is significant in our conclusion on the strength of the relationship.

A test of hypothesis whether the population correlation coefficient ρ is not zero is now presented.

The null hypothesis and the alternative hypothesis are

$H_0: \rho = 0$ versus $H_1: \rho \neq 0$

We then compute the test statistic

$$t_{calc} = \frac{r\sqrt{n-2}}{1-r^2} = \frac{0.94 \times \sqrt{5}}{0.1164} = 18.06$$

The null hypothesis is rejected if $t_{calc} > t_{0.05}(5)$. Otherwise, we cannot reject the null hypothesis at the 0.05 level of significance.

From the statistical table of the Student's t-distribution with 5 degrees of freedom, $t_{0.05}(5) = 2.015$. It is thus seen that the calculated value of the test statistic is greater than the tabulated value, hence reject the null hypothesis and conclude that the Pearson correlation coefficient ρ is significantly different from zero.

Summary

Correlation, as a measure of the strength of the linear relationship between two variables, was discussed in this unit. The assumptions and limitations of correlation were provided. The Pearson correlation coefficient, estimated from data indicates the strength of the linear relationship. The coefficient of determination is the square of the correlation coefficient and gives the proportion of the variation in the dependent variable that is accounted for by the linear regression equation.

Self-Assessment Questions

1. Define correlation?
2. Give the formula for the Pearson correlation coefficient.
3. Give the expression for the coefficient of determination. What is its interpretation?
4. Mention three limitations of correlation.
5. Give examples of two variables that are positively correlated and two that are

negatively correlated.

6. Given the bivariate data below

X	120	118	128	145	116	160	167	159	139	147	150	145	140	190	188
Y	60	40	72	63	36	47	55	49	44	42	68	59	64	50	63

(a) Estimate the Pearson correlation coefficient. (b) Examine whether the correlation coefficient is significantly different from zero at $\alpha = 0.05$.

7. Obtain the Pearson correlation coefficient for the data below

x	1.2	0.8	1.0	1.3	0.7	0.8	1.0	0.6	0.9	1.1
y	101	92	110	120	90	82	93	75	91	105

8. Obtain the Pearson correlation coefficient for the data below

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]62	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Tutor Marked Assignment

1. Obtain the Pearson's correlation coefficient for the data below

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

2. A physician wishes to know whether there is a relationship between a father's weight (in pounds) and his newborn son's weight (in pounds). (a) Determine the correlation coefficient. (b) Test whether $\rho = 0$ at $\alpha = 0.05$.

[illegible] x	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible] y	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

3. It is desired to find the relationship between verbal aptitude (x) of students and their corresponding math scores (y). Obtain the correlation coefficient.

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

4. A sample of 10 billionaires is selected, and the person's age (x) and net worth (y) are compared. The data are given below.

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Compute the Pearson correlation coefficient.

5. An architect wants to determine the relationship between the heights (in feet)(y) of a building and the number of stories in the building (x). The data for a sample of 10 buildings are shown. Compute the Pearson correlation coefficient.

∏	□□	□□	□□	□□	□□	□□	□2	□□	□□	□□
π	□□□	□□□	□□□	□□□	□□□	□□□	□□□	□□□	□□□	□□□

Further Reading

- https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
- https://www.spss-tutorials.com/pearson-correlation-coefficient/
- https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Service statistics concept

Photo: freepik.com

UNIT 3

Rank Correlation

Introduction

Welcome to Module Six, Unit 3. Our focus here is Rank correlation. There are several rank correlation coefficients, but we shall focus mainly on the Spearman's rank correlation coefficient, state its properties and demonstrate its application to data.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define rank correlation;
2. Define the Spearman's rank correlation coefficient;
3. Identify various scenarios where the Spearman's rank correlation coefficient could be applied;
4. Compute the Spearman's rank correlation coefficient given a bivariate dataset; and
5. Carry out a test of hypothesis on the significance of the Spearman's rank correlation coefficient.

Main Content

Rank Correlation

SAQ 1

2 mins

i. It is possible that some data cannot be quantified or are difficult to quantify. When this happens, the rank correlation procedure is used to determine the correlation coefficient. Such ranking is done in ascending order. The ranking procedure is also adopted when data assume large values to avoid rigorous

ii. calculations.

Rank correlation is any of several statistics that measure an ordinal association—the relationship between rankings of different ordinal variables or different rankings of the same variable, where a "ranking" is the assignment of the ordering labels "first", "second", "third", etc. to different observations of a particular variable. You should also bear in mind that a **rank correlation coefficient** measures the degree of similarity between two rankings and can be used to assess the significance of the relation between them.

Some of the more popular rank correlation statistics include

i. Spearman's ρ_s

ii. Kendall's τ

iii. Goodman and Kruskal's γ

iv. Somers' D

An increasing rank correlation coefficient implies an increasing agreement between rankings.

The coefficient is inside the interval $[-1, 1]$ and assumes the value:

(a) 1 if the agreement between the two rankings is perfect; the two rankings are the same.

(b) 0 if the rankings are completely independent.

(c) −1 if the disagreement between the two rankings is perfect; one ranking is the reverse of the other.

Also note that a ranking can be viewed as a permutation of a set of objects. Thus we can look at observed rankings as data obtained when the sample space is (identified with) a symmetric group. We can then introduce a metric, making the symmetric group into a metric space. Different metrics will correspond to different rank correlations.

Spearman's Rank Correlation Coefficient

SAQ 2-8

2 mins

Spearman's rank correlation coefficient ρ_s, named after Charles Spearman, is a non-parametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function.

Given a bivariate data set containing n pairs of observations $(x_i, y_i), i = 1,2, \dots, n$. Suppose that the observations are now ranked from 1 to n for both the x and y variables. We then have the bivariate ranks (r_{x_i}, r_{y_i}) . The ranking of each of the n values of the variables x and y of the bivariate data could be rearranged as a sequence of n natural numbers 1,2,...,n.

Therefore

$$\sum r_{x_i} = \sum r_{y_i} = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

Similarly,

$$\sum r_{x_i}^{\;2} = \sum r_{y_i}^{\;2} = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

Let $D = (D_1, D_2, \dots, D_n)$ be the differences between the corresponding ranks of the (x_i, y_i). That is

$$D_i = rank(x_i) - rank(y_i) = r_{x_i} - r_{y_i}, \qquad i = 1,2, \dots, n$$

Using this categorization, it can be shown that

$$\rho_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

which is the Spearman's rank correlation coefficient with n observations.

The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables. While Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank between the two variables and low when observations have a dissimilar rank between the two variables.

Spearman's coefficient is appropriate for both continuous and discrete ordinal variables. Both Spearman's ρ and Kendall's τ can be formulated as special cases of a more general correlation coefficient.

One assumption for testing the hypothesis that $\rho = 0$ for the Pearson coefficient is that the populations from which the samples are obtained are normally distributed. If this requirement cannot be met, the nonparametric equivalent, called the Spearman rank correlation coefficient (denoted by ρ_S), can be used when the data are ranked.

The computations for the rank correlation coefficient are simpler than those for the Pearson coefficient and involve ranking each set of data. The difference in ranks is found, and the sample Spearman's rank correlation coefficient, r_s will be close to zero.

Unit 3 | Module 6 | STA 124 Rank Correlation

Tie in Rank

| 5 mins

At times, two or more values of a variable may be equal. When such a situation arises, the affected values are assigned the mean rank of the consecutive rankings that the values would have taken had there been no tie. The subsequent value is however assigned the appropriate ranking under a situation where no tie occurs.

For example, for the given values 5,8,8,11,15 of a variable, the two values 8 (which fall in rank 2 and rank 3) are both assigned the rank (2+3)/2=2.5 while the value 1 is assigned the rank 4.

Activity 1: Ten drivers took part in a driving competition. The orders in which two judges placed them on their performances are as follows:

*µ§ß• 8	□	□	□	□	□	□	□□	□	□	□
*µ§ß• 9	□	□	□	□	□	□	□	□	□	□□

Use Spearman's method to calculate the rank correlation coefficient and briefly comment on your result.

Solution

The data already is ranked, hence there is no need to rank the observations. The next step is to compute the rank differences D_i, thereafter get their squares D_i^2 and $\sum D_i^2$. The computations are shown in the table below:

X	Y	D_i	D_i^2
1	2	-1	1
6	7	-1	1
9	8	1	1
5	5	0	0
7	4	3	9
3	1	2	4
10	9	1	1
2	3	-1	1
4	6	-2	4
8	10	-2	4
55	**55**	**0**	**26**

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)} = 1 - \frac{6(26)}{10(100-1)} = 0.842$$

Since r_s is very close to 1, there is a high correlation or agreement between the rankings of the two judges.

Activity 2: A researcher wishes to see if there is a relationship between the number of branches a bank has and the total amount of deposits (in billions of Naira) the bank receives. A sample of eight commercial banks is selected, and the numbers of branches and the amounts of deposits are shown in the table. Use Spearman's method to compute the correlation coefficient.

Bank	Number of Branches(x)	Deposits (in billions) (y)
A	209	23
B	353	31
C	19	7
D	201	12
E	344	26
F	132	5
G	401	24
H	126	5

Solution

Obtain the ranks for both the number of branches (x) and the deposits (y), thereafter obtain the rank differences D, the square rank differences (D^2) and sum the square rank differences up.

In ranking variable y, a tie in rank for the value 5 is observed, hence give both the mean rank (1+2)/2=1.5.

It will be seen that $\sum D_i^2 = 12.5$. Thus, the Spearman's rank correlation coefficient is given as

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)} = 1 - \frac{6(12.5)}{8(64-1)} = 0.85$$

The value of the Spearman's rank correlation coefficient implies a strong positive relationship between the number of branches of a bank and the amounts of deposits.

The value of the rank correlation coefficient is very close to the value of Pearson's correlation coefficient r=0.868, but Spearman's method involved less tedious computations.

Test of Hypothesis Using Spearman's Method

Although the Spearman's method is a non-parametric approach, a parametric test procedure is available for testing the significance of the Spearman's rank correlation coefficient. Indeed, this test is not different from the test of hypothesis for the Pearson's correlation coefficient.

The null hypothesis and the alternative hypothesis are

$H_0: \rho_s = 0$ versus $H_1: \rho_s \neq 0$.

The test statistic is

$$t_{calc} = \frac{r_s\sqrt{n-2}}{1-r_s^2}$$

where r_s is the sample Spearman's rank correlation coefficient and n is the number of observations. Since the alternative hypothesis is $\rho_s \neq 0$, we can neglect the negative sign for cases where r_s is negative and use the absolute value of r_s at all times for the test.

We reject the null hypothesis at the α level of significance if the computed value of the test statistic $t_{calc} > t_\alpha(n-2)$, where $t_\alpha(n-2)$ is the tabulated value of the Student's t-distribution with n-2 degrees of freedom.

If the null hypothesis is rejected, then it can be concluded that the Spearman's correlation coefficient is significantly different from zero.

Activity 3: Let us test the significance of the Spearman's rank correlation coefficient at $\alpha = 0.05$ if $r_s = 0.85$ and n=8.

Solution

$$t_{calc} = \frac{r_s\sqrt{n-2}}{1-r_s^2} = \frac{0.85 \times \sqrt{6}}{1-0.7225} = 7.503$$

The null hypothesis is rejected if $t_{calc} > t_{0.05}(6)$. Otherwise do not reject the null hypothesis at the 0.05 level of significance.

From the statistical table of the Student's t-distribution, $t_{0.05}(6) = 1.943$. It is seen that the calculated value of the test statistic is greater than the tabulated value, hence the null hypothesis is rejected and it is concluded that the Spearman's rank correlation coefficient ρ_s is significantly different from zero.

Summary

The rank correlation approach was introduced to explain the strength of association or agreement between two ordinal variables. The Spearman's rank correlation coefficient was presented as a quick alternative to the Pearson's correlation coefficient. Several illustrations were provided to establish the application of the concept of rank correlation.

Self-Assessment Questions

1. Define rank correlation.

2. Define the Spearman's rank correlation coefficient and describe how it is computed from data.

3. What should be done when we observe tie in the ranks using Spearman's method?

4. Explain the way to test the significance of the Spearman's rank correlation coefficient.

5. Given the data below, obtain the Spearman's rank correlation coefficient.

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]71	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

6. Given the bivariate data below

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(a) Estimate the Spearman's rank correlation coefficient. (b) Examine whether the rank correlation coefficient is significantly different from zero at $\alpha = 0.05$.

7. Obtain the Spearman's rank correlation coefficient for the data below

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

8. Obtain the Spearman's rank correlation coefficient for the data below.

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Tutor Marked Assignment

1. Obtain the Spearman's rank correlation coefficient for the data below

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

2. Given the data below, (a) determine the Spearman's rank correlation coefficient. (b) Test whether $\rho_s = 0$ at $\alpha = 0.05$.

x	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

3. Determine the Spearman's rank correlation coefficient for the data given in the table below.

8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

4. Given the data below, compute the Spearman's rank correlation coefficient.

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

5. Given the data below, compute the Spearman's rank correlation coefficient.

∏	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
π	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Further Reading

- https://en.wikipedia.org/wiki/Rank_correlation
- https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA

Picture: Statistics Presentation
Photo: freepik.com

UNIT 4

Association of Attributes

Introduction

Welcome to the last unit of the course. In this unit I will introduce to you the Association of the attribute, as a statistical concept. I focus my explanation on the basic approach of analyzing the association of attributes via the contingency table. I also present some measures of attributes together with hints on how to calculate these measures from data.

Learning Outcomes

When you have studied this unit, you should be able to:

1. Define a contingency table for categorical data;
2. Define the Chi-square statistic;
3. Identify scenarios where contingency tables could be applied;
4. Obtain the expected frequencies from a contingency table; and
5. Carry out a test of association between two categorical variables, using the contingency table.

Main Content

Association of Attributes

3 mins

Let me tell you that a **contingency table** (also known as a **cross-tabulation**) is a type of table in a matrix format that displays the frequency distribution of the variables. Contingency tables are used extensively in survey research, business intelligence, engineering and scientific research. They provide a basic picture of the interrelation or association between two **categorical variables** and can help to establish interactions between these variables.

It is imperative to note that a *categorical variable* is a variable that can take on one of a limited, and usually fixed number of possible values, assigning each individual or other unit of observation to a particular group or nominal category based on some qualitative property, like the choice of a political party, gender, employment status, state of origin, etc. The term *contingency table* was first used by Karl Pearson.

Now when data are available in tabular form in terms of frequencies, several types of hypotheses can be tested by using the chi-square test. Two such tests are the test of association or independence of attributes and the homogeneity of proportions test.

Furthermore, the test of independence is used to determine whether two variables are independent of or related to each other when a single sample is selected.

You should also note that the test of homogeneity of proportions is used to determine whether the proportions for a variable are equal when several samples are selected from different populations. Both tests use the chi-square distribution and a contingency table, and the test value is found in the same way.

Hypothesis tests may be performed on contingency tables in order to decide

whether or not the effects are present. Effects in a contingency table are defined as relationships between the row and column variables; that is, are the levels of the row variable differentially distributed over levels of the column variables? Significance in this hypothesis test means that interpretation of the cell frequencies is warranted. Non-significance means that any differences in cell frequencies could be explained by chance.

A contingency table which contains r rows and c columns is called a r x c contingency table and is used in testing hypothesis relating to association/ independence of two categorical variables.

Activity 1: A survey of political parties' members in a Nigerian city to determine whether there is any association between the political affiliation and the nature of employment of the respondents is given in the table below.

Political affiliation of employees

	Manual Employment	**Non-Manual Employment**	**Tota (n_i))**
APC	18	19	**37**
PDP	46	14	**60**
Labour	11	13	**24**
Tota (n_j)	**75**	**46**	**$n_{..}$=121**

This is an example of a 3 x 2 contingency table. Each observation is classified in two ways according to (a) political party or (b) type of employment.

The contingency table provides a means of testing if there is any association between the characteristics upon which the classification is based. To test the hypothesis whether there is any association between political party and type of employment, the null hypothesis, H_0 will be

H_0: There is no relationship between the characteristics, that is, there is no association between political party and type of employment. Or the political party and type of employment are independent.

The alternative hypothesis is H_1: There is a relationship between the characteristics. Or The characteristics are not independent.

The test statistic that can be used for this test of association is the X^2 statistic, given by

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

That is, the test statistic, X^2 is distributed as chi-square distribution with (r-1)(c-1) degrees of freedom when the null hypothesis is true.

O_{ij} is the observed value of the i^{th} row and the j^{th} column from the data, while E_{ij} is the corresponding expected value, and it is obtained using the row $(\boldsymbol{n}_{i.})$ and column totals $(\boldsymbol{n}_{.j})$:

$$E_{ij} = \frac{row\ sum \times column\ sum}{grand\ total} = \frac{n_{i.}n_{.j}}{n..}$$

The decision rule is to reject H_0 when $X^2 > \chi^2_{(r-1)(c-1),\alpha}$ and conclude that there is no association between the attributes. Otherwise, we do not reject the null hypothesis. α is called the significance level of the test and is mostly taken to be 0.05. Thus we could check the tabulated value of $\chi^2_{(r-1)(c-1),\alpha}$ from the chi-square table and compare with the value of the test statistic X^2.

$$\chi^2_{2,0.05} = 5.991$$

The calculated value of the statistic $X^2 = 10.9383$ which is greater than the tabulated value of 5.991, hence the hypothesis of independence is rejected and it is concluded that the two characteristics (political party and type of employment) are associated.

SAQ 1-8

Steps in Testing the Hypothesis of Independence

7 mins

These are the following steps we need in the analysis of a $r \times c$ contingency table.

Step 1: We state the null hypothesis and the alternative hypothesis.

H_0: Factor (or characteristic) A is dependent on Factor (or characteristic) B.

Step 2: We will also obtain the row and column totals, which are then used to compute the expected cell counts E_{ij} via the formula

$$E_{ij} = \frac{i^{th}\ row\ sum \times j^{th}\ column\ sum}{grand\ total} = \frac{n_{i.}n_{.j}}{n..}$$

Step 3: The test statistic, X^2 is then computed via the formula:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

Step 4: At the specified level of significance (α), you compare the calculated value of the X^2 statistic with the corresponding statistical table value with the degrees of freedom (r-1)(c-1) (i.e $\chi^2((r-1)(c-1))$.

Step 5: Decision Rule: You reject the null hypothesis at the α (=0.05 or 0.01) significance level if the calculated value of the test statistic is greater than the table value.

That is, reject the null hypothesis H_0, if $X^2 > \chi^2((r-1)(c-1))$. Otherwise, do not reject the null hypothesis.

Step 6: Decision: Based on step 5 above, we can conclude that

Factor (or characteristic) A is independent of Factor (or characteristic) B (if the null hypothesis is rejected)

OR

Factor (or characteristic) A is dependent on Factor (or characteristic) B (if the null hypothesis is not rejected).

Activity 2: Suppose a new postoperative procedure is administered to a number of patients in a large hospital. The researcher can ask the question: do the doctors feel differently about this procedure from the nurses, or do they feel basically the same way? Note that the question is not whether they prefer the procedure but whether there is a difference of opinion between the two groups.

To answer this question, a researcher selects a sample of nurses and doctors and tabulates the data in table form, as shown below.

Group	0≤•¶≤Æ∑ procedure	0≤•¶≤Ø§ procedure	. Ø preference
. µ≤•≥	[illegible]	[illegible]	[illegible]
$ Ø£¥Ø≥	[illegible]	[illegible]	[illegible]

Test the hypothesis whether the opinion about the procedure is independent of the profession.

Solution

The null hypothesis and the alternative hypothesis for this test are

H_0: The opinion about the procedure is independent of the profession.

H_1: The opinion about the procedure is dependent on the profession.

In order to accomplish this test, the expected values E_{ij} are computed and these are given in brackets in the table below.

Group	0≤•¶≤Æ∑ procedure	0≤•¶≤Ø§ procedure	. Ø preference	Total
. µ≤•≥	[illegible]	[illegible]	[illegible]	[illegible]
$ Ø£¥Ø≥	[illegible]	[illegible]	[illegible]	[illegible]
4Ø¢º¨	[illegible]	[illegible]	[illegible]	[illegible]

r=2, c=3, then the degrees of freedom is (2-1)(3-1)=2, and $\chi^2_{2,0.05} = 5.991$.

The calculated value of the statistic $X^2 = 26.6667$ which is greater than the tabulated value of 5.991, hence the hypothesis of independence is rejected.

The conclusion is that there is enough evidence to support the claim that opinion is related to (dependent on) profession—that is, that the doctors and nurses differ in their opinions about the procedure.

Activity 3: In a volunteer group, the members volunteer from one to nine hours each week to spend time with a disabled senior citizen. The program recruits among university students, polytechnic students, and nonstudents. The table below is a sample of the volunteers and the number of hours they volunteer per week. Is the number of hours volunteered independent of the type of volunteer?

Type of Volunteer	1-3 hours	4-6 hours	7-9 hours
University Students	111	96	48
Polytechnic Students	96	133	61
Nonstudents	91	150	53

Solution

The observed table and the question at the end of the problem, "Is the number of hours volunteered independent of the type of volunteer?" gives us the understanding that this is a **test of independence**. The two factors are the number of hours volunteered and the type of volunteer. This test is always **right-tailed**.

H_0: The number of hours volunteered is independent of the type of volunteer.

H_1: The number of hours volunteered is dependent on the type of volunteer.

Distribution for the test: $\chi_2(4)$ since df = (3 columns – 1)(3 rows – 1) = (2)(2) = 4

Calculate the test statistic: X^2 = 12.99

Type of Volunteer	1-3 hours	4-6 hours	7-9 hours	Total
University Students	111	96	48	
Polytechnic Students	96	133	61	
Nonstudents	91	150	53	
Total				

Summary

The concept of association of attribute is used to analyze contingency tables containing categorical data. The Chi-square test of independence of contingency tables was explained and several illustrations provided.

Self-Assessment Questions

1. Obtain the expected cell count e_{12} for the 3×3 contingency table given below

$$\begin{pmatrix} 15 & 20 & 18 \\ 22 & 25 & 21 \\ 17 & 18 & 16 \end{pmatrix}$$

2. Obtain the expected cell count e_{22} for the 3×3 contingency table given below

3. Use the following 2 x 2 contingency tables for both observed and expected cell counts respectively to obtain the value of Pearson X^2 Statistic

$$\begin{pmatrix} 15 & 20 \\ 22 & 25 \end{pmatrix} \text{ and } \begin{pmatrix} 15.79 & 19.21 \\ 21.21 & 25.79 \end{pmatrix}$$

4. A researcher wishes to see if there is a relationship between the hospital and the number of patient infections. A sample of 3 hospitals was selected, and the number of infections for a specific year has been reported. The data are shown below.

Hospital	[illegible] infections	[illegible] infections	[illegible] infections	Total
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(a) what are the expected cell counts (b) calculate the degrees of freedom for this test (c) compute the Pearson X^2 statistic (d) at $\alpha = 0.05$ significance level, can it be concluded that the number of infections is related to the hospital where they occurred?

5. A researcher wishes to determine whether there is a relationship between the gender of an individual and the amount of alcohol consumed. A sample of 68 people is selected, and the following data are obtained.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(a) Compute the X^2 value for the test and obtain the corresponding χ^2 from the statistical table at $\alpha = 0.10$. (b) Can the researcher conclude that alcohol consumption is related to gender?

6. (a) compute X^2 and the degrees of freedom for the test (b) Can you conclude a relationship between the class of vertebrate and whether it is endangered or threatened? Use the 0.05 level of significance. (c) Is there a different conclusion for the 0.01 level of significance?

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]atened	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

7. Is the size of the population by age-related to the state that it's in? Use a significance level of 0.05. (Population values are in thousands).

	[illegible]	[illegible]-17	[illegible]-24	[illegible]-44	[illegible]-64	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

8. Five states were randomly selected, and their members in the state or federal parliament are noted below. At $\alpha = 0.10$, can it be concluded that there is a dependent relationship between the state and the political party affiliation of their representatives?

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Tutor Marked Assignment

1. Given the contingency table below, obtain the expected values (E_{ij})

$$A = \begin{pmatrix} 33 & 10 & 14 & 12 & 10 \\ 19 & 15 & 10 & 20 & 20 \end{pmatrix}$$

2. Given the contingency table below, obtain the expected values (E_{ij})

$$A = \begin{pmatrix} 721 & 2140 & 1025 & 3515 & 2702 & 1899 \\ 740 & 2104 & 1065 & 3359 & 2487 & 1501 \end{pmatrix}$$

3. Given that $A = \begin{pmatrix} 15 & 20 & 18 \\ 22 & 25 & 21 \\ 17 & 18 & 16 \end{pmatrix}$, compute X^2

4. Given that $A = \begin{pmatrix} 15 & 20 \\ 22 & 25 \end{pmatrix}$, compute X^2

5. Given that $A = \begin{pmatrix} 15 & 25 & 38 \\ 20 & 14 & 19 \\ 17 & 13 & 12 \end{pmatrix}$, compute X^2

6. Find (E_{ij}) for $\begin{pmatrix} 15 & 25 & 38 \\ 20 & 14 & 19 \\ 17 & 13 & 12 \end{pmatrix}$

Further Reading

- https://en.wikipedia.org/wiki/Contingency_table
- https://study.com/academy/lesson/contingency-table-statistics-probability-examples.html

References

- Bluman, A. G. (2012). Elementary Statistics: A Step by Step Approach. McGraw-Hill, New York.
- Jaisingh, L. R. (2000). Statistics for the Utterly Confused, McGraw-Hill, USA.
- Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, USA