Statistics

Third form secondary

Student Book

Name : ....

Class : ....

School: ....

Authors

Mr. Kamal Younis Kabsha

Prof.Dr. Ahmed Kamal El-kholy

Revision, perparing and modifications

Prof. Shaban Ebrahim Abo Yousif

Mr. Sherif Atef El-Borhamy

Mr. Mohamed Ali Kasim

Mr. Amr Farouk Mohamed

Mr. Amgad Samir Gebrael

Mr. Maged Mohamed Hassen

Mr. Othman Mostafa Othman

Dr. Mohamed Mohy Ebrahim Abdelsalam

Scientific Supervision

Ms. Manal Azqoul

Pedagogical Supervision

Dr. Akram Hassan

First eddition 2016 / 2017

Deposit number 8701 / 2016

I.S.B.N 5 - 029 - 706 - 977 - 978

SAKKARA PUBLISHING

Introduction

بسم الله الرحمن الرحيم

يسعدنا ونحن نقدم هذا الكتاب أن نوضح الفلسفة التى تم فى ضوئها بناء المادة التعليمية ونوجزها فيمايلى:

1 تنمية وحدة المعرفة وتكاملها فى الرياضيات، ودمج المفاهيم والترابط بين كل مجالات الرياضيات المدرسية.

2 تزويد المتعلم بما هو وظيفى من معلومات ومفاهيم وخطط لحل المشكلات.

3 تبنّى مدخل المعايير القومية للتعليم فى مصر والمستويات التعليمية وذلك من خلال:

أ) تحديد ما ينبغى على المتعلم أن يتعلمه ولماذا يتعلمه.

ب) تحديد مخرجات التعلم بدقة، وقد ركزت على مايلى:

أن يظل تعلم الرياضيات هدف يسعى المتعلم لتحقيقه طوال حياته - أن يكون المتعلم محبًّا للرياضيات ومبادرًا بدراستها - أن يكون المتعلم قادرًا على العمل منفردًا أو ضمن فريق - أن يكون المتعلم نشطًا ومثابرًا ومواظبًا ومبتكرًا - أن يكون المتعلم قادرًا على التواصل بلغة الرياضيات.

4 اقتراح أساليب وطرق للتدريس وذلك من خلال كتاب (دليل المعلم).

5 اقتراح أنشطة متنوعة تتناسب مع المحتوى ليختار المتعلم النشاط الملائم له.

6 احترام الرياضيات واحترام المساهمات الإنسانية منها على مستوى العالم والأمة والوطن، وتعرف مساهمات وإنجازات العلماء المسلمين والعرب والأجانب.

وأخيرًا ..نتمنى أن نكون قد وفقنا فى إنجاز هذا العمل لما فيه خير لأولادنا، ولمصرنا العزيزة.

والله من وراء القصد، وهو يهدى إلى سواء السبيل

طبعة ٢٠٢٥ - ٢٠٢٦م

Contents

Contents

Unit 1

Correction and Regression

introduction

Statistics is an important branch of mathematics which has various applications since it concerns with gathering, representing and reducing the data in the form of digital indicators to describe, measure and analyze the basic features of such data in order to take the right decisions. Statistics has wide applications in all fields of physical, human, economic and social sciences. In this unit, we are going to focus on analyzing the bivariate data and studying the degree of the direction of the relation between the two variables and the shape of this relation. Initially, we study the correlation which reveals the degree and the power of the relation between the two variables. This relation may be in the form of a direct or inverse form. It is worth mentioning that the correlation studies the relation and its direction between a variable and another. It is necessary to realize that the relation does not refer to causation because it does not refer to the existence of an effect to a variable on the other as it will be shown in Lesson 1 in this Unit. In this unit also, you are going to study the simple linear regression which concerns with estimating the form of this relation through which you can predict the value of the dependent variable if the value of the independent variable is known. It gets more accurate, the more the sample randomly chosen. You will study some modern technologies about the scientific calculators and the computer statistical programs such as(SPSS) to perform the calculations and to graph the data related to the correlation and the linear regression between two phenomena.

Unit objectives

By the end of the unit and carrying out the involved activities, the students should be able to

- ß Identify what is meant by correlation between two variables.
- ß Calculate the correlation coefficient between two variables in different ways (Pearson's method and Spearman's method) and interpret them mathematically.
- ß Understand the meaning of the regression line and estimate its importance to study the relation between two variables.
- ß Represent the relation between two variables in a Cartesian diagram and judge from the graph the existence of the relation and the power of its degree.
- ß Identify the meaning of the liner regression coefficient and interpret what can be deduce by knowing the value of such a coefficient.
- ß Find the regression line equation of any of the two variables on the other using the least squares method.
- ß Use the calculator and computer to perform the mathematical calculations and graph the data of the correlation and the linear regression between two phenomena.
- ß Use a given regression line equation to predict the value of one of the two variables in terms of the corresponding value of the other variable.
- ß Apply the correlation and the linear regression in research situations.
- ß Appreciate contributions of using the correlation and the linear regression in solving social and daily life problems.

Key Terms

- *Correlation*
- *Regression*
- *Linear correlation*
- *Correlation Coefficient*
- *Direct Correlation*
- *Inverse Correlation*
- *Scatter diagram*
- *Pearson Correlation coefficient*
- *Spearman Correlation coefficient*
- *Regression Line*
- *Least Square*

Lessons of the unit

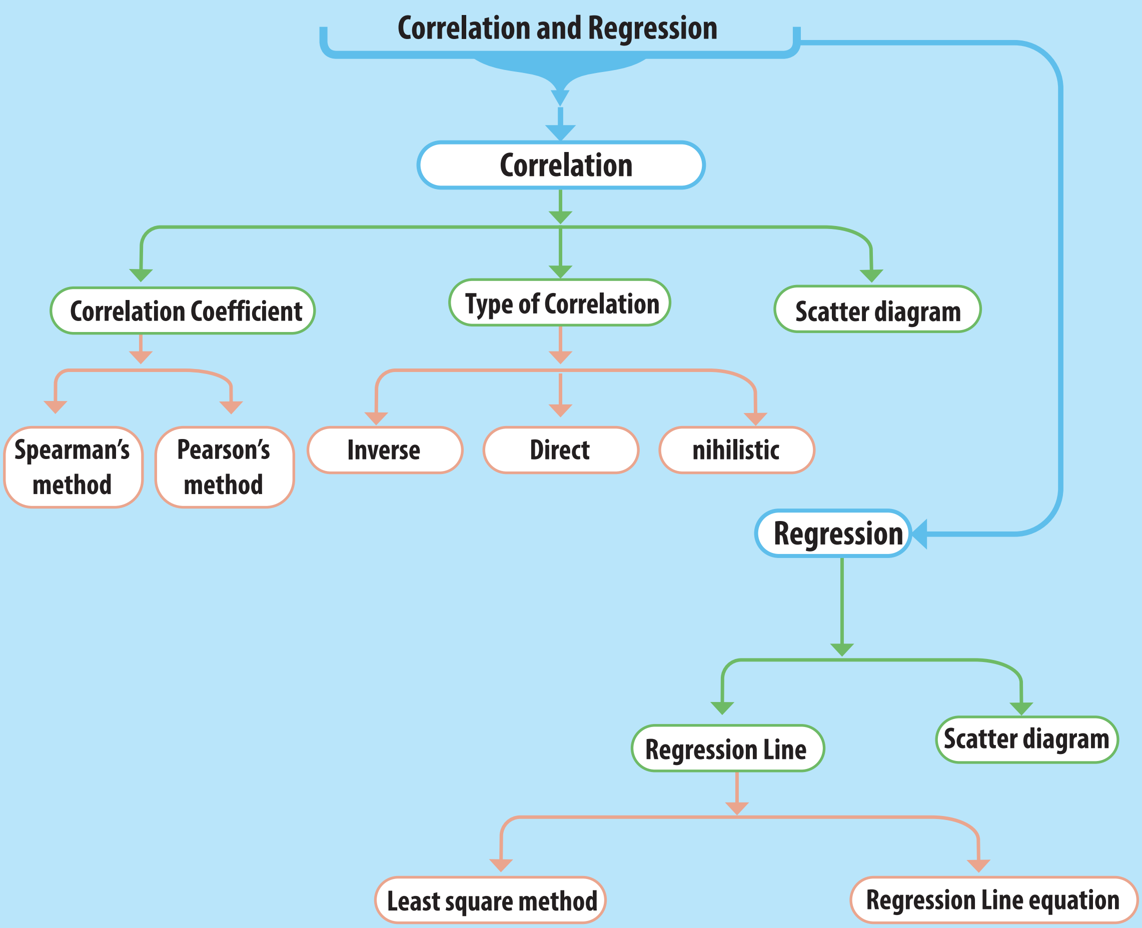
Lesson (1-1) Correlation

Lesson (1-2) Regression

Materials

Scientific calculator - excel program - Spss program

Chart of the unit



Unit one

1 - 1

Correlation

You will learn

- Definition of correlation
- Scatter diagram
- Direct correlation and inverse correlation
- Linear correlation coefficient
- Pearson's linear correlation coefficient
- Spearman's rank correlation coefficient

Key terms

- Correlation
- Linear Correlation
- Correlation Coefficient
- Direct Correlation
- Inverse Correlation
- Scatter diagram
- Pearson Correlation Coefficient
- spearman`s rank correltion coefficient

Introduction:

You have previously learned in Statistics how to describe a set of data representing a phenomenon using some statistical scales such as the scales of the central tendency, dispersion and the coefficient of variation. In this lesson, you are going to learn how to describe the items of two different phenomena in accordance to the relation between both. In other words, if one of the two variables has changed in a certain direction(by increasing or decreasing), the other variable leans to change in a certain direction by increasing or decreasing, too. In this case, the correlation is called a direct correlation. If one of the two variable changes in the direction of increasing and the other changes in the direction of decreasing and vice versa, the correlation in this case is called the inverse correlation.

Correlation:

Think and discuss

meditate the following examples and record your observations:

1- The relation between the side length and the area of the square.

2- The relation between the blood pressure and the age.

3- The increase in the price of an item and the demands of buying it.

4- The decrease of temperature and the demand to consume the fuel.

5- The relation between rising of the sea level and the rise of the temperature.

From the examples above, we observe that:

- The two correlated variables change in the same direction. I.e. the increasing or decreasing of a variable leads to the increasing or decreasing of the other as in examples 1,2 and 3. It is said the correlation between both variables is positive (direct).

materials - Scientific calculator

In examples 4 and 5, we notice that two correlated variables inversely change. In other words, the increasing or decreasing of one variable leads to decreasing or increasing of the other variable. hence, it is said that the correlation between them is negative (inverse).

Definition Correlation is a statistical method by which the degree and the type between two variables can be determined.

The relation between the two variables ranges from a strong degree to a weak one. When the relation between the two variables is strong, it means that knowing the value of one variable will help to predict the value of the other variable. On the contrary, when the relation is weak, it means that knowing the value of one variable will not help to predict the value of the other variable.
One of the most important methods helping us identify the degree and type of the relation between the two variables is the Scatter Diagram.

Scatter diagram

Definition Scatter Diagram is a graphical representation of a number of the ordered pairs (x,y) to describe the relation between two variables.

If we denote the first phenomenon by the symbol (x) and the second phenomenon by the symbol (y), then the next diagrams illustrate the relation between x and y which show the scatter diagram.

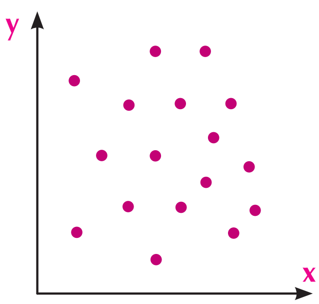


Figure 1
There is no correlation

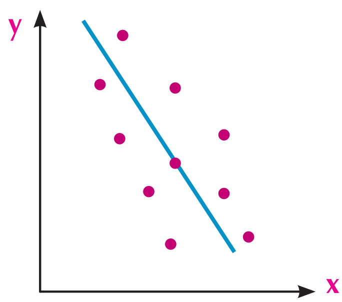


Figure 2
There is an inverse linear correlation

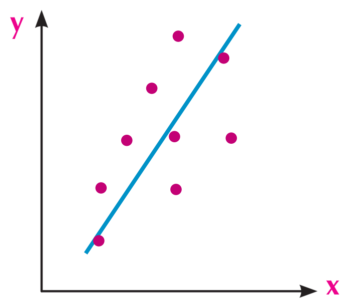


Figure 3
There is a direct linear correlation

Linear Correlation

Definition The simple linear correlation is known as the measure of the degree of the relation between two variables.

Activity

graph the scatter diagram for each of the following data mentioning the type of the relation expressing such data.

(1)

x	7	8	9	10	11	12
y	13	14	17	18	21	23

(2)

x	3	4	7	8	11	15
y	23	22	20	18	17	16

3

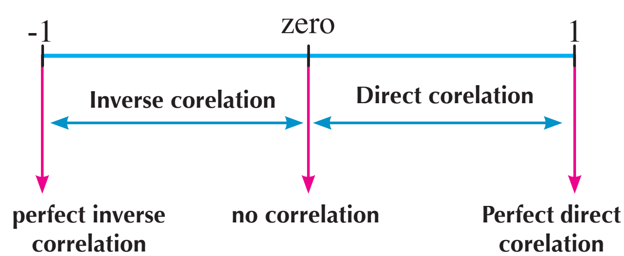
x	7	9	11	13	15	16
y	14	7	20	6	12	10

Correlation Coefficient

The correlation coefficient is denoted by the symbol (r). It is a relatively quantitative scale measuring how much the two variables are correlated where $-1 \leqslant r \leqslant 1$. It is said the correlation is perfect direct if the correlation coefficient r = 1, is perfect inverse if the correlation coefficient r = -1 and the correlation is nihilistic when r = 0.

We notice that:

The nearer the value of the correlation coefficient to 1 is, the stronger the direct correlation between the two variable is and the nearer its value to zero is, the weaker the direct correlation is and we will say the same in the inverse correlation. The opposite figure shows that.



Oral expression: Multi -choice:

The strongest correlation coefficient of the following is:

a – 0.8   b – 0.5   c 0.4   d 0.7

Pearson`s Correlation coefficient

Suppose we have a set of (n) people and we get data from those people about the values of two variables x and y. The data we get ought to be on the form:

The values of first variable x: $x_1 , x_2 , x_3 , \ldots\ldots\ldots\ldots , x_n$

The values of second variable y: $y_1 , y_2 , y_3 , \ldots\ldots\ldots\ldots , y_n$

If we denote the correlation coefficient by the symbol (r), then Pearson's correlation coefficient between the two variables x and y or the linear correlation coefficient can be found from the relation:

$$r = \frac{n \Sigma x y -(\Sigma x \times \Sigma y)}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \sqrt{n \Sigma y^2 -(\Sigma y)^2}}$$

Where: "Σ" is the symbol of summation notation and read as the sum.

n denotes the number of the items.

$\Sigma x = x_1 + x_2 + x_3 \ldots\ldots + x_n$,

$\Sigma y = y_1 + y_2 + y_3 \ldots\ldots + y_n$,

$\Sigma x y = x_1 y_1 + x_2 y_2 + x_3 y_3 \ldots\ldots + x_n y_n$

$\Sigma x^2 = x_1{}^2 + x_2{}^2 + x_3{}^2 \ldots\ldots + x^2_n$,

$\Sigma y^2 = y_1{}^2 + y_2{}^2 + y_3{}^2 \ldots\ldots + y_n^2$

Example

1 The next table shows the marks of 10 students in the history and geography subjects.

History X	75	80	93	65	87	71	98	69	84	78
Geography Y	82	78	86	72	91	80	95	73	89	74

Required is to calculate the Pearson's correlation coefficient between the two variables x and y and identify the type of the correlation.

Solution

We form the next table:

X	Y	X^2	Y^2	X Y
75	82	5625	6724	6150
80	78	6400	6084	6240
93	86	8649	7396	7998
65	72	4225	5184	4680
87	91	7569	8281	7917
71	80	5041	6400	5680
98	95	9604	9025	9310
69	73	4761	5329	5037
84	89	7056	7921	7476
78	74	6084	5476	5772
$\Sigma X = 800$	$\Sigma Y = 820$	$\Sigma X^2 = 65014$	$\Sigma Y^2 = 67820$	$\Sigma XY = 66260$

$$\because r = \frac{n \Sigma x y - (\Sigma x \times \Sigma y)}{\sqrt{n \Sigma x^2 - (\Sigma x)^2}\sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

$$\therefore r = \frac{10 \times 66260 - (800 \times 820)}{\sqrt{10 \times 65014 - (800)^2}\sqrt{10 \times 67820 - (820)^2}}$$

$$= \frac{6006}{\sqrt{10140}\sqrt{5800}} \simeq 0.8606$$

The correlation is direct .

Try to solve

1 **From the data of the next table:**

X	20	23	24	25	28	30
Y	35	31	30	27	29	28

Calculate the Pearson's correlation coefficient (linear) between the two variables x and y and identify its type

Using the scientific calculator:

A lot of scientific calculators in the markets are supporting to find the sums of the columns in the table above and calculate the correlation coefficient as follows:

- **Preparing the calculator for the statistics system:**

  By pressing MODE then 3

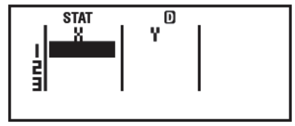
Statistical and regression calculations	MODE 3 (STAT)

  Select from the drop-down list:

Paired-variable (X, Y), linear regression (y = A + Bx)	2 (A+BX)

- **Entering the data:**

  MODE 3 (STAT) 2 (A+BX)



  Fill the shown table in the figure in all the values of (x and y) by typing the number existed in a table After typing the number, press = until you finish typing all the values of (x and y).

- **Recalling the products:**

  Press the buttons. SHIFT 1 (STAT) to get 3:sum

  And select from this list each of n:

  1 : Σx^2 , 2 : Σx , 3 : Σy^2 , 4 : Σy , 5 : Σxy

  By pressing the buttons from 1 to 5 severally liable.

To find the correlation coefficient (r) press the following buttons:

(STAT), from the drop-down list, press 5 : Reg

And from the drop-down list, press 3: r to get the value of the correlation coefficient required between the two variables x and y.

Activity

Use the calculator to check the answer of the example above

The program (SPSS)

The program (SPSS) is the abbreviation of (Statistical Package for Social Science). It is a set of comprehensive mathematical packs or data to analyze such data. This program is used in the scientific researches involving digital data.

The program can read all the data of different files, analyzing them, getting the results and giving statistical reports. The programs also allows the users to edit the data in the form of variables and new data using the equation. You can also use the program to save the data in files and naming or editing the names of the data files. You can also pack up the data and files through controlling the order and selection list available in this program. This can be done to include all the stages of analyzing the data and the statistical process via four main steps:

1 - Coding the data.
2 - Placing the data in the program.
3 - Selecting the proper form and testing and analyzing the data.
4 - Identifying the variable data to be analyzed and to fulfill the statistical process.

Operating the SPSS program:

You can start the SPSS program by pressing the window (START) in the main list, then go to the (program) list and search for the SPSS program within two times to open the program.

Components of the programs and their function:

Command list:

It is a special task bar for this program where the user can select the order which he /she wants through pressing the icon of each statistical order and then the results are viewed in the report board. The command list contains nine main orders. When you press any of them, they get branched into sub- orders besides, the (help) icon.

Data view:

It is a screen where the user can control adding or cancelling the dependent data for each variable. The user can add any independent variable in a column on the data view where the user can convert to view or watch the variables by pressing and moving between the two orders (Variable View) and (Data View) placed on the left of the variable view.

Variable view:

The view of defining the variable data which contains parallel columns and each column contains special data for each variable. To view the definition of each variable, the user ought to use the mouse to do (Double Click) or to press the order (Variable View) placed under the definition view. At this very minute, the shape of the view get changed and the definition bar appears:

Name	Type
Width	Values

When you press the definition bar, the coding (denotation)appears. After that, press the button (Add) to view the value of the symbol and position.

Steps you can control:

(1) **The ability to recall the previous data:** The user can control the process of recalling the data and files by pressing the button (File), then the button (Open). The user selects the file containing the data wanted to be recalled which contain the statistical reports done before and finally presses the button (Save).

(2) **Saving the new changes in a file:** The user can save the new changes in a file by pressing the order (Save) or the order (Save as) to save and name the new selected file.

(3) Adding edits and managing the variables: The user goes to the (Data Editor) and add the data he /she wants to be able to:

- edit the value of the data.
- Defining the variables; identifying the type of the added data, the economic indicators and all the variables.

(4) The user can add a new variable view and watch the arrange of the changes happened by using the main order (Data), then follow each variable that the user wants such as adding a variable or adding a new view or editing the arrangement of the data.

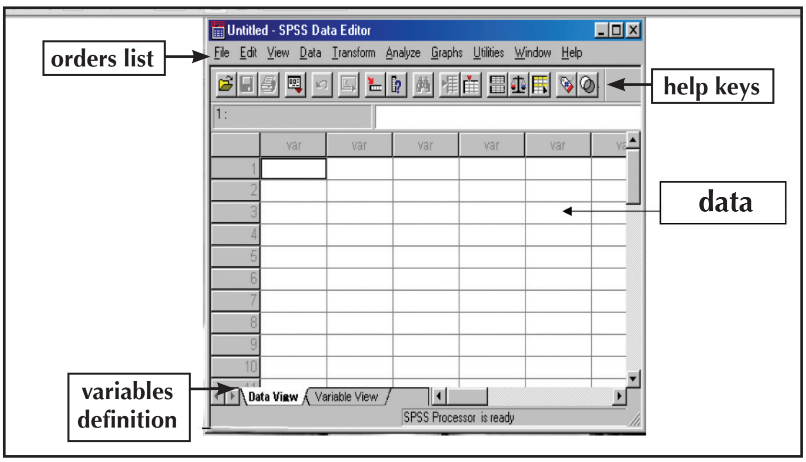
(5) Forming a fully new variable using an equation where the user goes to (Transform), then moving to (Compute) and identifies the name of the new variable in the list (Targer Variable).

(6) The ability to cancel any variable or view.

(7) Ordering the view where the program sets up a new variable containing a serial number in order to order the views ascendingly or descendingly.

(8) Performing a statistical process identifying and graduating the statistical description and the repetition of data.

(9) The ability of graphing the variables through setting up a graph to view the variable analysis and interpret what happened to the new variables.



Activity

You can visit the next site to download the SPSS: *http://www-01.ibm.com/software/analytics/spss/* to check the example above.

Example

(2) Find the pearson's correlation coefficient between the two variables x and y and identify its type if.

$\Sigma x = 68$ $\Sigma y = 36$ $\Sigma xy = 348$

$\Sigma x^2 = 620$ $\Sigma y^2 = 204$ $n = 8$

Solution

$$\because r = \frac{n\,\Sigma xy - (\Sigma x \times \Sigma y)}{\sqrt{n\,\Sigma x^2 - (\Sigma x)^2}\,\sqrt{n\,\Sigma y^2 - (\Sigma y)^2}}$$

$$\therefore r = \frac{8 \times 348 - (68 \times 36)}{\sqrt{8 \times 620 - (68)^2}\,\sqrt{8 \times 204 - (36)^2}} = \frac{336}{\sqrt{336}\,\sqrt{336}} = 1$$

The value of the correlation coefficient (+1) means that the relation between the two variables x and y is perfect direct.

Try to solve

2. Find the pearson's correlation coefficient between the two variables x and y and identify its type if.

$\Sigma\, x = 92$ $\quad\Sigma\, y = 36$ $\quad\Sigma\, x\, y = 372$

$\Sigma\, x2 = 1100$ $\quad\Sigma\, y2 = 204$ $\quad n = 8$

Spearman's Rank Correltion Coefficient

Think and discuss

A statistician has studied the relation among the degrees of two school subjects for seven students and has recorded the results in the next table:

Subject 1	Weak	Pass	Weak	good	Weak	excellent	Very good
Subject 2	Weak	Pass	good	Pass	Weak	Very good	Pass

can you help the statistician , to realize the relation between those two subjects and find the correlation coefficient between them?

We cannot use Pearson's correlation coefficient in "Think and discuss" since it depends on the quantitative (numerical)data only. In this case, the data are descriptive (as in the previous title). another correlation coefficient can be used _ Spearman's rank correlation coefficient. It gives a measure to the correlation in both the descriptive and quantitative data which are ordered as in the previous title. This coefficient relies on the order of the variable values taking into consideration the ascending and descending order. Then, we use the next relation:

$$r = 1 - \frac{6\Sigma\, D^2}{n(n^2 - 1)}$$

Where D is the difference between the order of the two variables x and y and n is the number of the values of each variable.

Notice that:

- Spearman's correlation coefficient can be calculated whether the data are quantitative or descriptive, whereas Pearson's correlation coefficient can be calculated only if the variables are quantitative.
- Spearman's rank correlation coefficient is characterized with its easiness even the data are not ordered.
- The Spearman's correlation coefficient does not regard the difference of the numbers when the ranks are calculated and though, it is less accurate.

Example

3 Find Spearman's rank correlation coefficient in the previous "Think and discuss" and identify its type.

Solution

In this example, we order the two phenomena a regular ascending order by giving each student a rank for a subject so does the other subject for the same student as shown in the next table.

Subjects 1	Weak	Pass	Weak	good	Weak	excellent	Very good
Ordering with repetition	1	4	2	5	3	7	6
Final repetition	2	4	2	5	2	7	6

We notice that the case "weak" has repeated three times and occupied the places 1,2 and 3.

hence, the rank of each = $\frac{1+2+3}{3}$ = 2 (it is the arithmetic mean of the numbers 1,2 and 3) , and similarly,

Subjects 2	Weak	Pass	good	Pass	Weak	Very good	Pass
Ordering with repetition	1	3	6	4	2	7	5
Final repetition	1.5	4	6	4	1.5	7	4

We notice that the level "weak" has repeated twice and occupied the places 1 and 2.

hence, the rank of each = $\frac{1+2}{2}$ = 1.5 it is the arithmetic mean of the numbers 1 and 2).

Also, the level"pass" has repeated three times and occupied the places 3,4 and 5.

hence, the rank of each = $\frac{3+4+5}{3}$ = 4 We summarize the solution in the table as follows:

x	y	Ranks of x	Ranks of y	D	D^2
Weak	Weak	2	1.5	0.5	0.25
Pass	Pass	4	4	zero	zero
Weak	good	2	6	– 4	16
good	Pass	5	4	1	1
Weak	Weak	2	1.5	0.5	0.25
excellent	Very good	7	7	zero	zero
Very good	Pass	6	4	2	4
					21.5

$\because$ r $= 1 - \dfrac{6\Sigma D^2}{n(n^2-1)}$

$= 1 - \dfrac{129}{336} \simeq 0.6161$

$\because$ r $= 1 - \dfrac{6 \times 21.5}{7(49-1)}$

It is a direct correlation.

Try to solve

3 In a study about the relation between the students' levels in statistics and mathematics, the degrees of six students have been as follows:

Degrees of statistics(x)	Pass	Very good	excellent	Very good	Pass	Pass
Degrees of mathematics (y)	good	good	Very good	excellent	good	Weak

calculate the Spearman's rank correlation coefficient among the degrees and identify its type.

Example

4 calculate the Spearman's rank correlation coefficient between x and y through the data of the next table:

x	4	7	8	5	8	12
y	7	6	6	4	6	10

Solution

Form the next table:

x	y	ranks x	ranks y	D	D^2
4	7	6	2	4	16
7	6	4	4	0	0
8	6	2.5	4	–1.5	2.25
5	4	5	6	–1	1
8	6	2.5	4	–1.5	2.25
12	10	1	1	0	0
					21.5

$\because$ r $= 1 - \dfrac{6\Sigma D^2}{n(n^2-1)}$ $\therefore 1 = \dfrac{6 \times 21.5}{6(36-1)}$ $\simeq 0.3857$ The correlation is direct.

Critical thinking: Does ΣD^2 vary if we order the two phenomena x and y ascendingly? explain your answer

Try to solve

4 Calculate the Spearman's rank correlation coefficient between x and y and identify its type from the data shown in the next table:

x	10	7	8	7	6	4
y	5	8	7	9	9	10

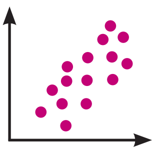
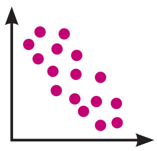
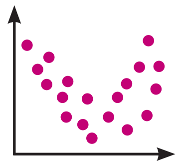
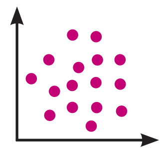
Exercises 1 - 1

First: Choose the correct answer:

1. The strongest correlation coefficient of the following is:
   a) – 0.94   b) zero   c) 0.5   d) 0.85

2. The strongest inverse correlation coefficient of the following is:
   a) – 0.2   b) – 0.5   c) – 0.7   d) – 0.8

3. the scatter diagram representing inverse correlation is:

   a)    b)    c)    d) 

4. The weakest correlation coefficient of the following is:
   a) - 1.2   b) - 0.7   c) 0.12   d) 0.9

5. One of the next numbers can represent the strongest inverse correlation coefficient between two variables:
   a) 0.3   b) 0.9   c) - 1.1   d) - 0.95

Answer the following questions

6. From the data of the next table

x	12	10	14	11	12	9
y	18	17	23	19	20	15

First: Calculate Spearman's rank correlation coefficient between the two variables x and y.

Second: Calculate Pearson's linear correlation coefficient between x and y.

7. From the data of the next table

x	7	7	8	3	7	11
y	8	4	12	2	10	11

Calculate Spearman's rank correlation coefficient between the two variables x and y.

8. From the data of the next table

x	1	3	4	6	7	9
y	6	4	4	3	2	1

Calculate Pearson's correlation coefficient between the values of x and y and show its type.

9. From the data of the next table

x	6	5	7	8	10	6	7
y	4	7	5	6	8	7	8

Calculate Pearson's correlation coefficient between the values of x and y and show its type.

10 From the data of the next table

x	3	1	6	4	3	8
y	7	4	5	8	6	7

Calculate Spearman's rank correlation coefficient between x and y and show its type.

11 From the data of the next table

x	Very good	Very good	good	Weak	Pass	Very good
y	good	Pass	good	excellent	Very good	Pass

Calculate Spearman's rank correlation coefficient between x and y.

12 Find Pearson's correlation coefficient between the values of x and y and show its type if:

$\Sigma x = 220$ $\Sigma y = 140$ $\Sigma xy = 2658$

$\Sigma x^2 = 5486$ $\Sigma y^2 = 2292$ $n = 10$

13 Trade: The following table shows a 6 - book group in regard to its price (x) and the sales (y).

Price (x)	Low	Very low	Average	Very high	high	Very high
Sales (y)	high	high	Very high	Low	Average	Low

Calculate Spearman's rank correlation coefficient between the price and sale of the book.

14 Advertisement: An advertising company has decided to study the relation between its expenditure on advertisement x (in thousand L.E) and its sales y (in thousand units). If the data of the company's eight branches are as follows:

x	19	18	7	10	4	13	15	5
y	12	10	7	9	6	13	14	12

Find the rank correlation coefficient between the expenditure and sales and show its type.

15 Education The next data represent the marks of ten students in chemistry and biology.

chemistry	60	85	55	90	65	50	80	70	95	75
biology	55	75	50	95	60	65	85	80	90	70

Calculate Pearson's linear correlation coefficient and determine its type.

16 Birth: In a study to determine the relation between the mother's age and the number of her kids, the following data have been determined:

mother's age	18	20	23	27	29	32	33	35
number of kids	2	1	1	2	3	4	3	5

Calculate spearman's rank correlation relation and determine its type.

Unit one

1 - 2

Regression

Regression

You will learn

- Definition of regression
- Types of regression
- Regression line equation
- Method of least squares
- Activities on finding regression line equation

Key terms

- Regression
- Regression line
- Least squares

Introduction:

You have previously studied the function and identified its graph. In the previous lesson, you learned the scatter diagram and you also realized that the objective of graphing it is to identify the nature of the relation between two variables x and y through the data related to both variables. You also learned that the properties of the correlation between two phenomena can be in the form of the following:

Linear Relation negative Linear Relation

non-Linear Relation no Relation

In this lesson, you are going to learn how to identify the regression line equation. The purpose of this study is to help the researcher know the type of the given data and perform correct predictions through them.

Remember

- **The function is a relation between the two sets x and y where each element of set x elements has only one element of set y elements**
- **The function is defined whenever each domain, Codomain and rule are known.**

Definition Regression is a statistical method by which the value of a variable can be estimated in terms of the other variable.

It has several types:

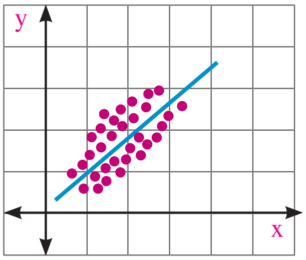
a **simple linear regression:** in which, the dependent variable (y) depends on one variable (x) through a linear relation.

b **multiple regression:** in which, the dependent variable (y) depends on more than an independent variable.

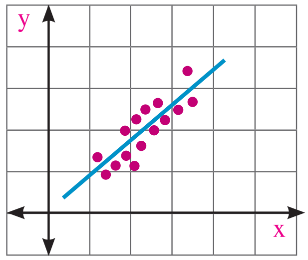
c **Non-linear regression:** if the relation between the dependent variable (y) an nonlinear independent variables (of second degree, third degree, exponential, or logarithmic....).

In this lesson, we concentrate only on the simple linear regression. The following figures illustrate the relation between the value of the correlation relation and the difference of the point positions on the regression line. The closer the points are congruent to this line, the more the value of (r) increases or decreases till all the points get congruent on the line. In this case, the value of (r) is either (+1) or (-1).

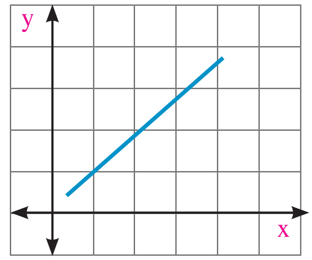
materials: scientific calculator Microsoft Excel



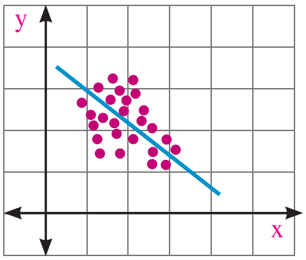
(1) moderate direct correlation



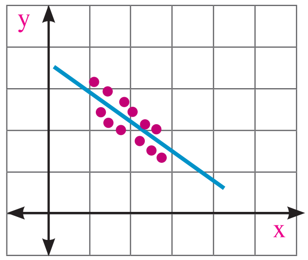
(2) strong direct correlation



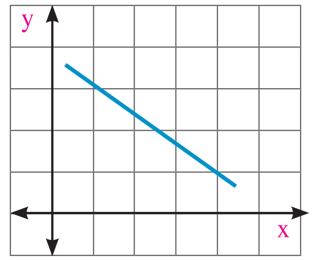
(3) perfect direct correlation



(4) moderate inverse correlation



(5) strong inverse correlation correlation



(6) perfect inverse correlation

Regression line equation:

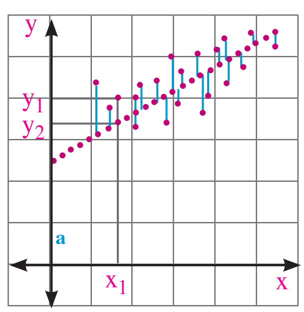
You have previously learned in the analytic geometry the straight line equation whose slope is m and y-intercept of c is y = mx + c.

In turn to the scatter diagrams shown previously, we find that if the scatter diagram starts as in figures (2) or (5), this indicates that the relation between the two variables are linear. Since we can imagine in the existence of a straight line, the points are located on and near the straight line even all the points do not lie on it. If the scatter diagram starts as in figures (1) or (4), we doubt about the linearity of the relation between the two variables. As a result, our main mission here is to use the seen pairs of values (x_r, y_r) to find the best straight line matching the points of the set of the sample. Let its equation be:

$$y = a + bx$$

The most common method to find the best values of a and b is called the least square method.

The least square method:



Of the mentioned previously, we knew that in case of correlation, it is not necessary all the points lie on the regression line. Therefore, there is an error ratio for the points which do not lie on the regression line. To get the best regression line, the least possible valued deviations should be reduced (the proper regression line should pass or approach to the greatest number of scattering points). if (x, y) is a true point of data and $(x, \hat{y})$ is the point locating off on the regression line ($\hat{y}$ is read as y hat), then the proper regression line when $|\hat{y} - y|$ is the least for the values of x or when $\Sigma(\hat{y} - y)^2$ is the least and by supposing the regression line equation is $\hat{y} = a + bx$.

The absolute difference = $|(a + bx)-y|$

The required is to determine the values of a,b such that the absolute difference is minimum by solving the following two equations

$\Sigma y = n a + b \Sigma x$ (1) $\Sigma x y = a \Sigma x + b \Sigma x^2$ (2)

From equation (1) $a = \dfrac{\Sigma y - b \Sigma x}{n}$ And by substituting in (2).

$b = \dfrac{n \Sigma x y - (\Sigma x)(\Sigma y)}{n \Sigma x^2 - (\Sigma x)^2}$ is called the regression coefficient of y on x. It expresses the slope of the regression line on the positive direction of x axis.

The regression line equation of y on x is used for:

1- predicting the value of y if the value of x is known.

2- identifying the error which can be identified by the relation:

Error = | Table value - the value satisfying the regression equation |

Note: when the regression equation is used for prediction (estimation), it is favorable not to exceed the range of the variable x used in calculating the regression equation much.

Critical thinking: the value of the regression coefficient refers to the correlation. Explain.

Example

1 The next table illustrates the production of a summer crop (y) of the cultivated land (x) in feddan.

Cultivated land (x) in feddan	50	200	110	80	120	74.5	88.9	5.7	11	3.2
Production of (y) in kg	140	500	400	300	356	240.5	200.6	33.5	69.8	18.7

First: find the regression line equation.

Second: predict the value of the production in kg if the cultivated land is 100 feddans.

Third: find the error in the production if you know that the cultivated land is 120 feddans.

Solution

Solution by using the scientific calculator:

1- enter the data:

Follow up the same method explained previously in Example (1) in the previous lesson (correlation) to enter the data.

2-recall the products(results):

Press the following buttons:

Use the following buttons to find the results of the following operations: SHIFT 1 (STAT) from the drop-down list and press the button Select 3

A new list from 1 to 8 (sums of results) appears and you select as follows:

$2 : \Sigma x = 743.3$

$4 : \Sigma y = 2259.1$

$1 : \Sigma x^2 = 89017.19$

$5 : \Sigma xy = 254489.18$

First: we calculate the value of the constant b from the relation:

$$b = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{10 \times 254489.18 - 743.3 \times 2259.1}{10 \times^2 (743.3) - 89017.19} \simeq 2.5637$$

we calculate the value of the constant a from the relation: $a = \bar{y} - b\,\bar{x}$

Where : $\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$

$\therefore \bar{x} = \frac{743.3}{10} = 74.33$, $\bar{y} = \frac{2259.1}{10} = 225.91$

$\therefore a = 225.91 - 2.5637 \times 74.33 \simeq 35.35$

Note:

The constant can be directly calculated as follows:

$\because a = \frac{\Sigma y - b \Sigma x}{n}$ $\quad \therefore a = \frac{2259.1 - (2.5637 \times 743.3)}{10} \simeq 35.35$

$\therefore$ The regression line equation is : $\hat{y} = 2.564x + 35.35$

Second: From the regression line equation: $\hat{y} = a + bx$

$\because \hat{y} = 2.564x + 35.35$, $x = 100$

$\therefore \hat{y} = 2.564 \times 100 + 35.35 = 291.72$ Kg

The result can be checked using the calculator as follows:

100 SHIFT 1 (STAT) 5 (Reg) 5 : $\hat{y}$ =

Third: To find the error in the production if you know that x= 120 feddans:

$\because \hat{y} = 2.564x + 35.35$

$\therefore \hat{y} = 2.564 \times 120 + 35.35 \simeq 343$

$\because$ Error = |Table value - the value satisfying the regression equation|

$\therefore$ Error = $|356 - 343| = 13$

Activity

First: check the solution of the example above using (Microsoft Excel).

Second: check the solution of the example above using the statistical program (SPSS).

First: using (Microsoft Excel)

1- start (Microsoft Excel) and enter the previous data in the cells of the two columns a and b under the name x and y as two real variables or the real name of such data as shown in figure 1.

2- from the tool bar, press chartwizard to get chart type, then from the list xy scatter, press finish as in figure 2.

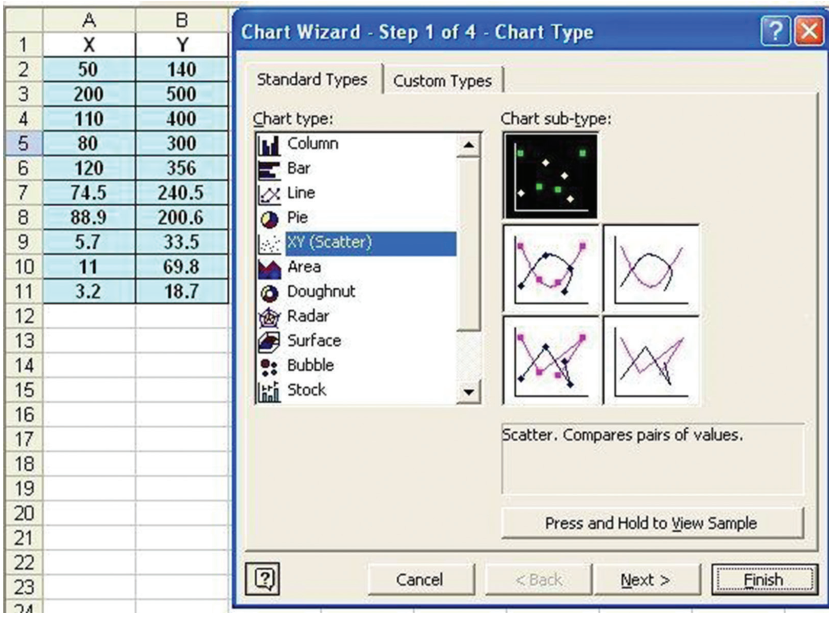


figure (2)

3- figure 3 illustrates the graphical representation of the points listed in the table and is called scatter diagram. We select the figure shaded in black which appears here after changing the background as shown in the figure.

4- the values on the horizontal axis represent the values of X for the data and the vertical axis of the values Y. were, we are going to find the regression line equation which is in the form:

$y = a + bx.$

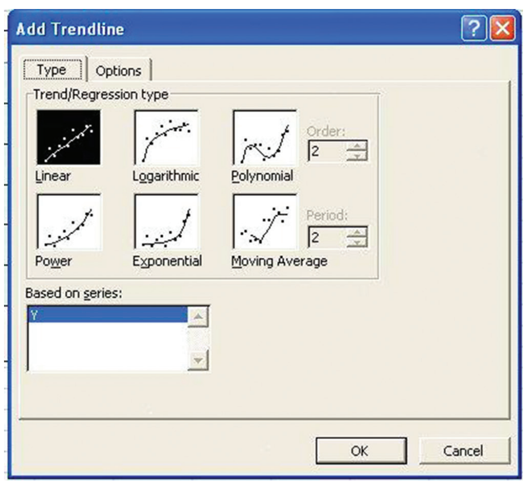


figure 3

5- do a mouse right click on a point (in figure 4). the shown list appears where we select Add Trendline and by doing a mouse click, we get the following figure which appears in six scatter diagrams. We have selected the first scatter diagram as shown by shading in black as a proper choice. Since we need the straight line, we use Option to identify the required by doing a mouse click where the following box appears:

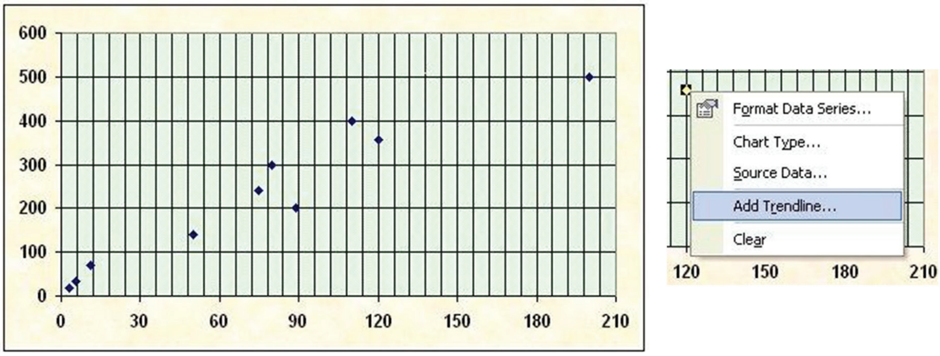


figure (4)

6 - label Display equation on chart as shown in figure 5.

7 - press OK to get the required:

a the figure showing the regression line is the moderate of the points representing the data pairs.

b the regression line equation in figure 6. were, we have moved the equation from its place in the figure above and changed the line to illustrate the matter.

The following figure is the result of the operation and it shows the required especially the following equation:

$Y = 35.35 + 2.5637x$

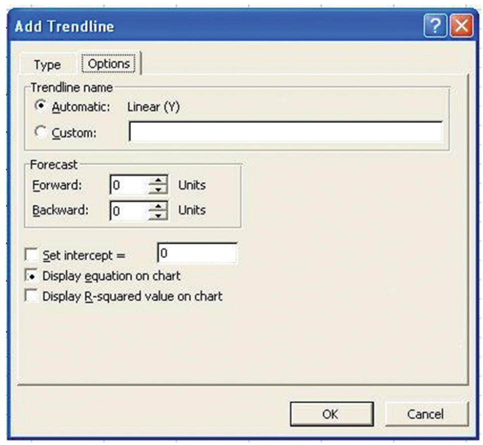


figure (5)

It is a regression line equation and it is the same equation which we have got in the previous solution.

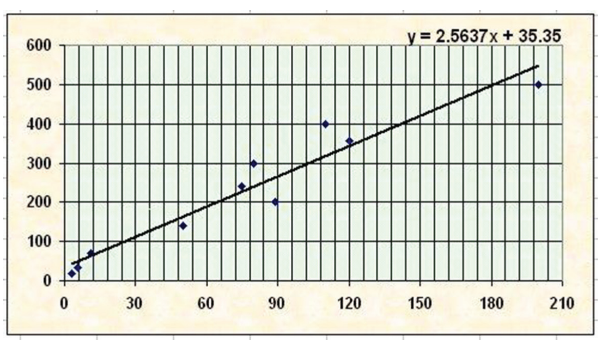
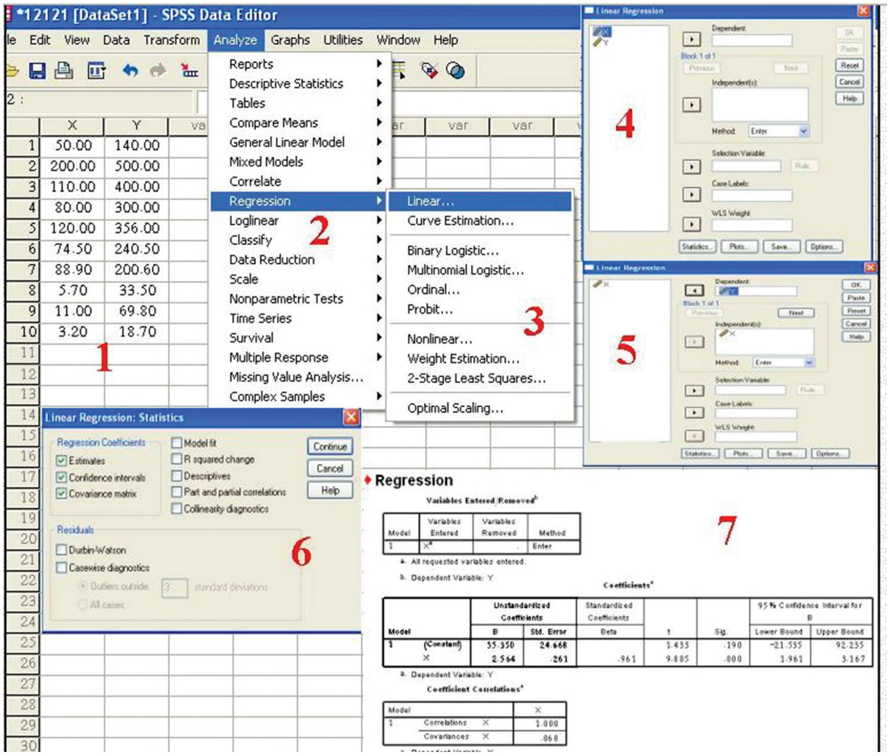


figure (6)

Using SPSS program



(figure 7)

Example

(2) Mining: the following table shows the data about the average price of the oil barrel and the rates of the economic growth in a country within eight years. Required is to:

Price of the oil barrel (x)	36	40	36.2	31.1	29.7	16.3	18.7	14.6
Rates of the economic growth(y)	0.91	3.5	3.2	2.7	2.3	– 1	– 0.9	– 1.6

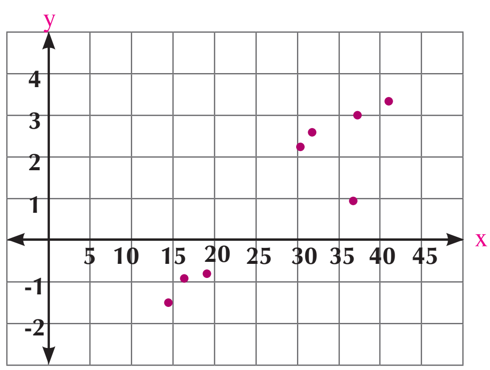
First: graph the scatter diagram and show the type of the correlation.

Second: find the regression line equation of the given data.

Third: predict the economic growth when the price of the oil barrel is \$15, then when the price is \$35.

Solution

First: the figure opposite represents the scatter diagram and it shows also the correlation is direct.



x	y	x2	y2	x y
36	0.91	1296	0.8281	32.76
40	3.5	1600	12.25	140
36.2	3.2	1310.44	10.24	115.84
31.1	2.7	967.21	7.29	83.97
29.7	2.3	882.09	5.29	68.31
16.3	– 1	265.69	1	– 16.3
18.7	– 0.9	349.69	0.81	– 16.83
14.6	– 1.6	213.16	2.56	– 23.36
222.6	9.11	6884.28	40.2681	384.39

From the table data:

$\Sigma y = 9.11$ $\qquad \Sigma x = 222.6$

$\Sigma\, x\, y = 384.39$ $\qquad \Sigma x^2 = 6884.28$

Second: we calculate the value of the constant b from the relation:

$$b = \frac{n\,\Sigma x\,y - \Sigma\,x\,\Sigma\,y}{n\,\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{8 \times 384.39 - (222.6 \times 9.11)}{8 \times^2 (222.6) - 6884.28} \simeq 0.1896$$

$\because a = \dfrac{\Sigma y - b\Sigma\, x}{n}$ $\qquad \therefore a = \dfrac{9.11 - (0.1896 \times 222.6)}{8} \simeq -4.1368$

$\because$ The regression line equation is: $\hat{y} = a + b\,x$

$\hat{y} = 0.1896 \;\therefore\; x - 4.1368$

Third:

When $x = 15$ $\qquad$ then : $\hat{y} = 0.1896 \times 15 - 4.1368 \simeq -1.2928$

When $x = 35$ $\qquad$ then : $\hat{y} = 0.1896 \times 35 - 4.1368 \simeq 2.4992$

Try to solve

1 in a study to find the relation between the income (x) and consumption (y) in thousand pounds, the results show that:

$\Sigma x = 120$, $\quad \Sigma y = 100$, $\quad \Sigma xy = 516$

$\Sigma x^2 = 720$, $\quad \Sigma y^2 = 410$, $\quad n = 40$

a find the linear correlation coefficient between x and y using the Pearson's method and identify its type.

b find the regression line equation.

c predict the value of consumption (y) when the income is 10000 LE

Exercises 1 - 2

First: choose the correct answer:

1. the statistical equation of the regression line equation where b is the regression coefficient is:

   a) $\hat{y} = a\,x + b$ b) $\hat{y} = a + b\,x$

   c) $\hat{y} = a\,y + b$ d) $\hat{y} = a + b\,y$.

2. if the regression line equation is: $\hat{y} = 2 + 0.5x$ then the value of y expected when x = 6 is:

   a) 4 b) 5 c) 7 d) 8

3. if the two points (11.5 , 10) and (6.5 , 5) lie on the regression line y on x, then the correlation between x and y is:

   a) direct b) inverse c) perfect d) nihilistic

4. if the two points (5 , 13) ,and (14 , 4) lie on the regression line y on x, then all the following points lie on the same line except:

   a) (15 , 5) b) (10 , 8) c) (6 , 12) d) (5 , 13)

5. if all the points in a scatter diagram lie on a straight line whose slope is negative, then the correlation coefficient between x and y equals:

   a) 1 b) zero c) - 0.5 d) -1

6. if all the points in a scatter diagram lie on a straight line whose slope is positive, then the correlation coefficient between the two variables equals:

   a) - 1 b) zero c) $\frac{1}{2}$ d) 1

Second: Answer the following questions:

7. the next table illustrates the relation between the two variables x and y:

x	5	8	10	14	16	20
y	4	6	9	11	12	15

   a) graph the scatter diagram b) find the regression line equation

   c) predict the value of y when x = 12

8. from the data in the next table:

x	20	33	30	40	13	15	26	25
y	7	8	9	11	4	5	8	9

   a) predict the value of y when x = 35
   b) find the error in y if x= 30

9 in a statistical study to find the relation between two variables x and y, we have got the next data:

$n = 10$, $\overline{x} = 8$, $\overline{y} = 10$, $\Sigma xy = 870$, $\Sigma x^2 = 665$, $\Sigma y^2 = 1400$ find:

a linear correlation coefficient

b regression line equation

10 if : $\Sigma x = 30$, $\Sigma y = 40$, $\Sigma xy = 162$

$\Sigma x^2 = 210$, $\Sigma y^2 = 304$, $n = 6$ find :

a regression line equation

b linear correlation coefficient between x and y and identify its type

11 sales: In a market for selling used cars, the sales were as follows:

car age (x)	3	2	1	1	5	6	1	4
Sale price (y)	54	80	74	98	45	40	85	60

Find:

a Pearson’s linear correlation coefficient

b Regression line equation

12 Economy: the next table represents the monthly income (x) and the expenditure (y) for a group of families in hundred pounds:

Income (x)	38	27	39	40	56	66	42	44
expenditure (y)	19	25	20	28	31	38	27	22

a find Pearson’s rank correlation coefficient and identify its type.

b find the regression line equation

c Estimate the value of the expenditure (y) if the income (x) is 5000 LE

d find the error in (y) if (x)= 40

13 Family to study the relation between the income (y) and consumption (x) in hundred pounds monthly in a city, a 40- family sample has been taken to give the following results:

$\Sigma x = 100$, $\Sigma y = 120$, $\Sigma xy = 516$, $\Sigma x^2 = 410$, $\Sigma y^2 = 720$.

a find the regression line equation

b predict the family income whose monthly consumption is 700 LE.

Unit
2

Advanced Measurements in Statistics

introduction

Statistical measures are an essential part of applied science in tools used to measure various phenomena and variables. These measures help us in summarizing and analyzing data, understanding the relationships between variables, deducing results, and predicting the occurrence of some phenomena. Statistical standards vary according to the type and characteristics of the data you are working on, such as displaying data using the stem-and-leaf method, calculating quadrilaterals for a group of data, and representing them graphically. Calculating the semi-interquartile range for a set of data using a frequency table and using the stem and leaf method, all of this through life applications in various fields such as computer science, medicine, industry, agriculture, etc., which makes the student appreciate the importance of studying metrics. Statistics in life.

Unit objectives

It is expected that after the student studies this unit and carries out the activities:

- Display a data set using the “Stem-and-Leaf” method.
- Compare two data sets using the “Stem-and-Leaf” method.
- Understand the advantages and disadvantages of using the “Stem-and-Leaf” method for displaying data.
- Calculate and graph quartiles for a data set.
- Calculate the semi-interquartile range using a frequency table and the “Stem-and-Leaf” method.
- Appreciate the importance of statistics in daily life.

Key Terms

- *Representation by stem and leaves*
- *Stem and Leaves*
- *Double representation of stem and leaves*
- *Semi interquartile range*
- *Frequency table*
- *Lower quartile (first)*
- *Middle quartile (second)*
- *Upper quartile (third)*
- *Ascending cumulative frequency*

Lessons of the unit

Lesson (2 - 1): Displaying Data Using the Stem-and-Leaf Method.

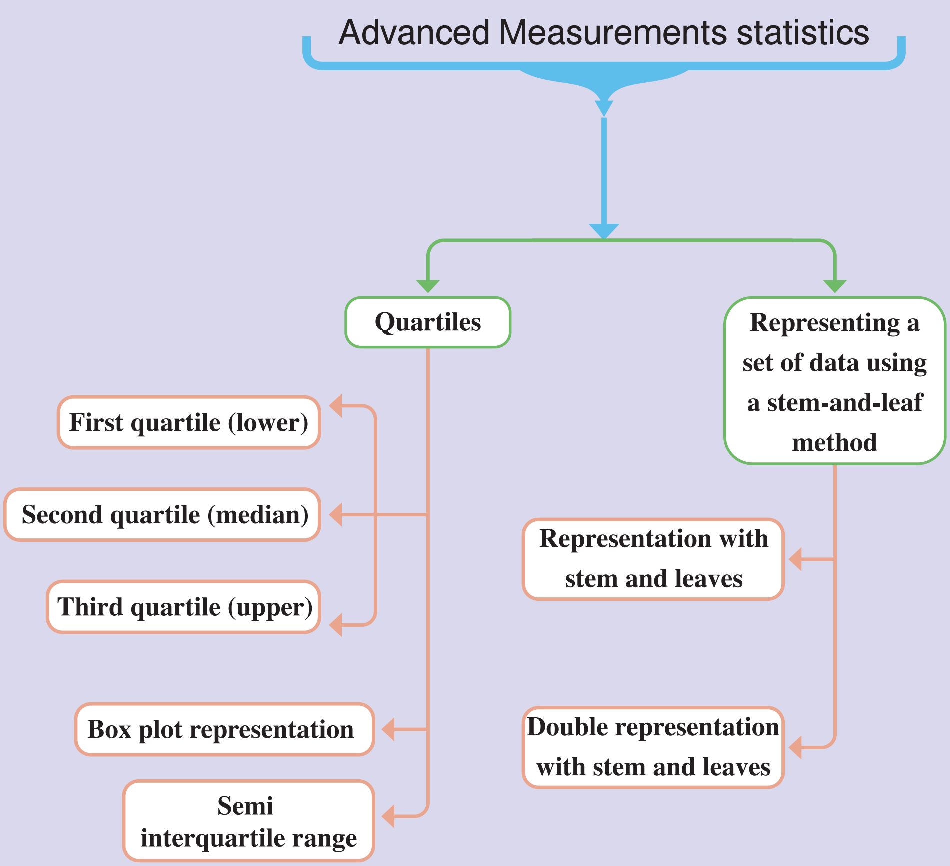
Lesson (2 - 2): Quartiles and its graphic representation.

Lesson (2 - 3): Semi-interquartile Range.

Materials

Scientific Calculator

Unit planning organizer



Unit Two

2 - 1 Stem-and-Leaf Data Representation

You will learn

- Representing a set of data using a stem-and-leaf method
- Using the stem and leaves method to compare between a set of data

Key terms

- Stem
- Leaves
- Double representation with stem and leaves

Think and discuss

points			
19	6	7	10
11	13	18	25
21	12	5	12
20	21	11	12

The following data represents the points scored by 16 basketball players in one team of the school teams

Find

(1) The largest number of points scored by one of the players.

(2) Number of players who scored more than 10 points.

Learn

Represent this data using the stem-and-leaf method.

When representing data 8, 135, 71, 3452 using the stem-and-leaf method, we arrange the data in ascending order. The digit in the smallest place value (ones) represents the leaf, and the remaining part of the number represents the stem, as shown in the table:

Number
8
71
135
3452

Stem	Leaf
0	8
7	1
13	5
345	2

Example

1. Represent the data in think and discuss using stem and leaves method.

Solution

Step 1: Find the largest and smallest values in the data, then determine the tens digit for each.
The smallest value is 5, with a tens digit of 0.
The largest value is 25, with a tens digit of 2.

Stem	leaves								
0	7	6	5						
1	0	9	8	3	1	2	2	2	1
2	5	1	1	0					

Key 25 = 2|5

Step 2: Draw a vertical line, and another horizontal line then record the stem on the left side of the line.

Step 3: Write the leaves corresponding to each stem on the opposite side of the line. For example, for the number 19, write 9 to the right of the number 1 (the stem), and for the number 6, write 6 to the right of the number 0 (the stem), and so on, until all the data are recorded.

Repeat the leaf according to how many times it appears in the data.

Step 4: Arrange the leaves in ascending order, then create a key that explains how to read the data.

The leaves are written, no matter how many times they are repeated.

Represents tens →

stem	Leaves
0	5 6 7
1	0 1 1 2 2 2 3 8 9
2	0 1 1 5

← Represents ones

Key 13 =1|3

It is noted that

The Greatest number of points scored by one of the players = 25 points.

The number of players who scored more than 10 points = 12 players.

Remember

For any set of values it is

Arithmetic mean = $\frac{\text{sum of values}}{\text{their number}}$

Median : It is the value that is in the middle of a group of visual values in ascending or descending order

Mode : It is the most frequent or common value

Try to solve:

1 The following data shows the marks of some students in a mathematics exam.

92	78	73	89	86
85	76	81	73	88
83	75	83	83	71
86	82	94	100	98

Required:

a Representation of data using a stem-and-leaf plot (method):

b Calculate the median of these marks.

c If an "Excellent" grade is given to students who score 85 or more, how many students their grade is an "Excellent" grade?

Sports

Example

2 The following data represents time of the cycling race in one of the Olympic Games, measured in seconds:

89.4	90.4	87.5	84.3	89.7	90.3	91.4
91	86.7	84.1	89.2	86	89.1	88.2
89.5	90.5	90.2	89.2	91.1	88.9	–

Required:

a Representation of the data using a stem-and-leaf plot

b How long did it take the last competitor to reach the end of the race?

Solution

a Representation of the data using a stem-and-leaf plot

The data contains decimal numbers, which represent the smallest place (the leaves), and integer numbers represent the tens (the stem). The smallest integer is 84, and the largest integer is 91, so the stem is the numbers 84 to 91.

b The last competitor took 91.4 seconds

Bike racing time

Stem	Leaves
84	3 1
86	7 0
87	5
88	2 9
89	4 7 2 1 5 2
90	4 3 5 2
91	4 0 1

Leaves order →

Key 88.2 = 88 |2

Bike racing time

Stem	Leaves
84	1 3
86	0 7
87	5
88	2 9
89	1 2 2 4 5 7
90	2 3 4 5
91	0 1 4

Try to solve

weight

2 The opposite representation shows the average weights of chicks in grams.

chick weights

Stem	Leaves
5	0 9
6	1 5 7 8
8	3 3 3 5 7 8
9	0 1 5 5 9

Key 83 = 8|3

a What is the minimum and maximum weight?

b What is the median of these weights?

c What is the mode of these weights?

Learn

Double representation with stem and leaves

You can compare two sets of data using a double stem-and-leaf plot, where the stem is the same for both data sets. The leaves for the first data set are on the right side of the stem, while the leaves for the second data set are on the left side of the stem.

Example

3 The following data represents the maximum and minimum temperatures for Alexandria city over a period of two weeks.

Maximum temperature	19	28	22	29	25	29	32	35	36	34	37	39	41	42
Minimum temperature	13	22	19	18	16	20	21	22	23	23	21	30	32	31

Required: Represnt the temperatures by stem and leaves with description of this date and which of these data is more dispersion

Solution

The greatest maximum temperature is 42 and the smallest maximum temperature is 19.

The stem is from 1 to 4

From the corresponding figure we find that the maximum temperatures range between (19 - 42) while we find that all of the minimum temperature scores range between (13 - 32)

Maximum	stem	Minimum
9	1	3 6 8 9
9 9 8 5 2	2	0 1 1 2 2 3 3
9 7 6 5 4 2	3	0 1 2
2 1	4	

32 = 3|2   Key   13 = 1|3

Maximum temperature range = 23, minimum temperature range = 19

Thus: We find that maximum temperatures are more dispersion than minimum temperatures

Advantages of the method of representing data using stem and leaves

The original data are preserved, unlike in frequency tables where it is not possible to return to the original data after representing it in frequency tables, as you have previously studied.

Remember

Range

the differerence between the greatest and smallest values

Disadvantages

It is not suitable for large data set.

Try to solve

3) Health: The following table represents the number of males and female patients visiting a hospital during a week.

Represent the data using a stem-and-leaf plot with a description of these data, and determine which of these datasets is more diversion.

number of patients		
Section	males	females
General Surgery	52	47
ENT (Ear, Nose, and Throat)	61	42
Internal Medicine	42	42
Cardiology	60	17
Ophthalmology	44	42
Nephrology	50	54
Obstetrics and Gynecology	42	52
Pediatrics	55	42
Urology	49	29
Orthopedics and Fractures	46	37

Exercises 2 - 1

1. Put a mark (✓) in front of the correct statement and a mark (✗) in front of the incorrect statement for each of the following

The opposite representation represents Heights of a group of Trees in meter

Stem	Leaves
0	1 2 4 5 6 8 9
1	0 1 1 5 7
2	2 5
3	6

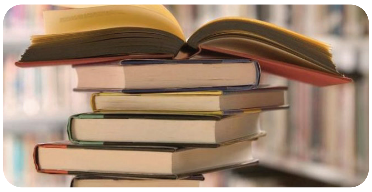
Key ⟶ 15 = 1|5

a. Most trees are less than 20 meters high. ( )
b. The median tree height is 11 meters. ( )
c. The range for tree height is 35 meters. ( )
d. The mode height of the trees is 11 meters ( )

2. The following data represents the number of mathematics books in the libraries of 15 schools

Stem	Leaves
0	1 1 1 2
1	0 1 1 1 2 2 3 3 4
2	1 1

Key ⟶ 13 = 1|3



It is required to write the original data of the number of books for each school.

Lengths:

3. The following data represents the heights of 30 students in a secondary school, measured in centimeters

161	175	174	165	167	177	180	182
170	157	170	176	185	188	162	165
159	172	158	171	169	173	175	178
164	158	172	170	178	181		

It is required to represent data using the stem and leaf method.

4. Represent each of the following data sets using the stem-and-leaf method:

First group	10	26	9	12	27	13	19	15	27	12	29	22
Second group	11	12	10	15	30	9	29	35	11	34	11	12
Third group	1.1	2.4	3	2	6.6	5.8	0.5	2.5	4.1	2.2		

5. Choose the correct answer from the given answers

Stem	Leaves
23	4 5
24	4 7 9
25	0 4 8 8
26	3 8 9
27	1 2 5

Key ⟶ 24.7 = 24|7

(1) In the opposite representation: the largest number is ....

a. 2.71 b. 23.5
c. 27.5 d. 275

(2) The median of the previous representation ....

a. 25.4 b. 25.8
c. 254 d. 258

Linking Temperatures :

6 The following data represents the maximum and minimum temperatures for some governorates of the Arab Republic of Egypt:

a Represent data using stem-and-leaf method (double representation)

b Find the median for each group separately.

c Which of these data is more dispersion?

Cities	Maximum temperature	Minimum temperature
Cairo	27	22
Giza	26	22
Fayyum	30	25
Alexandria	25	17
Damietta	26	18
Luxor	36	22
Aswan	41	32
Beni Suef	30	24

Unit Two

2 - 2 Quartiles and their graphic representation

Real Functions

You will learn

- Quartiles and their graphic representation.
- Finding quartiles from the frequency tables.
- Finding quartiles from stem and leaf method .
- Box plot diagram

Key terms

- Lower quartile
- Middle quartile
- upper quartile
- Box plot representation
- The stem
- The leaves
- Frequency table
- Ascending cumulative frequency

Think and discuss

Mathematics teachers in school ran out For a (midterm) mid-semester test for 200 students, and the results were recorded in the Excel sheet and arrange students using the program the students were divided into two equal groups by: A statistical measure is the median (one of the measures of central tendency) between underachievers and outstanding achievers to create appropriate strengthening programs for each level, but this division was not sufficient the course instructor asked students to be divided into the following levels:

(Weak - Acceptable - Good - Excellent). Therefore, it is necessary to divide the data into four equal sections.

How do you implement this, whether the data is single or represented by a frequency table or the stem method

What do we call the values that divide this data?to describe the level

Students' achievement.

Learn

After arranging the data in ascending or descending order, the values divide the data into four sections Equals are called quadrilaterals, and their number is three values

First quartile Q_1: It is the value preceded by a quarter of the data (25%) and followed by three-quarters of the data

Second quartile Q_2: It is the median, meaning the value that is preceded by $\frac{1}{2}$ of the data (50%) and followed by the other half

Third quartile Q_3:

Its the value preceded by $\frac{3}{4}$ the data (75%) and followed by quarter the data (25%).

Tools

- Scientific Calculater.
- Computer graphic programs.

Assigning quartiles from single (ungrouped) data

There are two cases

First case : If the number of data n is odd and (n + 1) is divisible by 4, then the quadrilaterals are one of values of the given data and are determined directly from them as follows:

The order of the first(lower) quartile $(Q_1) = \frac{n+1}{4}$

The order of the second(middle) quartile $Q_2 = \frac{n+1}{2}$ (median)

The order of the third (upper) quartile $Q_3 = \frac{3(n+1)}{4}$

Notes

The difference between the order of quartile and its value.

Example

1 Find the three quartiles of the following set of data :

23 , 7 , 16 , 2 , 4 , 5 , 1 , 21 , 15 , 22 , 14 , 11 , 10 , 9 , 3

Solution

1 ، 2 ، 3 ، 4 ، 5 ، 7 ، 9 ، 10 ، 11 ، 14 ، 15 ، 16 ، 21 ، 22 ، 23

4 → Q_1, 10 → Q_2, 16 → Q_3

First : Arrange the data in ascending order

Second : numbers of data n = 15(odd number)

n + 1 = 16 (divisible by 4) ∴ quartiles are one of the data values

The order of the first quartile $Q_1 = \frac{15+1}{4} = \frac{16}{4} = 4$ its value = 4

The order of the second quartile $Q_2 = \frac{15+1}{2} = \frac{16}{2} = 8$ its value = 10

The order of the third quartile $Q_3 = \frac{3(15+1)}{4} = \frac{48}{4} = 12$ its value = 16

Second case

If the number of data n is even or odd, (n + 1) is not divisible by 4, then the quartiles are assigned according to the following law:

The value of the required quartile =

its previous value + (its next value - **its previous value**) × (its order - its previous order)

Example

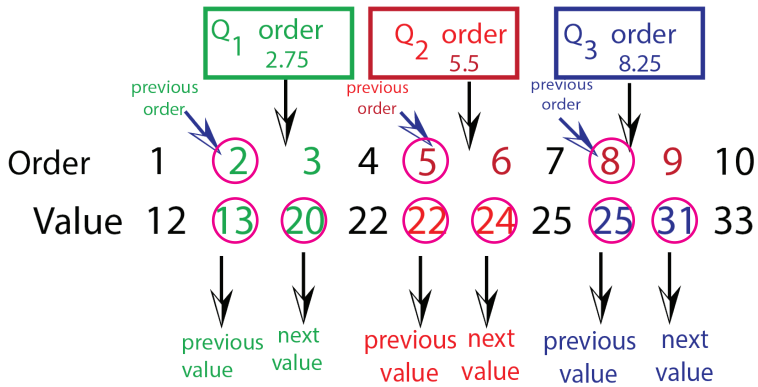
2 The data represented by the stem and leaves method represents the ages of 10 persons who visit a library on one day.

Find the three quartiles of these data.

Stem	Leaves
1	2 3
2	0 2 2 4 5 5
3	1 3

Key ⟶ 24 = 2|4

Solution



The order of the first quartile $Q_1 = \frac{(n+1)}{4} = \frac{(10+1)}{4} = 2.75$

The order of the second quartile $Q_2 = \frac{(n+1)}{4} = \frac{(10+1)}{2} = 5.5$

The order of the third quartile $Q_3 = \frac{3(n+1)}{4} = \frac{33}{4} = 8.25$

The value of Q1 $= 13 + (20 - 13)(2.75 - 2) = 18.25$

The value of Q2 $= 22 + (24 - 22)(5.5 - 5) = 23$

The value of Q3 $= 25 + (31 - 25)(8.25 - 8) = 26.5$

Try to solve

1. In the previous example find the median by two different methods then compare the two results ?
2. Find the three quartiles (lower -middle – upper) of the opposite data

Stem	Leaves								
0	6	7	5						
1	9	0	1	3	8	2	2	1	2
2	5	1	0	1					

Key ⟵ 19 = 1|9

Finding quartiles from frequency tables:

You previously learned how to find the median by plotting the intersection of the ascending cumulative frequency curve and the descending cumulative curve

It represents the middle (the second). You will now learn how to find quartiles algebraically as follows:

The first step: We create the ascending cumulative table

The second step: We determine the orders of the quartiles

(First quartile order $= \frac{n}{4}$, second quartile order $= \frac{2n}{4}$, third quartile order $= \frac{3n}{4}$)

Third step : determine the interval (class) which the required quartile lies in (it called the quartile interval) from this interval we determine the start of the interval , its width Interval frequency, ascending cumulative frequency at previous of interval

Forth step : Use the flowing formula to calculate the required quartile

The value of the required quartile =

$$\text{Quartile interval start} + \frac{\text{Quartile order - cum.freq.at start of the interval}}{\text{frequency correspoding to the quartile interval}} \times \text{interval width}$$

Industry

Example

3 In a factory, if the following frequency table represents the number of working hours in a week for 50 workers, find the three quartiles.

Solution

Number of working hours	22–	27–	32–	37–	42–	47–	total
Number of workers (frequency)	9	3	10	8	12	8	50

form the opposite ascending cumulative table

Frequency ascending cumulative table	
upper boundaries	ascending cumulative frequency
Less than 22	Zero
Less than 27	9
Less than 32	12
Less than 37	22
Less than 42	30
Less than 47	42
Less than 52	50

(1) Assigning the first quartile Q_1:

order of $Q_1 = \frac{50}{4} = 12.5$

$\therefore$ Q_1 lies in the interval 12 , 22

from the ascending cumulative frequency column

$\therefore$ lower boundary of the first quartile interval = 32

First quartile interval width = 5

Frequency corresponding to the quartile interval = 10

Ascending accumulative frequency corresponding to the first quartile interval = 12

Substituting in the rule to determine the first quartile vale

$$Q_1 = 32 + \frac{12.5 - 12}{10} \times 5 = 32 + \frac{0.5 \times 5}{10} = 32 + 0.25$$

$\therefore$ $Q_1 = 32.25$

2) Assigning the second quartile (median) Q_2

order of $Q_2 = \frac{50}{2} = 25$

$\therefore$ order of Q_2 lies in the interval 22,30

$\therefore$lower boundary of the second quartile interval = 37

second quartile interval width = 5

Frequency corresponding to the quartile interval = 8

Ascending accumulative frequency corresponding to the second quartile interval = 22

Substituting in the rule to assign the second quartile vale

$$Q_2 = 37 + \frac{25 - 22}{8} \times 5 = 37 + \frac{15}{8}$$

$$= 37 + 1.875 = 38.875$$

3) Assigning the third quartile Q_3

Order of $Q_3 = 50 \times \frac{3}{4} = \frac{150}{4} = 37.5$

$\therefore$ Order of Q_3 lies in the interval 30 , 42

$\therefore$ lower boundary of the third quartile interval = 42

third quartile interval width = 5

Frequency corresponding to the quartile interval = 12

Ascending accumulative frequency corresponding to the third quartile interval = 30

$$Q_3 = 42 + \frac{37.5 - 30}{12} \times 5 = 42 + \frac{7.5 \times 5}{12} = 45.125$$

Second : Finding quartiles graphically:

You have previously learned how to find the median graphically from the ascending cumulative or descending cumulative frequency curve, and the same method can be applied To find quartiles, follow these steps:

First step: form the ascending cumulative frequency table

Second step: draw the ascending cumulative frequency curve

Third step: find the position of the quartiles ($\frac{n}{4}$, $\frac{n}{2}$, $\frac{3n}{4}$) and determine it on the vertical axis (cumulative frequency)

Forth step : at each position of the quartiles draw a horizontal line to intersect the curve at a point then The value of the quartile is the projection of this point on the horizontal axis

Example

4 If the frequency distribution of temperatures during 60 consecutive days In the spring season in the Arab Republic of Egypt, as follows:

temperature	16	18	20	22	24	26	28	Total
Number of days	4	7	10	18	9	7	5	60

Find the quartiles graphically

Solution

n = 60

The order of the first quartile Q_1

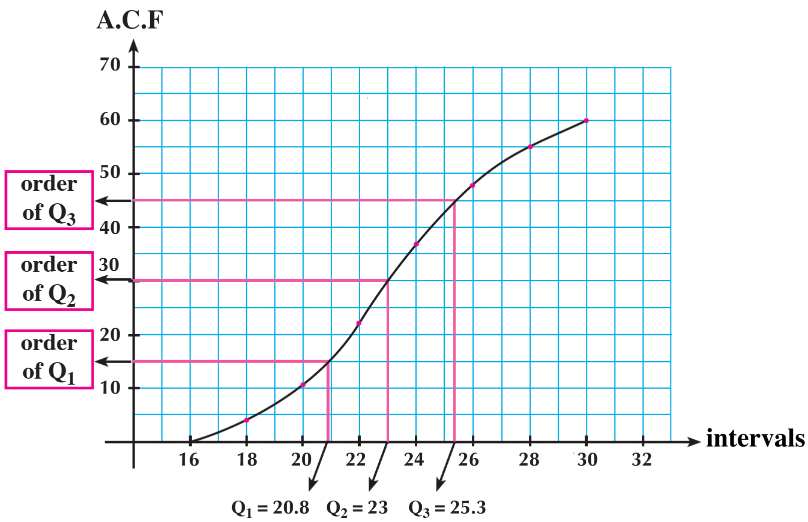
$= \frac{60}{4} = 15$

The order of the second quartile

$Q_2 = \frac{60}{2} = 30$

The order of the third quartile $= \frac{3 \times 60}{4} = \frac{180}{4} = 45$

Frequency ascending cumulative table	
upper boundaries	ascending cumulative frequency
Less than 16	Zero
Less than 18	4
Less than 20	11
Less than 22	21
Less than 24	39
Less than 26	48
Less than 28	55
Less than 30	60



From the graph we find:

$Q_1 = 20.8$

$Q_2 = 23$

$Q_3 = 25.3$

Try to solve

3 a In the previous example, verify algebraically the values of the quartiles that you obtained graphically.

b Medicine: If the following table represent the frequency table of the average of the hemoglobin level in blood in a sample of 50 persons find the quartiles algebraically and graphically

Hemoglobin level	13–	14–	15–	16–	17–	18–	Total
Frequency	3	5	15	16	10	1	50

4 If the following table represents the exam results of 200 students in mathematics, assuming that the lowest mark is 10 and the final mark is 50 find the three quartiles

Class	10–	15–	20–	25–	30–	35–	40–	45–	Total
Frequency	12	17	20	35	58	38	11	9	200

Learn

Finding quadrilaterals for data represented using the stem-and-leaf method

We have previously studied that the median (second quartile) in single data after ordering it

(1) If n is odd then : the median = the value of the term whose order $\frac{n+1}{2}$

(2) If n is even then : the median = $\frac{1}{2}$ (the value of term whose order $\frac{n}{2}$ + the value of term whose order $\frac{n}{2}+1$)

Generally

If the number of data is n and n +1 is a number divisible by 4, then the quartiles is one of the values in the given table and we get it directly from the following relationship:

order of the first quartile Q_1(lower quartile) = $\frac{n+1}{4}$

order of the first quartile Q_2(median) = $\frac{n+1}{2}$

order of the third quartile Q_3(upper quartile) = $\frac{3\,(n+1)}{4}$

Example

5 The following data represents the scores of 15 students on a test The monthly period is represented by the stem and leaves method, given that final mark of 30 find the three quartiles.

Stem	Leaves						
0	1	1	1	2	2	3	3
1	0	1	1	1	4		
2	1	2	2				

Key ← 10 = 1|0

Solution

∵ n = 15 ∴ n + 1 = 16 A number divisible by 4

∵ The data in the table is arranged in ascending order

So we find the order of the quartiles and assign them directly from the table data

Stem	Leaves						
0	1	1	1	2	2	3	3
1	0	1	1	1	4		
2	1	2	2				

lower quartile Q_1

middle quartile Q_2

upper quartile Q_3

1) The first quartile Q_1 its order = $\frac{n+1}{4} = \frac{16}{4} = 4$

The value of the first quartile **(the fourth element from the first row)**

∴ $Q_1 = 2$

2) The second quartile Q_2 its order = $\frac{n+1}{2} = \frac{16}{2} = 8$

The value of the first quartile **(the first element from the second row)**

∴ $Q_2 = 10$

3) The third quartile Q_3 its order = $\frac{3(n+1)}{4} = \frac{3 \times 16}{4} = \frac{48}{4} = 12$

The value of the third quartile **(the fifth element from the second row)**

∴ $Q_3 = 14$

Box blot

Learn

Quartiles are called ordinal position scales and are used to identify ... the spread of the data distribution in quartiles box plot uses these values to describe the data by draw a rectangle with the beginning of the lower quartile and the end of the upper quartile after representing the data next on the same line is arranged

(minimum value - lower quartile - median - upper quartile - maximum value)

The resulting shape is called a two-sided box. (box whisker plot)

Add to your knoledege

There are other position measurements like deciles and perciltiles

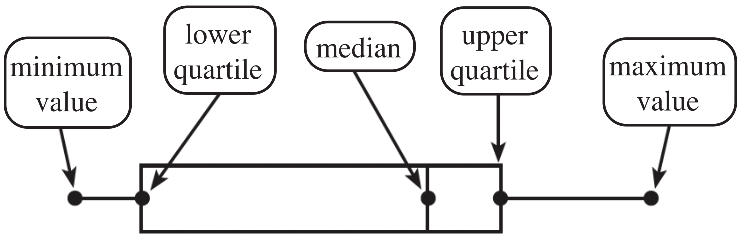
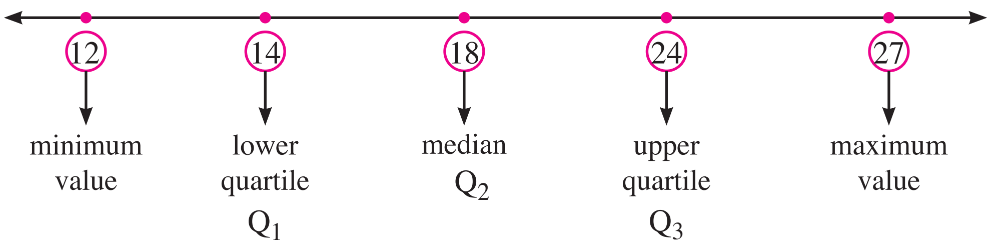


figure (1)

Example

6 Represent following data: 14, 24, 16, 12, 18, 20, 24, 16, 26, 13 and 27, using box plot.

Solution



The box plot corresponding to the previous data is as follows:

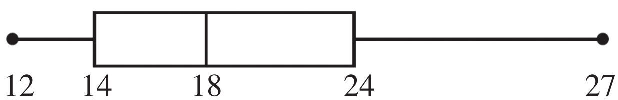


figure (2)

(1) We note that 50% of the data is between the lower quartile and the upper quartile

(2) The box representation can be drawn in a vertical way

Try to solve

5 Draw the box blot of the following data :

a 27, 24, 20, 18, 17, 15, 13

b

Stem	Leaves
4	0 3 3 6 7
5	1 8 9
6	2 3 4

Key ⟵ 51 = 5|1

Example

7 The following marks represent the marks of 15 students in the statistics exam

37	40	45	23	18
44	53	38	49	55
15	58	35	32	42

Draw the box plot these marks

Solution

The ascending order of the marks

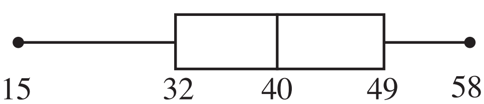
15, 18, 23, 32, 35, 37, 38, 40, 42, 44, 45, 49, 53, 55, 58

The minimum value = 15, the maximum value = 58

First quartile (lower) = 32

Second quartile (median) = 40

Third quartile (upper) = 49



Exercises 2 - 2

1 find the lower , middle , upper quartiles of the following set of values :

a 70, 81, 82, 58, 88, 90, 93

b 7, 5, 2, 7, 6, 12, 10, 4, 8, 9

c

Stem	Leaves
4	0 3 3 6 7
5	1 8 9
6	2

Key ⟵ 58 = 8|5

2 Energy : In a study of the consumption of a group of gasoline-powered cars, the results were as follows :

Number of kilometers per each liter	20-	25-	30-	35-	40-	45-
Number of cars	7	11	12	7	6	8

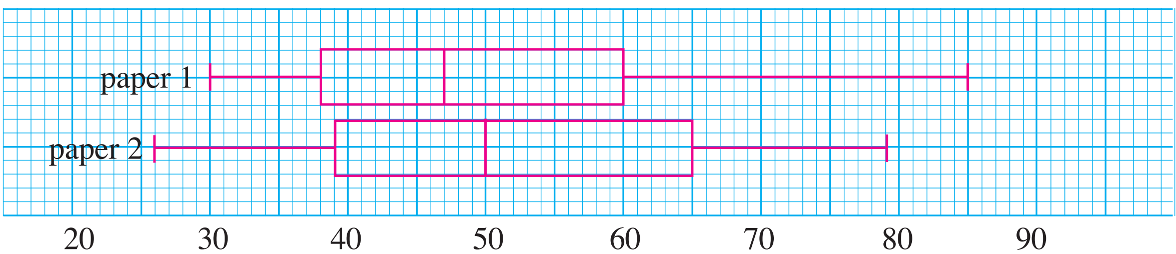
form the ascending cumulative frequency table and then find the quartiles in two different ways.

3 The following representation represents data on the grades of students in two different classes in science. Draw the box representation for each of the two classes, then calculate the quartiles

First class	Stem	Second class
4	3	2 3
3 3 2 2 1 1 0	4	0 2 3
1 0 0	5	0 0 2 3 3 4 5
4	6	1 3

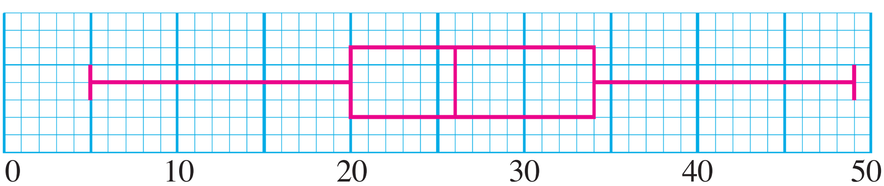
50 = 5|0 → Key ← 42 = 4|2

4 The following figure shows the distribution of marks for two exams for a group of students: determine quartiles for each of them and write two sentences explaining the comparison between the marks

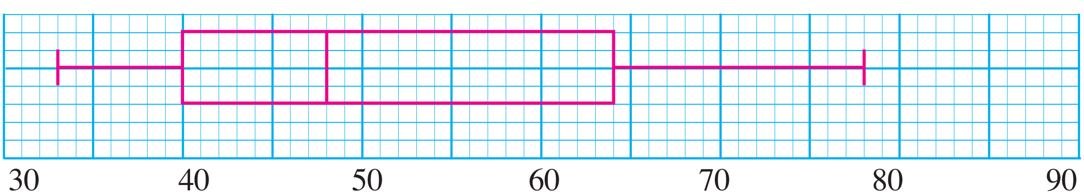


5 Describe each following Box polt, indicating the lowest value - largest value - lower quartile - median - upper quartile - for each

a



b



Unit Two

2 - 3

Semi-interquartile range

You will learn

- semi- interquartile range

Key terms

- Range
- The first quartile
- The third quartile
- Semi-inter quartile range

Think and discuss

The following data shows the marks of 7 groups in one of the subject competitions Mathematics under the supervision of the class teacher, knowing that the maximum mark of mathemtics = 50 marks

1) Find the range of these marks

2) Find the three quartiles of these marks

3) Draw a box plot of the data

What does the length of the box plot and how much of the original data does it contain?

Group	Marks
First	27
Second	23
Third	45
Fourth	30
Fifth	38
Sixth	48
Seventh	41

Remember

Some of the dispersion measurments you studied before.
1) The range.
2) The standard diviation.
3) The variance

Learn

Since the box does in the box plot not contain outliers of the data and represents 50% of the values, the half of the interquartile range is defined as :

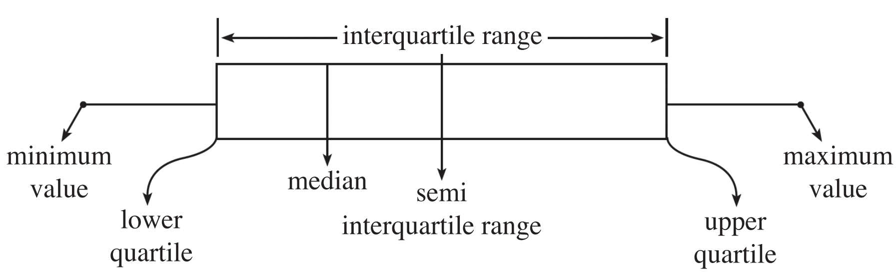
$$\text{Semi -interquartile range} = \frac{\text{upper quartile - lower quartile}}{2}$$

$$Q = \frac{Q_3 - Q_1}{2}$$

where : Q " semi interquartile range "

Q_1 lower quartile

Q_3 upper quartile



Advantages and disadvantages of semi- interquartile range:

Advantages : **It is better to use as a dispersion measurement in case of exists of outliers values, and it is easy and simple to calculate.**

Disadvantages : **does not take all values in consider**

Tools

- Scientific Calculater.
- Computer graphic programs.

Agriculture

Example

1. The following frequency table shows the distribution of 60 farms

Area	15–	20–	25–	30–	35–	40 – 45
Number of farms	3	9	15	18	12	3

Calculate

a. Area planted with corn in hectares

b. Semi-interquartile range of area planted with corn

Add to knowledge

Hectar is a measuring unit for area and equals 10000 square meter

Solution

Follow the following steps to calculated semi-interquartile range:

a. Find the order of the data inside the table

(1) order of the first quartile $= \frac{n}{4} = \frac{60}{4} = 15$

$\therefore$ lower boundary of the first quartile class = 25

Frequency of Q_1 , interval = 15

First quartile interval width = 5

Frequency previous to Q_1 interval = 12

$Q_1 = 25 + \frac{15 - 12}{15} \times 5 = 26$

(2) order of the third quartile $= \frac{3n}{4} = \frac{3 \times 60}{4} = 45$

$\therefore Q_3 = 35$

upper limit	ascending cumulative frequency
less than 15	0
less than 20	3
less than 25	12
less than 30	27
less than 35	45
less than 40	57
less than 45	60

b. We find the semi- interquartile range

$Q = \frac{Q_3 - Q_1}{2} = \frac{35 - 26}{2} = 4.5$

semi- interquartile range for the area = 4.5 hectare

= 45 thousands square meter

Try to solve

1. The following data shows a frequency table for the ages of 20 teachers

Ages	33–	38–	43–	48–	53–	Total
Number of teachers	3	7	4	2	4	20

Calculate the semi interquartile range for these ages

Example

2 The following data shows the marks of a group of students on a test. Find the semi interquartile range for these marks

Stem	Leaves
5	6 9
6	4 5 9
7	0 1 3 6 7 8
8	0 2 2 5

Solution

n = 15 (where n represent the number of data)

the order of the first quartile $= \frac{n+1}{4} = \frac{15+1}{4} = 4$

$Q_1 = 65$

the order of the third quartile $= \frac{3(n+1)}{4} = \frac{48}{4} = 12$

$Q_3 = 80$

Semi- interquartile range is $Q = \frac{Q_3 - Q_1}{2} = \frac{80 - 56}{2} = \frac{15}{2} = 7.5$

Try to solve

2 The following is the amount of daily milk production in liters for a sample of cows selected from a farm:

30, 27, 18, 20, 29, 34, 25, 32, 29, 21, 23, 28, 25, 19

Represent the data using the stem and leaf method and calculate the semi-interquartile range

Exercises 2 - 3

1 Find the range and the semi- interquartile range of the following data

a 46 , 64 , 52 , 61 , 56 , 55 , 43 , 62 , 60 , 51 , 54 , 51

b 4.3 , 5.9 , 4.1 , 1.5 , 0.8 , 1.1 , 1.2 , 2.4 , 1.5

Stem	Leaves
0	3
1	0 0 2
2	3 8 9
3	0 2 5 5 7 9
4	1

Key ← 28 = 2|8

lengths

2 The following table shows the heights of 240 female students at a university

Length in cm	140–	145–	150–	155–	160–	165–	170–	175–	180–	Total
Number of students	3	10	21	54	72	48	25	5	2	240

Find semi- interquartile range and represent the data by a box plot

Health

3 The following frequency table shows the weights of a number of births during 14 days in a hospital

Birth weights in kilograms	2–	2.5–	3–	3.5–	4–	4.5–	Total
Number of births	3	7	10	8	4	2	34

Find the inter-quartile range and the semi- interquartile range

4 If the following data represents the marks of 14 students on two mathematics tests during two consecutive month

First exam	17	18	5	4	11	14	18	18	6	15	14	15	11	10
Second exam	5	4	8	10	8	18	12	12	13	13	18	18	17	16

Required

a Find the quartiles of the two exams and there semi-interquartile rang

b Compare between the marks of the students in the two exams using the median and the semi-interquartile range. Determine which of the two tests students performed better on and why?

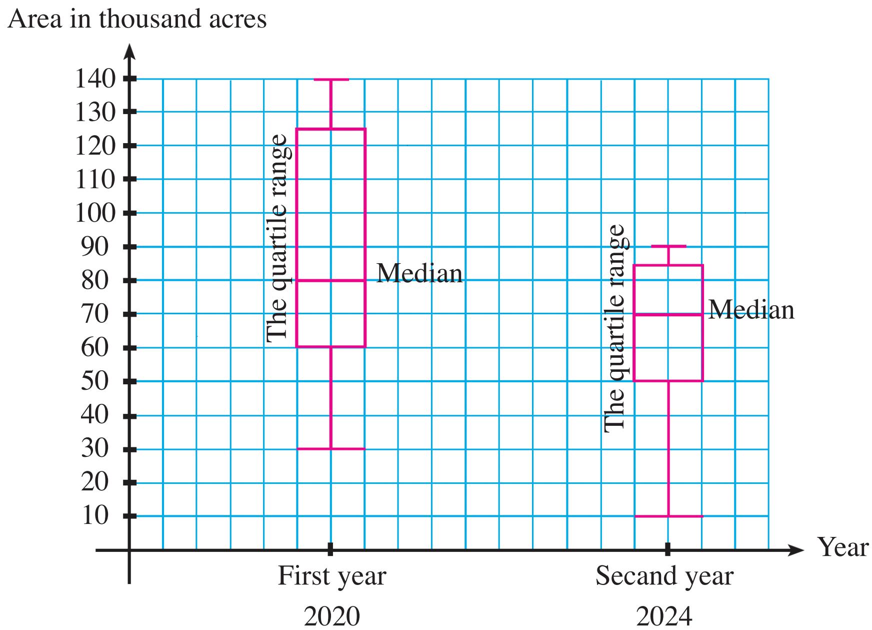
Agriculture

5 The following graph represents the cultivated area in thousand acres in 25 villages during two different years.

Required

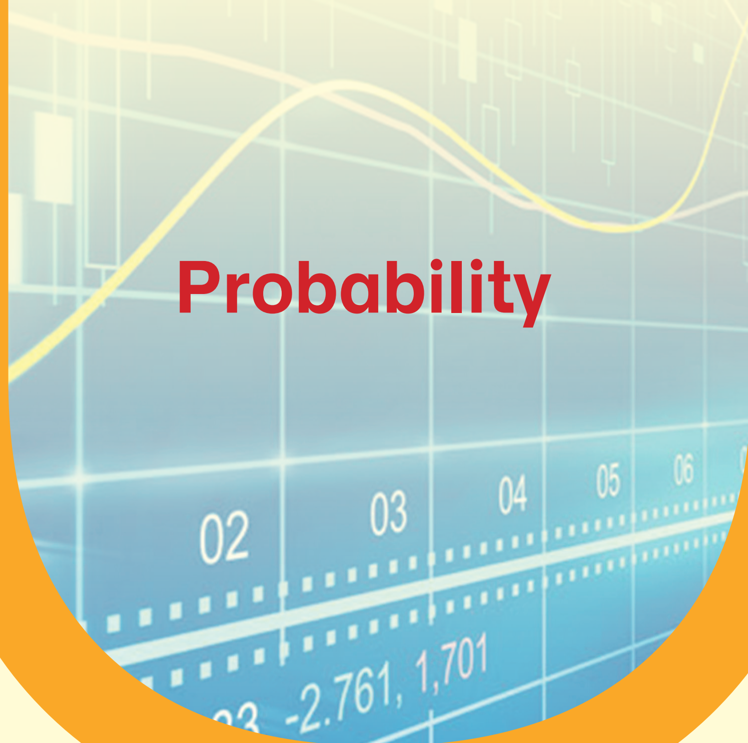
a Find the upper quartile, the lowest quartile, and the median. And the semi- interquartile range of the two years

b What do you conclude from this data?



Unit 3

Probability



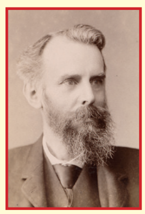
introduction

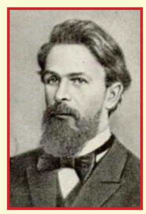
We have previously known that statistics is a branch of mathematics concerning with collecting, ordering and interpreting data in order to get the ability to take the proper decisions for a certain phenomenon. The probabilities are considered the mathematical background for the statistical methods. Long ago, the researchers used the probabilities for social, economic and healthy reasons.

The probability has been founded as it is today by a great number of scientists such as the French scientist (Pierre Simon Laplace- 1749- 1827), the English scientists (De Morgan 1806- 1871), (John Vin 1834-1923) and the Russian scientist (Andrei Markov 1856-1922) and others.


Pierre Simon Laplace


De Morgan


John Vin


Andrei Markov

It is worth mentioning that the applications of statistics and probability are much in the educational, social and economic fields. In this unit, we are going to learn the conditional probability between two events, its theory and its applications in the different daily life situations. Besides, we will learn the independent and dependent events.

Unit objectives

:**At the end of the unit and carrying out the involved activities, the students should be able to**

- Identify operation on events.
- Identify the concept of probability.
- Use axioms of probability.
- Solve problems on axioms of probability.
- Solve life situations using probability.
- Identify the mutually exclusive events and events are not mutually exclusive
- Identify the conditional probability
- Deduce theories on the conditional probability
- Identify the independent and dependent events
- Apply the conditional probability in the different daily life situations

Key Terms

- *Mutually Exculusive events*
- *Events are not mutually exclusive*
- *Conditional probability*
- *Independent Events*
- *Dependent Events*

Lessons of the unit

Lesson (3 - 1): Calculating probability

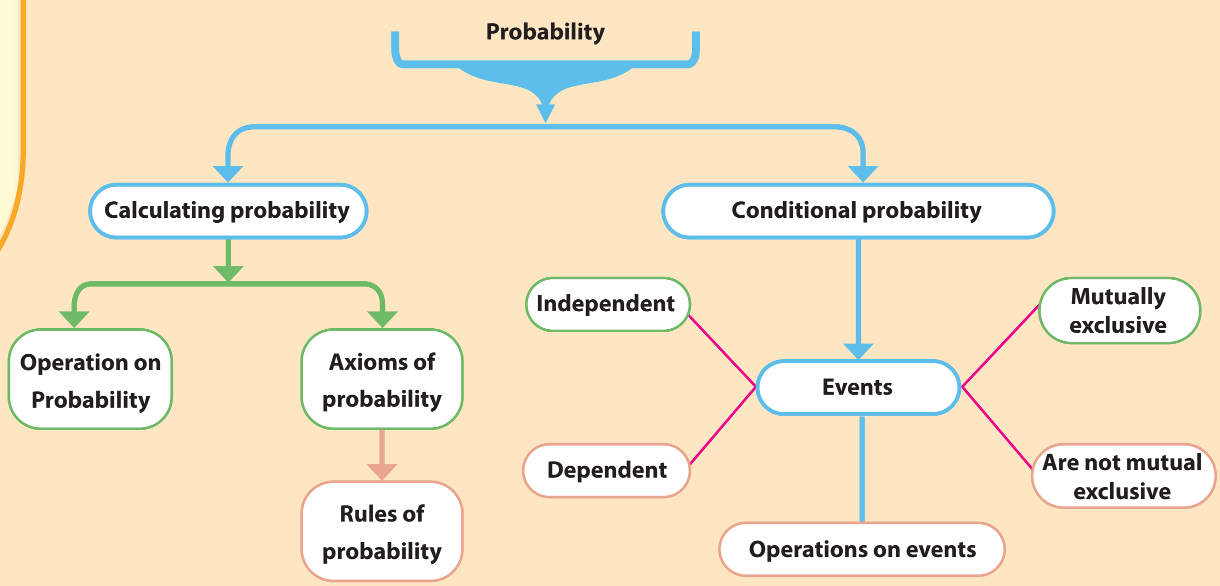
Lesson (3 - 2): Conditional probability.

Lesson (3 - 3): Independent events.

Materials

Scientific Calculator

Unit planning organizer



Unit Three

3 - 1

Calculating Probability

Real Functions

You will learn

- The concept of the random experiment and the sample space.
- The concept of the event –simple event-sure event-impossible event.
- Operations on events (union - intersection - difference - complement)
- Mutually exclusive events .
- De' Morgan's laws.
- Concepts of probability
- Calculating probability
- Probability axioms and its life applications.

Key terms

- random experiment
- sample space
- event
- simple event
- certain event
- impossible event
- mutually exclusive events
- probability
- probability axioms

Introduction:

In our previous study, we learned the probability in a simple way. So we will complete the study of these concepts and the operations on events while calculating the probability of the occurrence of an event through examples and different life applications.

Basic terms and concepts

Learn

The random experiment: It's an experiment we know all of its outcomes before we do it but we cannot predict which of these outcomes will occur when we do the experiment.

Example

1 Show which of the following experiments represent a random experiment?

a Rolling a regular die and observe the number written in its upper face.

b Draw a color ball from a bag including a set of color balls (without determine their color) and recognize the color of the drawn ball.

c Throw a coin and observe what appears in its upper face.

d Draw a ball from a bag including four balls identical in volume and weight. The first is white, the second is black, the third is red and the fourth is green. Recognize the color of the drawn ball.

Solution

The experiments (a),(c),(d) are random experiments because we know all of their outcomes before we do each of them but we cannot predict which of these outcomes will occur when we do these experiments.

The experiment (b) is not a random experiment because we cannot determine the outcomes of the experiment before we do it.

Matrials
- Scientific calculator.

Try to solve

1. Show which of the following experiments represent a random experiment?
   - a) Throw a coin twice and observe the sequence of heads and tails.
   - b) Draw a numbered card from a bag contains a set of numbered cards (we do not know their numbers) and recognize the written number on the drawn card.
   - c) Draw a card from a bag contains a set of 20 identical cards numbered from 1 to 20 and observe the written number on the drawn card.

Learn

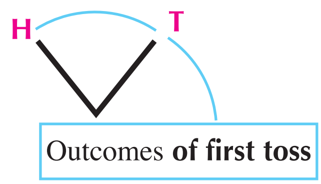
definition

Sample space (outcomes space)

- The sample space of a random experiment is the set of all possible outcomes for this experiment, it denoted by (S)

Remarks:
- The number of elements in the sample space of a random experiment is denoted by n(S).
- The sample space will be finite if the number of its elements is finite and it will be infinite if the number of its elements is infinite. We will deal with the finite sample space.

The sample space for some famous random experiments:

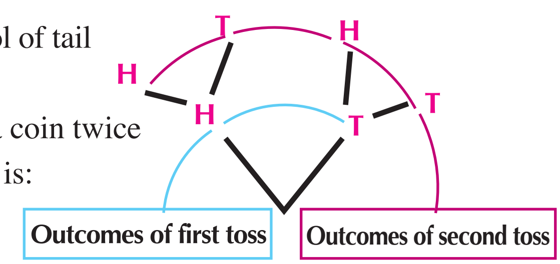


First: Tossing a coin

1- The sample space of the experiment of tossing a coin once and observing the shown face is : S = { H, T}

Where: H is the symbol of head, T is the symbol of tail

Where: n(S) = 2



2- The sample space of the experiment of tossing a coin twice and observing the sequence of heads, and tails is:

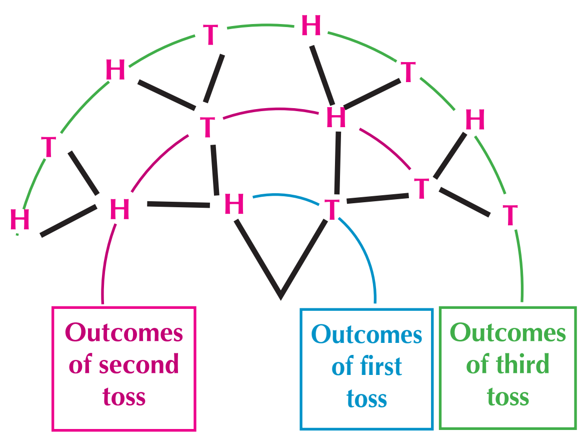
S = { (H, H), (H, T), (T, H), (T, T)}

Where: $n(S) = 2 \times 2 = 4 = 2^2$

3- The sample space of the experiment of tossing a coin three respective times and observing the sequence of heads and tails (could be as shown in the opposite tree diagram) is:

S = { (H , H , H) , (T , T , T) ,
(H , H , T) , (T , T , H) ,
(H , T , H) , (T , H , T) ,
(H , T , H) , (T , H , H)}

Where: $n(S) = 2 \times 2 \times 2 = 8 = 2^3$



Note that

1- On tossing a piece of coin m times, then $n(S) = 2^m$ **2-** (H, T) $\neq$(T, H) why?

3- The sample space of the experiment of tossing two different coins (different in shape and volume) simultaneously (at the same time) is the same sample space of tossing a coin two successive times, and each result of the experiment results is written in the form of an ordered pair (the face of the 1st coin, the face of the 2nd coin).

Second: Rolling a die

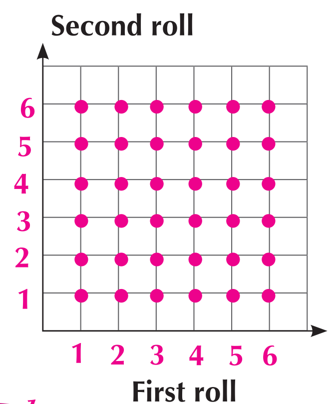
1- The sample space of the experiment of rolling a die once and observing the number shown on the upper face is:
$S = \{ 1, 2, 3, 4, 5, 6 \}$ where: $n(S) = 6$

2- The sample space of the experiment of rolling a die two successive times and observing the number shown each time on the upper face is the group of ordered pairs. The first co-ordinate is the result of the first roll and the second co- ordinate is the result of the second roll. i.e.: $S = \{ (x, y) : x\in \{ 1, 2, 3, 4, 5, 6\}, y \in \{ 1, 2, 3, 4, 5, 6\}\}$ and the following figures illustrate this.

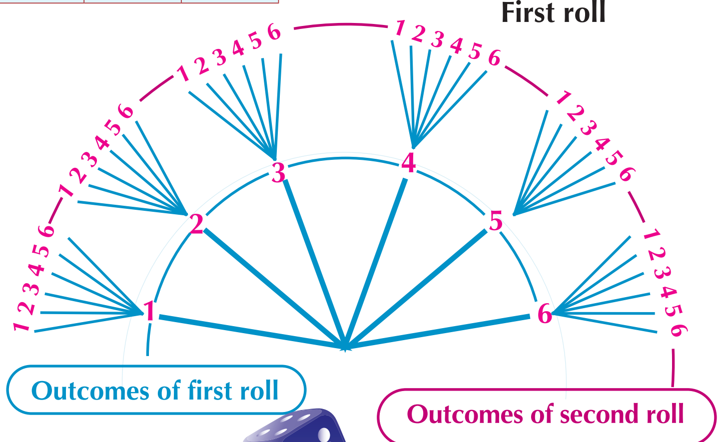
a Tabulated representation:

Toss First / Second	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

b Geometrical representation:



c Tree diagram



Note that:

1- $n(S) = 6 \times 6 = 36 = 6^2$

2- $S = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$

3- The sample space of the experiment of rolling two different dies at the same time (simultaneously) is the same sample space of rolling a die two successive times.

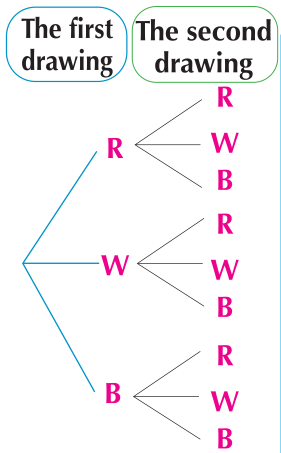
Example

2 A bag contains three identical balls: the first is red, the second is white and the third is yellow. Write down the sample space if you draw two balls, one after the other, while reputing the drawn ball before drawing the other one (with replacement) and observing the succession of colors.

Solution

Denoting the red ball with the letter (R), the white ball with the letter (W) and the yellow ball with the letter (B):

First : When the ball is returned back to the bag before the second ball is drawn, each ball has the chance of appearance in the second drawing and it is then possible to draw the same ball twice. The opposite fig. shows the tree diagram of the sample space where n (S) = 3^2 = 9



S
(R , R)
(R , W)
(R , B)
(W , R)
(W , W)
(W , B)
(B , R)
(B , W)
(B , B)

S = {(R, R), (R, W), (R, B), (W, R), (W, W), (W, B), (B, R), (B, W), (B, B)}

Add to your information

If a ball is drawn without replacement, that means: not to reputing the drawn ball before drawing the other one ,so there is no possible chance for that ball to appear in the second drawing.

Try to solve

2 A box contains three identical balls numbered from 1 to 3. Two balls are drawn one after another with replacement and observing the number on the ball. Write down the sample space and find the number of its elements.

Learn

definition

The event

- An event is a subset from the sample space.

The simple event

- Is a subset from the sample space that contains only one element.

definition

The certain event

It is an event whose elements are the elements of the sample space S.
And it is an event that must occurs in each time we do the experiment.

The impossible event

It is an event that has no elements and is denoted by the symbol ϕ
And it is an event that must not occur each time we do the experiment.

Example

(3) In the experiment of throwing a coin several times, the experiment will stop if a head or three tails appear.

Write down the sample space, and then determine the following events:

A The appearance of head at most

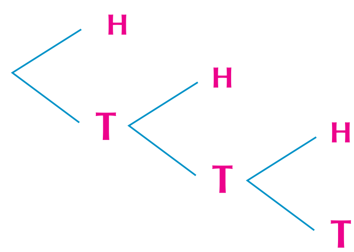
B The appearance of head at least

C The appearance of two tails at least

D The appearance of two heads at least

Solution

From the drawing, we find:



S $= \{H, (T, H), (T, T, H), (T, T, T)\}$

A $= \{H, (T, H), (T, T, H), (T, T, T)\} = S$

B $= \{H, (T, H), (T, T, H)\}$

C $= \{(T, T, H), (T, T, T)\}$

D $= \{ \quad \} = \phi$ the impossible event

Try to solve

(3) In the experiment of throwing a coin several times, the experiment will stop if two heads or two tails appear.

Write down the sample space, and then determine the following events:

A The appearance of head at least

B The appearance of two tails at most

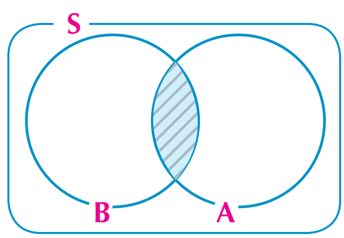
C The appearance of tail at most

Operation of the events

Learn

First: Intersection

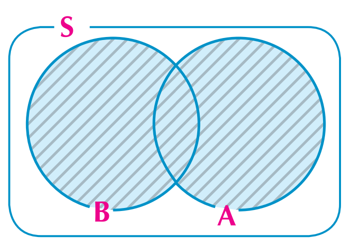
The intersection of the two events A and B is the event $A \cap B$ which contains all elements of the sample space that belong to both A and B and **means the occurrence of A and B (the occurrence of the two events at the same time).**



$A \cap B$

Second: Union

The union of the two events A and B is the event $A \cup B$ which contains all elements of the sample space that belong to A or B or both of them and **means the occurrence of A or B (the occurrence of one of them at least).**

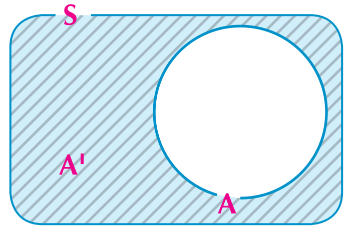


$A \cup B$

Third: Complement

The event A` is called the complementary event of the event A, where A` contains all elements of the sample space that does not belong to the event A, and **means non - occurrence of the event A** .

Note : $A \cup A` = S$ , $A \cap A` = \phi$

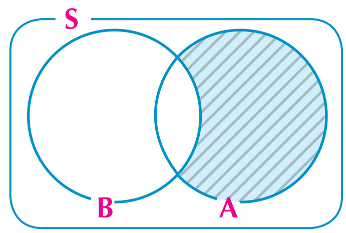


Fourth: Difference

The event A - B contains all elements of the sample space that belong to A and does not belong to B it also contains the same elements of the event $A \cap B`$

and it means the occurrence **of A and non-occurrence of B (occurrence of A only).**

$A - B = A \cap B` = A - (A \cap B)$

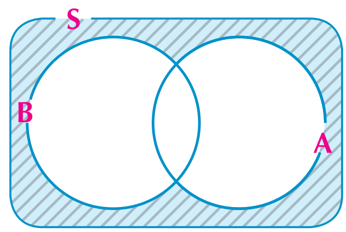


Fifth: De morgan`s laws

A and B are the two events from the sample space S, then:

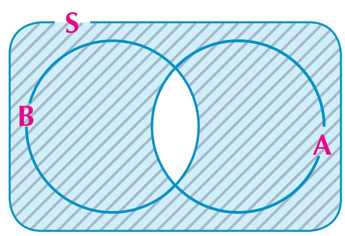
(first) $A` \cap B` = (A \cup B)`$

which means the event of (non-occurrence of any of the two events) or the event of (the non-occurrence of A and the non-occurrence of B)



(second) $A` \cup B` = (A \cap B)`$

which means the event of (non-occurrence of any of the two events all together) or the event of (the occurrence of one of the events at most)



Learn

Mutually exclusive events

Two events A and B are mutually exclusive events if the occurrence of one of them prevent the occurrence of the other

For example: **1-** If A" event of success in an exam" , B "event of failure in the same exam", then The occurrence of one of them prevent the occurrence of the other.

2- In experiment of rolling a die once, observing the number on the upper face then $S = \{1, 2, 3, 4, 5, 6\}$

If A: appearance of an odd number $A = \{1, 3, 5\}$

B: appearance of an even number $B = \{2, 4, 6\}$

then $A \cap B = \phi$ so, the occurrence of one of them prevent the occurrence of the other.

Definition

- Two events A and B are said to be mutually exclusive if $A \cap B = \phi$
- Several events are said to be mutually exclusive if and only if each two by two are mutually exclusive events.

Notice that :

1- If $A \cap B = \phi$, then A and B are mutually exclusive events.

If A, B and C are three events in a sample space S and: $A \cap B = \phi$, $B \cap C = \phi$, $C \cap A = \phi$ then: A , B, C are said to be mutually exclusive events and vice versa.

2- Simple events (primary) in any random experiment are mutually exclusive.

3- Any event A and its complement A\` are mutually exclusive events.

Example

4 Two distinct dice are tossed and observing the numbers on the upper faces.

First: represent the sample space geometrically, and then write down the following two events.

A "appearance of the same numbers on the two faces"

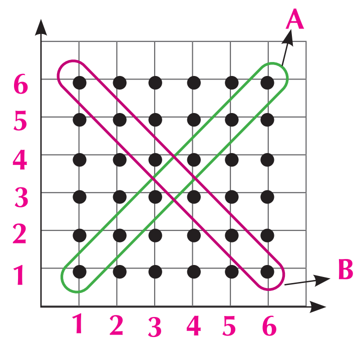
B "appearance of two numbers their sum equals 7"

Second : Are A , B mutually exclusive? Justify your answer

Solution

First : The elements of the sample space are ordered pairs, their number = $6^2 = 36$

the opposite figure is the geometrical representation of the sample space where every element is represented by a point



A = {(1, 1), (2, 2), (3, 3), (4, 4) , (5, 5) , (6, 6) }

B = { (6, 1) , (5, 2), (4, 3) , (3, 4) , (2, 5) , (1, 6) }

Second: $\because A \cap B = \phi$ $\therefore$ A, B are mutually exclusive events

Try to solve

4 In the previous example: write each of the following events:

C "appearance of two numbers their sum equal 5"

D "appearance of two numbers one of them is twice the other"

Are C , D mutually exclusive? Justify your answer.

Propability

Learn

Calculation of probability :

If A is an event in the sample space S for a random experiment all its outcomes (primary events) are equal possibility i.e. $A \subset S$, number of elements of event A equals n (A), number of elements of S equals n(S), if we denoted the probability of occurrence of A by P(A):

$$P(A) = \frac{n(A)}{n(S)} = \frac{\text{The number of the outcomes leads to the occurence of the event A}}{\text{The number of all possible outcomes for the random experiment}}$$

Example

5 If a ball drawn from a box contains 10 identical balls, 5 of them are white, 2 are red and the rest are green .Find the probability of the following events:

A the event “the drawn ball is red”

B the event “the drawn ball is red or green”

C the event “the drawn ball is not green”

Solution

The probability that the drawn ball is red $= P(A) = \frac{\text{The number of red balls}}{\text{The number of all balls}} = \frac{2}{10} = 0.2$

The probability that the drawn ball is red or green $= \frac{\text{The number of red balls + green balls}}{\text{The number of all balls}}$

$$= \frac{2+3}{10} = \frac{5}{10} = 0.5$$

The probability that the drawn ball is not green = P(C)

= The probability that the drawn ball is red or white $= \frac{2+5}{10} = 0.7$

Think: Can you obtain P(C) with another method? Explain that.

Try to solve

5 In the previous example: find the probability of the following events:

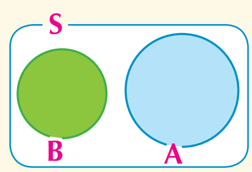
D: the event “the drawn ball is red or white”

E: the event “the drawn ball is red or white or green”

Learn

Axioms of probability

1- For every event $A \subset S$ there exists a real number called probability of event A, and denoted by P(A) Where : $0 \leqslant P(A) \leqslant 1$



2- $P(S) = 1$

3- If $A \subset S$, $B \subset S$

and A , B are mutually exclusive events, then : $P(A \cup B) = P(A) + P(B)$

From the previous axioms we notice that:

The first axiom means that the probability of the occurrence of any event is a real number belongs to the interval [0, 1]

The second axiom means that the probability of the sure event = 1

The third axiom is called the sum of probabilities of mutually exclusive events rule which is circulating for a finite number of mutually exclusive events.

$P(A_1 \cup A_2 \cup A_3 \cup ... \cup A_n) = P(A_1) + P(A_2) + P(A_3) + ... + P(A_n)$

where A_1, A_2, A_3 , ... $,A_n$ each two of them are mutually exclusive events

Important Corollaries

(1) $P(\phi) = 0$

(2) $P(A`) = 1 - P(A)$

(3) $P(A - B) = P(A) - P(A \cap B)$

(4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Add to your information

If $A \subset B$
then $P(A) \leqslant P(B)$

Example

6 If A , B are two events in a sample space S of a random experiment, such that:

$P(A) = \frac{3}{8}$, $P(B) = \frac{3}{4}$, $P(A \cap B) = \frac{1}{4}$, Find:

a $P(A \cup B)$ **b** $P(A`)$ **c** $P(A - B)$ **d** $P(A` \cap B`)$

Solution

a $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{8} + \frac{3}{4} - \frac{1}{4} = \frac{7}{8}$

b $P(A`) = 1 - P(A) = 1 - \frac{3}{8} = \frac{5}{8}$

c $P(A - B) = P(A) - P(A \cap B) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$

d $P(A` \cap B`) = P(A \cup B)` = 1 - P(A \cup B) = 1 - \frac{7}{8} = \frac{1}{8}$

Try to solve

6 In the previous example, find the following probabilities:

a $P(B`)$ **b** $P(B - A)$ **c** $P(A` \cup B`)$

Example

7 If A , B are two events in a sample space S of a random experiment, such that $P(A) = \frac{5}{8}$,

$P(B) = \frac{1}{2}$, $P(A - B) = \frac{3}{8}$ Find :

a $P(A \cap B)$ **b** $P(A \cup B)$ **c** $P(A` \cap B`)$ **d** $P(A` \cup B)$

Solution

a $P(A \cap B) = P(A) - P(A - B) = \frac{5}{8} - \frac{3}{8} = \frac{2}{8} = \frac{1}{4}$

b $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{5}{8} + \frac{1}{2} - \frac{1}{4} = \frac{7}{8}$

c $P(A` \cap B`) = P(A \cup B)` = 1 - P(A \cup B) = 1 - \frac{7}{8} = \frac{1}{8}$

d $P(A` \cup B) = P(A \cap B`)` = 1 - P(A \cap B`) = 1 - P(A - B)$
$= 1 - \frac{3}{8} = \frac{5}{8}$

Think : Can you obtain $P(A` \cup B)$ with another method? Explain that.

Try to solve

7 In the previous question, find the following probabilities:

a $P(A`)$ b $P(A` \cup B`)$ c $P(B \cap A`)$

Example

8 If A and B are two events in a sample space S of a random experiment, such that $P(A`) = \frac{1}{3}P(A)$, $P(B) = \frac{1}{2}$, $P(A` \cup B`) = \frac{5}{8}$ Find :

a The probability of occurrence of one of the two events at least.

b The probability of occurrence of one of the two events at most.

c The probability of occurrence of the event B only

d The probability of occurrence of only one of the two events.

Solution

$\because P(A` \cup B`) = \frac{5}{8}$ $\therefore P(A \cap B)` = 1 - P(A \cap B) = \frac{5}{8}$ $\therefore P(A \cap B) = \frac{3}{8}$

$\because P(A`) = \frac{1}{3}P(A)$ $\therefore 1 - P(A) = \frac{1}{3}P(A)$ $\therefore \frac{4}{3}P(A) = 1$ $\therefore P(A) = \frac{3}{4}$

a The probability of occurrence of one of the two events at least $= P(A \cup B)$
$= P(A) + P(B) - P(A \cap B) = \frac{3}{4} + \frac{1}{2} - \frac{3}{8} = \frac{7}{8}$

b The probability of occurrence of one of the two events at most $= P(A \cap B)`$
$= P(A` \cup B`) = \frac{5}{8}$

c The probability of occurrence of the event B only $= P(B - A)$
$= P(B) - P(A \cap B) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$

d The probability of occurrence of only one of the two events $= P(A \cup B) - P(A \cap B)$
$= \frac{7}{8} - \frac{3}{8} = \frac{1}{2}$

Think: Can you find the probability of occurrence of only one of the two events with another method? Explain this.

Try to solve

8 If A and B are two events in a sample space S of a random experiment, such that $P(A) = 0.8$, $P(B) = 0.6$, $P(A \cup B)` = 0.1$ Find The probability of the following events:

a The occurrence of one of the two events at least.

b The occurrence of the event A only

c The occurrence of only one of the two events

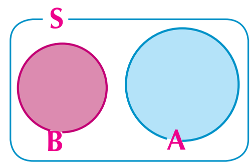
d The occurrence of one of the two events at most.

Example

9 A and B are two events in a sample space S of a random experiment, where:

$P(B) = 3P(A)$, $P(A \cup B) = 0.72$, find: $P(A)$, $P(B)$

First: if A , B are mutually exclusive events. **Second:** if $A \subset B$



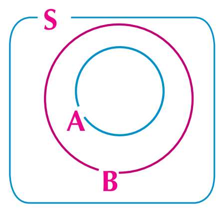
Solution

Let $P(A) = x$ $\therefore P(B) = 3x$

First: $\because$ A , B are mutually exclusive events.

$\therefore P(A \cup B) = P(A) + P(B)$ $: 0.72 = 3x + x$

$\therefore x = 0.18, P(A) = 0.18$, $P(B) = 0.54$



Second: $\because A \subset B$ $\therefore A \cup B = B$

$P(A \cup B) = P(B) = 3x = 0.72$

$\therefore P(A) = 0.24$, $P(B) = 0.72$

Try to solve

9 If A and B are two events in a sample space of a random experiment , Where:

$P(B) = \frac{1}{5}$, $P(A \cup B) = \frac{1}{3}$ Find $P(A)$

a If A , B are mutually exclusive events. **b** if $B \subset A$

Critical thinking:

Explain how to calculate $P(A)$ if $A \subset S$, S is a sample space of a random experiment, if $\frac{P(A')}{P(A)} = \frac{3}{7}$

Try to solve

10 IF S is a sample space of a random experiment where S = {A , B , C}, and $\frac{P(A')}{P(A)} = \frac{2}{3}$, $\frac{P(B')}{P(B)} = \frac{5}{2}$ Find P(C)

Example

10 Join with the school inviroment: If the probability of success of a student in the physics exam equals 0.85 and the probability of success in mathematics exam equals 0.9 and the probability of success in both subjects equals 0.8. Find the probability of:

a The success of the student in at least one of the two subjects.

b The success of the student in mathematics only.

c The non-success of the student in both subjects together.

Solution

Let A denoted the event of success of the student in physics , B: denoted the event of success of the student in mathematics.

then: $P(A) = 0.85$, $P(B) = 0.9$, $P(A \cap B) = 0.8$

a Probability of success of the student in at least one of the two subject = P (A ∪ B)
$\therefore$ P (A ∪ B) = P (A) + P (B) - P (A ∩ B) = 0.85 + 0.9 - 0.8 = 0.95

b Probability of success of the student in mathematics only means probability of success in mathematics and not success in physics P (B - A)
$\therefore$ P (B - A) = P (B) - P (B ∩ A) = 0.9 - 0.8 = 0.1

c The event of the student will not success in both subjects = (A ∩ B)' which is the complement of (A ∩ B)
$\therefore$ P(A ∩ B)` =1 - P (A ∩ B) = 1 - 0.8 = 0.2

Life applications:

Try to solve

11 To get a job in a company, a person has to pass two exams: theoretical and practical. If the probability to succeed in the theoretical exam is 0.75, the probability to succeed in the practical exam is 0.6 and the probability to succeed in both of them is 0.5. If a person applies to this job for the first time. Find the probability of:

a Success in the theoretical exam only. **b** Success in at least one of the two exams.

Critical thinking:

Join with sport: A coach of one of the sports teams says in a news briefing that the probability that his team wins in the away match is (0.7), the probability of winning in the rematch is (0.9) and the probability of winning both matches is 0.5. Does the concept of probability agree with the words of the coach? Justify your answer.

Example

11 A die is rolled two consecutive time. The number on the upper face is observed in each time. Determine each of the following events:

First : A The appearance of two numbers their sum is less than or equal 4.
Second: B One of the two numbers is twice the other.
Third: C The asbsolute difference between the two numbers equals 2
Fourth: D The appearance of two numbers their sum is more than 12

Solution

n(S) = 36

First: A = {(1, 1), (1 , 2) , (1, 3), (2, 1), (2, 2), (3, 1)} $\therefore$ n (A) = 6, $\therefore$P (A) $=\frac{6}{36} = \frac{1}{6}$

Second: B = { (1, 2), (2, 1), (2, 4), (4, 2), (3, 6) , (6, 3)} $\therefore$ n (B) = 6 $\therefore$P (B) $= \frac{6}{36} = \frac{1}{6}$

Third: C = {(1, 3), (3, 1), (2, 4), (4, 2), (3, 5), (5, 3) , (4, 6) , (6, 4)} $\therefore$ P(C) $= \frac{8}{36} = \frac{2}{9}$

Fourth: Its impossible to get two numbers their sum is more than 12, $\therefore$ D $=\phi$, P (D) = 0

Try to solve

(12) In the previous example, calculate the following probabilities:

First: A the event "the two appearance numbers are equal".

Second: B the event "the number in the first roll is odd and the number in the second roll is even"

Example

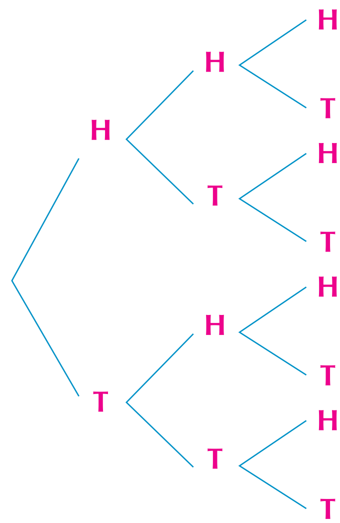
(12) A coin is tossed three consecutive times, the sequence of heads and tails is observed. Find the probability of each of the following:

First: A appearance of only one head.

Second: B appearance of at least two heads.

Third: C appearance of exactly two heads.

Solution



S = { (H, H, H), (H, H, T), (H, T, H), (H, T, T) , (T, H, H), (T, H, T) , (T, T, H), (T, T, T)},

n (S) = 8

First: $\because$ A appearance of only one head.

$\therefore$ A = { (H, T, T) , (T, H, T), (T, T, H)} ,

$\because$ n(A) = 3 $\therefore$ P (A) = $\frac{3}{8}$

Second : $\because$ B appearance of at least two heads means either two or more heads

$\therefore$ B = { (H, H, T) , (H, T, H), (T, H, H), (H, H, H)}

$\because$ n (B) = 4 $\therefore$ P (B) = $\frac{4}{8} = \frac{1}{2}$

Third: $\because$ C appearance of exactly two heads

$\therefore$ C = { (H, H, T), (H, T, H), (T, H, H)} $\because$ n (C) = 3 $\therefore$ P (C) = $\frac{3}{8}$

Try to solve

(13) In the previous example, calculate the probability of:

First : A appearance of the same face in the three tosses **Second :** B appearance of at most one head.

Third: C appearance of odd number of heads **Fourth:** D appearance of at least one tail.

Fifth: E appearance of number of heads equals number of tails.

Example

(13) Join with community: In one of the conferences, 200 persons from different nationalities participated in the conference and their data is represented by the following table:

	Speak Arabic	Speak English	Speak French	Total sum
Man	50	45	25	120
Woman	45	30	5	80
Sum	95	75	30	200

If one of the participants is chosen randomly, then find the probability that the chosen person is:

a) A woman speaks Arabic.
b) A man speaks English.
c) Speaks Arabic or French.
d) Speaks Arabic and English.
e) A woman does not speak English and does not speak Arabic.

Solution

a) The probability that the chosen person" A woman speaks Arabic" = $\frac{45}{200}$ = 0.225

b) The probability that the chosen person" A man speaks English" = $\frac{45}{200}$ = 0.225

c) The probability that the chosen person" Speaks Arabic or French" = $\frac{95+30}{200}$ = 0.625

d) The probability that the chosen person" Speaks Arabic and English" = p (ϕ) = 0

e) The probability that the chosen person" A woman does not speak English and does not speak Arabic" = $\frac{5}{200}$ = 0.025

Try to solve

(14) In the previous example, find the probability that the chosen person:

a) Does not speak English.
b) Speaks German.
c) A woman speaks French or English.
d) A man speaks Arabic or a woman speaks English.

Exercises (1 - 4)

(1) A student wants to buy a bag. It is possible to choose from three types. Each one has two sizes and the color of the bag is either black or brown. Represent the sample space by a tree diagram.

(2) In an experiment of tossing a coin once, then a die is rolled, observing the upper faces.

a) Write down the sample space of this experiment, then determine the following events.

- A: appearance of a head and an odd number.
- B: appearance of a tail and an even number
- C: appearance of a prime number 2 <
- D: appearance of a number divisible by 3

3 A die is rolled two consecutive times, the number on the upper face is observed in each time Determine each of the following events:

- A: The appearance of two equal numbers
- B: The appearance of two numbers their sum equals 9
- C: The appearance of two numbers their sum equals 13
- D: The appearance of number 3 one time at least.

4 From the set of numbers {1, 2, 3, 4} we need to form a two different digit number. Represent the sample space in a tree diagram, and then determine the following events:

- A: The event "the unit digit is an odd number ".
- B: The event "the tens digit is an odd number".
- C The event "the two digits are odd numbers".
- D The event "the unit digit or the tens digit is an odd number.

5 A bag contains 20 identical cards numbered from 1 to 20, If a card is selected randomly and the number written on it is recorded. Write the following events:

A the event "the recorded number is even and greater than 10

B the event "the recorded number is a factor of 12"

C the event "the recorded number is odd and divisible by 3

D the event "the recorded number is a multiple of the two numbers 2, 5

E the event "the recorded number is prime"

F the event "the recorded number satisfying the inequality $5x - 3 \leqslant 17$

6 Two cards are drawn one after the other from a set of 8 identical cards numbered from 1 to 8 and the drawn card must returned before drawn another card. What is the number of the elements in the sample space? and if:

A: is the event "the number in the second draw is three times the number in the first draw"

B: is the event "the sum of the two numbers is more than 13"

Write the events A, B. Are there two mutually exclusive events? Explain that.

7 In the experiment of tossing a coin three consecutive times and observing the sequence of heads and tails .represent the sample space with tree diagram, then determine the following events:

A the event "appearance of two tails at least" B the event "appearance of two tails at most"

C the event " appearance of a head in the first toss"

A the event " non-appearance of a head in the three tosses"

8 In an experiment of tossing a coin once, then a die is rolled, observing the upper faces Represent the sample space of this experiment by a probability tree diagram, and then determine the following events:

A appearance of a Tail and an even number"

B appearance of a head and an odd number"

C non-occurrence of A or non- occurrence of B"

D occurrence of the event A only

E occurrence of the event A and occurrence of the event B

Choose the correct answer from those given :

9 If a regular die is rolled once, then the probability of the appearance of an odd number less than 5 in the upper face equals:

a $\frac{2}{5}$ b $\frac{1}{2}$ c $\frac{1}{3}$ d $\frac{1}{6}$

10 If a regular die is rolled twice, then the probability of the appearance of an even number in the first roll and a prime number in the second roll equals:

a $\frac{1}{3}$ b $\frac{1}{6}$ c $\frac{1}{9}$ d $\frac{1}{4}$

11 If a ball is drawn randomly from a box contained 3 white balls, 5 red balls and 7 green balls, then the probability that the selected ball is white or green equals:

a $\frac{1}{5}$ b $\frac{2}{3}$ c $\frac{7}{15}$ d $\frac{1}{2}$

12 A card is drawn from a set of 9 identical cards numbered from 1 to 9. What is the probability that the drawn card carrying a divisor of (factor of) 9 or an odd number equals:

a $\frac{1}{3}$ b $\frac{7}{9}$ c $\frac{1}{2}$ d $\frac{5}{9}$

13 If A , B are two events in a sample space of a random experiment $B \subset A$, $P(A) = 2P(B) = 0.6$ then $P(A - B)$ equals:

a 0.6 b 0.3 c 0.4 d 0.2

14 A uniform die, the numbers 8, 9, 10, 11, 12, 13 written in its faces. If the die is rolled once, observing the number appearing on its upper face

a Find the probability of each of the following events:

- A: "appearance of an odd number."
- B "appearance of a prime number."
- C: "appearance of an even number."
- D "appearance of a number great than 12."
- E: "appearance of a number consists of two digits."
- F "appearance of a number consists of only one digit."

b Calculate: $P(A \cup C)$, $P(E \cup F)$, $P(B \cap D)$.

15 If is a sample space of a random experiment, where S = {A , B, C, D}, find: P(A) , P(B) , given that $P(A) = 3\,P(B)$, $P(C) = P(D) = \frac{7}{18}$

16 If A, B are two mutually exclusive events, S is a sample space of its random experiment, If $P(A \cup B) = 0.6$, $P(A - B) = 0.25$ find, P(A), P(B).

17 If A, B is a sample space of a random experiment, and $P(A) = \frac{1}{3}$, $P(B) = \frac{3}{8}$, $P(A \cap B) = \frac{1}{4}$ find:

a $P(A`)$ b $P(A \cup B)$ c $P(A - B)$ d $P(A` \cap B`)$

18 If A, B are two events, of a sample space of a random experiment, where: $P(A) = 0.4$, $P(B`) = 3P(B)$, $P(A \cap B) = 0.2$ find the probability of:

a Occurrence of A only. b Occurrence of A or B

c Occurrence of A and non-Occurrence of B.

19 A box contains colored identical balls: 4 red, 6 blue, and 5 yellow. A ball is selected at random from the box, find the probability that the drawn ball is:

a Red. b Blue or yellow. c Not blue. d Not red and not yellow.

20 One card is selected at random from 30 identical cards numbered from 1 to 30. Find the probability that the selected card is carring a number:

a Divisible by 3 b divisible by 5

c Divisible by 3 and 5 d divisible by 3 or 5

21 Three distinct coins are tossed once. Observing the upper faces, find the probability of the following events:

- A: appearance of a head or two heads.
- B: appearance of at least one head.
- C: appearance of a head at most.
- D: appearance of at two consecutive tails at least.

22 A die is rolled two consecutive times, the number on the upper face is observed in each time. Find the probability of each of the following events:

- Appearance of the number 4 in the first roll.
- Appearance of two numbers, their sum equals 8
- Appearance of two numbers, their sum is less than or equal 5

23 Join with sport : A random sample consists of 60 persons in a survey, it is found that 40 of them encourage Al Hilal club, 28 of encourage El negma club and 8 of them don't encourage any of them. A person is chosen at random from the sample. Find the probability that the chosen person encourages:

a At least one of the two clubs. b Both clubs.

c Al Hilal club only. d Only one of the two clubs.

24 In an experiment of tossing a coin once, then a die is rolled once , observing the upper faces. If A is the event of the appearance of a head and a prime number, B is the event of the appearance of an even number. Find the probability of the occurrence of each of the two events, and then calculate the probability of the following events:

- **a** The occurrence of one of the events at least
- **b** The occurrence of the two events together
- **c** The occurrence of only the event B
- **d** The occurrence of only one of the two events

25 a card is selected randomly from 50 identical cards numbered from 1 to 50, if the number written on it is recorded. Find the probability that the number written on the selected card is:

- **a** A multiple of number 7
- **b** A perfect square number
- **c** A multiple of number 7 and a perfect square number
- **d** Not a perfect square number and not a multiple of 7

26 If A , B are two events, in a sample space of a random experiment, where: $P(B) = \frac{4}{5} P(A)$, $P(A - B) = 0.24$ $P(B \cap A`) = 0.15$ then find: $P(A)$, $P(B)$, $P(A \cup B)$, $P(A` \cup B`)$

27 Tarek wrote 75 letters on the typewriter ,he found that 60% of them are without mistakes and Zead wrote 25 letters on the typewriter ,he found that 80% of them are without mistakes. If a letter is selected randomly from all letters written by both Tarek and Zead ,then find the probability that the selected letter is:

- **a** Without mistakes.
- **b** Written by Zead.
- **c** Written by Zead without mistakes.
- **d** Written by Tarek with mistakes.

28 If A , B are two events, in a sample space of a random experiment, where: $P(A) = 0.6$, $P(B) = 0.8$, $P(A` \cup B`) = 0.5$ then find $P(A` \cap B)$

Unit Three

3 - 2 Conditional Probability

You will learn

- Mutually exclusive events
- Events are not mutually exclusive
- Conditional probability

Key terms

- Mutually Exclusive Events
- Events are not Mutually Exclusive
- Conditional probability

Introduction:

You have previously learned to calculate the probability of an event (let it be A) of a random experiment by knowing the relation between the number of the elements of this event n(A) and the number of the elements of the sample space of the random experiment n(S) through the relation

$$P(A) \text{ (the probability of occurring even A)} = \frac{\text{the number of outcomes in event } n(A)}{\text{the number of outcomes in sample space } n(s)}$$

Mutually Exclusive Events

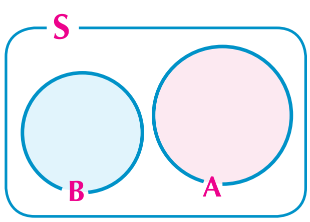
Of your study to the probability, you have learned that the mutually exclusive events are the events which cannot occur at the same time because when an event occurs, the other events are hindered to occur. This means that there are not common elements of the elements forming them.

Mutually exclusive events:

They are the two events which do not share in any element and their intersection is the null set ϕ.

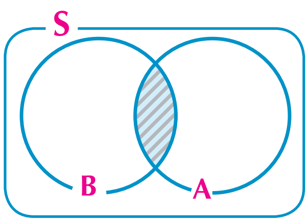
If A and B are two mutually exclusive events, then $A \cap B = \phi$

$\therefore P(A \cap B) = \text{Zero}$ and $P(A \cup B) = P(A) + P(B)$



Events are not Mutually Exclusive

They are the two events in which if an event occurs, it does not trap the occurrence of the other(there are common elements between them).Then,



(1) $P(A \cup B) = B(A) + P(B) - P(A \cap B)$

(2) $P(A') = 1 - P(A)$

(3) $P(A - B) = P(A) - P(A \cap B)$

(4) $P(A \cap B') = P(A - B) = P(A) - P(A \cap B)$

(5) $P(A' \cap B) = P(B - A) = P(B) - P(A \cap B)$

Materials

- Scientific calculator - Computer graph program.

Conditional Probability

If A and B are two events from S, sometimes the information is available that an events such as B has been occurred P (B). in this case, the occurrence of the event B may affect the probability of occurring the event A. The probability of occurring the event A in condition of occurring the event B can be calculated by knowing the relation among the outcomes of the event A or the outcomes of the event B.

Introductory example: In an experiment of rolling a regular dice once, the sample space S is : $S = \{1, 2, 3, 4, 5, 6\}$, If the event $A = \{1, 2, 3\}$ is the event of the appearance of a number less then 4

It becomes clear that: $P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$

If the event $B = \{2, 4, 6\}$ is the event of the appearance of an even number.

We ask: What is the probability of occurring the event A, if we know that the event B has already occurred?

In other words: what is the probability of getting an even number less than 4?

We notice that the given condition reduces the sample space into set $B = \{2, 4, 6\}$

Then, the event matching the appearance of an even number is $A \cap B = \{2\}$

And the probability required is: $\frac{P(A \cap B)}{P(B)} = \frac{1}{6} \div \frac{1}{2} = \frac{1}{3}$

This example shows us how the probabilities of some events differ in regard to the difference of the sample space.

Learn

Conditional Probability

If S is the sample space of a random experiment and A and B are two events of this sample space, then the probability of occurring the event A in condition of occurring the event B is denoted by the symbol $P(A|B)$ and read as the probability of occurring the event A in condition of occurring the event B. it can be determined by the next relation:

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) > 0}$$

Notice that: the conditional probability has the same properties of the unconditional one.

1- $0 \leqslant P(A|B) \leqslant 1$

2- $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

3- If $A_1 \cap A_2 = \phi$ then, $P[(A_1 \cup A_2)|B] = P(A_1|B) + P(A_2|B)$

Regard the following:

- $P(A|B) \neq P(B|A)$
- $P(A'|B) = 1 - P(A|B)$
- $P(A \cap B) = P(A|B) \times P(B)$ in a condition $P(B) > 0$
- $P(A \cap B) = P((B|A) \times P(A)$ in a condition $P(A) > 0$

Example the conditional probability

14 A regular die has been rolled once. calculate the probability of appearing the number 2 known that the number appeared is even.

Solution

Let the sample space S= { 1, 2, 3, 4, 5, 6} , A = {2} , B = { 2 , 4, 6}

Then, $P(B) = \frac{3}{6} = \frac{1}{2}$, $P(A \cap B) = P(A) = \frac{1}{6}$

$\because P(A | B) = \frac{P(A \cap B)}{P(B)}$

$\therefore P(A | B) = \frac{1}{6} \div \frac{1}{2} = \frac{1}{6} \times 2 = \frac{1}{3}$

The probability of appearing the number 2 known that the number appeared is even is $\frac{1}{3}$

Notice

In the conditional probability, notice that the event after the phrase (What is the probability of) is the events which we start with and the event next to the phrases (known that A, if A, if known that a and so on) is the conditional event.

Try to solve:

1 A regular die has been rolled twice. What is the probability the number of points in the first roll is not more than 4 if you know that the absolute difference between the two number appeared equals 2?

Example Doing the operations

15 If A and B are two events of the sample space where P(A) = 0.45 , P (B) = 0.6 , P (B | A) = 0.8 find:

a $P(A \cap B)$ b $P(A \cup B)$ c $P(A | B)$ d $P(B' | A)$

Solution

a $\because P(B | A) = \frac{P(B \cap A)}{P(A)}$

$\therefore 0.8 = \frac{P(A \cap B)}{0.45}$ $\therefore P(A \cap B) = 0.8 \times 0.45 = 0.36$

b $\because P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\therefore P(A \cup B) = 0.45 + 0.6 - 0.36 = 0.69$

c $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0.36}{0.6} = 0.6$

Notice that: $P(A | B) \neq P(B | A)$

d $P(B' | A) = \frac{P(B' \cap A)}{P(A)} = \frac{P(A - B)}{P(A)}$

$= \frac{P(A) - P(A \cap B)}{P(A)}$

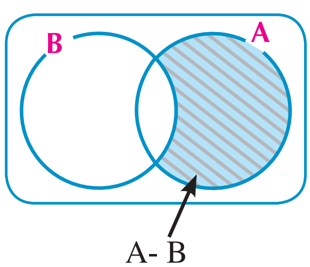
$= \frac{0.45 - 0.36}{0.45} = 0.2$

Remember

$P(B \cap A) = P(A \cap B)$

$P(A \cup B) =$
$P(A) + P(B) - P(A \cap B)$

$P(A - B) =$
$P(A) - P(A \cap B)$



Try to solve

2 If A and B are two events of the sample space of a random experiment S where $P(A) = 0.7$, $P(B) = 0.25$, $P(A - B) = 0.45$ find:

a $P(A \mid B)$

b $P(B \mid A)$

c $P(A' \mid B)$

d $P(A' \mid B')$

Example **Harmonic tables**

16 From the data in the next table:

Case	Number of people	
	Wear glasses	Do not wear glasses
Man	800	600
Woman	400	200

Find the probability of a woman if she is wearing glasses has been randomly chosen

Solution

Let n(S)= Number of people under study=2000,

A is the event that the person chosen is a woman,

B is the event that the person chosen wears glasses.

$$P(A \cap B) = \frac{400}{2000} = \frac{1}{5}$$

$$P(B) = \frac{1200}{2000} = \frac{3}{5}$$

Required is to find the probability of A known that B has already occurred. i.e. $P(A \mid B)$

$$\because P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore P(A \mid B) = \frac{1}{5} \div \frac{3}{5} = \frac{1}{3}$$

The probability of a woman wearing glasses has been randomly chosen is $\frac{1}{3}$

Try to solve

3 In the previous example, find the probability that:

a A man if he does not wear glasses is randomly chosen.

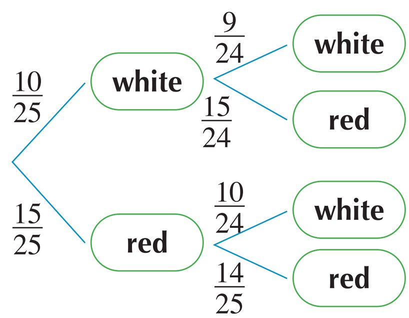
b A woman or man gevin that they are wearing glasses is randomly chosen.

Example Tree diagram

17 A bag contains 10 white balls and 15 red balls. Two balls have been consecutively drawn without replacing. What is the probability the two drawn balls are white?

Solution

In this example, we notice that drawing the balls has been conducted consecutively. As a result, it is subjected to the order. In other words, the second drawing is conditioned by the occurrence of the first drawing. This example can be represented by the tree diagram as shown in the figure opposite.



Let A denote the event that the first drawn ball is white and B denote the event that the second drawn ball is white.

Let (B | A) denotes the event of drawing the second ball in a condition the first ball has already been drawn .

and **let** (A $\cap$ B) denotes the event of drawing two white balls..

$\because$ P(B | A) $= \dfrac{P(B \cap A)}{P(A)}$

$\therefore \dfrac{9}{24} = \dfrac{(A \cap B)}{\frac{10}{25}}$

$\therefore$ P (A $\cap$ B) $= \dfrac{9}{24} \times \dfrac{10}{25} = \dfrac{3}{20}$

The probability the two drawn balls are white is $\dfrac{3}{20}$

Try to solve

4 In the previous example, find the probability the two drawn balls are red.

Example Education

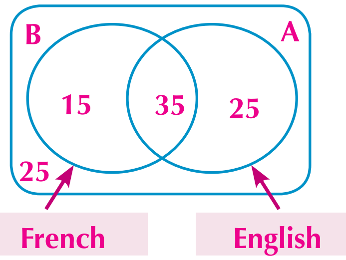
18 100 learners are studying in a language institute. The number of learners studying English is 60, the number of the learners studying French is 50 and the number of the learners studying both languages is 35. If a learner has been randomly chosen from the institute, find the probability the learner studies:

a a language at least.

b English if he (she) studies French.

c French if he (she) studies English.

Solution:



The data of this problem can be explained using the Venn diagram as shown in the figure opposite.

Let the following events be as follows:

The learner studies English = A

The learner studies French = B, then:

$P(A) = \frac{60}{100} = 0.6$, $P(B) = \frac{50}{100} = 0.5$, $P(A \cap B) = \frac{35}{100} = 0.35$

a The probability the learner studies a language at least is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\therefore P(A \cup B) = 0.6 + 0.5 - 0.35 = 0.75$

I.e. the probability the learner studies a language at least is 0.75

b the probability the learner studies English if he (she) studies French = $P(A \mid B)$

$\because P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

$\therefore P(A \mid B) = \frac{0.35}{0.5} = 0.7$

I.e. the probability the learner studies English if he (she) studies French is 0.7

c the probability the learner studies French if he (she) studies English $= P(B \mid A)$

$\because P(B \mid A) = \frac{P(B \cap A)}{P(A)}$

$\therefore P(B \mid A) = \frac{0.35}{0.6} \simeq 0.583$

I.e. the probability the learner studies French if he (she) studies English is 0.583 approximately.

Try to solve

5 Two players A and B shoot at the same time towards the goal. If the probability the player A scores a goal $= \frac{2}{5}$, the probability the player B scores a goal $= \frac{1}{4}$. and the probability both players A and B score together $= \frac{1}{6}$, Find the probability of:

a Scoring a goal.

b Scoring a goal by player A if player B scored a goal.

c Scoring a goal by player B if player A scored a goal.

Exercises 3 - 1

First: Choose the correct answer:

1) in an experiment for tossing a regular coin twice, the probability of appearing Tail in the second toss if Head appears in the first toss equals:

a) $\frac{1}{4}$ b) $\frac{1}{2}$ c) $\frac{3}{4}$ d) 1

2) in an experiment for rolling a regular die once, the probability of appearing a prime even number greater than 1 is:

a) $\frac{1}{5}$ b) $\frac{2}{5}$ c) $\frac{3}{5}$ d) $\frac{4}{5}$

3) in an experiment for rolling a regular die once, the probability of appearing the number 3 known that the appearing number is odd is:

a) $\frac{1}{4}$ b) $\frac{1}{3}$ c) $\frac{1}{2}$ d) $\frac{3}{4}$

4) If $P(A \cap B) = \frac{2}{5}$ ، $P(A) = \frac{4}{5}$, then $P(B \mid A) =$

a) $\frac{1}{2}$ b) $\frac{8}{25}$ c) $\frac{1}{4}$ d) $\frac{2}{5}$

5) If $P(A \mid B) = \frac{1}{3}$ ، $P(B) = \frac{12}{25}$ then $P(A \cap B) =$

a) $\frac{4}{25}$ b) $\frac{1}{4}$ c) $\frac{25}{36}$ d) $\frac{16}{25}$

Second: Answer the following questions:

6) If A and B are two events of the sample space of a random experiment S where $P(A) = 0.4$ ، $P(B) = 0.7$, $P(A \mid B) = 0.3$ find:

a) $P(A \cap B)$ b) $P(A \cup B)$ c) $P(B \mid A)$ d) $P(A \mid B`)$

7) If $P(A`) = 0.4$ ، $P(B) = 0.5$ ، $P(A \cup B) = 0.8$ find $P(A \mid B`)$

8) If $P(B \mid A) = \frac{2}{3}$, $P(B \mid A`) = \frac{4}{7}$ ، $P(A) = \frac{3}{5}$ find

a) $P(A \cap B)$ b) $P(A \cup B)$

9) A die has been rolled once. Calculate the probability the appearing number is a prime number in a condition the appearing number is an odd number.

10) in an experiment of rolling two different dices once, find the probability:

a) the appearing number on the second die equals 4 known that the appearing number on the first die equals 2.

b) the sum of the two appearing numbers is even known that the appearing number on the first die equals 6.

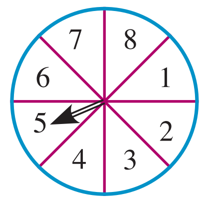
11) If the probability of a student to succeed in an exam is 0.7 and the probability to travel abroad if he succeeded is 0.6. What is the probability of his success and traveling abroad?

12 A 45- student class; 27 students study French, 15 students study German and 9 students study both languages. A student is randomly chosen, calculate the probability the chosen student studies:

a a language at least.
b French if he (she) studies German.
c German if he (she) studies French.

13 two different dices have been rolled once, find the probability of the following events:

a the appearance of number 2 on the two faces together known that the same number appeared on both of them.
b the appearance of number 5 on the two faces together known that each of the appearing number is greater than 4.
c the non appearance of number 3 on any of the two faces known that the two numbers appeared are odd..



14 Spinning wheel: A spinning wheel has been divided into 8 equal circular sectors from 1 to 8. What is the probability the gage lands on the number 5 if known that it has previously landed on an odd number?

15 the next table illustrates the numbers of the sports teams participating in different games:

Sports game	Hand ball	Soccer	Volleyball	Basketball	Hockey ball
Numbers of participating teams	4	10	6	7	3

A game is randomly chosen. What is the probability the game is:

a Hockey ball known that it is not of Volleyball games?
b Basketball known that it is neither of soccer nor Hand ball?

16 a random sample made up of 30 male students and 20 female students to answer some questions related to economy and consuming the energy has been chosen and their answers have been as follows:

Answer	Yes	No	Uncertain	Total
Male students	20	6	4	30
Female students	15	3	2	20

A student is randomly chosen. What is the probability the student chosen is a female and her answer is “Yes“?

17 A box contains 5 white balls and 7 black balls. If 2 balls have been consecutively drawn without replacing, find the probability::

a the second ball is white if the first one is white.
b the first ball is white and the second one is white.
c the second ball is black and the first one is white.

18 Karim and Ziad compete for the presidency of the student Union at their school in three classes, and the following table illustrates the votes obtained by each of them:

	First class	Second class	Third class	Total
Karim	196	174	130	500
Ziad	240	165	135	540

A student is randomly chosen. What is the probability the student chosen::

a voted for Karim known that the student is from the third class.

b voted for Ziad known that the student is from the second class.

19 A job has been announced for and 100 people have applied for this job. If their data have been ordered as follows::

Qualified			Unqualified		
	Married	**Single**		**Married**	**Single**
Male	40	10	**Male**	3	12
Female	10	10	**Female**	10	5

a calculate the probability the employee chosen is married in a condition he (she) is qualified.

b calculate the probability the employee chosen is married and qualified.

c calculate the probability the employee chosen is married in a condition he(she)is unqualified.

20 In the final year exam, 30% of the students have failed in chemistry, 20% have failed in physics and 15% have failed in both chemistry and physics. A student has been randomly chosen.

a what is the probability the student fails in physics if he (she) has already failed in chemistry?

b what is the probability the student fails in chemistry if he (she) has already failed in physics?

c find the probability the student fails in chemistry in a condition he (she) fails in physics.

d find the probability the student succeeds in physics in a condition he (she) succeeds in chemistry.

21 Activity: Use Venn diagram:

A and B are two events of the sample space S where $P(A) = 0.7$ ، $P(B) = 0.4$ ، $P(A \cap B) = 0.2$

a represent the previous sets using Venn diagram and write down the probability of these events to occur on the graph.

b find the probability of the following events:

First: the occurrence of event A in a condition event B do not occur.

Second: the occurrence of event B in a condition event A do not occur..

Independent Events

Unit Two
3 - 3

You will learn

- Independent events
- Dependent events

Key terms

- Independent Events
- Dependent Events

Think and discuss

Meditate the following examples

1- tossing a coin and rolling a die once.
2- a student has succeeded in mathematics and chemistry.
3- a ball has been randomly drawn from a box containing 10 balls, then it has been turned back to the box and another ball has been drawn.
4- a student has succeeded in the physics lab-exam and in physics.
5- a ball has been randomly drawn from a box containing 10 balls without turning it back, then another ball has been drawn.

What do you notice?

From the first three examples, we notice that:

1- the outcomes of the coin do not affect the outcomes of the dice .
2- the success or failure of the student in mathematics does not affect his (her) success or failure of the student in chemistry
3 - returning the first ball to the box after drawing it does not change the number of the balls, so the first draw does not affect the second one.

As a result, the events in each example of the first previous three examples are called the independent events.

4- the success of the student in the physics lab-exam affects the success in physics.
5- when a ball has been drawn from the box without returning to the box it affects the number of the balls in the box and so the first draw affects the second draw.

As a result, the events in examples (4) and (5) are called the dependent events.

Learn **independent events**

Definition **It is said that A and B are two independent events if and only if $P(A \cap B) = P(A) \times P(B)$.**

I.e. the probability of occurring two independent events together equals the probability of occurring the first event multiplied by the probability of occurring the second event

Materials: Scientific calculator- computer graphs.

It is noticed that if the two events A and B are independent and $P(B) \neq 0$

then $P(A \mid B) = P(A)$ **I.e.** the occurrence of an event does not affect the occurrence of the other event.

For example: a regular coin has been tossed twice and the frequency of occurring the Tail and Head has been noticed , then: S = { (H, H), (H, T), (T, H) , (T, T)}

So the probability of any of the outcomes $= \frac{1}{4}$

Let event A represent the appearance of tail in the second time = {(H , T) , (T , T) }

and event B represent the appearance of Head in the first time = {(H , H) , (H , T)}

then $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} = P(A)$

I.e. the occurrence of event B does not affect the probability of occurring event A. In other words, the probability of event A does not depend on knowing that the event B is occurred or not. Thus, we say that the two events A and B are independent.

Notice that: The mutually exclusive events A and B are independent if and only if $P(A) \times P(B) = 0$

In other words, if and only if the probability of A or the probability of B equals zero.

Example

1 what is the probability of appearing Head and number 5 in an experiment of tossing a coin once, then rolling a dice?

Solution

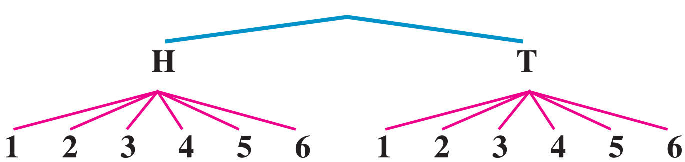
The tree diagram can be used to write down the sample space. We notice that when the coin is tossed, it does not affect the sample outcomes of rolling the dice. This means that the two events are independent.

Let A event of appearing Head, then $P(A) = \frac{1}{2}$, B = event of appearing number 5, then $P(B) = \frac{1}{6}$

$\because P(A \cap B) = P(A) \times P(B)$ $\therefore P(A \cap B) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

$\therefore$ The probability of appearing Head and number 5 is $\frac{1}{12}$

Note: The probability of appearing Head and number 5 can be found directly by writing down the sample space as shown in the figure opposite.



S = { (H , 1) , (H , 2) , (H ,3) , (H , 4) , (H , 5) , (H , 6) , (T , 1) , (T, 2) , (T, 3) , (T , 4), (T , 5) , (T , 6) }

The event of appearing Head and number 5 = {(H , 5)}

and The probability of appearing Head and number 5 $= \frac{1}{12}$

Try to solve:

1 in the previous example, find the probability of appearing Tail and a prime number.

Example

(2) If A and B are two events of a sample space of a random experiment S and P(A)=0.5, P(B)=0.6 and $P(A \cup B) = 0.8$. Explain if A and B are two independent events.

Solution

$\because P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$\therefore P(A \cap B) = 0.5 + 0.6 - 0.8 = 0.3$ **(1)**

$\because P(A) \times P(B) = 0.5 \times 0.6 = 0.3$ **(2)**

From **(1)** and **(2)**, A and B are two independent events.

Notice that: To show the difference between two mutually exclusive events and two independent events, we study the next example:

We know that when we toss a regular coin once, the sample space S={H, T}

We also know that $P(H) = \frac{1}{2}$ and $P(T) = \frac{1}{2}$

Furthermore, we know that the two events H and T are two mutually exclusive events because the occurrence of an event negates the occurrence of other event.

$\therefore P(H \cap T) = \text{zero}$, $\therefore P(H \cap T) \neq P(H) \times P(T)$

I.e. H and T are two mutually exclusive events but they are dependent

Try to solve

(2) If A and B are two events of a sample space of a random experiment S where S= {1,2,3,4,5,6} A = {2, 3, 5, 6} and B {1, 4, 5, 6}. Are A and B two dependent events? Explain.

Example

(3) Insurance: a man and his wife have insured their life at a life insurance company. If the company has estimated the probability that the man will live more than 20 years to be 0.2 and the probability that his wife will live more than 20 years to be 0.3, find the probability that:

a) The man and his wife will live more than 20 years together.

b) At least one of them will live more than 20 years.

c) Only one of them will live more than 20 years.

Solution

Let A be the event that the man will live more than 20 years $\therefore$ P(A) = 0.2,

B the event that the wife will live more than 20 years $\therefore$ P(B) = 0.3

a) The probability that the man and his wife will live more than 20 years together = $P(A \cap B)$

$\because P(A \cap B) = P(A) \times P(B)$ $\therefore P(A \cap B) = 0.2 \times 0.3 = 0.06$

b) The probability that at least one of them will live more than 20 years = $P(A \cup B)$

$\because P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $\therefore P(A \cup B) = 0.2 + 0.3 - 0.06 = 0.44$

c) The probability that only one of them will live more than 20 years = P (A ∪ B) - P (A ∩ B)

$\therefore P(A \cup B) - P(A \cap B) = 0.44 - 0.06 = 0.38$

Try to solve

3) **Shooting:** two soldiers A and B have shot a missile in the direction of a target. If the probability soldier A shot the target is 0.6 and the probability soldier B shot the same target is 0.5, find the probabilities of the following events:

a) Shooting the target by soldier A and soldier B together.

b) Shooting the target by a missile at least.

c) Shooting the target by one missile only. d) No one Shot the target.

Example

4) **Drawing with replacing:** A bag contains 6 blue balls and 4 red balls. A ball is randomly drawn, then it is turned back to the bag, then another ball is drawn. What is the probability

a) the two balls are red in the two times b) the two balls are blue in the two times?

c) the first ball is red and the second is blue? d) one ball is red and the other is blue?

Solution

a) As long as the process of drawing the ball is accompanied with replacing, the two events are **independent.**

Let S= Sample space, A= drawing the ball first time and B= drawing the ball second time.

$\therefore n(S) = 10\ ,\ P(A) = \frac{4}{10},\ P(B) = \frac{4}{10}$ **(since the drawing is with replacing)**

$\because P(A \cap B) = P(A) \times P(B)$ $\therefore P(A \cap B) = \frac{4}{10} \times \frac{4}{10} = \frac{16}{100} = \frac{4}{25}$

Similarly:

b) The probability the two balls are blue in the two times $= \frac{6}{10} \times \frac{6}{10} = \frac{36}{100} = \frac{9}{25}$

c) The probability the first ball is red and the second is blue $= \frac{4}{10} \times \frac{6}{10} = \frac{24}{100} = \frac{6}{25}$

d) The probability one ball is red and the other is blue = the probability the first ball is red and the second is blue + the probability the first ball is blue and the second is red

$$= \frac{4}{10} \times \frac{6}{10} + \frac{6}{10} \times \frac{4}{10} = \frac{12}{25}$$

Try to solve

4) The probability of rising the stock market index in a country (A) equals 0.84 and the probability of rising the stock market index in a country (B) equals 0.75. What is the probability the index of the two stock markets of the two countries A and B rises?

Learn Dependent events

A and B are two dependent events if: $\boxed{P(A \cap B) \neq P(A) \times P(B)}$

because we know from the conditional probability that:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) \neq 0$$

$$,P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad P(A) \neq 0$$

I.e. it can be written as $P(A \cap B) = P(A \mid B) \times P(B)$

$= P(B \mid A) \times P(A) \qquad P(A) \neq 0, P(B) \neq 0$

In other words, the two events A and B are dependent if the probability of the occurrence of one of them affects in a way the probability of the occurrence of the other event.

Probability of dependent events:

Example

5 If S is the sample space of a random experiment where S = {1, 2, 3, 4, 5, 6, 7, 8}, A = {1, 2, 4, 8} and B ={2, 5, 6, 7}. Are A and B two independent events? Explain.

Solution

$\because n(A) = 4 \qquad \therefore P(A) = \frac{4}{8} = \frac{1}{2}, \qquad \because n(B) = 4 \qquad \therefore P(B) = \frac{4}{8} = \frac{1}{2}$

$\because A \cap B = \{2\} \qquad \therefore P(A \cap B) = \frac{1}{8} \qquad$ (1)

$\because P(A) \times P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \qquad$ (2)

From (1) and (2), $P(A \cap B) \neq P(A) \times P(B)$ so, A and B are two dependent events.

Try to solve:

5 In the previous example If C = {2, 3, 4, 7}. Are B and C two independent events? Explain.

Drawing without replacing:

Example

6 A bag contains 6 blue balls and 4 red balls. If two balls are drawn one after another without replacing. What is the probability:

a the two balls are red?

b the two balls are blue?

c the first ball is red and the second is blue?

Solution

This example is similar to example (3) but the difference is that drawing the balls is conducted without replacing. So, the two events are dependent.

a if the two drawn balls are red, then

The probability the first drawn ball is red and the second is red =

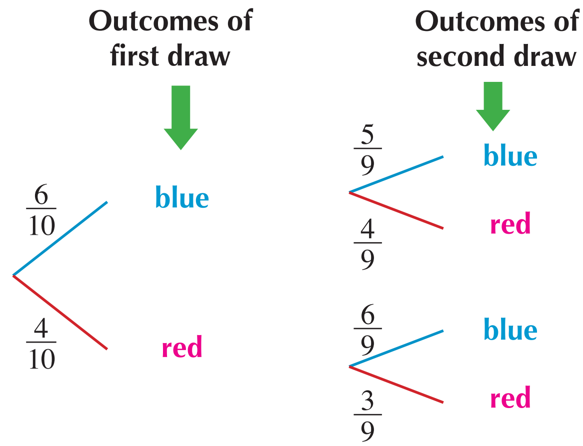
the probability the second drawn ball is red after drawing the first red ball × The probability the first drawn ball is red

$= \frac{4}{10} \times \frac{3}{9} = \frac{2}{15}$

b if the two drawn balls are blue, then: The probability the first drawn ball is blue and the second is blue = $\frac{6}{10} \times \frac{5}{9} = \frac{1}{3}$

c The probability the first drawn ball is red and the second is blue =
The probability the first drawn ball is red × The probability the second drawn ball is blue in a condition the first drawn ball is red
$= \frac{4}{10} \times \frac{6}{9} = \frac{4}{15}$

The tree diagram can be used as shown in the figure opposite to find the outcomes of the dependent events.



Try to solve

6 a bag contains 3 red balls and 5 black balls. Two balls are drawn one after another without replacing. What is the probability:

a the two balls are black?

b the first ball is black and the second is red?

c one of the two balls is red and the other is black?

Exercises 3 - 2

1 which of the following events is independent and which is dependent?

a tossing a coin, then rolling a die once.

b a card is drawn from a bag without replacing, then another card is drawn from the same bag

c a ball is drawn from a bag with replacing, then another ball is drawn from the same bag.

d a football team reaches the semi-final match. If it wins, it will play the championship match.

e choosing a name by lot without replacing, then choosing another name.

f choosing a ball from a bag and placing it in another place, then choosing another ball from the Same bag.

g karem participated and succeeded in the cultural competition on Monday and participated and succeeded in the scientific competition on Thursday, too.

Choose the correct answer:

2 If A and B are two independent events, $P(A) = 0.2$ ، $P(B) = 0.6$, then $P(A \cup B) =$

a 0.12 b 0.32 c 0.68 d 0.8

3 If A and B are two independent events $P(A) = 0.25$ ، $P(B) = 0.4$, then $P(A - B) =$

a 0.1 b 0.15 c 0.3 d 0.65

4. If A and B are two independent events $P(A) = 0.3$ ، $P(B) = x$ ، $P(A \cup B) = 0.72$ then x equals:

   a) 0.24   b) 0.28   c) 0.4   d) 0.6

Answer the following questions:

5. What is the probability of appearing Head and number 3 if a coin is tossed then a die is rolled once?

6. What is the probability of getting Tail 4 times if a coin is tossed four consecutive times?

7. A regular die is rolled once. And A is the event of appearing an even number and B is the event of appearing a squared number. Are A and B two independent events? Explain.

8. If A and B are two events of a sample space of a random experiment and $P(B) = 0.3$, $P(A \cup B) = 0.5$ find the value of P(A) if a and b are:

   a) Two mutually exclusive events.   b) Two independent events .

9. A bag contains a set of marbles distributed as follows: 2 red, 3 green and one blue. A marble is randomly chosen with replacing, then another marble is chosen. Find the probability the two marbles chosen are green.

10. In the previous question, if a marble is randomly chosen without replacing, then another marble is chosen, find the probability the first marble chosen is blue and the second is green.

11. A bag contains the following balls: 6 red, 4 orange, 3 yellow, 2 blue and 5 green. A ball is randomly chosen without replacing, then another ball is chosen.
    Find the probability the balls chosen are:

    a) Red and blue.   b) Red and yellow.   c) Red and red.   d) Orange and blue.

12. Two soldiers A and B have shot a bullet in the direction of a target. If the probability the first soldier shot the target is 0.4 and the probability the second soldier shot the same target is 0.7,
    First: find the probabilities of the following events::

    a) The two soldier shot the target together.   b) At least one soldier shot the target.
    c) Only one soldier shot the target..   d) At most one of them Shot the target.

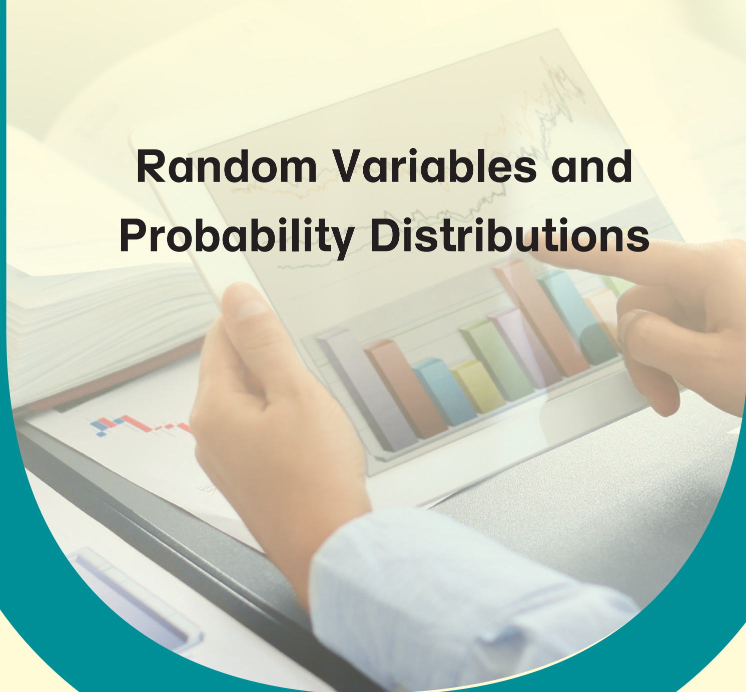
    Second: If you know that one soldier at least shot the target, find the probability the soldier A only shot the target.

13. If A and B are two independent events, prove that all the following event pairs are also independent.

    a) A', B'   b) A', B   c) A, B'

Unit
4

Random Variables and Probability Distributions



introduction

You have previously learned the random experience and some probability concepts.

You sometimes like to deal with numerical values related to the results of the random experience which may be characteristics that are mathematically difficult to deal with. In such a case, we convert these descriptive values into real numerical values called a random variable, which is used to express the results of the random experience.

In this unit, we are going to learn two types of random variables which are:

- Discrete Random Variables
- Continuous Random Variables

Furthermore, we will learn the probability distribution functions of the random variables which are divided into:

- Probability distribution function of discrete random variable.
- Probability density function.

Unit objectives

By the end of the unit and carrying the involved activities, the students should be able to:

- Identify the concept of the random variable and distinguish the discrete random variable and the continuous one.
- Identify the probability density function of a continuous random variable and use it to calculate the probability of occurring the value of the random variable within a certain interval.
- Identify the concept of the mean (expectation) and variance.
- Deduce the standard deviation of a random variable.
- Identify the coefficient of variation
- Identify the continuous distributions.
- Find probability of Geometric distribution
- Calculate expectation، variance and S.D of G.distribution
- Find prob. of binomial dist.
- Calculate Mean، variance of binomial distribution.

Key Terms

- *Random Variable*
- *Discrete Random Variable*
- *Probability Distributions*
- *Expectation(Mean)*
- *Variance*
- *Coefficient of Variation*
- *Probability Density*

Lessons of the unit

Lesson (4 - 1): Discrete random variable.

Lesson (4 - 2): Expectation (mean) and variance of the discrete random variable.

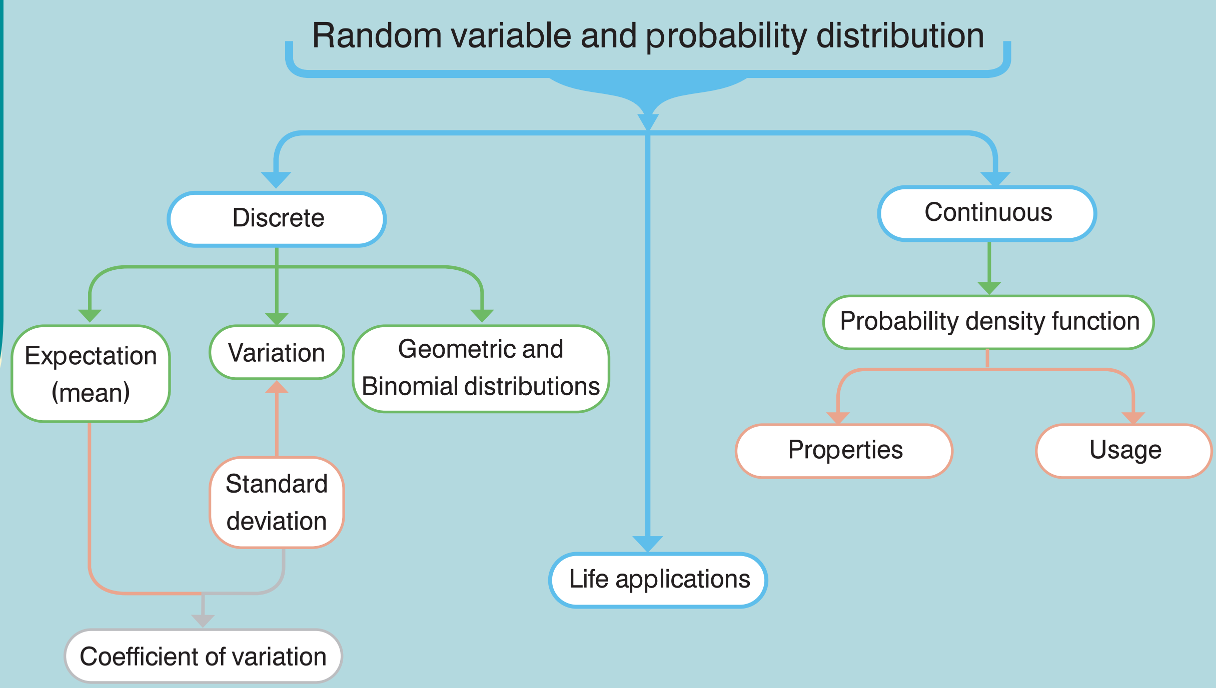
Lesson (4 - 3): Geometric and Binomial Distributions

Lesson (4 - 4): The probability density function of the continuous random variable

Materials

Scientific Calculator

Unit planning organizer



Unit Four

4 - 1

Discrete random variable

You will learn

- Random variable
- Discrete random variable
- Continuous random variable
- Probability distribution

Key terms

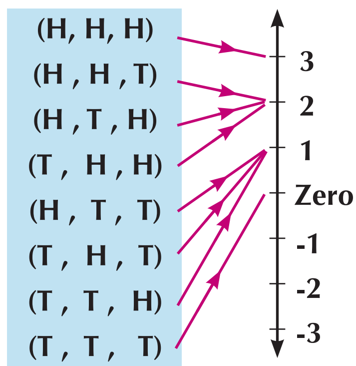
- Random Variable
- Discrete Random Variable
- Continuous Random Variable
- Probability Distributions

Introduction: : you have learned the random experiment and found the sample space for this experiment . In this lesson, you will identify a new variable related to such an experiment - The random variable .

You will learn in this lesson how to describe the items of two different phenomena with respect to the relation between them.

Random variable:

In an experiment for tossing a coin three consecutive times, the sample spaces S can be identified as shown in the opposite figure. If you are asked to find " the number of Heads" which appear in the sample space S, you draw a diagram to show the relation between S (as an independent variable) and the number of the image which is real number R (as a dependent variable). This relation expresses a function and can be symbolically written as : X: S $\rightarrow$ R where X is denoted the random variable .



Definition

The random variable is a function whose domain is the set of the elements of the sample space S and Co-domain is the set of real numbers R.

Then, the random variable range X in the previous example = {0 , 1 , 2 , 3 }

Note: The random variable divides the sample space S into mutual exclusive events; each event is linked with a real number , this link expresses a function X from the sample space S to the set of the real number R.

Remember

The function is defined by:

- **Domain**
- **Co-domain**
- **Rule of the function**

The function range is the set of the images of the domain elements in the co-domain.

Discrete Random Variable

Definition

The range of the discrete random variable is a finite set (i.e. it can be counted) of the real numbers.

Materials

- Scientific Calculator , Computer graphics.

For examples :

- the number of stocks specialized for an individual in an underwriting of a joint-stock company.
- The number of car accidents on a high way within a week.
- The number of the outcoming telephone calls for a family within a month.

Example **Discrete random variable**

1. In an experiment for tossing a coin three consecutive times , If the random variable X expresses " number of Heads - number of Tails" , write down the random variable range.

Solution

S = { (H , H , H) , (H , H , T) , (H , T , H) , (H , T , T) , (T , H , H) , (T , H , T) , (T , T , H) , (T , T , T) }

Sample space S	X: number of tails – number of heads
(H , H , H)	$3 - 0 = 3$
(H , H , T)	$2 - 1 = 1$
(H , T , H)	$2 - 1 = 1$
(H , T , T)	$1 - 2 = -1$
(T , H , H)	$2 - 1 = 1$
(T , H , T)	$1 - 2 = -1$
(T , T , H)	$1 - 2 = -1$
(T , T , T)	$0 - 3 = -3$

The random variable range = {- 3 , - 1 , 1 , 3 }

Try to solve

1. in the previous example, find the random variable range expressing : number of Heads × number of Tails.

Example **Discrete random variable**

2. A regular die has been tossed two consecutive times . Find the random variable expressing the sum of the two shown numbers .

Solution

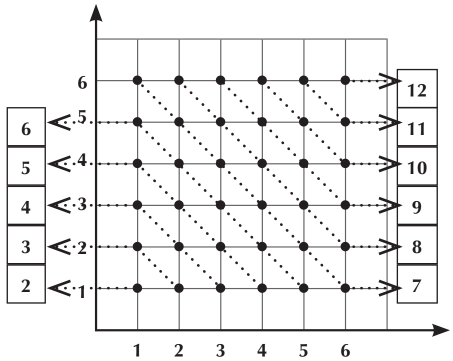
Sample space S	x: sum of the two numbers
(1 , 1)	2
(1 , 2) , (2 , 1)	3
(3 , 1), (2 , 2), (1 , 3)	4
(4 , 1) , (3 , 2) , (2 , 3) , (1 , 4)	5
(5 , 1), (4 , 2), (3 , 3), (2 , 4), (1. 5)	6

Sample space S	x: sum of the two numbers
(4 , 3) ,(3 , 4) ,(2 , 5) ,(1 , 6), (6 , 1) , (5 , 2)	7
(6 , 2), (5 , 3), (4 , 4), (3 , 5), (2 , 6)	8
(6 , 3), (5 , 4), (4 , 5), (3 , 6)	9
(6 , 4) , (5 , 5) , (4 , 6)	10
(6 , 5) , (5 , 6)	11
(6 , 6)	12

From the previous table , the random variable range is

x = {2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10 , 11 , 12 }

we can use the lateral shape to find the random variable range X .



Try to solve

2 In the previous example, find the random variable range expressing : (the greater of the two shown numbers).

Probability distribution

Probability Distribution Function of Discrete Random Variable

Definition

If X is a discrete random variable whose range is the set:{ x_1 , x_2 , x_3, ... , x_r } then the function f defined as follows: $f(x_r)$ = P (X =x_r) for each r = 1 , 2 , 3 , ...

Identifies what is called the probability distribution function of the discrete random variable X which is expressed by a set of ordered pairs specifying the function f.

i.e. the probability distribution of the random variable X = { (x_1 , $f(x_1)$) , (x_2 , $f(x_2)$) , (x_3 , $f(x_3)$) , ... , (x_n , $f(x_n)$) }

Note : the probability distribution of the random variable X can be written in the form of the following table:

x_r	x_1	x_2	x_3	......	x_n
$f(x_r)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	......	$f(x_n)$

You can notice that the function f in the previous definition satisfies the following two conditions:

1- $f(x_r) \geqslant 0$ for each **r** = 1 , 2 , 3 , ... , n

2- $f(x_1) + f(x_2) + f(x_3) + \ldots\ldots + f(x_r) = 1$

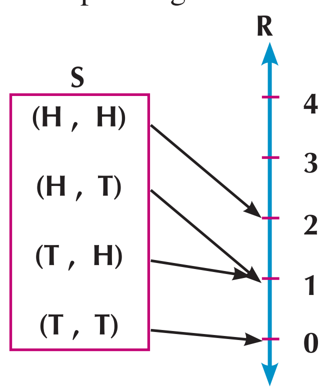
Example probability distribution function

3 A coin has been tossed three consecutive times and the upper face has been seen. Write down the probability distribution function of the discrete random variable x expressing the number of appearing the Head face.

Solution

S = { (H , H) , (H , T) , (T , H) , (T , T) }

From the lateral shape , we find that the random variable range expressing the number of appearing the head = { 0 , 1 , 2 }



$f(0) = p(X = 0) = \dfrac{n(x_1)}{n(s)} = \dfrac{1}{4}$

$f(1) = p(X = 1) = \dfrac{n(x_2)}{n(s)} = \dfrac{2}{4}$, $f(2) = p(X = 2) = \dfrac{n(x_3)}{n(s)} = \dfrac{1}{4}$

The probability distribution function is:

x_r	0	1	2
$f(x_r)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Try to solve

3 In the previous example, write down the probability distribution function of the discrete random variable X expressing:(number of appearance of Head - number of appearance of Tail).

Example

Drawing without replacing

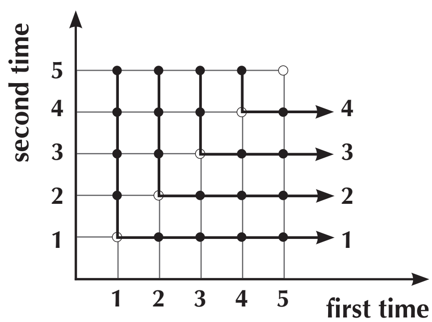
4 A box contains 5 identical cards numbered from 1 to 5. Two cards have been drawn one after another without replacing , find the probability distribution function for the random variable expressing the least number of the two numbers on the two drawn cards..

Solution

As long as the cards have been drawn without replacing to the box, the cards drawn is not repeated more, in other words, the pairs of cards carrying the digits (1 , 1) , (2 , 2) , (3 , 3) , (4 , 4) , (5 , 5) are not included in the sample space as shown in the opposite figure.

n(S) = 20

From the figure opposite, the range of random variable X =: { 1 , 2 , 3 , 4 } and:



$f(1) = p\ (X = 1) = \frac{8}{20}$

$f(2) = p\ (X = 2) = \frac{6}{20}$

$f(3) = p\ (X = 3) = \frac{4}{20}$

$f(4) = p\ (X = 4) = \frac{2}{20}$

the probability distribution function of the discrete random variable X is given as shown in the following table :

x_r	1	2	3	4
$f(x_r)$	$\frac{8}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	$\frac{2}{20}$

Try to solve

4 In an experiment of rolling a regular die two consecutive times and observing the seen face in each time, find the probability distribution function of the discrete random variable expressing the greatest number of the two numbers appeared on the two upper faces.

Example

Using the rule of the function

5 If X is a discrete random variable and its probability distribution function is determined by the relation :

$f(x) = \frac{k + 2x}{24}$ where x = 0 , 1 , 2 , 3 find the value of k, then write down the probability distribution function.

Solution

$\because f(0) = p(X = 0) = \frac{k}{24}$, $f(1) = p(x = 1) = \frac{k+2}{24}$,

$f(2) = p(X = 2) = \frac{k+4}{24}$, $f(3) = p(x = 3) = \frac{k+6}{24}$

$\because p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) = 1$

$\frac{k}{24} + \frac{k+2}{24} + \frac{k+4}{24} + \frac{k+6}{24} = 1$

$\therefore \frac{k + k + 2 + k + 4 + k + 6}{24} = 1$ $\therefore 4k + 12 = 24$

$\therefore 4k = 24 - 12$ $\therefore 4k = 12$ $\therefore k = 3$

To find the probability distribution function , you should find:

$p(X = 0) = \frac{k}{24} = \frac{3}{24}$, $p(X = 1) = \frac{k+2}{24} = \frac{5}{24}$

$p(X = 2) = \frac{k+4}{24} = \frac{7}{24}$, $p(X = 3) = \frac{k+6}{24} = \frac{9}{24}$

$\therefore$ The probability distribution function is:

x_r	0	1	2	3
$f(x_r)$	$\frac{3}{24}$	$\frac{5}{24}$	$\frac{7}{24}$	$\frac{9}{24}$

Try to solve

5 If X is a discrete random variable whose range = { 1 , 2 , 3 } and its probability distribution function is determined by the relation $f(x) = \frac{ax}{9}$. Find the value of a , then write down the probability distribution function.

Exercises 4 - 1

First : choose the correct answer:

1. Which function of the following represents a probability distribution function of the discrete random variable X:

a)

x_r	1	2	3	4
$f(x_r)$	0.06	0.15	0.42	0.26

b)

x_r	0	1	3	5
$f(x_r)$	0.5	0.3	0.4	–0.2

c)

x_r	–2	–1	1	2
$f(x_r)$	0.32	0.14	0.23	0.31

d)

x_r	3	4	5	6
$f(x_r)$	0.23	0.32	0.17	0.18

2. If X is a random variable whose range is {0, 1, 2} then all the following functions do not represent its probability distribution function except :

a) $f(x) = \frac{x^2 + 1}{8}$ b) $f(x) = \frac{2x + 1}{3}$ c) $f(x) = \frac{1}{x + 2}$ d) $f(x) = \frac{3x - 1}{6}$

3. If X is a random variable whose range is {1 , 2 , 3}, $p(X = 1) = 0.3$ and $p(X = 2) = 0.5$ then $p(X = 3)$ is equal:

a) 0.1 b) 0.2 c) 0.7 d) 0.8

4. If X is a random variable whose range is {1 , 2 , - 1 , 0 }, $p(X = - 1) = 0.2$ and $p(X = 0) = 0.4$, $p(X = 1) = 0.1$ then $p(X > 1)$ equal :

a) 0.3 b) 0.4 c) 0.5 d) 0.6

5. in an experiment for tossing a coin three consecutive times and X is the discrete random variable expressing (number of Heads - number of Tails) , then the range of X is :

a) {1 , 3} b) {0 , 1 , 3} c) { 0 , 1 , 2 , 3} d) {-3 , -1 , 1 , 3}

6. If X is a discrete random variable whose range = { 0, 1, 2} and its probability distribution function is determined by the relation $f(x) = \frac{ax}{6}$ then the value of a is equal:

a) $\frac{1}{2}$ b) 1 c) $\frac{3}{2}$ d) 2

Second : Answer the following questions:

7. The two following tables show the probability distribution of a discrete random variable X. Find the value of a in each table:

a)

x_r	1	2	3
$f(x_r)$	a	2a	3a

b)

x_r	-2	–1	0	1	2
$f(x_r)$	a	0.2	0.3	3a	a

c)

x_r	0	1	3	4
$f(x_r)$	a	$2a^2$	$3a^2$	3a

8. If X is a discrete random variable whose range = { 0 , 1 , 2 , 3 } and the values of $p(X = 0) = 0.2$, $p(X = 1) = 0.33$ and $p(X = 2) = 0.37$ find the probability distribution of the discrete random variable X.

9. If the values of the discrete random variable X in a random experiment are: -2 , 0 , 2 , 4 with probabilities of $\frac{m}{5}$, $\frac{m+1}{5}$, $\frac{2m-1}{5}$ and $\frac{3m-2}{5}$ respectively , find the value of m , then write down the probability distribution function of the variable X.

10. If X is a discrete random variable whose probability distribution function is determined by the relation : $f(x) = \frac{2a + 3x}{54}$ and the range of $X = \{1, 2, 3, 4\}$ find the value of a and write down the probability distribution function of the variable X.

11. If X is a discrete random variable whose probability distribution is determined by the function $f(x) = \frac{k + 3x}{50}$ where x = 1 , 2 , 3 , 4 find the value of k, then write down the probability distribution of the variable X.

12. In an experiment for tossing a coin three consecutive times, if the discrete random variable X expresses " number of Heads - number of Tails" , write down the probability distribution of the variable X.

13. Two boxes each contains three numbered balls from 3 to 5. A ball has been randomly drawn from each box and the discrete random variable X is defined as " the sum of two numbers" existed on the two drawn balls. Find the probability distribution of the discrete random variable X.

14. In an experiment for tossing a die two consecutive times and observing the number on the upper face in each time , write down the probability distribution of the discrete random variable X expressing " the least number of the two observed numbers".

15. A box contains 4 balls numbered from 1 to 4, two balls have been drawn one after another (with replacing), write down the probability distribution of the discrete random variable X expressing " the mean of the two numbers on the two drawn balls".

16. If X is a discrete random variable expressing the number of girls in a three- kid family, write down the range of the discrete random variable X . If we suppose that the probability of giving a birth of a boy is equal to the probability of giving a birth of a girl and disregarding having twins, find the probability distribution of the random variable X (the order of boys and girls is taken into acount)

Expectation (mean) and Variance of a Discrete Random Variable

Unit Four

4 - 2

You will learn

- Expectation (Mean)
- Variance
- Standard deviation
- Coefficient of variation

Key terms

- Expectation (Mean)
- Variance
- Coefficient of Variation

Introduction : To identify the characteristics of the probability distribution (identifying the characteristics of the original community or comparing different communities) , the basic (Parametars) are necessarily needed to measure its mean value which the possible values of the discrete random value accumulate around it and is called expectation (mean), There are also other values measuring the dispersion of the discrete random variable values from the value of the mean and is known as variance, Therefore, expectation and variance summarize the most important characteristics of the discrete random variable.

Expectation (Mean)

Expectation is the value at which the most values of the discrete random variable are centralized and sometimes is called " mean" and is denoted by the symbol (μ) and read as (Mu).

If X is a discrete random variable whose probability distribution function is f and its range is: $\{x_1, x_2, x_3, \dots\dots, x_n\}$ by probabilities of $f(x_1), f(x_2), f(x_3), \dots\dots, f(x_n)$ respectively, then the expectation is given by the relation:

$$\text{Expectation } (\mu) = \sum_{r=1}^{n} x_r \times f(x_r)$$

i.e.: Expectation $(\mu) = x_1 \times f(x_1) + x_2 \times f(x_2) + x_3 \times f(x_3) + \dots\dots + x_n \times f(x_n)$

Example

1 If X is a discrete random variable whose probability distribution is shown in the below table:

x_r	-1	0	1	2	3
$f(x_r)$	0.3	0.1	0.1	**a**	0.2

First : Find the value of a **Second :** Find the expectation (Mean)

Solution

First : you know that the sum of probabilities is equal one

$\because p(X = -1) + p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) = 1$

$\therefore 0.3 + 0.1 + 0.1 + \mathbf{a} + 0.2 = 1$

$\therefore \mathbf{a} + 0.7 = 1$ $\quad\quad \therefore \mathbf{a} = 1 - 0.7 = 0.3$

Materials

- Scientific calculator - Computer graph programs.

Second :

$\because$ Expectation (μ) = $\sum_{r=1}^{n} x_r \times f(x_r) = -1 \times 0.3 + 0 \times 0.1 + 1 \times 0.1 + 2 \times 0.3 + 3 \times 0.2$

$= -0.3 + 0 + 0.1 + 0.6 + 0.6 = 1$

Try to solve

1. If X is a discrete random variable whose range = {0 , 1 , 2 , 3 , 4} ,
   $p(X = 0) = p(X = 4) = \frac{1}{16}$, $p(X = 1) = p(X = 3) = \frac{1}{4}$
   Find: **first:** p (X = 2) **second:** Expectation

Example

2. If x is a discrete random variable whose probability distribution is shown as follows:

x_r	0	1	2	b	6
$f(x_r)$	0.1	0.1	0.3	**a**	0.3

Find the value of a and b if the expectation $\mu = 3.5$

Solution

From the properties of the probability distribution: f(0) + f(1) + f(2) + f(b) + f(6) = 1

$\therefore 0.1 + 0.1 + 0.3 + \mathbf{a} + 0.3 = 1$ $\therefore \mathbf{a} = 1 - 0.8$ **a = 0.2**

$\because$ expectation (μ) = $\sum_{r=1}^{n} x_r \times f(x_r) = 3.5$

$\therefore 0 \times 0.1 + 1 \times 0.1 + 2 \times 0.3 + b \times 0.2 + 6 \times 0.3 = 3.5$

$\therefore 0 + 0.1 + 0.6 + 0.2\,b + 1.8 = 3.5$ $\therefore 0.2\,b = 3.5 - 2.5$

$\therefore b = 1 \div 0.2 = 5$ **b = 5**

Try to solve

2. If X is a discrete random variable whose probability distribution is shown as follows:

x_r	0	2	3	4
$f(x_r)$	$\frac{3}{16}$	2 L	$\frac{1}{16}$	L

First : find the value of L. **Second :** find the expectation

Variance

The variance of a discrete random variable measures the amount of dispersion of the discrete random variable of its expected value and is denoted by the symbol (σ^2) and read as (sigma squared) and is given by the relation:

$$\sigma^2 = \sum_{r=1}^{n} x_r^2 \times f(x_r) - \mu^2$$

Note: The standard deviation of a discrete random variable X is the square root of the variance and is denoted by the symbol σ . It's noticed that the variance and the standard deviation are always positive quantities .

Example

3 If X is a discrete random variable whose probability distribution function is $f(x) = \frac{x+4}{16}$ where x = -2 , m , 1 , 2 find the value of m, then find the mean and the variance of the discrete random variable X.

Solution

From the properties of the probability distribution:

$\because p(X = -2) + p(X = m) + p(X = 1) + p(X = 2) = 1$

$\therefore \frac{2}{16} + \frac{m+4}{16} + \frac{5}{16} + \frac{6}{16} = 1$

$\therefore \frac{17+m}{16} = 1$ $\qquad \therefore 17 + m = 16 \qquad \therefore m = -1$

x_r	$f(x_r)$	$x_r \cdot f(x_r)$	$x^2_r \cdot f(x_r)$
-2	$\frac{2}{16}$	$\frac{-4}{16}$	$\frac{8}{16}$
-1	$\frac{3}{16}$	$\frac{-3}{16}$	$\frac{3}{16}$
1	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{16}$
2	$\frac{6}{16}$	$\frac{12}{16}$	$\frac{24}{16}$
		$\frac{5}{8}$	$\frac{5}{2}$

The expectation $(\mu) = \sum_{r=1}^{n} x_r \times f(x_r) = \frac{5}{8}$

The variance $(\sigma^2) = \sum_{r=1}^{n} x_r^2 \times f(x_r) - \mu^2 = \frac{5}{2} - (\frac{5}{8})^2 = \frac{135}{64}$

Try to solve

3 If X is a discrete random variable whose probability distribution is defined by the function $f(x) = \frac{a}{x+1}$ where x = 0 , 1 , 2 , 3 find:

First: the value of a

Second: The expectation and the standard deviation of the discrete random variable x.

Coefficient of Variation

As you were learning the standard deviation as a measure for the dispersion of the values of the discrete random variable of its expectation, you knew that it is measured by the same measuring units of the variable investigated whether these units are in degrees, meters or kilograms... i.e. it is proper to compare two sets of the same measuring units and the same means. If the measuring units or means are different between the two sets, the standard deviation is no longer proper to be used as a measure for the comparison, As a result, the need of a relative measure for dispersion can help us get rid of such different units. The coefficient of variation represents a proper solution to such a problem.

The coefficient of variation for any set of items is known as the percentage between the standard deviation of the set and its expectation (mean) and it can be defined as in the following relation :

$$\textbf{Coefficient of variation} = \frac{\textbf{standard deviation}}{\textbf{mean}} \times 100\ \% = \frac{\sigma}{\mu} \times 100\%$$

This coefficient forms the dispersion of the set in the form of a percentage abstracted from distinction where the measured units are not affected by the phenomenon.

Example

4 If the expectation and the standard deviation for the marks of a group of students in history and Geography are as follows , known that the full mark is 100 marks.

Subject	history exam	geography exam
Expectation	70	96
Standard deviation	7	8

Find the coefficient of variation for each subject - What do you notice?

Solution

$\because$ Coefficient of variation $= \dfrac{\text{standard deviation}}{\text{mean}} \times 100\ \%$

$\therefore$ The coefficient of variation of history $= \dfrac{7}{70} \times 100\ \% = 10\ \%$,

The coefficient of variation of geography $= \dfrac{8}{96} \times 100\ \% \simeq 8.3\ \%$

From the Solution: we observe that the relative dispersion of the history exam is greater than of geography. This means the geography exam is more homogeneous than the history exam.

Try to solve

4 If a factory produces two types of electric lamps A, B and their average of use is 1850, 1580 hours and the standard deviation of both is 250, 230 respectively, find the coefficient of variation of each type. What do you notice.

Example

5 A box contains 6 cards; 2 cards carry the number two, 3 cards carry the number 3 and a card carries the number 11, if a card has been randomly drawn and the discrete random variable X is defined as " the appeared number on the drawn card"

Find:

a The probability distribution function of the variable X.

b Expectation and standard deviation of the variable X c the coefficient of variation.

Solution

a x takes the values 2 ، 3 ، 11 where: $f(2) = p\,(X = 2) = \frac{2}{6} = \frac{1}{3}$

$f(3) = p\,(X = 3) = \frac{3}{6} = \frac{1}{2}$ ، $f\,(11) = p(X = 11) = \frac{1}{6}$

the table below illustrated the probability distribution function of the discrete random variable X.

x_r	2	3	11
$f(x_r)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

To calculate the expectation and the standard deviation, you ought to form the table below:

x_r	$f(x_r)$	$x_r \cdot f(x_r)$	$x^2_r \cdot f(x_r)$
2	$\frac{2}{6}$	$\frac{4}{6}$	$\frac{8}{6}$
3	$\frac{3}{6}$	$\frac{9}{6}$	$\frac{27}{6}$
11	$\frac{1}{6}$	$\frac{11}{6}$	$\frac{121}{6}$
Total		4	26

b Expectation $(\mu) = \sum_{r=1}^{n} x_r \times f(x_r) = 4$

Variance $(\sigma^2) = \sum_{r=1}^{n} x^2\, f(x_r) - \mu^2 = 26 - (4)^2 = 10$

standard deviation $\sigma = \sqrt{10} = 3.16$

c $\because$ Coefficient of variation $= \frac{\text{standard deviation}}{\text{mean}} \times 100\ \%$

$\therefore$ Coefficient of variation $= \frac{3.16}{4} \times 100\ \% = 79\ \%$

Try to solve

5 A box contains 10 cards ; a card carries the number 1 , two cards each carries the number 2, 3 cards each carries the number 3, 4 cards each carries the number 4. If a card is randomly drawn and the discrete random variable X expresses the number on the drawn card, find the probability distribution function of this variable and calculate each of the expectation , its standard deviation and the coefficient of variation.

Exercises 4 - 2

First : Choose the correct answer:

1. If the probability distribution of the discrete random variable X is {(0 , 0.25) , (1 , 0.5) , (2 , 0.25) } then its expectation equals:

   a 0.5 b 1 c 1.25 d 1.5

2. If X is a discrete random variable and the expectation equals 0.6 , $\sum_{r=1}^{n} x_r^2 \times f(x_r) = 4.36$ then its standard deviation equals:

   a 1.94 b 2 c 3.76 d 4

3. If X is a discrete random variable and the expectation equals 0.4 , $\sum_{r=1}^{n} x_r^2 \times f(x_r) = 6.16$ then its variance equals :

   a 2.4 b 5.76 c 6 d 6.56

Second : Find the expectation and the standard deviation for the probability distribution for each of the following:

4.

x_r	2	3	9
$f(x_r)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

5.

x_r	–5	–4	1	2
$f(x_r)$	$\frac{1}{24}$	$\frac{3}{8}$	$\frac{5}{12}$	$\frac{1}{6}$

6.

x_r	–3	–1	0	1	2	3
$f(x_r)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$

Third : Answer the following questions:

7. If X is a discrete random variable whose probability distribution is shown in the table below:

x_r	1	2	4	6
$f(x_r)$	0.2	0.3	a	0.1

First : find the value of a. Second : find the mean and the standard deviation

8. If the range of the discrete random variable X is { 1 , 2 , 3 , 4 }, $p(X = 1) = \frac{4}{25}$, $p(X = 2) = \frac{7}{25}$ and $p(X = 4) = \frac{1}{5}$, find the expectation and variance of X .

9. If X is a discrete random variable whose range { 0 , 1 , 2 , 3 , 4 } , $p(X = 0) = p(X = 4) = \frac{1}{16}$ and $p(X = 1) = p(X = 3) = \frac{1}{4}$ Find: first : p (X = 2) second: the mean and the variance of the variable X .

10. If X is a discrete random variable whose probability distribution function is shown in the table below where $0 < h < 1$

x_r	–3	zero	3	6
$f(x_r)$	h	h^2	$2h^2$	h

Find:

a the value of h

b The probability distribution of the variable X.

c The mean and variance of the variable X

11 If X is a discrete random variable whose probability distribution is shown in the table below:

x_r	1	2	4	a
$f(x_r)$	0.2	0.3	0.4	0.1

Calculate the value of a if the expectation $\mu = 3$, then find the standard variation of the discrete random variable X

12 If the probability distribution of a discrete random variable X is defined by the function f where : $f(x) = \frac{a\,x}{9}$, where x = 1, 2, 3 . find:

a the value of a b calculate the expectation and the variance of the variable X.

13 If X is a discrete random variable whose probability distribution is defined by the function: $f(x) = \frac{x^2 + 1}{a}$ where x = 0, 1 , 2 , 3

find: a the value of a b calculate the coefficient of variation of the variable X.

14 If X is a discrete random variable whose probability distribution is defined by the function: $f(x) = \frac{x + 4}{16}$ where x = - 2, m, 1, 2

find: a the value of m b the mean and variance of the variable x.

15 If X is a discrete random variable whose probability distribution is defined by the function f where: $f(x) = \frac{a}{x + 3}$, x = 0 , 1 , 2 , 3

a find the value of a b the expectation and variance.

16 If the range of the discrete random variable X is { -1 , 0 , 2 } , $p(X = -1) = \frac{1}{4}$ and the expectation equals 1 , find:

a $p(X = 0)$, $p(X = 2)$ b The coefficient of variation.

17 If x is a discrete random variable whose mean $\mu = 3$ and its probability distribution is as follows :

x_r	0	2	k	4
$f(x_r)$	1	2a	$\frac{1}{4}$	5a

a Calculate the values of a and k

b Find the standard deviation of the variable X.

Unit Four

4 - 3

Geometric and Binomial Distributions

You will learn

- geometric probability experiment
- Binomial distribution
- Probability distribution of a binomial random variable

Key terms

- expectation , variance, and standard deviation of a geometric distribution

Bernoulli experiment

It is a random experiment that has only one of two outcomes, one of them is expressed as success, and the other is expressed as failure . For example, the experiment of tossing a coin once and noticing the heads up represents Bernoulli's experiment

Because it has one of two results: a head or a tail. In this experiment , the head is success, and the tail is failure, or vice versa

Another example: When rolling a dice whose faces are numbered: $\{1, 2, 3, 4, 5, 6\}$, this experiment can be considered Bernoulli's experiment is based on the fact that the appearance of a number greater than 3 is success, and any other number is failure

Geometric probability experiment

The repetition of Bernoulli's experiment a number of independent times until the first success is achieved is called Geometric probability experiment

Conditions for the geometric probability experiment

If the following four conditions are satisfied in a random experiment, it is considered a geometric probability experiment

(1) The experiment includes repeated and mutually exclusive trails

(2) Each trail has two independent outcomes (success or failure)

(3) The probability of success in each trail is constant

(4) Stop at the first success

Geometric random variable

In a probability experiment, if the random variable X symbolizes reaching the first success, then x is called a geometric random variable, and it will be symbolized by the symbol $X \sim G(S)$ to indicate that X is a geometric variable and S represents the probability of success.

Materials

- Scientific calculator, computer graph programs.

Geometric probability distribution function

If $X \sim G(S)$ then

$$P(X = n) = S(1 - S)^{n-1} \qquad n = 1, 2, 3, \ldots$$

Where S is the probability of success, n is the number of trails until reach to the first success

Example

1 Ahmed tossed a coin and the success was the appearance of a head. What is the probability of the head appearing on the fourth trail?

Solution

Assuming that X is a random variable that symbolizes the first successful trail,

then $X \sim G(\frac{1}{2})$

$P(\text{Head}) = \frac{1}{2}$, $n = 4$

$P(X = 4) = \frac{1}{2}(1 - \frac{1}{2})^{4-1} = \frac{1}{2} \times (\frac{1}{2})^3 = (\frac{1}{2})^4 = \frac{1}{16}$

Example

2 If $X \sim G(0.4)$ find each of the following

a $P(X = 2)$ b $P(X > 4)$ c $P(X = 6)$

Solution

a $P(X = 2) = S(1 - S)^{n-1} = 0.4 \times (1 - 0.4)^{2-1} = 0.24$

b $P(X > 4) = 1 - P(X \leq 4)$

$= 1 - [P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$

$= 1 - [0.4(1 - 0.4)^0 + 0.4(1 - 0.4)^1 + 0.4(1 - 0.4)^2 + 0.4(1 - 0.4)^3]$

$= 1 - [0.4 + 0.4 \times 0.6^1 + 0.4 \times (0.6)^2 + 0.4 \times (0.6)^3]$

$= 0.1296$

c $P(X = 6) = S(1 - S)^{n-1} = 0.4 \times (1 - 0.4)^{6-1} = 0.031104$

Try to solve

1 If $X \sim G(0.8)$ find each of the following

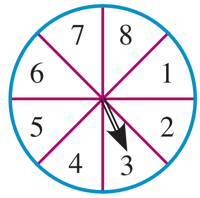
a $P(X = 3)$ b $P(X \leq 3)$ c $P(X > 3)$

d $P(X < 2)$ e $P(X \leq 2)$ f $P(2 < X \leq 4)$

Example

3 A spinner disk has eight equal sections numbered from 1 to 8. If the disk is rolled several times, find the probability that it will take more than four times for its pointer to indicate a prime number for the first time.

Solution



Assuming that $X \sim G(S)$

Prime numbers are 2, 3, 5, 7 $\quad \therefore S = \frac{4}{8} = 0.5$

$P(X > 4) = 1 - P(X \leq 4)$

$= 1 - [P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$

$= 1 - [0.5(1 - 0.5)^0 + 0.5(1 - 0.5)^1 + 0.5(1 - 0.5)^2 + 0.5(1 - 0.5)^3]$

$= 1 - [0.5 + 0.5 \times 0.5^1 + 0.5 \times (0.5)^2 + 0.5 \times (0.5)^3]$

$= 0.0625$

Another solution :

Calculating the probability that it will take more than four cycles to see a prime number for the first time, this means that we fail to get a prime number in each of the first four cycles. The probability is calculated as

$P(x > 4) = (0.5)^4 = 0.0625$

Expectation and variance of geometric distribution

Mean (expectation)	μ	$= \frac{1}{S}$
Variance	σ^2	$= \frac{1 - S}{S^2}$
Standard deviation	σ	= the positive square root of the variance

Try to solve

1. In example 3 calculate the expectation and the standard deviation

Binomial distribution

Repeating a Bernoulli experiment a specific number of independent times is called a probability experiment

Binomial probability experiment

If the following four conditions are satisfied in a randomized trial, it is considered a binomial probability trial:

(1) The experiment includes repeated and independent trails

(2) each trail has two mutually exclusive outcomes (success or failure)

(3) The probability of success in each trail is constant

(4) There is a specific number of trails in the experiment.

Note : we denote the distribution of a binomial random variable by If $X \sim B(n,S)$

Probability distribution of a binomial random variable

If $X \sim B(n, S)$ then

$P(X = r) = {}^{n}C_{r} \times S^{r} \times (1 - S)^{n - r}, \quad r = 1, 2, 3, \ldots n$

n: number of trails in the experiment

S:probability of success in each trail

r :is the number of trails (required number)

For example: tossing 7 regular coins and then writing the number of heads appeared on the tupper face (binomial experiment). Because it satisfies the previous four conditions

Example

4 In the experiment of tossing a regular coin 15 times, the random variable expressed the number of heads. Find the probability that the head appears 5 times.

Solution

Let $X \sim B(15, \frac{1}{2})$

$P(X = r) = {}^{n}C_r \times S^r \times (1 - S)^{n - r}$

$n = 15 \quad , r = 5 \quad , S = \frac{1}{2}$

$P(X = 5) = {}^{15}C_5 \times (\frac{1}{2})^5 \times (1 - \frac{1}{2})^{15 - 5} \simeq 0.0916$ approximately

Example

5 A statistics test consists of 50 questions, all of them are multiple choice questions , each one has with 4 choises, only one is a correct choise . If all of these questions are answered randomly, what is the probability that the answers to only 10 questions are correct?

Solution

Let $X \sim B(50, \frac{1}{4})$

$P(X = r) = {}^{n}C_r \times S^r \times (1 - S)^{n - r}$

$n = 50 \quad , \quad r = 10 \quad , \quad S = \frac{1}{4}$

$P(X = 10) = {}^{50}C_{10} \times (\frac{1}{4})^{10} \times (1 - \frac{1}{4})^{50 - 10} = 0.09852$ approximately

Example

6 If the probability of a team to win a football match is 0.6, If the team plays 7 matches, find

a The probability of winning only 4 matches

b The probability of winning at least 6 matches

c The probability of winning two matches at most

Solution

Let $X \sim B(7, 0.6)$

a $P(X = 4) = {}^{7}C_4 \times (0.6)^4 \times (1 - 0.6)^3 = 0.290304$

b $P(X \geq 6) = P(X = 6) + P(X = 7)$

$= {}^{7}C_6 \times (0.6)^6 \times (1 - 0.6)^1 + {}^{7}C_7 \times (0.6)^7 \times (1 - 0.6)^0 = 0.1586$

c $(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$= {}^{7}C_0 \times (0.6)^0 \times (0.4)^7 + {}^{7}C_1 \times (0.6)^1 \times (0.4)^6 + {}^{7}C_2 \times (0.6)^2 \times (0.4)^5$

$= 0.096256$

Example

(7) If X is a binomial random variable $X \sim B(3,S)$, $P(X \geq 1) = \frac{19}{27}$ find , $P(X = 2)$

Solution

$1 - P(X = 0) = \frac{19}{27}$,

$1 - {}^3C_0 \times (S)^0 \times (1 - S)^3 = \frac{19}{27}$ $\qquad \therefore (1 - S)^3 = \frac{8}{27}$

$\therefore 1 - S = \frac{2}{3}$ $\qquad \therefore S = \frac{1}{3}$

$P(X = 2) = {}^3C_2 \times (\frac{1}{3})^2 \times (\frac{2}{3})^1 = \frac{2}{9}$

Expectation and variance of binomial distribution

If X is a random variable $X \sim B(n, S)$

> Mean (expectation) $\mu = n \times S$
>
> Variance $\sigma^2 = n \times S \times (1 - S)$
>
> Standard deviation σ = the positive square root of the variance
>
> Where n is the number of trails in the experiment, S is the probability of success in each trail

Example

(8) life : A study was conducted on the side effects seen in children after taking a new medicine. The study concluded that 10% of the children who took this medicine showed side effects. If a doctor gave this medicine to 150 children, how many children would be expected to show this side effects ?

Solution

$X \sim B(150, 0.1)$

Expectation $= n \times S = 150 \times 0.1 = 15$

So, it is expected that side effects of the new medicine will appear on 15 children

Example

(9) Ahmed tossed an irregular coin 200 times, so the number of times the tails appeared was 140 times. If Ahmed tossed the coin 20 again, and find each of the following

(a) The expected number of times the tails will appear when Ahmed throws the coin is 20 times.

(b) The variance of times the tails appeared when Ahmed threw the coin varied 20 times

Solution

$S = \frac{140}{200} = 0.7$ $\qquad \therefore$ expectation $= n \times S = 20 \times 0.7 = 14$

Variance $(\sigma^2) = n \times S \times (1 - S) = 20 \times 0.7 \times 0.3 = 4.2$

Exercises 4 - 3

1. If the probability of success in a single trial is 0.3, then the probability that the first success occurs on the third trail equals

   a) 0.147 b) 0.21 c) 0.343 d) 0.09

2. If the probability of failure in a certain trial is 0.8, then the expected number of trails before the first success is...

   a) 3 b) 4 c) 5 d) 6

3. The expected value (mean) of a geometric distribution with success probability of $\frac{1}{4}$ is..

   a) 3 b) 4 c) 5 d) 6

4. If the probability of success in a single trial is 0.2, then the probability that it will take more than four trails to see the first success equals ……

   a) 0.4096 b) 0.4915 c) 0.5904 d) 0.6723

5. The variance of a geometric distribution with a probability of success of 0.4 is equal to...

   a) 0.25 b) 1.25 c) 3.75 d) 2.75

6. If the probability of success on one trial is 0.25, then the probability that first success will occur before or on the third trail equals …..

   a) $\frac{15}{64}$ b) $\frac{37}{64}$ c) $\frac{7}{16}$ d) $\frac{69}{64}$

7. If the probability of success on one trial is 0.2, then the probability that first success will occur after 3 failure trails equals

   a) 0.1024 b) 0.251 c) 0.512 d) 0.215

8. If the probability of success of one trial is S = 0.4 and the number of trials is n = 10 then the probability of getting 4 successes is equal to...

   a) 0.2508 b) 0.4 c) 0.0537 d) 0.0124

9. If the probability of success of one trial is S = 0.5 and the number of trials is n = 5 then the probability of getting of 3 successes at least is equal to...

   a) 0.5 b) 0.1825 c) 0.15625 d) 0.84375

10. If the probability of success of one trial is S = 0.3 and the number of trials is n = 7 then the probability of getting no success occurring is approximately equal to...

    a) 0.001 b) 0.2187 c) 0.5041 d) 0.082

11. If the probability of success of one trial is n = 0.75 and the number of trials is n = 12 then the probability of getting 11 or more successes is equal……

    a) 0.1584 b) 0.1454 c) 0.1234 d) 0.2668

12. If the probability of success of one trial is S = 0.7 and the number of trials is n = 10 the probability getting exactly 4 successes equals ....

a) 0.0368 b) 0.2001 c) 0.4787 d) 0.2668

13. If $X \sim B(5, \frac{2}{3})$, then P(X = 4) equals ....

a) $\frac{80}{81}$ b) $\frac{10}{243}$ c) $\frac{80}{243}$ d) $\frac{16}{243}$

14. If X~B(n,S) , expectation equals 8 and variance = $\frac{20}{3}$ then the value of n = ...

a) 48 b) 56 c) 64 d) 32

15. In the experiment of throwing a regular coin on the ground 4 times, the probability of a heads appearing only 3 times it is equal to ....

a) $\frac{1}{16}$ b) $\frac{1}{2}$ c) $\frac{1}{8}$ d) $\frac{1}{4}$

16. Gana rolled an irregular dice 100 times and the number of times the number 2 appears is 10 times. If Gana rolled the dice 30 times, then the expected number of times the number 2 would appear is equal to ....

a) 2 b) 3 c) 6 d) 9

17. A calculator contains 16 buttons for numbers from 0 to 9, in addition to the basic operations, the equal sign, and the decimal point. If Ahmed closes his eyes and then presses the buttons on this calculator 20 times randomly, then the probability that he will press the arithmetic operations buttons only 3 times .... Approximately

a) 0.134 b) 0.139 c) 0.239 d) 0.245

18. If a player wins 75% of the matches, he plays during his sports career, then the probability that he win 3 matches out of the next 5 matches is equal to ....

a) $\frac{135}{512}$ b) $\frac{45}{512}$ c) $\frac{5}{1024}$ d) $\frac{47}{512}$

19. If the probability of success of a surgery is 90%, then the probability of success of at least one operation if the operation is performed three times, it is ....

a) 0.001 b) 0.1 c) 0.9 d) 0.999

20. Mona took a ten-question multiple-choice test, each questions has four alternatives, if Mona answers randomly, then probability that Mona will get 7 questions correct is equal to ....

a) 0.00308 b) 0.25 c) 0.0308 d) 0.0307

Secand: Answer the following questions.

1. A production company that manufactures electronic parts. The probability that the part is defective is 0.05. The company inspects 20 pieces of its production. randomly What is the probability that exactly 2 pieces will be defective?

2 In a customer service center, If the probability that a customer's problem will be resolved on the first call is 0.2. What is the probability that the problem will be resolved on the third call?

3 The probability that a person will accept a telemarketing offer is 0.1. What is the probability that the first person on the fifth call will agree?

4 The probability that the order will be delivered on time is 0.9. If 12 orders are delivered, what is the probability that 10 of them will be delivered on time?

5 Appliance repair shop, the probability of successfully repairing a particular appliance is 0.85. If 15 appliances are repaired, what is the probability that 13 of them will be successfully repaired?

6 The probability that a given email message is unwanted (spam) is 0.2. If you receive 25 emails, what is the probability that 5 of them are spam?

7 The probability that a particular seed will grow after being planted is 0.7. If a farmer plants 30 seeds, what is the probability that 20 of them will grow?

8 The probability that a given voter will vote for a given candidate is 0.6. If 10 voters are selected at random, what is the probability that at least 8 of them will vote for this candidate?

9 The probability that a person will donate to a particular campaign is 0.1. If 100 people are contacted, what is the probability that 2 at most of them will donate?

10 The probability that a person will find a parking spot on his first trail is 0.3. What is the probability that he will find the position on his fourth trail?

11 The probability that the technician will succeed in repairing the machine on the first trail is 0.6. What is the probability that the repair will be completed successfully on the second trail?

12 The probability that a customer will agree to a particular sales offer is 0.15. What is the probability that the first customer on the fourth call will agree?

13 The probability of discovering a malfunction in a particular device when inspected is 0.1. What is the probability of detecting the malfunction in the second inspection?

14 The probability that a company will obtain regulatory approval on the first trail is 0.3. What is the probability of getting approved at most on the third trail?

Unit Four

4 - 4 Probability density function of the continuous random variable

You will learn

- Probability density function

Key terms

- *Probability Density*

Continuous Random Variable

Definition

The continuous random variable : its range is an interval of the real numbers (closed or opened) . i.e. it's a non - counted set of real numbers .

For examples:

Ò The wage of a government worker randomly chosen.

Ò The temperature expected on a day.

Ò The length of a basketball candidate.

Example **Continuous random variable**

1 The point (x , y) is located inside or on the circle $x^2 + y^2 = 4$ whose center is the origin O and its radius length is 2 units. Required is to find the range of the random variable X expressing how distant the point from the center of the circle

Solution

$\because f = \{ (x , y): x^2 + y^2 \leqslant 4\}$

$\therefore 0 \leqslant a \leqslant 2$ where a is the distance between from point (x , y) to the center of the circle.

$\therefore$ the range of the random variable X = [0 , 2]

we can notice that each point in this interval is a possible value to the random variable X as shown in the figure

Try to solve

1 If the maximal life span of a type of cellular phones (X) is approximated by 18 working hours, write down the range of X.

Try to solve

2 Show which of the following represents a discrete random variable and which represents a continuous random variable.

a Number of the bread leaves produced by a bakery within an hour.

b The time kareem takes to wait his friend Ziad.

c Number of the goals scored by the team won the handball match.

d The number of the traffic violations recorded on Cairo - Alex desert rood on a day.

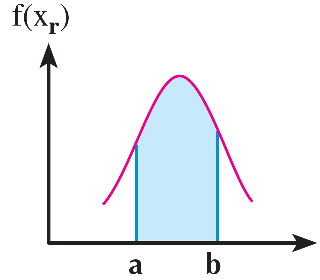
e The time taken by a teacher to explain the lesson of the random variable.

Materials

- Scientific calculator, computer graph programs.

Probability Density Function

For any continuous random variable X, there is a real function whose range is non- negative, and denoted by the symbol f(x) called the probability density function by which the probabilities of events expressing it by the random variable can be found throughout the area included beneath the function curve and on the x axis and p $(a<x<b)$ can be calculated by finding the area of the shaded part of the curve of the function f between the two values a and b as shown in the opposite figure .



This function satisfies the next conditions:

- $f(x) \geqslant 0$ for all the values of x belonging to the domain of the function.
- The area of the region located beneath the curve of the function f and on the x-axis equal 1.

Example

2 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{1}{6}(2x - 1), & 1 \leqslant x \leqslant 3 \\ \text{zero}, & \text{otherwise} \end{cases}$$

a **Prove that :** $p(1 < X < 3) = 1$

b **Find :** $p(X \leqslant 2)$, $p(X > 2.5)$, $p(2 \leqslant X \leqslant 2.5)$.

Solution

Remember

Area of the rectangle = length × width
Area of the triangle = $\frac{1}{2}$ **base length × height**
Area of the trapezoid = $\frac{1}{2}$ **the sum of the two parallel bases × height**

$$f(1) = \frac{1}{6} \times (2 - 1) = \frac{1}{6}$$

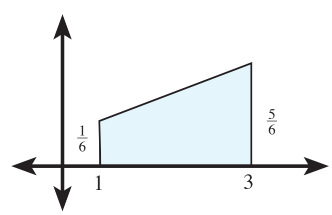
$$f(3) = \frac{1}{6} \times (6 - 1) = \frac{5}{6}$$

$$f(2) = \frac{1}{6} \times (4 - 1) = \frac{3}{6}$$

$$f(2.5) = \frac{1}{6} \times (5 - 1) = \frac{4}{6}$$

a $p(1 \leqslant X \leqslant 3) = \frac{1}{2}\left(\frac{1}{6} + \frac{5}{6}\right) \times 2$

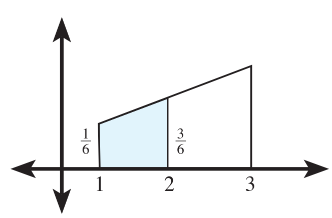
$= \frac{1}{2} \times \frac{6}{6} \times 2 = 1$



b $p(X \leqslant 2) = p(1 \leqslant X \leqslant 2)$

$= \frac{1}{2}\left(\frac{1}{6} + \frac{3}{6}\right) \times 1$

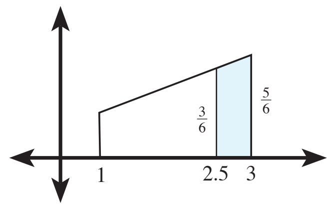
$= \frac{1}{2} \times \frac{4}{6} = \frac{4}{12} = \frac{1}{3}$



, $p(X > 2.5) = p(2.5 < X \leqslant 3)$

$$= \frac{1}{2}\left(\frac{4}{6} + \frac{5}{6}\right) \times \frac{1}{2}$$

$$= \frac{1}{2} \times \frac{9}{6} \times \frac{1}{2} = \frac{9}{24} = \frac{3}{8}$$

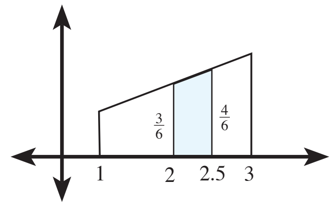


, $p(2 \leqslant X \leqslant 2.5) = \frac{1}{2}\left(\frac{3}{6} + \frac{4}{6}\right) \times \frac{1}{2}$

$$= \frac{1}{2} \times \frac{7}{6} \times \frac{1}{2} = \frac{7}{24}$$

Notice that : $p(2 \leqslant X \leqslant 2.5) = 1 - [\, p(X \leqslant 2) + p(X \geqslant 2.5)\,]$

$$= 1 - \left(\frac{1}{3} + \frac{3}{8}\right) = 1 - \frac{17}{24} = \frac{7}{24}$$



Try to solve

3 If X is a continuous random variable where :

$$f(X) = \begin{cases} \frac{1}{50}(17 - 2x) & \text{where } 1 < x < 6 \\ \text{zero} & \text{otherwise} \end{cases}$$

a **Prove that** f(X) is a probability density function of the random variable X .

b **Find** $p(X > 3)$

c **Find** $p(4 < X < 7)$

Example

3 If X is a continuous random variable whose probability density function is:

$$f(X) = \begin{cases} \dfrac{2x + k}{24} & \text{where } 1 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$

a Find the value of k.

b find $p(X > 3)$

Solution

$\because p(1 < X < 4) = 1$

$\therefore \frac{1}{2} \times 3 \times \frac{10 + 2k}{24} = 1$

$\therefore \frac{1}{2}\left(\frac{2 + k}{24} + \frac{8 + k}{24}\right) \times 3 = 1$

$\therefore k = 3$

, $f(3) = \frac{6 + 3}{24} = \frac{9}{24}$

$f(4) = \frac{8 + 3}{24} = \frac{11}{24}$,

$\therefore P(X > 3) = \frac{1}{6}\left(\frac{9}{24} + \frac{11}{24}\right) \times 1 = \frac{1}{6} \times \frac{20}{24} = \frac{5}{12}$

Try to solve

4 If X is a continuous random variable whose probability density function is :

$$f(X) = \begin{cases} \dfrac{2x + 1}{28} & 1 < x < 5 \\ \text{zero} & \text{otherwise} \end{cases}$$

a Find the value of a if $p(X < a) = \frac{1}{7}$

b Find the value of b if $p(b < X < b + 2) = \frac{1}{2}$

Exercises (4 - 4)

First : choose the correct answer :

1. If the probability distribution of the continuous random variable X is :

$$f(x) = \begin{cases} \frac{1}{2} & \text{where } 2 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$ then p (X > 3) =

a) $\frac{1}{4}$ b) $\frac{1}{2}$ c) $\frac{3}{4}$ d) 1

2. If the probability distribution of the continuous random variable X is :

$$f(x) = \begin{cases} k\,x & \text{where } 2 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$ then k =

a) $\frac{1}{6}$ b) $\frac{1}{3}$ c) $\frac{1}{2}$ d) $\frac{3}{4}$

3. If the probability distribution of the continuous random variable X is:

$$f(x) = \begin{cases} \frac{1}{6} & \text{where } -3 < x < 3 \\ \text{zero} & \text{otherwise} \end{cases}$$ then p (X = 3) =

a) zero b) $\frac{1}{6}$ c) $\frac{1}{3}$ d) $\frac{1}{2}$

Second : Answer the following questions :

4. If X is a continuous random variable where

$$f(x) = \begin{cases} \frac{x+3}{18} & \text{where } -3 < x < 3 \\ \text{zero} & \text{otherwise} \end{cases}$$

Find : **first**: p (X < 0) **second** : p (-1 < X < 2)

5. If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{2x+1}{24} & \text{where } 2 < x < 5 \\ \text{zero} & \text{otherwise} \end{cases}$$

find : **first** : p (3 < X < 5) **second** : p (X > 4)

6. If X is a continuous random variable where :

$$f(x) = \begin{cases} \frac{2(x+1)}{27} & \text{where } 2 < x < 5 \\ \text{zero} & \text{otherwise} \end{cases}$$

first : prove that f(x) is a density function to the random variable X.

second : Find p (X > 3)

7 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{2x+1}{18} & \text{where } 1 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$

find : **first** : p (X > 3) **second** : p (2 < X < 4)

8 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} ax & \text{where } 0 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$

find : **first** : the value of a **second** : p (1 < X < 3)

9 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{1}{8}x + \mathbf{a} & \text{where } 0 < x < 4 \\ \text{zero} & , \text{otherwise} \end{cases}$$

find : **first** : the value of a **second** : p (1 < X < 3)

10 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{a\,x}{2} & \text{where } 0 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$

find : **first** : the value of a **second** : p (1 < X < 3)

11 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{x-1}{k} & \text{where } 1 < x < 5 \\ \text{zero} & \text{otherwise} \end{cases}$$

find : **first** : the value of k **second** : p (2 < X < 3)

Critical thinking :

12 If X is a continuous random variable whose probability density function is:

$$f(x) = \begin{cases} \frac{x}{6} & \text{where } 0 < x < 2 \\ \frac{1}{3} & \text{where } 2 < x < 4 \\ \text{zero} & \text{otherwise} \end{cases}$$

Calculate : a $p(1 < x < 2)$ b the value of a making $p(2 < X < \mathbf{a}) = 0.5$

13 If X is a continuous random variable whose probability density function is:

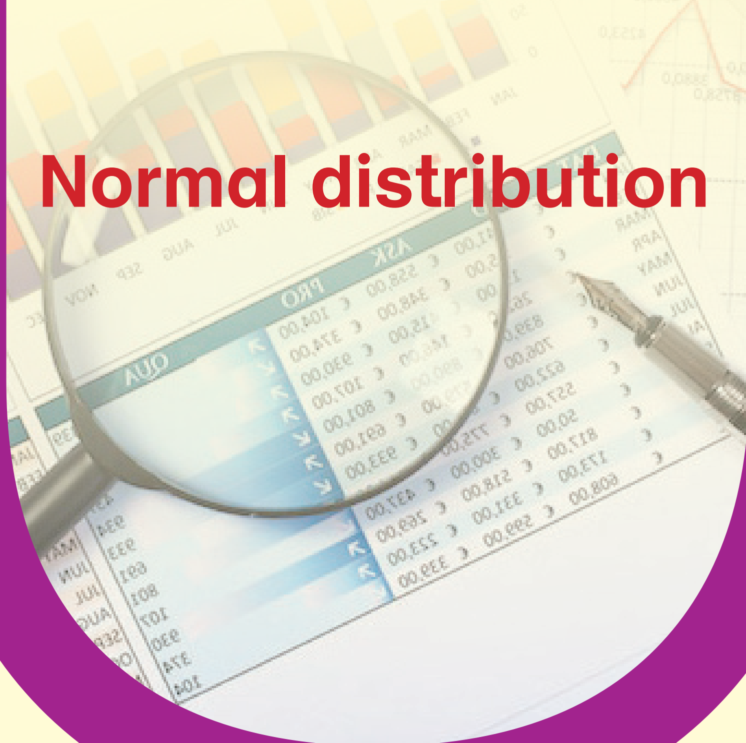
$$f(x) = \begin{cases} \frac{3x+1}{40} & \text{where } 1 \leqslant x \leqslant 5 \\ \text{zero} & \text{otherwise} \end{cases}$$

and if **a** b $\in [1, 5[$ find

a the value of a if $p(a < X < a + 2) = \frac{7}{20}$ b the value of b if $p(X > b) = \frac{69}{80}$

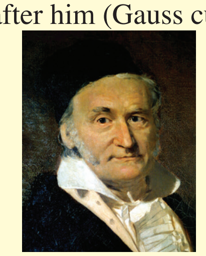
Unit
5

Normal distribution

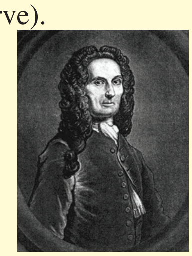


introduction

The normal distribution is considered one of the most important probability distributions learned in the statistics curricula due to its various uses to the outcomes of some processes in the physical, economic and social sciences and it deals with the phenomena of our daily life. The french scientist Abraham de Moivre was the first to use the normal distribution in 1756 in one of its prints. Some other scientists such as the German scientist Carl Friedrich Gauss(1777-1855) has participated in developing the normal distribution. The normal distribution is sometimes named after him (Gauss curve).



Carl Friedrich Gauss



Abraham de Moivre

The most well-known application of the normal distribution is the administrative evaluation of the subordinates in order to ensure the justice. The normal distribution is used to study the remains of analyzing the slope and it has a close relation with the control charts and so on.

Unit objectives

By the end of the unit and carrying out the involved activities, the student should be able to:

- Identify the normal distribution and its properties.
- Calculate the probability of the standard variable.
- Calculate the probability of the non-standard normal variable.
- Identify the standard normal random variable and the general form of the curve representing the density function of this variable.
- Convert any normal random variable into a standard normal variable.
- Find the values of the probabilities of a random variable with a standard normal distribution using the statistical tables.
- Explain the properties of the normal distribution curve and some phenomena it expresses.
- Explain the results which they have obtained through calculating the probability of the normal random variable.
- Estimate the mean of a community by a point.
- Estimate the mean of a community by a confidence interval.

Key Terms

- Normal distribution
- Normal random variable
- Normal curve
- Standard normal curve

Lessons of the unit

Lesson (5 - 1) Normal distribution

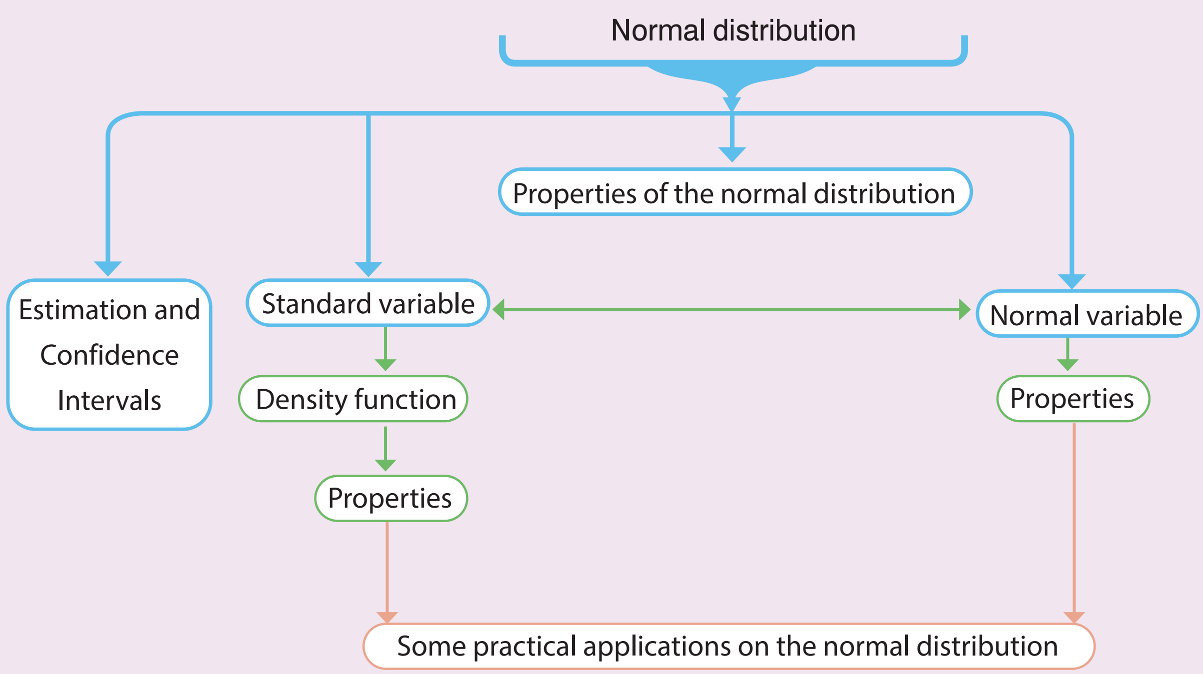
Lesson (5 - 2) Some practical applications on the normal distribution

Lesson (5 - 3) Estimation and Confidence Intervals

Materials

Scientific calculator

Unit planning organizer



Unit Five

5 - 1

Normal distribution

You will learn

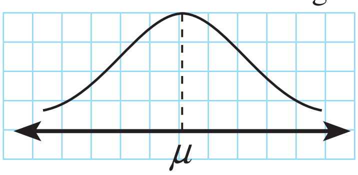
- Normal random variable
- Some properties of the normal distribution
- Standard normal distribution
- Properties of the density function of the standard normal distribution
- Calculating the probability of the standard normal distribution.

Key terms

- Normal distribution
- Normal random variable
- Normal curve
- Standard normal distribution

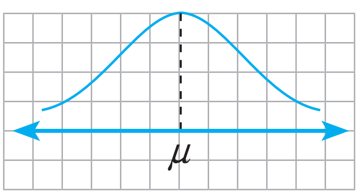
Introduction:

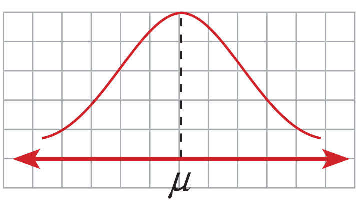
The normal distribution is one of the most important continuous probability distributions since it has important theoretical properties. Furthermore, the outcomes of the normal distribution can take any value within an interval of the real numbers. For example, the lengths of adults, the weights of babies as they were born and the intelligence degree of people and so on. The normal distribution can be described as a mathematical equation identifying its curve. This equation is identified thoroughly by knowing the expectation (mean) μ and the standard deviation σ . This curve is bell-like. It is symmetrical on the straight line x = μ and its two ends are converged from the horizontal axis where its two ends expand infinitely as shown in the opposite figure.



Normal random variable:

It is said that the continuous random variable x is "a normal random variable" if its range can be identified by the interval $]-\infty , \infty[$ and its probability density function is represented by a bill-like curve. The curve of the density function is called the normal curve or "Gauss curve". The value of the curve is identified by knowing two basic values: the mean μ and the standard deviation σ of the random variable x as shown in the next figures.





Some properties of the normal curve

(1) It has one top and its two ends extends to $-\infty$, ∞ .

(2) It has a symmetrical axis passing through the top and intersecting the horizontal axis when X = μ .

(3) The area of the region placed under the normal curve and above the X axis equals 1 .

(4) From symmetry, we find that the straight line X = μ divides the area placed under the curve and above the X axis into two regions; the area of each region = 0.5.

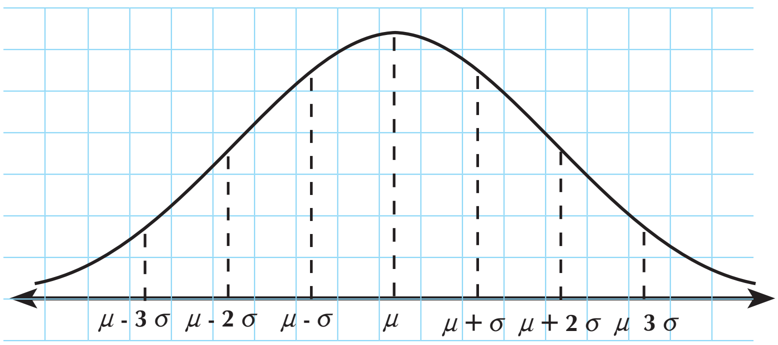
Materials — Scientific calculator.

(5) The approximate area of the region under the curve and above the X axis can be calculated in regard to the following intervals:

Ò From $\mu - \sigma$ to $\mu + \sigma = 68.26\,\%$ the total area .

Ò From $\mu - 2\sigma$ to $\mu + 2\sigma = 95.44\,\%$ the total area .

Ò From $\mu - 3\sigma$ to $\mu + 3\sigma = 99.74\,\%$ the total area .



Notice : The number of data should be big in order to make the normal distribution approximated

Example

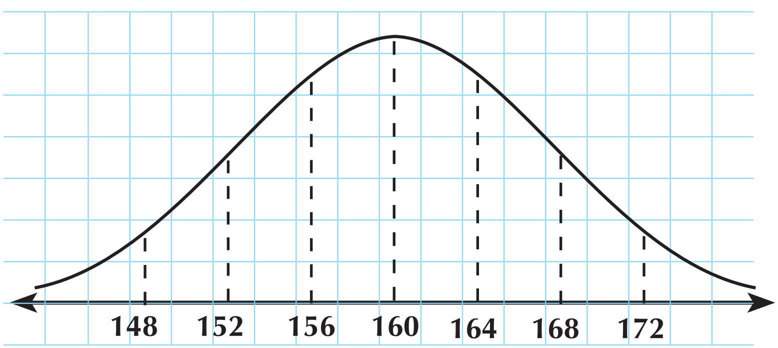
(1) If the lengths of students at a school follow a normal distribution of an average of 160 cm, the standard deviation is 4 cm. a student is randomly chosen, find the probability of the length of the student is:

a Taller than 172 cm

b Shorter than 156 cm

c Included between 156 cm and 168 cm

Solution



From the given data, we find that: the mean $\mu = 160$ and the standard deviation $\sigma = 4$

By comparing the data with the normal distribution curve, we find that: $\mu + 3\,x = 160 + 3 \times 4$ so,

a $P(x > 172) = P(x > \mu + 3\sigma)$

$\because$ The area from $\mu - 3\sigma$ to $\mu + 3\sigma = 0.9974$

$\therefore$ The area from μ to $\mu + 3\sigma = 0.9974 \div 2 = 0.4987$

$\therefore$ The area on the right of $\mu + 3\sigma = 0.5 - 0.4987 = 0.0013$

b $P(x < 156) = P(x < \mu - \sigma)$

$\because$ The area from $\mu - \sigma$ to $\mu + \sigma = 0.6826$

$\therefore$ The area from μ to $\mu - \sigma = 0.6826 \div 2 = 0.3413$

$\therefore$ The area on the left of $\mu - \sigma = 0.5 - 0.3413 = 0.1587$

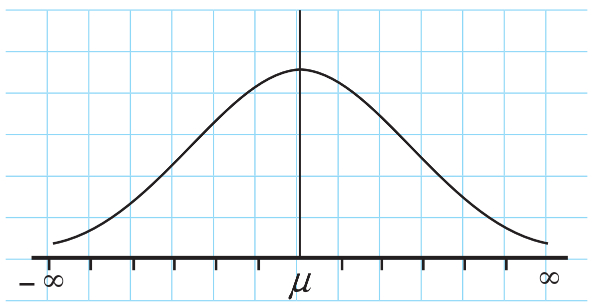
c $P(156 < x < 168) = P(\mu - \sigma < x < \mu + 2\sigma) = P(\mu - \sigma < x < \mu) + P(\mu < x < \mu + 2\sigma)$

$$= \frac{0.6816}{2} + \frac{0.9544}{2} = 0.3408 + 0.4772 = 0.818$$

Try to solve

1. If the weights of the students in a faculty follow a normal distribution μ of an average = 68 kg and its variance is 16 kg^2, find:
   - **a** The probability that the weight is greater than 72 kg
   - **b** The percentage of the students whose weights range between 64 kg and 72 kg “the weight of each student”.
   - **c** The number of the students whose weights are more than 64 kg if the total number of the college students is 2000 students.

The standard normal distribution

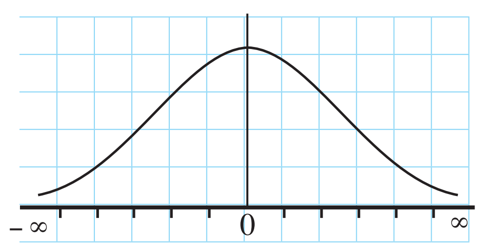


In the normal distribution, we have noticed that when you find the probability, the lengths of the intervals are multiples of the standard deviation in order to calculate the probability. Hence, it has been proper to convert the normal distributions into standard normal distributions by converting the values of (x) into standard values of (z) in terms of the mean μ and the standard deviation σ; at which $\mu = 0$, $\sigma = 1$

Definition

If the probability distribution of the random variable x is the normal distribution by a mean μ and standard deviation σ, then $z = \frac{x - \mu}{\sigma}$ is the standard normal distribution whose mean is μ = zero and standard deviation $\sigma = 1$.

Some properties of the density function of the standard normal distribution (z).



(1) The curve is located above the horizontal axis (X axis).

(2) It is symmetrical in terms of the vertical axis (Y axis).

(3) The two ends of the curve extends to tinfinitely without getting converged with the horizontal axis.

(4) The area of the region under the curve and above the horizontal axis = 1.

(5) From the symmetry, we find that the vertical axis divides the region under the curve and above the horizontal axis into two regions the area of each = 0.5.

(6) The approximate area of the region located under the standard curve only and above any interval]a , b[by special tables can be calculated.

The table of the area under the standard normal distribution curve:

To convert the normal distribution X into standard normal distribution Z you use the relation:

$Z = \frac{X - \mu}{\sigma}$, from the table of the standard normal distribution attached at the end of the book, you can find the required area .

Here, we explain how you can use the table of area under the standard normal distribution.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0						0.0199				
0.1										
0.2										
0.3										
0.4	0.1554									
0.5										
0.6				0.2357						
2.5								0.4949		
3.5										

$P(0 \leqslant Z \leqslant 0.05)$ = the area under the standard normal distribution is above the interval [0 , 0.05] i.e. z = 0.05, and so, we search in the table in raw 0.00 and under the column 0.05, to find the number is 0.0199

$\therefore P(0 \leqslant Z \leqslant 0.05) = 0.0199$

$P(0 \leqslant Z \leqslant 0.4)$ = the area under the standard normal distribution is above the interval [0 , 0.4] i.e. z = 0.4 , and so, we search in the table in raw 0.4 and under the column 0.00, to find the number is 0.1554.

$\therefore P(0 \leqslant Z \leqslant 0.4) = 0.1554$

$P(0 \leqslant Z \leqslant 0.63)$ = the area under the standard normal distribution is above the interval [0 , 0.63] I.e. z = 0.63 , and so, we search in the table in raw 0.6 and under the column 0.03, to find the number is 0.2357. 0.2357

$\therefore P(0 \leqslant Z \leqslant 0.63) = 0.2357$

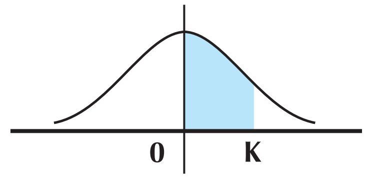
$P(0 \leqslant Z \leqslant 2.57)$ = the area under the standard normal distribution is above the interval [0 , 2.57] I.e. z = 2.57 , and so, we search in the table in raw 2.5 and under the column 0.07, to find the number is 0.4949

$\therefore P(0 \leqslant Z \leqslant 2.57) = 0.4949$

Calculating the probability of the standard normal variable:

(1) To find the area of the region under the curve in the interval [0 , K] from the table

The table of areas under the standard normal curve gives the approximate area above the interval [0 , K] and under the normal curve where $K \geqslant 0$, i.e. the table directly gives us $P(0 \leqslant Z \leqslant K)$



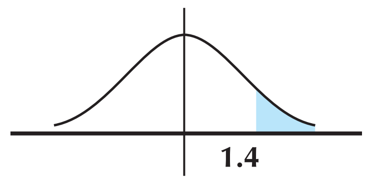
For example: $P(0 \leqslant Z \leqslant 0.3) = 0.1179$, $P(0 < Z \leqslant 0.64) = 0.2389$

$P(0 \leqslant Z \leqslant 1.7) = 0.4554$, $P(0 \leqslant Z < 2.45) = 0.4929$

Notice that: $P(Z \geqslant 1.4) = 0.5 - P(0 \leqslant Z \leqslant 1.4)$

$= 0.05 - 0.4192$

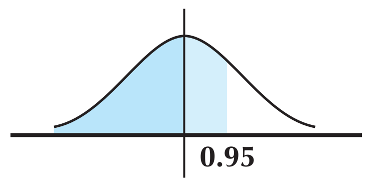
$= 0.0808$



Similarly: $P(Z \leqslant 0.95)$ $= 0.5 + P(0 \leqslant Z \leqslant 0.95)$

$= 0.5 + 0.3289$

$= 0.8289$



(2) Finding the area of the region under the curve in the interval [- K , 0] from the table

From the symmetry of the standard normal curve around the vertical axis, we find that :

$P(-K \leqslant Z \leqslant 0) = P(0 \leqslant Z \leqslant K)$

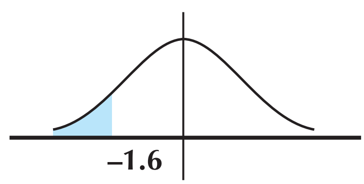
For example: $P(-1.25 \leqslant Z \leqslant 0) = P(0 \leqslant Z \leqslant 1.25) = 0.3944$

, $P(-2.24 \leqslant Z \leqslant 0) = P(0 \leqslant Z \leqslant 2.24) = 0.4875$

, $P(Z \leqslant -1.6) = 0.5 - P(-1.6 \leqslant Z \leqslant 0)$

$= 0.5 - P(0 \leqslant Z \leqslant 1.6)$

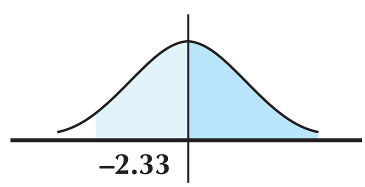
$= 0.5 - 0.4452 = 0.0548$



, $P(Z \geqslant -2.32) = 0.5 + P(-2.32 \leqslant Z \leqslant 0)$

$= 0.5 + P(0 \leqslant Z \leqslant 2.32)$

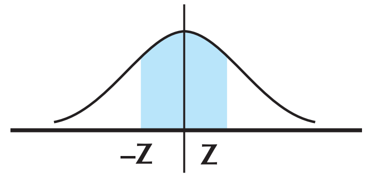
$= 0.5 + 0.4898 = 0.9898$



Notice that: $P(-K \leqslant Z \leqslant K) = 2 \times P(0 \leqslant Z \leqslant K)$

For example: $P(-1.4 \leqslant Z \leqslant 1.4) = 2 \times P(0 \leqslant Z \leqslant 1.4)$

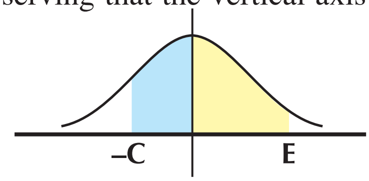
$= 2 \times 0.4192 = 0.8384$



, $P(-2.0 \leqslant Z \leqslant 2.0) = 2 \times P(0 \leqslant Z \leqslant 2.0) = 2 \times 0.4772 = 0.9544$

(3) Finding the area of the region under the curve in any interval [C , E]:

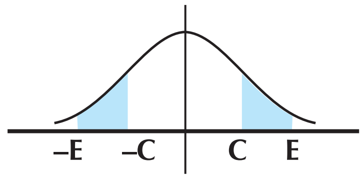
In this case, it is favorable to draw the standard curve and observing that the vertical axis divides the area under the curve and above the horizontal axis into two equal areas and the area of each = 0.5.



First: P (- C ⩽ Z ⩽ E) where C and E are positive.

= P (- C ⩽ Z ⩽ 0) + P (0 ⩽ Z ⩽ E)

= P (0 ⩽ Z ⩽ c) + P (0 ⩽ Z ⩽ E)



Second: P (C ⩽ Z ⩽ E) = P (- E ⩽ Z ⩽ -C)

= P (0 ⩽ Z ⩽ E) – P (0 ⩽ Z ⩽ C)

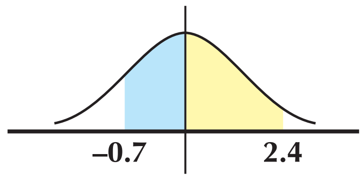
For example:

(1) P (-0.7 ⩽ Z ⩽ 2.4)

= P (-0.7 ⩽ Z ⩽ 0) + P(0 ⩽ Z ⩽ 2.4)

= P (0 ⩽ Z ⩽ 0.7) + P(0 ⩽ Z ⩽ 2.4) from symmetry

= 0.2580 + 0.4918 = 0.7498

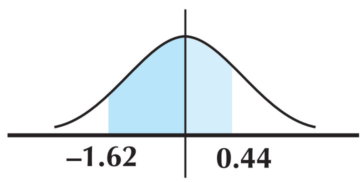


(2) P (-1.62 < Z ⩽ 0.44)

= P(-1.62 < Z ⩽ 0) + P(0 ⩽ Z ⩽ 0.44)

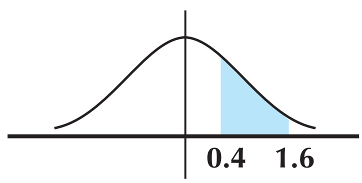
= P (0 ⩽ Z ⩽ 1.62) + P(0 ⩽ Z ⩽ 0.44) from symmetry

= 0.4474 + 0.1700 = 0.6174



(3) P (0.4 ⩽ Z < 1.6) = P(0 ⩽ Z < 1.6) - P (0 ⩽ Z ⩽ 0.4)

= 0.4452 – 0.1554 = 0.2898

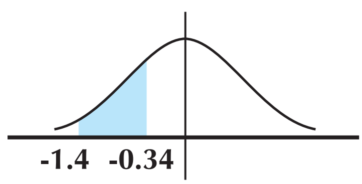


(4) P (- 1.4 < Z < -0.34)

= P(- 1.4 < Z ⩽ 0) - P(-0.34 < Z ⩽ 0)

= P (0 ⩽ Z ⩽ 1.4) + P(0 ⩽ Z ⩽ 0.34) from symmetry

= 0.4192 – 0.1331 = 0.2861



Example Finding the area under the standard normal curve

2 If Z is a standard normal random variable, find:

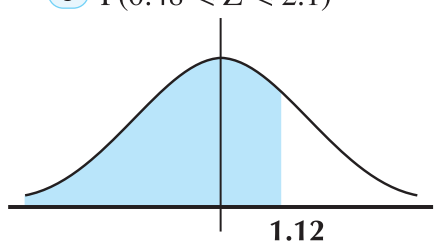
a P(Z ⩽ 1.12) **b** P(Z ⩾ 1.64) **c** P(0.48 ⩽ Z ⩽ 2.1)

Solution

a P(Z ⩽ 1.12)

= P(0 ⩽ Z ⩽ 1.12) + P(Z ⩽ 0)

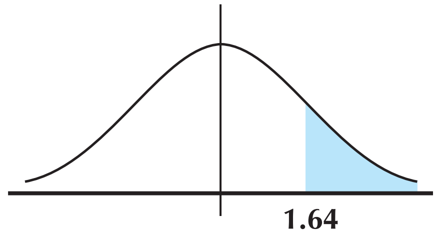
= 0.3686 + 0.5 = 0.8686



b $P(Z \geqslant 1.64)$

$= P(Z \geqslant 0) - P(0 \leqslant Z \leqslant 1.64)$

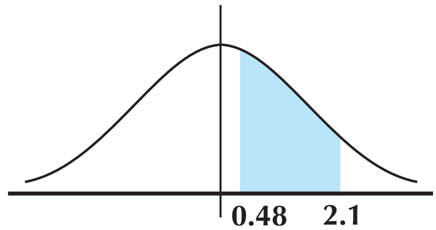
$= 0.5 - 0.4495 = 0.0505$



c $P(0.48 \leqslant Z \leqslant 2.1)$

$= P(0 \leqslant Z \leqslant 2.1) - P(0 \leqslant Z \leqslant 0.48)$

$= 0.4821 - 0.1844 = 0.2977$



Try to solve

2 If Z is a standard normal random variable, find:

a $P(0 \leqslant Z \leqslant 0.82)$ **b** $P(Z \geqslant 2.32)$

c $P(Z \leqslant 1.64)$ **d** $P(1.08 \leqslant Z \leqslant 3.12)$

Example

3 If Z is a standard normal random variable, find:

a $P(Z \leqslant -0.56)$ **b** $P(Z \geqslant -1.06)$

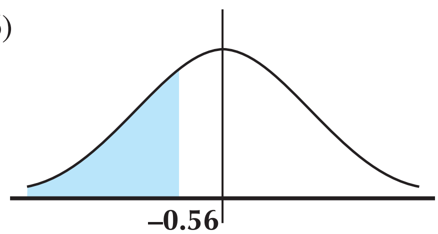
c $P(-1.2 \leqslant Z \leqslant 2.48)$ **d** $P(-2.2 \leqslant Z \leqslant -0.46)$

Solution

a $P(Z \leqslant -0.56)$

$= P(Z \geqslant 0.56)$

$= 0.5 - P(0 \leqslant Z \leqslant 0.56) = 0.5 - 0.2123 = 0.2877$

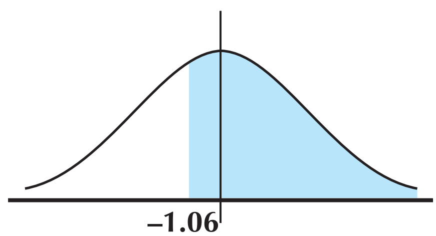


b $P(Z \geqslant -1.06)$

$= P(Z \leqslant 1.06)$

$= P(0 \leqslant Z \leqslant 1.06) + 0.5$

$= 0.0636 + 0.5 = 0.5636$

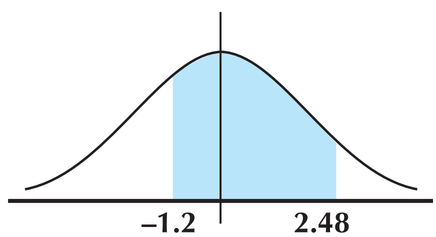


c $P(-1.2 \leqslant Z \leqslant 2.48)$

$= P(-1.2 \leqslant Z \leqslant 0) + P(0 \leqslant Z \leqslant 2.48)$

$= P(0 \leqslant Z \leqslant 1.2) + P(0 \leqslant Z \leqslant 2.48)$

$= 0.3849 + 0.4934 = 0.8783$

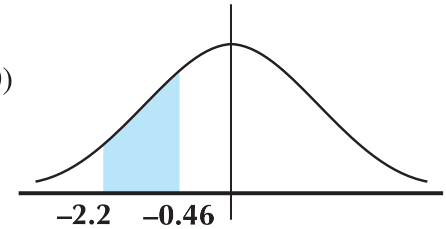


d $P(-2.2 \leqslant Z \leqslant -0.46)$

$= P(-2.2 \leqslant Z \leqslant 0) - P(-0.46 \leqslant Z \leqslant 0)$

$= P(0 \leqslant Z \leqslant 2.2) - P(0 \leqslant Z \leqslant 0.46)$

$= 0.4861 - 0.1772 = 0.3089$



Try to solve

3 If Z is a standard normal random variable, find:

a $P(Z \leqslant -0.56)$ **b** $P(Z \geqslant -1.06)$

c $P(-1.2 \leqslant Z \leqslant 2.48)$ **d** $P(-2.2 \leqslant Z \leqslant -0.46)$

Example **Converting from a normal variable into a standard normal variable**

4 If X is a normal random variable whose mean is μ and standard deviation is σ , find:

a $P(X > \mu - 1.5\sigma)$ **b** $P(X < \mu - 0.5\sigma)$

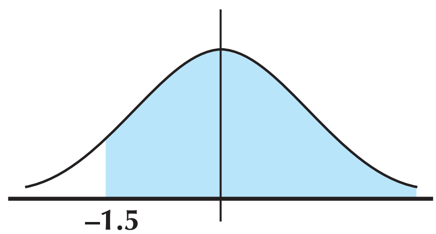
c $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma)$

Solution

a $P(Z > \frac{\mu - \sigma 1.5 - \mu}{\sigma}) = P(Z > -1.5)$

$= P(-1.5 < Z < 0) + 0.5$

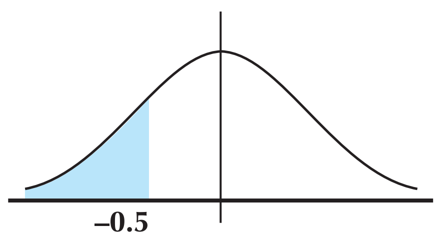
$= P(0 < Z < 1.5) + 0.5 = 0.4332 + 0.5 = 0.9332$



b $Z(X < \mu - 0.5\sigma) = P(Z < \frac{\mu - \sigma 0.5 - \mu}{\sigma})$

$= P(Z < -0.5) = P(Z > 0.5)$

$= 0.5 - P(0 < Z < 0.5) = 0.5 - 0.1915 = 0.3085$

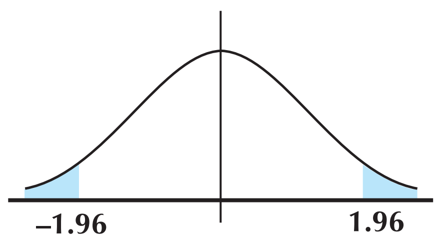


c $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma)$

$= P(\frac{\mu - \sigma 1.96 - \mu}{\sigma} < Z < \frac{\mu - \sigma 1.96 - \mu}{\sigma})$

$= P(-1.96 < Z < 1.96)$

$= 2\,P(0 < Z < 1.96) = 2 \times 0.4750 = 0.95$



Try to solve

4 If X is a normal random variable whose mean is μ and standard deviation is σ, find:

a $P(X < \mu - 2.1\sigma)$ **b** $(X > \mu + 0.8\sigma)$

c $P(\mu - 1.48\sigma < X < \mu + 1.48\sigma)$

Example

5 If Z is a standard normal random variable, find the value of k in each of the following cases :

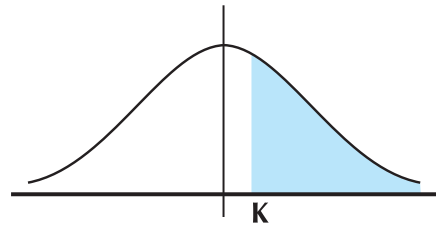
a $P(Z \geqslant K) = 0.1056$ **b** $P(Z \leqslant K) = 0.1151$

c $P(-0.44 \leqslant Z \leqslant K) = 0.5588$ **d** $P(K \leqslant Z \leqslant 2.1) = 0.2906$

Solution

a We notice that:

The area < 0.5 and the sign of the inequality “greater than“, so K lies at the positive interval as shown in the opposite figure.



$\because P(Z \geqslant K) = 0.1056$

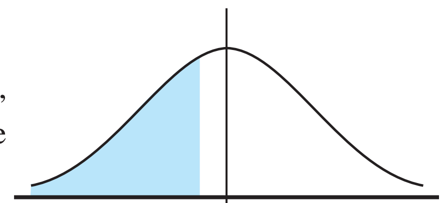
$\therefore 0.5 - P(0 \leqslant Z \leqslant K) = 0.1056$

$\therefore P(0 \leqslant Z \leqslant K) = 0.5 - 0.1056 = 0.3944$

In the area tables, we try to find the number (Z) or the nearest number corresponding the area 0.3944. We find that 1.2 is under the differences 0.05. **I.e.** **$K = 1.25$**

b We notice that:

The area < 0.5 , and the sign of the inequality “less than“, so k lies at the negative interval as shown in the opposite figure.



$\because P(Z \leqslant K) = 0.1151$

From the symmetry in the curve, we find that: $P(Z \geqslant K) = 0.1151$

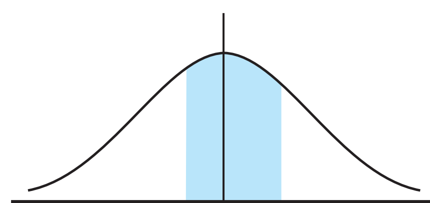
$\therefore 0.5 - P(0 \leqslant Z \leqslant K) = 0.1151$

$\therefore P(0 \leqslant Z \leqslant K) = 0.5 - 0.1151 = 0.3849$

$\therefore$ **$K = -1.2$** (**notice that k lies at the negative part**)

c We notice that :

The area > 0.5 and an end of the interval lie at the negative interval, so the other end of the interval z lies at the positive interval as shown in the opposite figure.



$\because P(-0.44 \leqslant Z \leqslant K) = 0.5588$

$\therefore P(-0.44 \leqslant Z \leqslant 0) + P(0 \leqslant Z \leqslant K) = 0.5588$

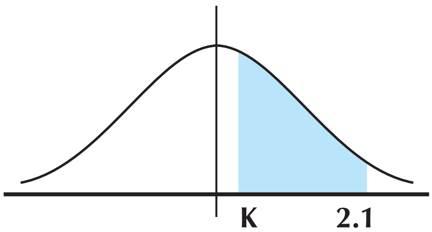
$\therefore P(0 \leqslant Z \leqslant 0.44) + P(0 \leqslant Z \leqslant K) = 0.5588$

$\therefore 0.1700 + P(0 \leqslant Z \leqslant K) = 0.5588$

$\therefore P(0 \leqslant Z \leqslant K) = 0.5588 - 0.1700 = 0.3888$

$\therefore$ **$K = 1.22$**

d We notice that :



The area < 0.5 and an end of the interval lie at the positive interval, so the other end of the interval g lies at the positive interval as shown in the opposite figure. (also $P(0 \leqslant Z \leqslant 2.1) > 0.2906$)

$\because P(K \leqslant Z \leqslant 2.1) = 0.2906$

$\therefore P(0 \leqslant Z \leqslant 2.1) - P(0 \leqslant Z \leqslant K) = 0.2906$

$\therefore P(0 \leqslant Z \leqslant K) = P(0 \leqslant Z \leqslant 2.1) - 0.2906$

$= 0.4821 - 0.2906 = 0.1915$ $\therefore$ **K = 0.5**

Try to solve

5 If Z is a standard normal random variable, find the value of k in each of the following cases:

a $P(Z \geqslant K) = 0.1980$ **b** $P(Z \leqslant K) = 0.1980$

c $P(-2.4 \leqslant Z \leqslant K) = 0.7970$ **d** $P(K \leqslant Z \leqslant 2.5) = 0.8238$

Example

6 If x is a normal random variable whose mean is μ and standard deviation is σ

a If: $P(X \geqslant 180) = 0.0062$, $\mu = 165$ calculate σ

b If: $P(X > 35) = 0.8643$, $\sigma = 5$ calculate μ

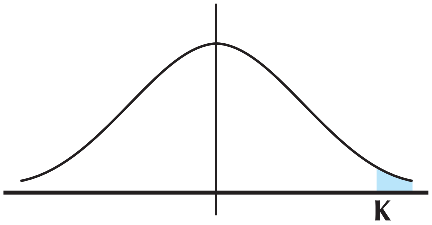
c If: $P(X \leqslant 170) = 0.0228$, $\sigma = 7$ calculate μ

d If: $P(X \leqslant K) = 0.8944$, $\mu = 125$, $\sigma = 8$ calculate K

e If: $P(X > K) = 0.9452$, $\mu = 50$, $\sigma = 5$ calculate K

Solution

a $P(X \geqslant 180) = P\left(Z \geqslant \frac{180 - 165}{\sigma}\right) = 0.0062$



$\therefore P(Z \geqslant K) = 0.0062$ where $K = \frac{15}{\sigma}$, $K > 0$

$\therefore P(0 \leqslant Z \leqslant K) = 0.5 - 0.0062 = 0.4938$

$\therefore K = 2.5$

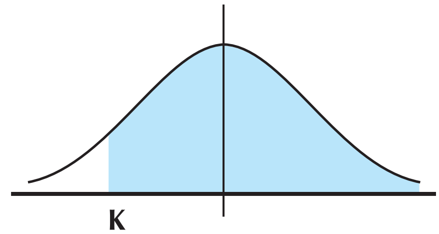
$\therefore \frac{15}{\sigma} = \frac{5}{2}$ $\therefore \sigma = \frac{2 \times 15}{5}$ $\therefore$ **$\sigma = 6$**

b $P(X > 35) = P(Z > \frac{35-\mu}{5}) = 0.8643$

$\therefore\ P(Z > K) = 0.8643$ where

$\therefore\ K = \frac{35-\mu}{5}$, $K < 0$

$\therefore\ P(0 \leqslant Z \leqslant -K) + 0.5 = -\ 0.8643$



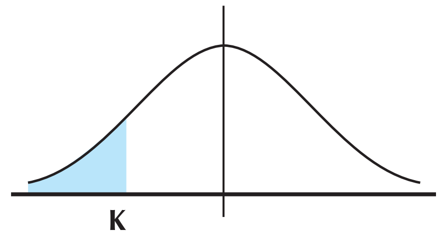
$P(K < Z \leqslant 0) = 0.8643 - 0.5 = 0.3643$ $\qquad \therefore K = -1.1$

$\therefore \frac{35-\mu}{5} = -1.1$ $\qquad \therefore 35 - \mu = -5.5$

$\therefore \mu = 35 + 5.5$ $\qquad \therefore \mu = \mathbf{40.5}$

c $P(X \leqslant 170) = P(Z \leqslant \frac{170-\mu}{7}) = 0.0228$

$\therefore P(Z \leqslant K) = 0.0228$ where $K = \frac{170-\mu}{7}$, $K < 0$



$\therefore P(K \leqslant Z \leqslant 0) = 0.5 - 0.0228 = 0.4772$

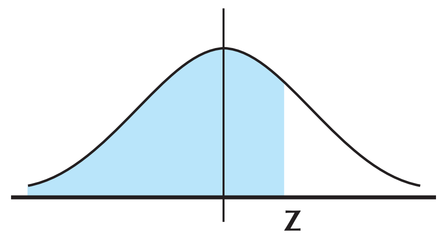
$\therefore K = -2$

$\therefore \frac{170-\mu}{7} = -2$ $\qquad \therefore 170 - \mu = -14$ $\qquad \therefore \mu = 170 + 14$ $\qquad \mu = \mathbf{184}$

d $P(X \leqslant K) = P(Z \leqslant \frac{K-125}{8}) = 0.8944$

$\therefore P(Z \leqslant a) = 0.8944$

where $a = \frac{K-125}{8}$, $a > 0$



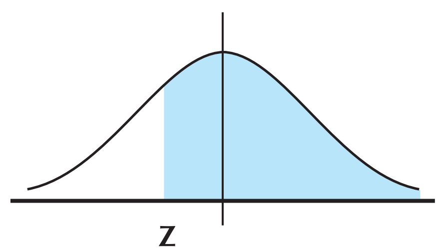
$\therefore P(0 \leqslant Z \leqslant a) = 0.8944 - 0.5 = 0.3944$ $\qquad \therefore a = 1.25$

$\therefore \frac{K-125}{8} = 1.25$ $\qquad \therefore K - 125 = 10$ $\qquad \therefore K = 125 + 10$ $\qquad \therefore K = 135$

e $P(X > K) = P(Z > \frac{K-50}{5}) = 0.9452$

$\therefore P(Z > a) = 0.9452$

where $a = \frac{K-50}{5}$, $a < 0$



$\therefore P(0 \leqslant Z < -a) = 0.9452 - 0.5 = 0.4452$ $\qquad \therefore a = -1.6$

$\therefore \frac{K-50}{5} = -1.6$ $\qquad \therefore K - 50 = -8$ $\qquad \therefore K = 50 - 8$ $\qquad \mathbf{K = 42}$

Try to solve

6 If X is a normal random variable whose mean is μ , standard deviation is σ, $P(X < 19) = 0.7734$ and $P(X > 10) = 0.9332$, calculate the value of μ and σ.

Exercises 5 - 1

1. If Z is a standard normal random variable, find :

a. $P(0 \leqslant Z \leqslant 1.15)$, $P(0 \leqslant Z \leqslant 2.42)$

b. $P(-0.04 \leqslant Z \leqslant 0)$, $P(-1.63 \leqslant Z \leqslant 0)$

c. $P(-0.7 \leqslant Z \leqslant 0.7)$, $P(-1.65 \leqslant Z \leqslant 1.65)$

d. $P(-2.42 \leqslant Z \leqslant 1.67)$, $P(-1.73 \leqslant Z \leqslant 0.64)$

e. $P(0.74 \leqslant Z \leqslant 1.02)$, $P(1.4 \leqslant Z \leqslant 2.2)$

f. $P(-2.1 \leqslant Z \leqslant -0.92)$, $P(-1.5 \leqslant Z \leqslant -0.84)$

g. $P(Z \leqslant 1.44)$, $P(Z \leqslant 2.05)$

h. $P(Z \leqslant -1.14)$, $P(Z \leqslant -2.32)$

i. $P(Z \leqslant 0.65)$, $P(Z \leqslant 1.42)$

j. $P(Z \leqslant -0.45)$, $P(Z \leqslant -1.6)$

2. If Z is a standard normal random variable, find the value of the real number k which satisfies:

a. $P(0 \leqslant Z \leqslant K) = 0.3554$

b. $P(K \leqslant Z \leqslant 0) = 0.4120$

c. $P(-K \leqslant Z \leqslant K) = 0.2206$

d. $P(Z \leqslant K) = 0.9754$

e. $P(Z \leqslant K) = 0.1977$

f. $P(Z \geqslant K) = 0.0934$

g. $P(Z \geqslant K) = 0.9955$

h. $P(K \leqslant Z \leqslant 1.11) = 0.6660$

i. $P(K \leqslant Z \leqslant 2.22) = 0.2446$

j. $P(-1.7 \leqslant Z \leqslant K) = 0.3261$

3. Z is a standard normal random variable, if :

a. $P(Z \leqslant K) = 0.1736$ Find: $P(K \leqslant Z \leqslant 1.7)$

b $P(Z \geqslant K) = 0.0207$ Find: $P(0.56 \leqslant Z \leqslant K)$

c $P(Z \leqslant K) = 0.8944$ Find: $P(-0.7 \leqslant Z \leqslant K)$

d $P(0.4 \leqslant Z \leqslant K) = 0.3110$ Find: $P(Z \leqslant K)$

e $P(1.4 \leqslant Z \leqslant K) = 0.0770$ Find: $P(-1.4 \leqslant Z \leqslant K)$

f $P(K \leqslant Z \leqslant 1.7) = 0.8586$ Find: $P(K \leqslant Z \leqslant 0.75)$

4 X is a normal random variable whose mean is μ, standard deviation is σ and

a $P(X \leqslant 90) = 0.0668$, $\mu = 102$ Calculate σ

b $P(X \geqslant 62) = 0.0548$, $\mu = 50$ Calculate σ

c $P(X \geqslant 48) = 0.0228$, $\sigma = 4$ Calculate μ

d $P(X > 68) = 0.1056$, $\sigma = 6.4$ Calculate μ

e $P(X \geqslant 42) = 0.8944$, $\sigma = 6.4$ Calculate μ

f $P(\mu - K\sigma \leqslant x \leqslant \mu + K\sigma) = 0.438$ Calculate K

g $P(X \leqslant K) = 0.2119$, $\mu = 42$, $\sigma = 5$ Calculate K

h $P(X \leqslant K) = 0.8413$, $\mu = 72$, $\sigma = 8$ Calculate K

i $P(X > K) = 0.9772$, $\mu = 60$, $\sigma = 4$ Calculate K

5 Answer the following questions

a If X is a normal random variable whose mean is 120, standard deviation is 10 and $p(X < K) = 0.9599$, find the value of k .

b If X is a normal random variable whose mean is μ, standard deviation is $\sigma = 5$, find the value of μ which makes $p(X \leqslant 35) = 0.0228$

c If X is a normal random variable whose mean is $\mu = 8$, standard deviation is $\sigma = 2$ and $P(K \geqslant X) = 0.1056$, find:

First : The value of k . **Second** : $P(X \leqslant 10)$

d If X is a normal random variable whose mean is μ and standard deviation is σ , find $P(\mu - \frac{1}{4}\sigma \leqslant X \leqslant \mu + \frac{1}{2}\sigma)$

e If Z is a standard normal random variable, find the value of k which satisfies :

First : $P(Z > K) = 0.0281$

Second : $P(-1 < Z < K) = 0.7918$

f If X is a normal random variable whose mean is 18 and standard deviation is 2.5, find :

First: $P(X < 15)$

Second: $P(17 < X < 21)$

g If X is a normal random variable whose mean $\mu = 24$ and standard deviation $\sigma = 5$, find:

First : $P(X \geqslant 32.5)$

Second : $P(14 < X < 29)$

g If X is a normal random variable whose mean $\mu = 48$ and standard deviation $\sigma = 5$, find:

First : $P(43 < X < 59)$

Second : The value of K if $P(x > K) = 0.1841$.

h If X is a normal random variable whose mean $\mu = 17$ and standard deviation $\sigma = 2$, find:

First : $P(16 \leqslant X \leqslant 20)$

Second : $P(X > 15)$

i If X is a normal random variable whose mean is 32 and variance is 16, find :

First : $P(X < 25)$

Second : $P(28 < X < 35)$

j If X is a normal random variable whose mean $\mu = 8$ and standard deviation $\sigma = 2$, find:

First : $P(X \leqslant 10)$

Second : If $P(X \geqslant k) = 0.1056$, find the value of k .

Unit Five
5 - 2 Some practical applications of the normal distribution

You will learn

- Practical applications on the normal distribution

Key terms

- Normal distribution
- Normal random variable
- Normal curve
- Standard normal distribution

Introduction:

In the previous lesson, you became familiar with the normal distribution and its properties and you also knew the standard normal random variable and how to find it in terms of the mean and the standard deviation. You also knew how to calculate the probabilities of a random variable with a standard normal distribution using the statistical tables.

In this lesson, you are going to learn some different uses of the normal random variable for studying some phenomena it may express.

Example

Industry

1 A machine in a factory produces cylinders of lengths follow a normal distribution whose mean is 56 cm and the standard deviation is 2 cm. The cylinder is only valid if its length ranges from 51 cm to 60 cm. A random sample has been randomly chosen out of 1000 cylinders. How many cylinders are expected to be valid?

Solution

Let X be a normal random variable expressing the cylinder length:

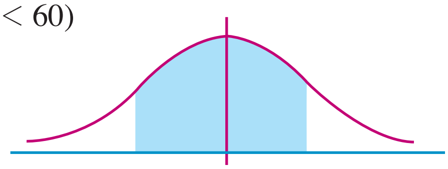
$\therefore$ The probability of (the cylinder is valid) = P (51 < X < 60)

$= P(\frac{51 - 56}{2} < Z < \frac{60 - 56}{2})$

$= P(-2.5 < Z < 2)$

$= P(-2.5 < Z \leqslant 0) + P(0 \leqslant Z < 2)$

$= 0.4938 + 0.4772 = 0.9710$



$\therefore$ The number of cylinders expected to be valid= $1000 \times 0.9710 = 971$ cylinders

Try to solve

1 Income: If the salaries of a 200- worker group at a factory follow a normal distribution whose mean is 175 LE and its standard deviation is 10 LE. How many workers do their salaries range from 170 LE to 180 LE?.

Materials

- Scientific calculator.

Example

2 **Education:** If the marks of the students at a school are a normal random variable whose mean $\mu = 44$ and its standard deviation is σ where 22.66% of students have got more than 50 marks, find the value of σ.

Solution

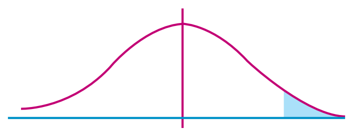
Let X be the normal random variable expressing the marks of the students

$\therefore P(X > 50) = \frac{22.26}{100}$

$\therefore P(Z > \frac{50 - 44}{\sigma}) = 0.2266$

$\therefore P(Z > K) = 0.2266$ where $K = \frac{6}{\sigma}$, $K > 0$

$\therefore P(0 \leqslant Z < K) = 0.5 - 0.2266 = 0.2734$



$\therefore K = 0.75$ $\qquad \therefore \frac{6}{\sigma} = 0.75$ $\qquad \therefore \sigma = \frac{6}{0.75} = 8$

Try to solve

2 If the marks of the students in an exam follow a normal distribution whose mean is 60 and its standard deviation is 12 and a student has been randomly chosen, find the probability that the mark of that student is between 66 , 75 marks. If 15% of the excellent students have got an Excellent rank, find the least mark of the student getting an Excellent rank.

Example

3 **Length:** If the lengths of the students at a high school follow a normal distribution whose mean $\mu = 160$ cm and its standard deviation $\sigma = 5$ cm, find the probability that the length of any student differs from μ not more than 8 cm .

Solution

Let X be a normal random variable expressing the lengths of the student and the length difference from $\mu = |X - \mu|$" i.e. the absolute difference between the length and the mean μ "

$\therefore P(|X - \mu| < 8) = P(|X - 160| < 8)$

$\therefore P(-8 < X - 160 < 8)$

$= P(152 < X < 168)$

$= P(\frac{152 - 160}{5} < Z < \frac{168 - 160}{5})$

$= P(-1.6 < Z < 1.6)$

$= 2 \times P(0 \leqslant Z < 1.6)$

$= 2 \times 0.4452 = 0.8904$

Remember

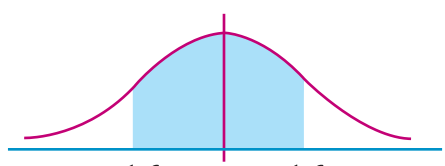
The expression:

$|X - a| < b$

is equivalent to the expression:

$-b < X - a < b$

I.e: $a - b < X < a + b$



Try to solve

3 **Weights:** If the weights of the students at a primary school follow a normal distribution whose mean is 30 kg and its standard deviation is 5 kg, calculate the percentage of the number of the students whose weights are more than 45 kg and so is the percentage of the number of the students whose weights range between 25 kg and 35 kg.

Example

4 **Work:** If the wages of the workers at a factory follow a normal distribution whose mean μ = 75 LE and its standard deviation σ = 10, find:

a The percentage of the workers whose wages are more then 90 LE.

b The percentage of the workers whose wages are less then 55 LE.

c The percentage of the workers whose wages range between 60 LE and 80 LE.

Solution

a $\because P(X > 90) = P(Z > \frac{90 - 75}{10})$

$= 0.5 - P(0 \leqslant Z \leqslant 1.5) = 0.5 - 0.4332 = 0.0668$

$\therefore$ The percentage of the workers whose wages are more then 90 LE = 6.68%

b $\because P(X < 55) \quad = P(Z < \frac{55 - 75}{10}) \quad = P(Z < -2)$

$= 0.5 - P(0 \leqslant Z \leqslant 2) \quad = 0.5 - 0.4772 = 0.0228$

$\therefore$ The percentage of the workers whose wages are less then 55 LE = 2.28% of the total number

c $\because P(60 \leqslant Z \leqslant 80) \quad = P(\frac{60 - 75}{10} \leqslant Z \leqslant \frac{80 - 75}{10})$

$= P(-1.5 \leqslant Z \leqslant 0.5) = P(0 \leqslant Z \leqslant 1.5) +$

$P(0 \leqslant Z \leqslant 0.5) = 0.1915 + 0.4332 = 0.6247$

$\therefore$ The percentage of the workers whose wages range between 60 LE and 80 LE = 62.47% of the total number of the workers at the factory.

Try to solve

4 Let the marks of an exam be a normal variable by expectation 76 and standard deviation 15 marks. By ordering the excellent students getting a mark higher than α mark, they represent 15% of the total students. By ordering the students getting marks lower than β mark, they represent 10% of the total students. Find:

a The least mark α to consider the student belonging to excellent .

b The failing mark β .

Exercises 5 - 2

1. If the salary of 1000 families in a city is a normal random variable whose mean is 170 LE, its standard deviation is 20 LE, and a family has been randomly chosen, find:
   a. The probability the salary is included between 160 LE and 200 LE .
   b. The number of families whose salaries are more than 150 LE.

2. If the weights of the students at a faculty follow a normal distribution whose mean is 68.5 kg and its standard deviation is 2.5 kg.
   a. calculate the percentage of the number of the students whose weights range between 67.5 kg and 71 kg.
   b. If the number of the students is 1000 students, calculate the number of the students whose weights are more than 71 kg .

3. A sample of 200 students at a school has been randomly taken. If their ages represent a normal random variable whose mean is 16.6 and its standard deviation is 1.2, find the number of the students whose ages are less than 16 years of this sample .

4. If the lengths of 2000 students at a faculty follow a normal distribution whose mean is 170 cm and its standard deviation is 8 cm, find the number of students whose lengths are less than 176 cm .

5. If the montly income of 300 families represents a random variable x following the normal distribution by expectation μ = 500 LE and standard deviation σ = 20 L.E, find
   a. The number of the families getting a salary more than 530 LE .
   b. The maximal salary of 4% of the families getting the lowest salaries .

6. If the montly income of 200 families is a random variable X following a normal distribution by expectation μ = 400, standard deviation σ = 80 LE and a family has been randomly chosen, find:
   a. The probability the salary of the family is more than 500 LE at maximum
   b. The number of the families getting a salary 500 LE at maximum .

7. If the life time (in hours) of a type of batteries is a random variable following a normal distribution whose mean is 2000 hours and its standard deviation is 120 hours; What is the probability the battery would keep working for more than 1800 hours?

8 If the salary of a 500 - employee group follows a normal distribution whose mean is 180 LE and its standard deviation is 15 LE, find the numbers of employees whose salaries are less than 198 LE.

9 If the rainwater rise during February follows a normal distribution whose mean $\mu = 3$ cm and its variance $\sigma^2 = 4$ cm^2, find the probability the rainwater rise in February next year:

a More than 1 cm
b Between 3.5 cm and 4 cm

10 If the temperature in August follows a normal distribution whose mean $\mu = 35$ degrees and its standard deviation $\sigma = 5$ degrees, find the probability that the temperature on a day during this month will be

a Between 28 degrees and 38 degrees.
b More than 39 degrees .
c Between 26 degrees and 32 degrees.

11 1000 adults have been applied to the enlisting management. If their lengths follow a normal distribution whose mean is 170 cm and its standard deviation is 10 cm, find the number of the adults:

a Whose lengths are less than 190 cm
b Unaccepted if the minimal required length is 155 cm

12 It is found that the lengths of a certain type of plants are distributed in regard to a normal distribution whose mean is 50 cm and its standard deviation is σ . If it is known that the lengths of 10.56% of these plants are less than 45 cm, find the variation of the lengths of these plants.

13 If the weights of the students in a faculty follow a normal distribution whose mean is 65 kg and its standard deviation is σ . If the weights of 33% if the students are more than 70 kg.

a Find the value of σ
b If the number of the students is 100 students, calculate the number of the students whose weights are less than 67.5 kg

14 If the weights of the students in a faculty follow a normal distribution whose mean is 68.5kg and its standard deviation is 2.5 kg:

a Calculate the percentage of the students whose weights range between 67.5 kg and 71kg .
b If the number of the students is 1000 students, calculate the number of the students whose weights are more than 71 kg.

15 If the marks of the students at a school are a normal random variable whose mean $\mu = 42$ and its standard variation is σ where 26.11% of the students got more than 50 marks, find the value of σ.

16 In a math exam, the marks of the students have a normal distribution whose mean is 70 and its standard deviation is 5, find the number of the students whose marks are more than 78 known that the number of the students did the exam are 100 students .

17 A factory produces cylinders of lengths follow a normal distribution whose mean is 56 cm and its standard deviation is 2 cm. The produced cylinders are valid if only their lengths are included between 51 cm and 60 cm. A random sample of 1000 cylinders has been taken. How many cylinders are expected to be valid?

18 If the radius lengths of the cylinders produced by a factory follow a normal distribution whose mean is 25 cm and its standard deviation is 20 cm. The cylinder is invalid if its radius length is less than 20 cm or more than 28 cm. If a cylinder has been randomly chosen, find the probability the cylinder is invalid.

19 If the weights of a group of the experimental animals follow a normal distribution whose mean is μ gm and its standard deviation is 10 gm and known that $P(x \geqslant 180) = 0.1587$, calculate the mean μ

20 If the marks of the students in an exam represent a random variable following a normal distribution whose mean is μ and its standard deviation is σ, find:

a The probability the students get a mark more than $(\mu - \sigma)$.

b The percentage of the students who get marks included between: $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$.

21 It is found that the lengths of a certain type of plants are distributed in regard to a normal distribution whose mean is μ and its standard deviation is 4 . If it is known that the lengths of 10.56% of these plants are less than 45 cm, find the mean μ of the lengths of these plants

22 If the temperature in January follows a normal distribution whose mean is 16 degrees and its standard deviation is 4 degrees, find the probability the temperature on a day during this month will be:

a Between 14 degrees and 20 degrees.

b More than 15 degrees.

23 In a community, it is found that the intelligence follows a normal distribution whose mean is 104.6 and its standard deviation is 6.25,

a Find the ratio of the individuals whose intelligence ranges between 90 and 120

b Find the ratio of the individuals whose intelligence is more than 110.

Unit Five

5 - 3 Estimation and Confidence Intervals

You will learn

- Estimating the mean of population with a point
- Estimating the mean of population with a Confidence Interval estimation

Key terms

- Critical Value
- Parameter
- Percentage of Error
- Statistics
- Interval Estimation
- Estimate
- Point Estimate
- Confidence Interval estimation
- Normal Distribution

Introduction:

Parameter :

A constant numerical value that characterizes a population and is often unknown.
Such as the mean population μ and is estimated by the sample mean $\overline{X}$

Estimation :

It is a statistic that depends on the sample values and reflects a value close to the parameter of the population as a whole and its distribution , It has two methods

(1) Point estimate

It is a single value calculated from the sample that is used to estimate an unknown parameter of the population.
Like the arithmetic mean of a random sample, $\overline{X}$, which used to estimate a mean for the population μ.

(2) Interval Estimation

It is to find a specific period within which the population parameter is expected to lie in by a certain percentage or with a certain probability. This period is called the confidence interval.

Confidence Interval An interval of values: used in statistics to estimate the value of an unknown parameter to the population.

Interpretation of confidence interval: A 95% confidence interval means that when an experiment of the same size is repeated 100 times, we are confident that 95 times out of 100 the parameter estimate lies inside the confidence interval.

level of confidence

It is the probability that the confidence interval contains the true value of the population parameter under study and its value is equal to $(1-\alpha)$, where α is the error in the estimate.

For example:

If $\alpha = 0.05$, then the confidence level $= (1 - \alpha) = 0.95 = 95\%$

If $\alpha = 0.01$, then the confidence level $= (1 - \alpha) = 0.99 = 99\%$

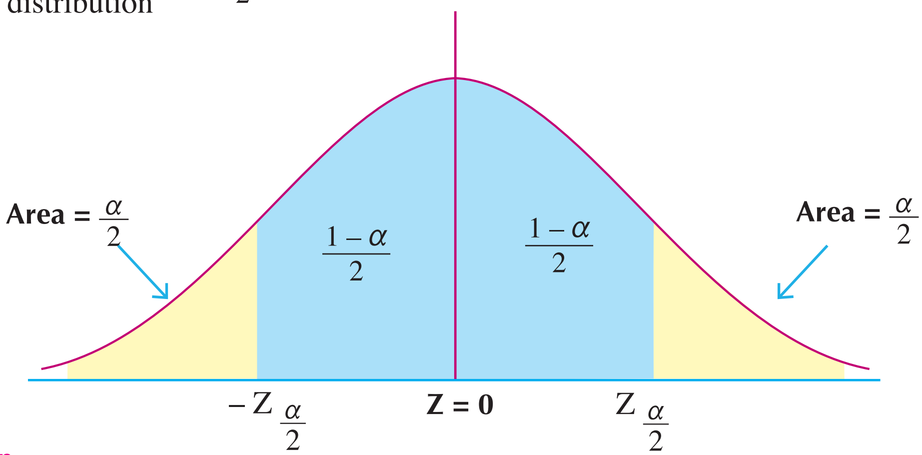
الأدوات المستخدمة آلة حاسبة علمية. برامج رسومية للحاسب.

Critical value: $Z_{\frac{\alpha}{2}}$

To find the critical value $Z_{\frac{\alpha}{2}}$, we calculate the area $\frac{1-\alpha}{2}$ from the table of areas under the standard normal distribution curve we get the value of $Z_{\frac{\alpha}{2}}$

Example

1. Find the critical value $Z_{\frac{\alpha}{2}}$ corresponding to the 95% confidence level using the standard normal distribution



Solution

$\because$ confidence level 95%

$\therefore$ $1-\alpha = 0.95$

$\therefore$ $\frac{1-\alpha}{2} = \frac{0.95}{2} = 0.475$

i.e $P(0 < Z < Z_{\frac{\alpha}{2}}) = 0.475$

By looking for this value in the standard normal distribution table of areas under the standard normal curve

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767

$\therefore$ $Z_{\frac{\alpha}{2}} = 1.96$

Try to solve

1. Find the critical value $Z_{\frac{\alpha}{2}}$ corresponding to the 99% confidence level using the standard normal distribution

Estimation Error

When a sample is used to estimate the population mean, the error in the estimate is represented by the symbol E at a confidence level of 1 - α, which is determined by the following relationship: $\boxed{E = \frac{\sigma}{\sqrt{n}} \times Z_{\frac{\alpha}{2}}}$

Where σ is the standard deviation and n is the sample size

Confidence interval for mean population μ

If a random sample of size n is taken from a population that follows a normal distribution with mean μ and variance σ^2

Then $\boxed{\mu \in \left]\overline{X} - E, \overline{X} + E\right[}$

Where $E = \frac{\sigma}{\sqrt{n}} \times Z_{\frac{\alpha}{2}}$ at confidence level (1 - α) , $\overline{X}$ is the arithmetic mean of the sample , E is the estimation error The two limits $\overline{X}$ -E, $\overline{X}$ +E are called the lower and upper bounds of the confidence interval

Remarks

(1) When finding the confidence interval, we will be satisfied with the 95% confidence level, which corresponds to the critical value $Z_{\frac{\alpha}{2}} = 1.96$ (From the standard normal distribution table ofareas under the curve)

(2) If the sample size is greater than 30, σ is unknown it can be considered that the population standard deviation σ is the standard deviation of the sample

Steps followed to find the confidence interval for the mean in the population μ

(1) We find the critical value $Z_{\frac{\alpha}{2}}$ corresponding to the 95% confidence level, which is 1.96

(2) We find the Estimation error $E = \frac{\sigma}{\sqrt{n}} \times Z_{\frac{\alpha}{2}}$ where σ is the population standard deviation , n is the sample size

(3) We find the confidence interval $\left]\overline{X} - E, \overline{X} + E\right[$

Example

2 A study was conducted on a sample of females about pulse rate. If the sample size was 49, the standard deviation for the female population was α = 12.5, and the arithmetic mean for the sample was $\overline{x} = 76.5$. Using a 95% confidence level.

a Find the Estimation error

b Find the confidence interval for the population mean μ

c Interpret the confidence interval

Solution

$\because$ confidence level 95% $\therefore$ critical value $Z_{\frac{\alpha}{2}} = 1.96$

a where $\overline{X} = 76.5$, $\sigma = 12.5$, $n = 49$, $Z_{\frac{\alpha}{2}} = 1.96$

Estimation error $E = \frac{\sigma}{\sqrt{n}} \times Z_{\frac{\alpha}{2}} = \frac{12.5}{\sqrt{49}} \times 1.96 = 3.5$

b confidence interval is

$]\ \overline{X} - E , \overline{X} + E\ [=]\ 76.5 - 3.5 , 76.5 + 3.5\ [=]\ 73 , 80[$

c Interpretation

When choosing 100 random samples of the same size (n = 49) and calculating the confidence interval for each sample, we expect that 95 intervals contain the true value of the mean of the population μ.

Try to solve

2 A study was conducted on a sample of females about pulse rate. If the sample size was 64, the standard deviation for the female population was $\sigma = 3.6$, and the arithmetic mean for the sample was $\overline{X} = 18.4$ Using a 95% confidence level.

a Find the Estimation error

b Find the confidence interval for the population mean μ

c Interpret the confidence interval

Example

3 Sample size is 49. If the arithmetic mean of the sample is 60 and its variance is 144, using a confidence level of 95%

a Find the Estimation error

b Find the confidence interval for the population mean μ

c Interpret the confidence interval

Solution

$\because$ confidence level 95% $\therefore$ critical value $Z_{\frac{\alpha}{2}} = 1.96$

a where $\overline{X} = 60$, $\sigma = 12$, $n = 49$, $Z_{\frac{\alpha}{2}} = 1.96$

Estimation error $E = \frac{\sigma}{\sqrt{n}} \times Z_{\frac{\alpha}{2}} = \frac{12}{\sqrt{49}} \times 1.96 = 3.36$

b confidence interval is

$]\ \overline{X} - E , \overline{X} + E\ [=]\ 60 - 3.36 , 60 + 3.36\ [=]\ 56.64 , 63.36\ [$

c Interpretation

When choosing 100 random samples of the same size (n = 49) and calculating the confidence interval for each sample, we expect that 95 intervals contain the true value of the mean of the population μ.

Exercises 5 - 3

First: choose the correct answer:

1. A sample of size (n). If the arithmetic mean of the sample is 13 and its standard deviation is 12 using a confidence level of 95% If the estimation error equals 2.352, then the sample size is equal to ....

   a 25 b 36 c 50 d 100

2. A sample size of 225, using a confidence level of 95%, If the estimation error equals 0.784, then the standard deviation of the sample is equal to ....

   a 25 b 5 c 6 d 36

3. If the upper limit of the 95% confidence interval for a sample mean is equal to 7.25 with an estimation error of 1.25, then the sample mean is equal to ....

   a 5 b 6 c 7 d 8

4. If the confidence interval for a sample mean is] 9.3 , 10.7 [then the arithmetic mean of the sample is equal to ....

   a 8 b 9 c 10 d 11

5. If the confidence interval for a sample mean is] 9.02 , 10.98 [, the standard deviation of the sample is equal to 4 using a confidence level of 95%, then the sample size is equal to ....

   a 30 b 49 c 225 d 64

6. If the lower limit of the confidence interval for the sample mean is equal to 23.04 with a confidence level of 95% and the sample size is 625 and the arithmetic mean of the sample is equal to 25, then the standard deviation of the data for this sample is equal to ....

   a 25 b 26 c 27 d 28

7. If the upper limit of the confidence interval for a sample mean is 31.96 with a confidence level of 95% and the arithmetic mean of the sample is equal to 30 and the standard deviation of the sample is 7, then the sample size is equal to ....

   a 25 b 36 c 49 d 64

8. If the mean of a statistical group μ satisfies the inequality in a sample size of 36, it satisfies the inequality $36 - 1.96 \times \frac{5}{6} < \mu < 36 + 1.96 \times \frac{5}{6}$ using a confidence level of 95%, then the standard deviation of the sample is equal to ....

   a 1.96 b 5 c 6 d 36

9. Using 95% confidence level we calculated for the mean of a sample of 100 people was (50 ± 2) kilograms, then the expected sample size if we want to reduce the error rate to 1 kilogram while maintaining the same level of confidence equals ....

a) 200 b) 250 c) 300 d) 400

Second: Answer the following

1. You have a sample of 50 students at a university who have scored on a particular test. The mean score in the sample is 75 and the standard deviation is 10. Calculate the 95% confidence interval for the mean score in the population

2. A sample of 100 employees was taken and the average weekly working hours were found to be 38 hours and the standard deviation was 4 hours. Calculate the 95% confidence interval for the average weekly hours worked.

3. A sample of 49 students was taken, and their average score was found to be 72 and the standard deviation was 6. Calculate the 95% confidence interval for the students' average scores.

4. A sample of 100 customers was taken, and the average bill value was found to be 250L.E and the standard deviation was 20 L.E. Calculate the 95% confidence interval for the average bill value.

5. The average sleep duration in a sample of 400 people is 7.2 hours and the standard deviation is 1.1 hours. Calculate the 95% confidence interval for the number of hours of sleep.

6. A sample of 15 companies was taken, and it was found that the average annual profits are 250,000 L.E and the standard deviation is 3000 L.E Calculate the 95% confidence interval for average annual profits.

Table of areas under the standard normal distribution curve

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0. 1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2160	0.2224
0.6	0.2259	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3815	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998