Benha University
Faculty of Science

الاحتمالات و الإحصاء

Probability and Mathematical Statistics

Presented by

Mathematics and Computer Science Department

كلية التربية شعبة الرياضيات عام

الفرقة الأولى (مميز)

الاحتمالات و الإحصاء

كود :

أ.م. د/ جمال احمد موسى

2024 - 2025

STATISTICS

Chapter 1

Data Collection, Organization and Presentation

Origin and Uses

The word statistics comes from the modern Latin phrase statisticum collegium (lecture about state affairs), from which came the Italian word statista, which means "statesman" or "politician" (compared to status), and the German Statistik, originally designating the analysis of data about the state. It acquired the meaning of the collection and classification of data generally in the early nineteenth century. Data collection about states and localities continues, largely through national and international statistical services; in particular, censuses provide regular information about the population.

The field of statistics includes methods for gathering data (sampling and experimental design), summarizing data (exploratory data analysis with both graphs and numbers), and statistical inference (making generalizations to a larger population based on observations from a sample).

We are bombarded by information in our everyday lives. Most of us associate “statistics” with the bits of information that appear in news reports. E.g., Imported car sales, the latest poll of the president’s popularity, and the average high temperature for today’s date. Advertisements often claim that data show the superiority of the advertiser’s product.

Research has revealed many fascinating pieces of information about the society and everyday life. E.g.:

1) Drunk driving is responsible for 15,000 to 25,000 deaths per year in the US, a rate that over two years equals the number of Americans killed throughout a decade of fighting in Vietnam.

2) Some theories have suggested that exposure to violence in the media may increase aggression in children and adults; whereas other theorists suggest that it may reduce aggression. Research tends to support the former situation. The more media 1 violence individuals watch as children, the more they will tend to like that individual.
3) Infertility in the US is more widespread than people believe. Estimation is that as many as 20% of American couples who want to have children have difficulty doing so.

The primary purpose of this course is to:

1) Provide medical researchers with sufficient understanding to enable them to read statistical methods and discussions appearing in medical journals intelligently.
2) To provide the means for researchers to undertake the simpler analyses independently if they wish.

3) To improve the common base of understanding which is necessary whenever medical researchers and statisticians interact.

❖ **Definition**

Statistics is the study of how to collect, organizes, analyze, and Interpret data.

❖ **Importance**

Statistics plays an important role in the research:

- Helps answer important research questions and the field of study.
- It is important to understand what tools are suitable for research study
- Statistics plays an important role in the decision-making process.

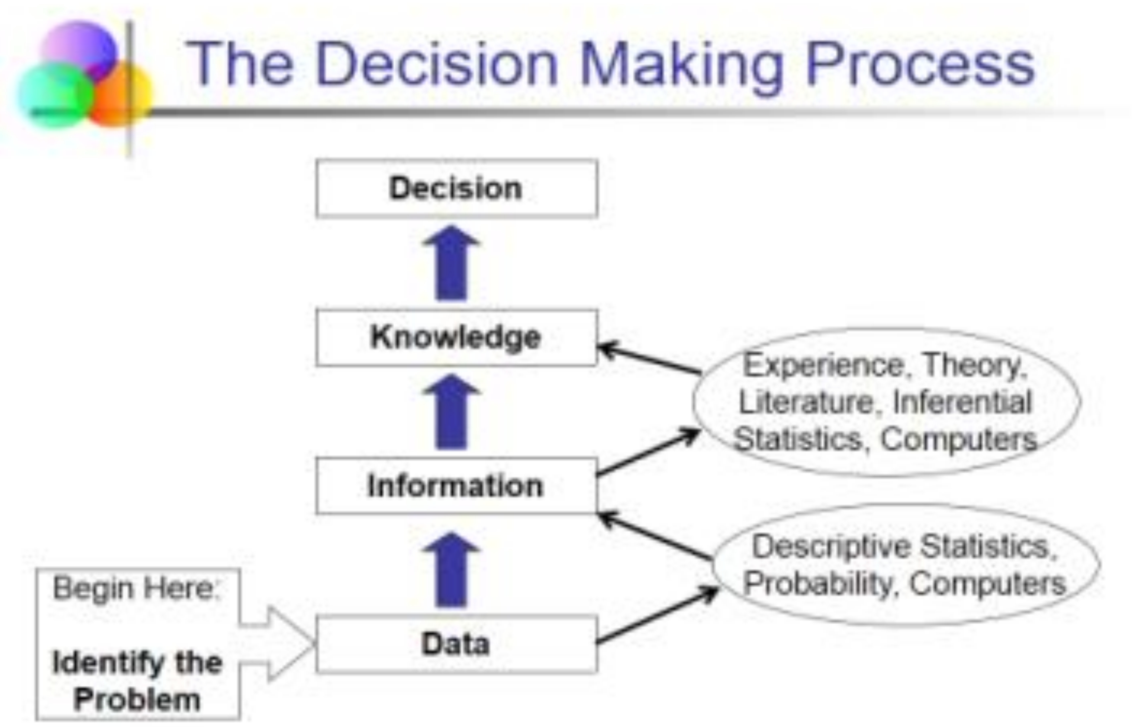


Fig.1The Decision Making Process

❖ **Types of statistics**

There are two approaches to the statistical analysis of data:

1. **Descriptive statistics:** deals with the enumeration, organization, and graphical representation of data.
2. **Inferential statistics:** concerned with reaching conclusions from incomplete information, that is, generalizing from the specific sample.

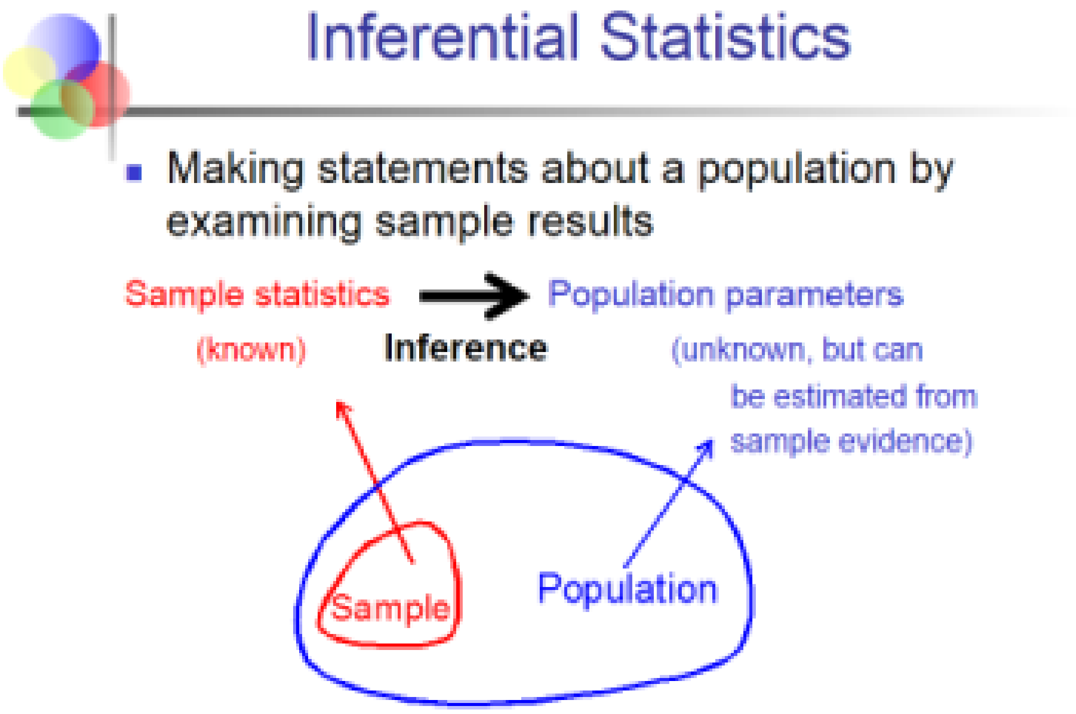


Fig. 2 Inferential Statistics

- ❖ **Basic Terms:**

<u>Biostatistics:</u>

The tools of statistics are employed in many fields – business, education, psychology, agriculture, and economics, to mention only a few. When the data being analyzed are derived from the biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

Population:

Population is the term statisticians use to describe a large set or collection of items that have something in common. A population consists of all elements: individuals, items, or objects whose characteristics are being studied. In medicine, population generally refers to patients or other living organisms. The population being studied is also called the target population.

Sample:

A portion of the population selected for study is referred to as a sample. Suppose our population consists of the weights of all the elementary school children enrolled in a certain county school system. If we collect for analysis the weight of only a part of our population of weights, that is, we have a sample.

Variable:

A variable is a characteristic under study that assumes different values for different elements. Some examples of variables include diastolic blood pressure, heart rate, the heights of adult males, the weights of preschool children, stage of bladder cancer patients, pain score of patients postoperatively, etc.…

Data:

Data are the values you get when you measure a variable. For example, 32 years (for the variable age), or female (for the variable sex) as shown in Table (1).

Table 1.1 Variables and data

The variables ...

... and the data.

	Mrs Brown	Mr Patel	Ms Manda
Age	32	24	20
Sex	Female	Male	Female
Blood type	O	O	A

Information:

Information is a translation of the data to meaningful knowledge Example: Your blood pressure is 190/95 ………. You are Hypertensive

Observation or measurement:

The variable's value for a single element is called an observation or measurement.

Data set:

A data set is a collection of observations on a variable. A typical data set is often represented with a matrix of information. Each row represents an individual or unit, while each column represents a variable

Parameter:

It is a descriptive measure derived from a population. Usually, we do not know the parameter's value; consequently, we estimate it from the sample.

For example, if we want to know the mean age of all Egyptian bladder cancer cases. This mean will be calculated from information collected from all bladder cancer cases in Egypt it is impossible to calculate)

<u>Statistic:</u>

It is a descriptive measure from a sample.

For example, if we want to know the mean age of all Egyptian bladder cancer cases. We take a portion of bladder cancer cases and calculate the mean age.

❖ **Types of samples:**

Broadly speaking there are two ways of sampling:

1. ***Probability sampling***

A probability sample uses a random process (a process governed by chance) to guarantee that each unit of the population has a specified chance of selection.

There are several versions of probability sampling:

a. **Simple random sampling:** A sample selected in such a way that each member of the population has an equal known chance of being selected in the population. It is the process of enumerating every unit of the accessible population and then selecting the sample at random.

b. Systematic random sampling: A sample involves selection by a periodic process.

c. Stratified random sampling: A sample involves dividing the population into subgroups (called strata) according to characteristics such as sex and race and taking a random sample from each of these "strata".

E.g.: In studying the incidence of bladder cancer in Egypt, we stratify the populations according to gender (male, female), then sample from each stratum separately. This will give us equal numbers in each stratum and will yield estimates of bladder cancer for males and females that have comparable precision.

d. Cluster sampling: taking a random sample of natural groupings (clusters) of individuals in the population. Cluster sampling is very useful when the population is widely dispersed, and it is impractical or costly to list and sample from all its elements.

2. ***Nonprobability sampling:***

Nonprobability sampling designs are more practical than probability designs for clinical research projects.

There are three main nonprobability sampling designs:

1) Consecutive sampling involves taking every patient who meets the selection criteria over a specified time interval or number of patients. Consecutive sampling is the best of the nonprobability techniques and is very often practical.

2) Convenience sampling is the process of taking those members of the accessible population who are easily available. Convenience sampling is widely used in clinical research because of its obvious advantages in cost and logistics.

3) Judgmental sampling involves hand-picking from the accessible population those individuals judged most appropriate for the study.

❖ Types of variables and scales of measurement

The purpose of this lecture is to introduce the different kinds of data collected in medical research. Investigators collect information and generally want to transform them into tables; graphs or summary numbers such as percentages or means. It does not matter whether the observations are made on people, animals, or objects. What does matter is the kind of observation made and how the characteristic observed is measured, because these features determine the types of tables, graphs, and summary statistics that best communicate the observations to someone else. It also decides on the type of test statistics to use in making conclusions about the observations.

There are several ways to classify the variables. They may be defined as **quantitative** (metric) variables or **qualitative** (categorical) variables.

Quantitative variables: A quantitative variable is a variable that can be measured in the usual sense, e.g. obtain measurements on the heights of adult males, the weights, and ages of patients seen in a clinic. Measurements made on quantitative variables convey information regarding the amount.

Quantitative variables are either:

> **Discrete** (only take values from some discrete set of possible values, e.g., number of patients admitted to the hospital) or

Continuous (take values from a continuous range of possible values, although the recorded measurements are rounded, e.g., weight, height, hemoglobin levels, etc..).

Qualitative variables:

Categorical variables can either be ordinal or nominal.

Ordinal variables: These are grouped variables that are ordered or ranked in increasing or decreasing order: e.g., severity of a disease: severe, moderate, mild

Nominal variables: The groups in these variables do not have an order or ranking in them e.g., Sex: Male and Female

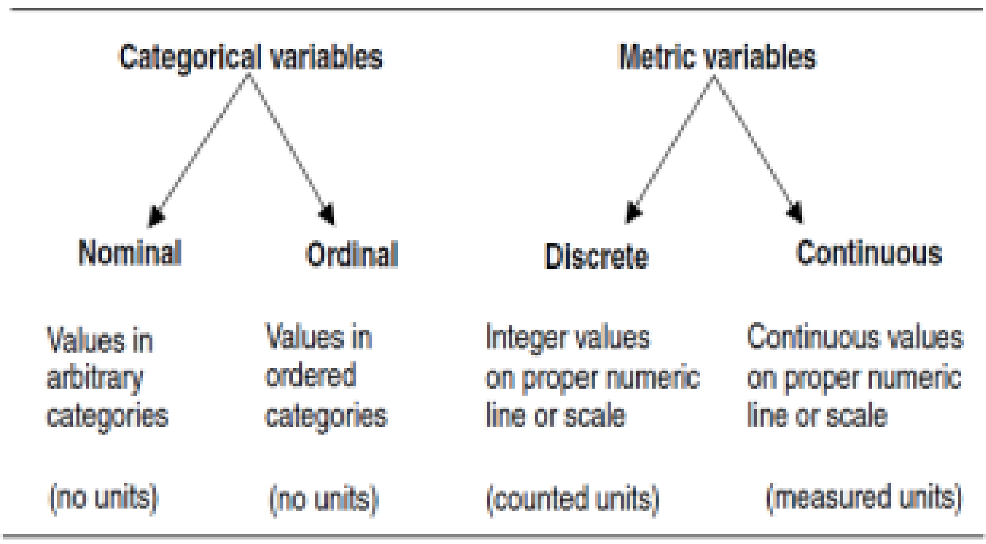
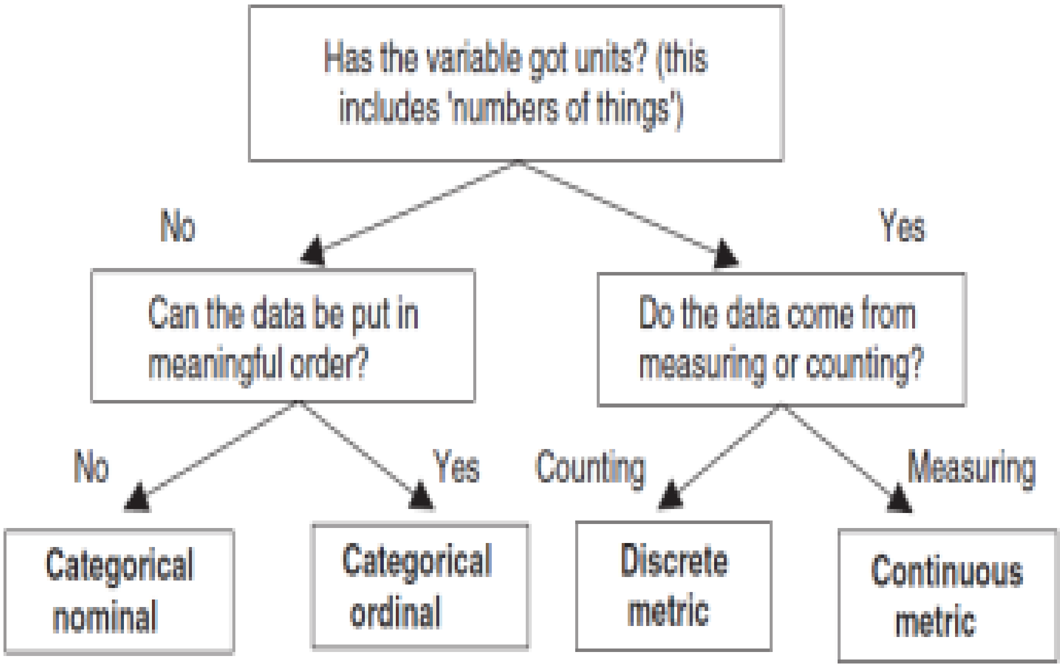


Fig.4.b Classification of variables

An algorithm to help identify variable type



Example

Number	SEX	AGE	Academic Performance	Pulse
1	Male	23	Good	70
2	Female	26	V. Good	68
3	Male	31	Good	78
4	Male	28	Good	72
5	Female	27	Excellent	66

- Sex is a qualitative nominal variable

- Age is a metric (quantitative) continuous variable

- Academic Performance is a qualitative ordinal variable

- Pulse is a metric (quantitative) discrete variable

❖ **Presentation of Data**

After data collection, computer data entry, and analysis, the data must be presented in an easy–to–understand way. The purposes of data presentation are:

- Find out the common finding,

- Find out group variations

<u>Proper data presentation</u>

Tables: When details of data are needed.

Graphs: When only impressions are needed.

Parameters: Precise mathematical summary, useful for comparison.

I. Tables

3. Simple tables showing single variable

a. Tables with data on qualitative variables (nominal) (e.g., percent distribution of the enrolled nursing students by sex) as shown in Table 1

b. Table with data on quantitative variable (continuous) (e.g. percent distribution of the enrolled nursing students by age)

Table 1: Percent distribution of the enrolled nursing students by sex

Sex	Number	Percent
Males Females	11 7	61.1 38.9
Total	18	100.0

4. Contingency tables or cross-tabulation of two variables.

In such tables, two variables are used. In the following example, two variables are presented: education and receiving antenatal care (ANC)

Table 2: **Percent distribution of mothers according to education and receiving ANC for live births born during the last 5 years**

Education	Received ANC		Did not receive ANC		Total	
	No.	%	No.	%	No.	%
Non-educated	64	40.0	96	60.0	160	100.0
Educated	216	90.0	24	10.0	240	100.0
Total →	280	70.0	120	30.0	400	100.0

The table should fulfill the following characteristics

1- No. of the table

2- Title of the table describing its contents

3- Suitable number of rows (4-12).

4- Title for each column and each row

5- Totals

6- Meaningful percent from row or column
7- Make sure that the tables and text refer to each other (through the table number)
8- Not everything displayed in the table needs to be mentioned in the text

II. Graphical representation

1. Pie chart

- A pie chart is a circular chart (pie-shaped); it is split into segments to show percentages or the relative contributions of categories of data.
- A pie chart gives an immediate visual idea of the relative sizes of the shares of a whole.

The following is an example of a pie chart of the Family planning methods used in the studied community

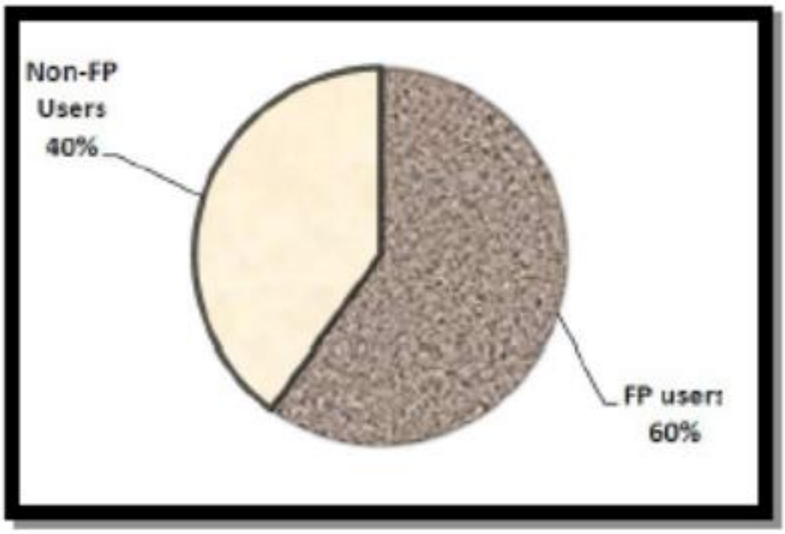


Fig.5 Contraceptive prevalence in the studied community

2. Bar Charts

A bar chart consists of parallel, usually vertical bars with their lengths corresponding to the frequency or percentage of each value. The bars are separated from each other by a space to reflect on the categorical aspect of the variable.

Example:

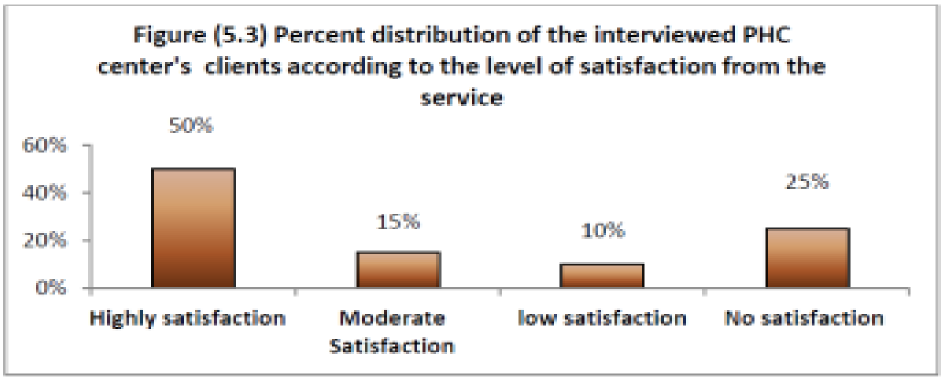


Fig.6 Percent distribution of the interviewed PHC centers clients according to the level of satisfaction with the service

N.B Bar charts and Pie charts are often used for qualitative (category) data

3. Map presentation of Data:

Map presentation of data can be applied at different levels, i.e., within different geographical zones or governorates of one country, within regions between countries in the same region, or global presentation between different worldwide countries.

Example

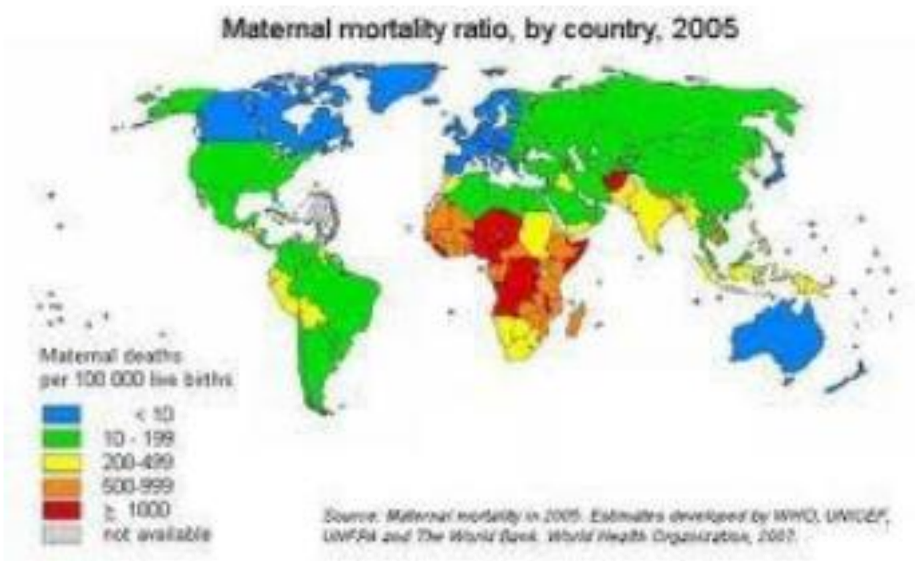


Fig.7 Maternal mortality ratio, by the country, 2005

4. Line charts

- Show values of one variable vs. time

- Time is traditionally shown on the horizontal axis

Example

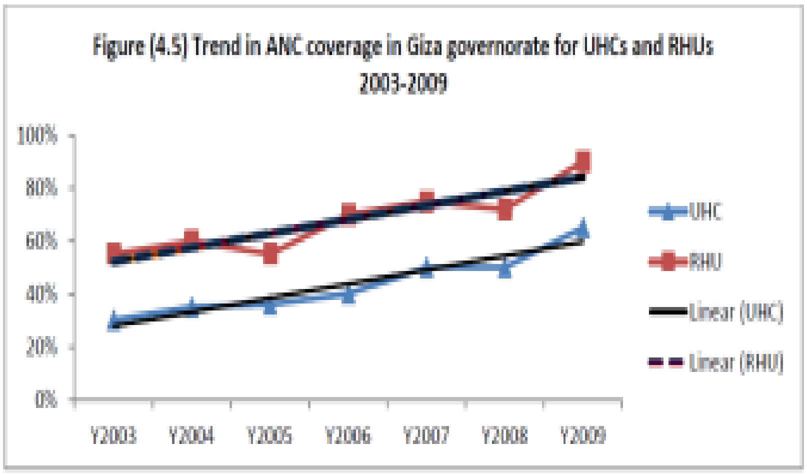
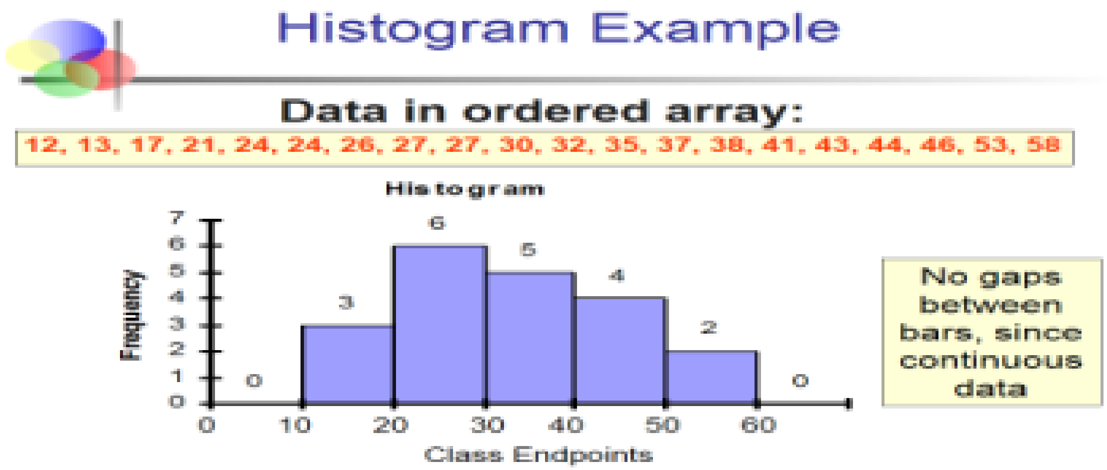


Fig. 8 Trend in antenatal care coverage in Giza for urban health care centers (UHC) and Rural Health care centers (RHC) (2003-2009)

5. Histogram

- It is appropriate for continuous variables (interval).
- It is like a bar chart, but in the histogram, the bars are placed side by side.
- The bar length represents the percent (frequency) falling within each interval.
- Each histogram has a total area of 100%.

Example:



6. Scatter Diagrams

- One variable is measured on the vertical axis and the other variable is measured on the horizontal axis

Example:

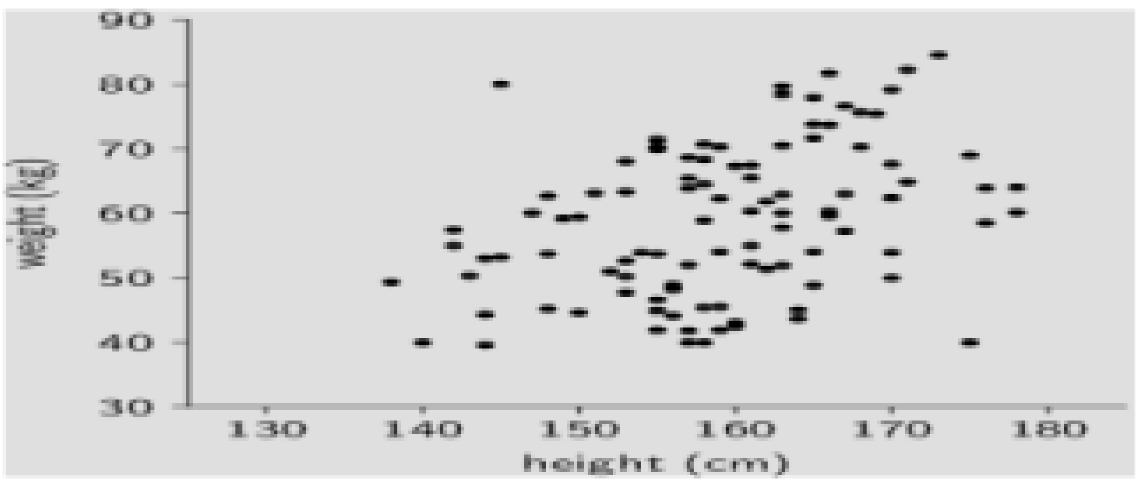


Fig.9 Scatter diagram showing the relationship between height in centimeters (cm) and weight in kilograms (kg).

III. Parameters

• A descriptive value for a population is called a **parameter** and a descriptive value for a sample is called a **statistic**.

• For Qualitative variables: proportion and ratios are used.

$$\text{Ratio} = \frac{\text{Part a}}{\text{Part b}}$$

$$\text{Male to Female Ratio} = \frac{11}{7} \cong 1.6$$

$$\text{Proportion} = \frac{\text{Part}}{\text{Total}} \times 100$$

$$\text{Proportion of Males} = \frac{11}{18} \times 100 = 61\%$$

• For Quantitative variables: **measures of central tendency** and **measures of dispersion** (variation) are used.

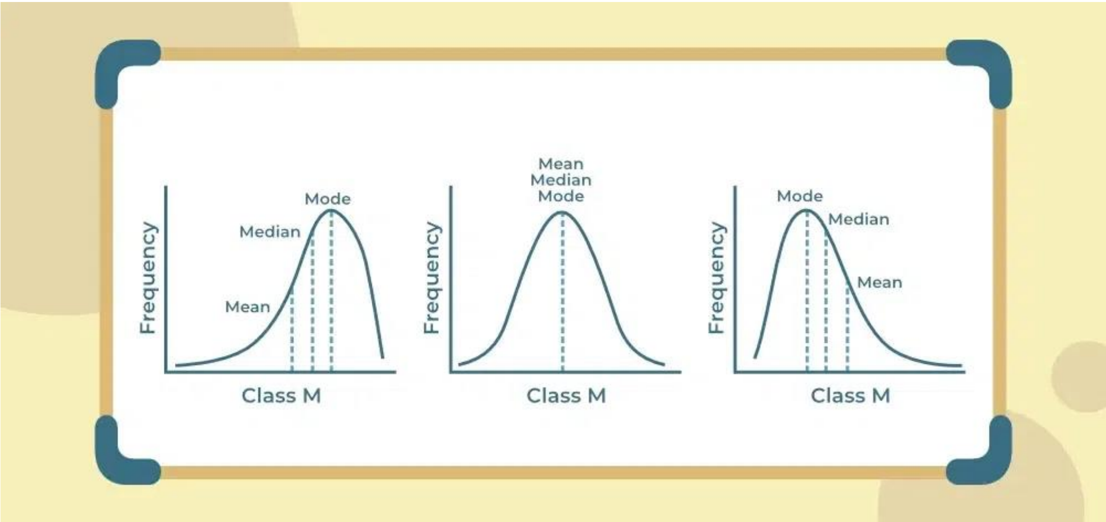
Chapter 2

Measures of Central Tendency and Measures of Dispersion.

- These represent a value around most of the data cluster around.

- The mean, median, mode, and midrange are statistical measures that show the average characteristics or tendency of the group.

A. Measures of Central Tendency



1. Mean:

a. Mean for Ungrouped Data

The Arithmetic mean is the sum of all the values, divided by the number of values.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example: Five women in a study on lipid-lowering agents are aged 52, 55, 56, 58, and 59 years.

Add these ages together:

52 + 55 + 56 + 58 + 59 = 280

Now divide by the number of women: 56

So, the mean age is 56 years.

b. Mean for Grouped Data

Mean ($\bar{x}$) is defined for the grouped data as the sum of the product of observations (x_i) and their corresponding frequencies (f_i) divided by the sum of all the frequencies (f_i).

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \, x_i}{\sum_{i=1}^{n} f_i}$$

Example: If the values (x_i) of the observations and their frequencies (f_i) are given as follows:

x_i	**4**	**6**	**15**	**10**	**9**
f_i	**5**	**10**	**8**	**7**	**10**

then Arithmetic mean ($\bar{x}$) of the above distribution is given by

$$\bar{x} = (4 \times 5 + 6 \times 10 + 15 \times 8 + 10 \times 7 + 9 \times 10)$$

$$\div (5 + 10 + 8 + 7 + 10)$$

$$\Rightarrow \bar{x} = (20 + 60 + 120 + 70 + 90) \div 40$$

$$\Rightarrow \bar{x} = 360 \div 40$$

$$\Rightarrow \bar{x} = 9$$

properties of Mean (Arithmetic)

There are various properties of Arithmetic Mean, some of which are as follows:

- The algebraic sum of deviations from the arithmetic mean is zero i.e., $\sum(x_i - \bar{x}) = 0$.

- If $\bar{x}$ is the arithmetic mean of observations and
  - **a** is added to each of the observations, then the new mean is given by $\bar{x} = \bar{x} + a$.
  - **a** is subtracted from each of the observations, then the new mean is given by $\bar{x} = \bar{x} - a$.
  - **a** is multiplied by each of the observations, then the new mean is given by $\bar{x} = \bar{x} * a$.
  - each of the observations is divided by **a**, then the new arithmetic mean is given by $\bar{x} = \bar{x} / a$
- Uniqueness: For a given set of data there is one and only one mean.
- Simplicity: Easy to understand and compute.
- The mean uses all the information available. Every value in the given set of data is used in the computation; it is therefore affected by every value. Extreme values have an influence on the mean and in some cases, can so distort it that it becomes undesirable as a measure of

location. It may not be "typical" when there are extreme values present.

2. Mode

a. Mode of Ungrouped Data

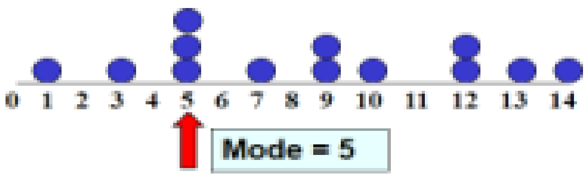
- The value that occurs most often

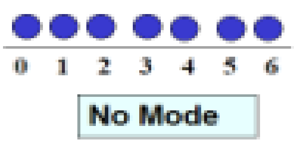
- Used for either numerical or categorical data

- There may be no mode

- There may be several modes

Example





b. Mode of Grouped Data

The formula to find the mode of the grouped data is:

$$Mode = L + \left[\frac{f1 - f0}{2\,f1 - f0 - f2}\right] * h$$

Where,

- **l** is the lower-class limit of the modal class,
- **h** is the class size,
- **f1** is the frequency of modal class,
- **f0** is the frequency of class which proceeds the modal class, and
- **f2** is the frequency of the class which succeeds the modal class.

Example: Find the mode of the dataset which is given as follows.

Class Interval	10-20	20-30	30-40	40-50	50-60
Frequency	5	8	12	16	10

Solution:

The class interval with the highest frequency is 40-50, which has a frequency of 16. Thus, 40-50 is the modal class.

Thus, $\boldsymbol{L} = 40$, $\boldsymbol{h} = 10$, $\boldsymbol{f1} = 16$, $\boldsymbol{f0} = 12$, $\boldsymbol{f2} = 10$

Plugging in the values in the formula $\boldsymbol{Mode} = \boldsymbol{L} + \left[\frac{\boldsymbol{f1 - f0}}{\boldsymbol{2\,f1 - f0 - f2}}\right] * \boldsymbol{h}$ *, we get*

Mode = 40 + (16 – 12)/(2 × 16 – 12 – 10) × 10

$\Rightarrow Mode = 40 + (4/10)\times 10$

$\Rightarrow Mode = 40 + 4$

$\Rightarrow Mode = 44$

3. Median:

a. Median of Ungrouped Data

- Median is the middle number or the number that divides the data into two halves. To calculate the median, the values of the continuous variables must be sorted in an ascending or descending order, after which the middle value is chosen.

- If the number of patients is even, then the median will be the average of the middle two values.

- One of the advantages of a median is that it is not sensitive to extreme values, but the disadvantage would be that all the patients will be ignored except the middle one.

Example 1: If the observations are 25, 36, 31, 23, 22, 26, 38, 28, 20, 32 then the Median is given by

Arranging the data in ascending order:

$$20, 22, 23, 25, \mathbf{26}, \mathbf{28}, 31, 32, 36, 38$$

N = 10 which is even then

Median = the mean of values at $(10 \div 2)^{th}$ *and* $[(10 \div 2) + 1]^{th}$ *position*

$\Rightarrow$ *Median = (Value at* 5^{th} *position + Value at* 6^{th} *position)* $\div 2$

$\Rightarrow$ *Median =* $(26 + 28) \div 2$

$\Rightarrow$ *Median = 27*

Example 2: If the observations are 25, 36, 31, 23, 22, 26, 38, 28, 20 then the Median is given by

Arranging the data in ascending order:

$$20, 22, 23, 25, \mathbf{26}, 28, 31, 36, 38$$

N = 9 which is odd then

Median = Value at $[(9 + 1) \div 2]^{th}$ *position*

$\Rightarrow$ *Median = Value at 5th position*

$\Rightarrow$ *Median = 26*

b. Median of Grouped Data

$$Median = L + \frac{N/2 - C_f}{f} \times h$$

Where,

- **L** is the lower limit of the median class,
- **N** is the total number of observations,
- C_f is the cumulative frequency of the preceding class,
- **f** is the frequency of each class, and
- **h** is the class size.

Example: Calculate the median for the following data.

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	10	12	8	5

Solution: *Create the following table for the given data.*

Class	*Frequency*	*Cumulative Frequency*
10 – 20	*5*	*5*
20 – 30	*10*	*15*

Class	Frequency	Cumulative Frequency
30 – 40	12	27
40 – 50	8	35
50 – 60	5	40

As N = 40 and N/2 = 20,

Thus, 30 – 40 is the median class.

$$L = 30, \quad C_f = 15, \quad f = 12, \quad and \;\; h = 10$$

Putting the values in the formula $\boldsymbol{Median = L + \frac{N/2 - C_f}{f} \times h}$

Median = 30 + ((20 – 15)/12) × 10

⇒ Median = 30 + (5/12) × 10

⇒ Median = 30 + 4.17

⇒ Median = 34.17

So, the median value for this data set is 34.17

Properties:

1) Uniqueness: As is true with the mean, there is only one median for a given set of data.

2) Simplicity: The median is easy to calculate.

3) It is not drastically affected by extreme values as in the case of the mean.

4) Since it uses the middle value of the data set. The median is not a very reliable measure.

B. Measures of Dispersion

They are used to measure the extent of variations between observations,

Example: Suppose two factories are producing the batteries. From each factory, 10 batteries are drawn to test for the lifetime (in hours). These lifetimes are

Factory 1: 10.1, 9.9, 10.1, 9.9, 9.9, 10.1, 9.9, 10.1, 9.9, 10.1

Factory 2: 16, 5, 7, 14, 6, 15, 3, 13, 9, 12.

The mean lifetimes of the two factories are both 10. However, looking at the data, it is obvious that the batteries produced by Factory 1 are much more reliable than those produced by Factory 2. This implies other measures for measuring the —dispersion or —variation of the data are required.

- Range
- Standard deviation
- Percentiles
- Quartiles and inter-quartile range.

B.1 Range

- Simplest measure of variation

- Difference between the largest and the smallest observation

Range = (largest value of the data) − (smallest value of the data).

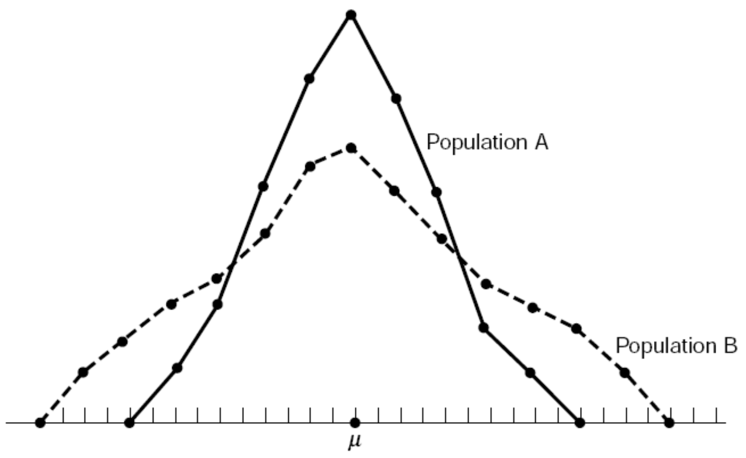
Example:

Range of lifetime data for factory 1 = 10.1 − 9.9 = 0.2

Range of lifetime data for factory 2 = 16 − 3 = 13

⇒ The range of battery lifetimes for Factory 1 is much smaller than for Factory 2.

$$R = x_L - x_S$$



B.2 Variance, Standard Deviation and Coefficient of Variation:

Population deviation about the mean: $x_i - \mu \qquad , i = 1,2,\dots,N$

Sample deviation about the mean: $X_i - \bar{x} \qquad , i = 1,2,\dots,n$

Intuitively, the population deviation and the sample deviation can measure how far the data is from the "center" of the data. Then, **Population variance** and **Sample variance** are the sum of the square of the population deviation and sample deviation,

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

and

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1},$$

respectively. The **population standard deviation** and **sample standard deviation** are the square root of population variance and sample variance:

$$\sigma = \sqrt{\sigma^2} \quad \text{and} \quad s = \sqrt{s^2}$$

respectively.

Large sample variance or sample standard deviation implies the data are "dispersed" or highly varied.

$$\text{Note: } \sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = \sum_{i=1}^{n} X_i - n\frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0$$

Example:

$$s^2(\text{ factory. } 1) = \frac{(10.1-10)^2 + (9.9-10)^2 + \cdots + (10.1-10)^2}{10-1} = 0.0111$$
$$s^2(\text{ factory. } 2) = \frac{(16-10)^2 + (5-10)^2 + \cdots + (12-10)^2}{10-1} = 21.1111$$

⇒ The sample variance of battery lifetimes for factory 2 is 190 times larger than the one for factory 1.

The sample standard deviation for the data from factories 1 and 2 are

$s1 = \sqrt{0.01111} = 0.1054$ and $s2 = \sqrt{21.1111} = 4.5946$

respectively.

Coefficient of Variation:

The coefficient of variation is another useful statistic for measuring the dispersion of the data. The coefficient of variation is

$$C.V. = \frac{\sigma}{\mu} * 100 \quad or \quad C.V. = \frac{s}{\bar{x}} * 100$$

Then

$$\boldsymbol{C.V.1 = \frac{0.1054}{10} * 100 = 1.054 \quad and\ C.V.2 = \frac{4.5946}{10} * 100 = 45.94}$$

However, the coefficient of variation for Factory 1, is much smaller than the one for Factory 2.

Standard deviation (SD) is used for “normally distributed” data to provide information on how much the data varies around their mean.

SD indicates how much a set of values is spread around the average/Mean. A range of one SD above and below the mean abbreviated to

Mean+/- 1 SD includes 68.2% of the data.

Mean+/-2 SD includes 95.4% of the data.

Mean +/-3 SD includes 99.7% of the data

EXAMPLE

Let us say that a group of patients enrolling for a trial had a normal distribution for weight. The mean weight of the patients was 80 kg. For this group, the SD was calculated to be 5 kg.

1 SD below the average is 80 – 5 = 75 kg.

1 SD above the average is 80 + 5 = 85 kg.

Mean+/- 1 SD will include 68.2% of the subjects, so 68.2% of patients will weigh between 75 and 85 kg.

95.4% will weigh between 70 and 90 kg (Mean+/-2 SD).

99.7% of patients will weigh between 65 and 95 kg (Mean +/- SD). 17

Fig. 10 Normal distribution of weights of patients in a trial with mean 80kg and SD 5kg

If we have two sets of data with the same mean but different SDs, then the data set with the larger SD has a wider spread than the data set with the smaller SD. For example, if another group of patients enrolling for the trial has the same mean weight of 80 kg but an SD of only 3, ±1 SD will include 68.2% of the subjects, so 68.2% of patients will weigh between 77 and 83 kg. Compare this with the example above.

Fig. 11 Normal distribution of weights of patients in a trial with mean 80kg and SD 3kg

SD should only be used when the data have a normal distribution. However, means and SDs are often wrongly used for data which are not normally distributed.

18

A simple check for a normal distribution is to see if 2 SDs away from the mean are still within the possible range for the variable. For example, if we have some length of hospital stay data with a mean stay of 10 days and a SD of 8 days then:

Mean – 2 × SD = 10 – 2 × 8 = 10 – 16 = -6 days.

This is clearly an impossible value for length of stay, so the data cannot be normally distributed. The mean and SDs are therefore not appropriate measures to use.

Examiners may ask what percentages of subjects are included in 1, 2 or 3 SDs from the mean.

B.3 Percentiles

• Percentile is a score value above which and below which a certain percentage of values in the distribution fall

• Percentiles are points that divide all the measurements into 100 equal parts. • The observations should be first arranged from the lowest to the highest values, just like when finding the median, which is the 50th percentile.

• Example: A sample of 100 newborn children was weighted, and the data were arranged from the lowest to the highest.

The value (score) of the 95th percentile was found to be 3.8 Kg. Such findings indicate that 95% of the children had a body weight of less than 3.8 Kg and 5% had a body weight of more than 3.8 kg.

The value (score) of the 5th percentile was found to be 2.5 Kg.

Such findings indicate that 5% of children had a body weight of less than 2.5 kg and 95% had body weight more than 2.5 kg.

19

Children are considered within normal weight if they are between 2.5 Kg - 3.8 Kg or the 5th and 95th percentiles.

B.4 Quartiles

Quartiles divide a rank-ordered data set into four equal parts. The values that separate parts are called the first, second, and third quartiles; and they are denoted by Q1 (lower quartile), Q2 (median), and Q3 (upper quartile)

 Quartiles:

• lower (1st)quartile (25%) = 25% lower & 75% greater

20

• upper (3rd)quartile (75%) = 75% lower & 25% greater

Chapter 3

Correlation and Regression

Definition: A **correlation** exists between two variables when one of them is related to the other in some way.

Linear correlation coefficient:

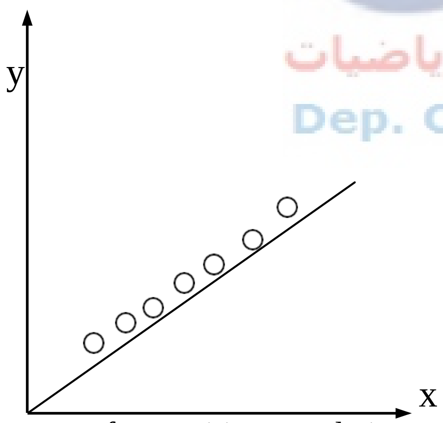
The linear correlation coefficient r measures the strength of the linear relationship between the paired x and y values in a sample.

It's value is computed by

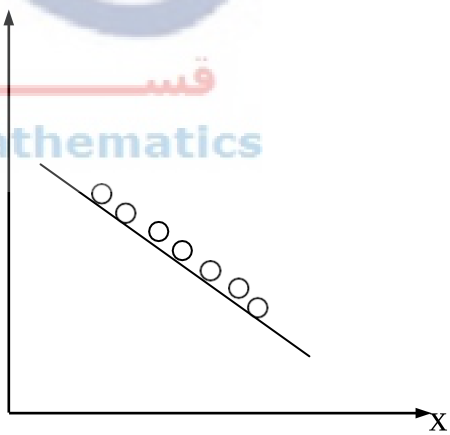
$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

Note that $-1 \le r \le 1$

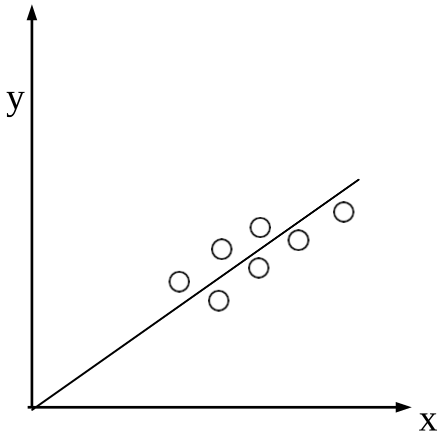
The following figures illustrates several cases



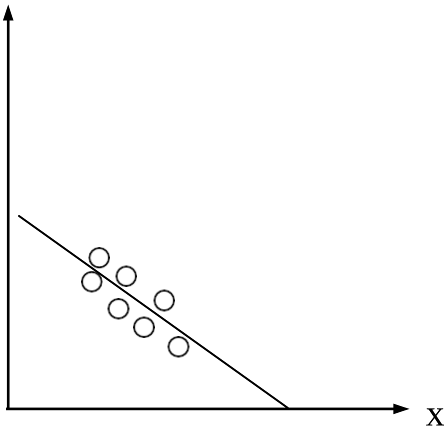
Perfect positive correlation
$r = 1$



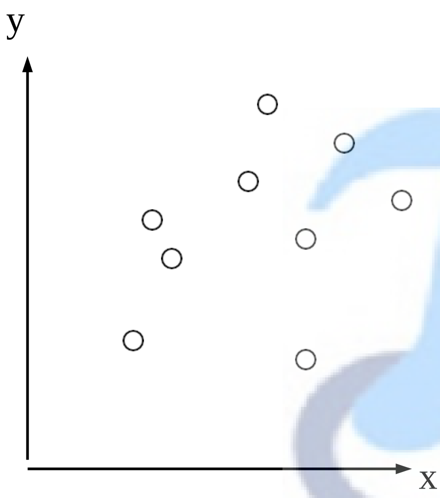
Perfect negative correlation
$r = -1$



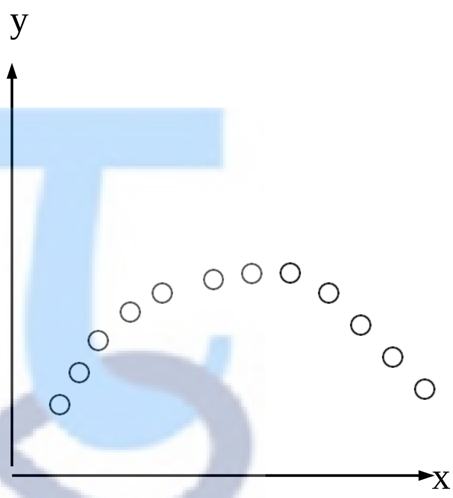
Strong positive correlation
r = .97



Weak negative correlation
r = - .61



No correlation



Nonlinear correlation

Spearman’s rank correlation coefficient

If we have data on two variables for a set of items and we want to see if these variables are related we can test them for correlation. Correlation comes in two forms:

Positive correlation – as one variable increases in value, so does the other
Negative correlation – as one variable increases in value, the other decreases in value

As an example we will use the Spearman‟s rank correlation coefficient to comment on the relationship between the size of a locust and the length of its wings. So in this example the set of items is the locusts in our sample and the two variables we are looking at for each locust are body length and wing length.
When we use the Spearman‟s rank coefficient to calculate a correlation, we first have to rank the data for each of the variables.

Locust	Body length x (mm)	Wing length y (mm)	Rank x	Rank y	*d*	d^2
1	15	7	9	10	-1	1
2	10	6	10	9	1	1
3	80	32	1	1	0	0
4	45	23	6	5	1	1
5	53	19	5	6.5	-1.5	2.25
6	62	29	2	2	0	0
7	35	18	8	8	0	0
8	41	19	7	6.5	0.5	0.25
9	58	28	4	3	1	1
10	61	27	3	4	-1	1

If two equal values appear e.g. for Rank y at rank 6, then both are given the rank 6.5 (halfway between rank 6 & 7) and no values are given rank 6 or 7. Next we calculate the difference between the ranks = *d* and then square this = d^2

Then we find the sum of all the d^2 values

$\sum d^2$ = 1+1+0+1+2.25+0+0+0.25+1+1 = 7.5

Now we can calculate the correlation coefficient using the formula:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 7.5}{7(7^2-1)} = 1 - \frac{45}{336} = 0.8661$$

An r_s value of +1 shows perfect positive correlation

An r_s value of -1 shows perfect negative correlation

An r_s value of 0 shows no correlation

Generalized correlation coefficient :

In the case of nonlinear regression curve, we use the generalized correlation coefficient r, where

$$r = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$$

where $\hat{y}$ is the estimated value of y for a given value of x obtained from the regression curve of y on x.

we use the above formula to obtain nonlinear correlation coefficient (which measure how well a nonlinear regression curve fits the data).

Sometimes r^2 is called the coefficients of determination

$$r^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \overline{y})^2} \ .$$

In the linear regression the formula is

$$r^2 = \frac{b^2 \left(\sum x^2 - n\,\overline{x}^2 \right)}{\sum y^2 - n\,\overline{y}^2}$$

Example : Fit a least-squares parabola having form $y = a + b\,x + c\,x^2$ to the following data, then find a nonlinear correlation coefficient between these variables, assuming the parabolic relationship obtained.

x	1.2	1.8	3.1	4.9	5.7	7.1	8.6	9.8
y	4.5	5.9	7.0	7.8	7.2	6.8	4.5	2.7

One can see that the required least - squares parabola has the form

equation $y = 2.588 + 2.065\,x - .211\,x^2$

For, x = 1.2 , $y = 2.588 + 2.064\,(1.2) - .211\,(1.2)^2 = 4.762$

y	$\hat{y}$	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$	$(y - \bar{y})$	$(y - \bar{y})^2$
4.5	4.762	-1.038	1.072	-1.3	1.69
5.9	5.621				
7.0	6.962				
7.8	7.640				
7.2	7.503				
6.8	6.613				
4.5	4.741				
2.7	2.561				
			21.02		21.40

Similarly other estimated values may be obtained as shown in the following table:

Thus $r^2 = \dfrac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2} = \dfrac{21.02}{21.40}$

So $\quad r = 0.99$

Which is strong nonlinear correlation coefficient.

Regression

Very often in practice a relationship is found to exist between two (or more) variables. And one whishes to express this relationship in mathematical form by determining an equation connecting the variables. A first step is the collection of data showing corresponding values of the variables.

A next step is to plot the points on a rectangular coordinate system. The resulting set of points is sometimes called a scatter diagram.

In Fig. (1), for example the data appear to the approximated well by a straight line. In Fig. (2) there exist a nonlinear relationship between the variables. In Fig. (3) there appears to be no relationship between the variables.

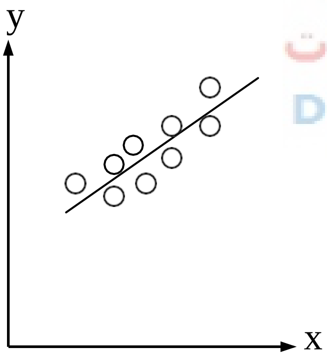


Fig. (1)

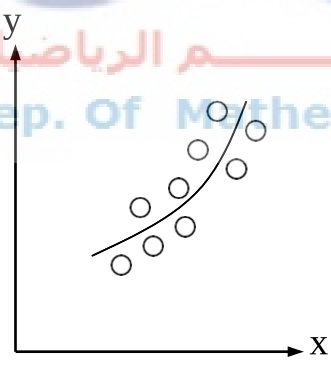


Fig. (2)

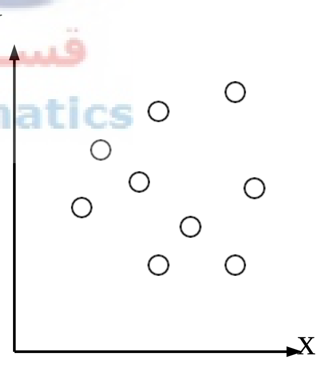


Fig. (3)

Definition : The general problem of finding equations of approximating curves which fit given data is called **curve fitting**.

In particular the type of equation is often suggested from the scatter diagram. Thus for Fig. (1) we could use a straight line $\quad y = a + b\,x$, while for Fig. (2) we could try a parabola

$$y = a + b\,x + c\,x^2$$

sometimes it helps to plot scatter diagram in terms of transformed variables. Thus for example if log y vs. x leads to a straight line we would try $\log y = a + b\,x$. as an equation for the approximating curve. Another possible equations (among many) used in practice we mention the following

- * - $\quad y = \dfrac{1}{a + b\,x} \quad$ or $\quad \dfrac{1}{y} = a + b\,x$

- ** - $\quad y = a_1\, a_2^x \quad$ or $\log y = \log a_1 + (\log a_2)\,x \quad$ or $\log y = a + b\,x$

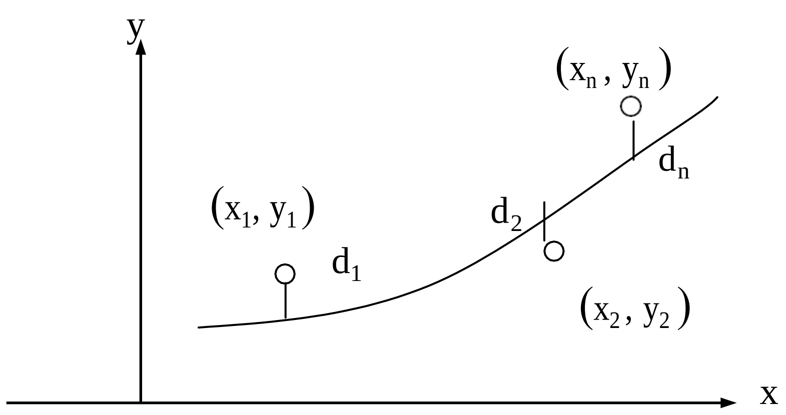
- *** - $\quad y = a_1\, x^b \quad$ or $\log y = \log a_2 + b \log x \quad$ or $\log y = a + b \log x$

to decide which curve should be used, it is helpful to obtain scatter diagram of transformed variables.

If a scatter diagram of $\dfrac{1}{y}$ versus x shows a linear relationship the equation has the form - * -, while if log y verses x shows a linear relationship the equation has the form - ** -, but if log y verses log x shows a linear relationship the equation has the form - *** -

The least-squares line :

Definition: Of all curves approximating a given set of data points, the curve having the property that $d_1^2 + d_2^2 + \ldots + d_n^2 = a$ minimum, is the best fitting curve.



By using the above definition we can show that the least-square line approximating the set of points $(x_1, y_1), \ldots, (x_n, y_n)$ has the equation $y = a + b\,x$ where the constants a and b are determined by solving the equations

$$\sum y = n\,a + b \sum x$$
$$\sum x\,y = n \sum x + b \sum x^2$$

which are called the normal equations for the least - square line.

The least-squares parabola :

The least square parabola which fits a set of samples points is given by

$$y = a + b\,x + c\,x^2$$

where a, b, c are determined from the normal equations.

$$\sum y = n\,a + b \sum x + c \sum x^2$$

$$\sum x\,y = a \sum x + b \sum x^2 + c \sum x^3$$

$$\sum x^2\,y = a \sum x^2 + b \sum x^3 + c \sum x^4$$

Example :

Fit a least - square line to the data

x :	1	3	4	6	8	9	11	14
y :	1	2	4	4	5	7	8	9

Solution

The equation of the line is $y = a + b\,x$

The normal equation are

$$\sum y = n\,a + b \sum x$$

$$\sum x\,y = a \sum x + b \sum x^2$$

we construct the following table to get the needed sums

x	:	1	3	4	……	$\sum x$	:	56
y	:	1	2	4	……	$\sum y$	:	40

$x\,y$	:	1	6	16	… …	$\sum x\,y$: 364
x^2	:	1	9	16	… …	$\sum x^2$: 524

then substituting in the normal equations we have

$$\begin{aligned} 8a + 56b &= 40 \\ 56a + 524b &= 364 \end{aligned} \quad \Rightarrow \quad a = \frac{6}{11}\ , \quad b = \frac{7}{11}$$

or directly from the calculator

$$a = .545 \qquad , b = .636$$

Then the required least - square line is

$$y = .545 + .636\ x$$

Example :

The following table gives experimental values of the pressure P of a given mass of gas corresponding to various values of V. According to thermodynamic principles a relationship having the form $P\,V^{\gamma} = c$, where γ and c are constants, should exist between the variables. Find the values of γ and c, estimate P when V = 100 cm^3.

(cm^3) V	:	890	1013	1186	1454	1944	3179
(kg/cm^3) P	:	4.303	3.480	2.644	1.997	1.350	0.710

Solution

$P\ V^{\gamma} = c \Rightarrow$ log p + γ log V = log c

or $\quad$ log P = log c - γ log V, $\quad$ setting

log V = x $\quad$, $\quad$ log c = a

log p = y $\quad$, $\quad$ - γ = b

then the last equation can be written in the form

$$y = a + b\,x$$

now, construct the following table :

V $\quad$: $\quad$ 890 $\qquad$ 1013 … …

P : 4.303 3.480 … …

$x = \log V$: 2.9494 3.0055 … …

$y = \log p$: 0.6338 0.5416 … …

Then using the calculator, we have

$$a = 4.756 \Rightarrow c = 57016.4$$

$$b = -1.4 \quad \Rightarrow \gamma = 1.4$$

$$\Rightarrow P\,V^{1.4} = 57016.4$$

$$\text{at } V = 100 \Rightarrow P = \frac{57016.4}{(100)^{1.4}} = 90.36$$

Example :

Fit a least square parabola having the form $y = a + b\,x + c\ x^2$ to the data in the following table :

x	1.2	1.8	3.1	4.9	5.7	7.1	8.6	9.8
y	4.5	5.9	7.1	7.8	6.8	4.5	4.5	2.7

Solution

The normal equation are

$$\sum y = n\,a + b \sum x + c \sum x^2$$

$$\sum x\,y^2 = a \sum x + b \sum x^2 + c \sum x^3$$

$$\sum x^2\,y = a \sum x^2 + b \sum x^3 + c \sum x^4$$

The work involved in computing the sums can be arranged as in the following table :

x	y	x^2	x^3	x^4	$x\,y$	$x^2\,y$
1.2	4.5	1.44	1.73	2.08	5.4	6.48
1.8	5.9	3.83	5.83	10.49	10.62	19.12
⋮						
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum x\,y$	$\sum x^2\,y$
42.2	46.4	291.2	2275.3	18971.9	230.42	1449.0

Then the normal equations becomes

$$8\,a + 42.2\,b + 291.2\,c = 46.4$$

$$42.2\,a + 291.2\,b + 2275.3\,c = 230.42$$

$$291.2\,a + 2275.3\,b + 18971.9\,c = 1449$$

solving $a = 2.588$,

$b = 2.065$,

$c = -.21$

or directly from the calculator, then the required least - square parabola has the form

$$y = 2.588 + 2.065\,x - .211\,x^2$$

Example :

For the following data, fit a least square curve having the form

$$y = \frac{1}{a + b\log x}$$

x	:	0.1	1	10	100
y	:	1	1/2	1/3	1/4

Solution

$$y = \frac{1}{a + b\log x} \Rightarrow \frac{1}{y} = a + b\log x \quad \text{, let } v = \log x \text{ , } u = \log y$$

we have $u = a + b\,v$

x	:	0.1	1	10	100
y	:	1	1/2	1/3	1/4
v = log x	:	-1	0	1	2
u = log y	:	1	2	3	4

using calculator, we have a = 2, b = 1, so

$$\therefore\ y = \frac{1}{2 + \log x}$$

There exist some nonlinear models which can be transformed into a straight line or a parabola, by using a suitable transformation. Some of these nonlinear models and the required transformation are listed in the following table :

Nonlinear model	If we take the transformation	Then we have
$y=\dfrac{1}{a+b\,x^2}$	$u=\dfrac{1}{y}$, $v=x^2$	$u=a+b\,v$
$y=\dfrac{1}{a+b\log x}$	$u=\dfrac{1}{y}$, $v=\log x$	$u=a+b\,v$
$y=a+(\log x)^2$	$v=(\log x)^2$	$y=a+b\,v$
$y=a+b\sqrt{x}$	$v=\sqrt{x}$	$Y=a+b\,v$
$y=a+b(\log x)+c(\log x)^2$	$y=\log x$	$y=a+b\,v+c\,v^2$
$y=\dfrac{1}{a+b(\log x)^2}$	$u=\dfrac{1}{y}$, $v=(\log x)^2$	$u=a+b\,v$
$y=\dfrac{1}{a+b\sqrt{x}}$	$u=\dfrac{1}{y}$, $v=\sqrt{x}$	$u=a+b\,v$
$y=\dfrac{1}{a+b(\log x)+c(\log x)^2}$	$u=\dfrac{1}{x}$, $v=\log x$	$u=a+b\,v+c\,v^2$

Exercises

(1) The following table shows the ages x and systolic blood pressures y of 12 women

Age (x)	56	42	72	36	63	47	55	49	38	42	68	60
Blood pressure (y)	147	125	160	118	149	128	150	149	115	140	152	155

determine the least squares regression line of y on x and estimate the blood pressure of a women whose age is 45 years.

(2) The number y of bacteria per unit volume present in a culture after x hours is given in the following table :

number of hours (x)	0	1	2	3	4	5	6

Number of bacteria per unit volume (y)	22	47	65	65	92	132	275

Fit a least-square curve having the form $y = a\,b^x$ to the data and estimate the value of y when x = 0.7.

(3) Fit a least squares parabola, $y = a + b\,x + c\,x^2$ to the following data

x	0	1	2	3	4	5	6
y	2.4	2.1	3.2	5.6	9.3	14.6	21.9

(4) Fit a least-squares curve, $y = \dfrac{1}{\alpha + \beta\,(\log x)^3}$ to the following data

x	.01	0.1	1	10
y	0.1	0.333	0.5	1

(5) Fit a least - squares curve, $y = \dfrac{1}{a + b\sqrt{x}}$ to the following data

x	0	1	4	9
y	-1/2	1/8	1/18	1/28

Introduction to Probability Theory

Prepared by

Department of Mathematics

Introduction

In Probability we commonly use the terms population and sample because these terms are central to our study, we define them now.

A population is the complete collection of all elements (scores, people, measurements, and so on) to be studied. The collection is competing in the sense that it includes all subjects to be studied.

A sample is a sub collection of elements drawn from a population.

The tools of Probability are employed in many fields, business, education, psychology, agriculture, and economics, to mention only a few. When the data being analyzed are derived from the biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

Contents

Chapter 1

Principles Theory of Probability

Statistical Experiment:

We use the word experiment to describe any process that generates a set of data.

A sample example of a statistical experiment is the tossing of a coin.

Sample Space:

The set of all possible out-come of a statistical experiment is called the sample space and is denoted by S.

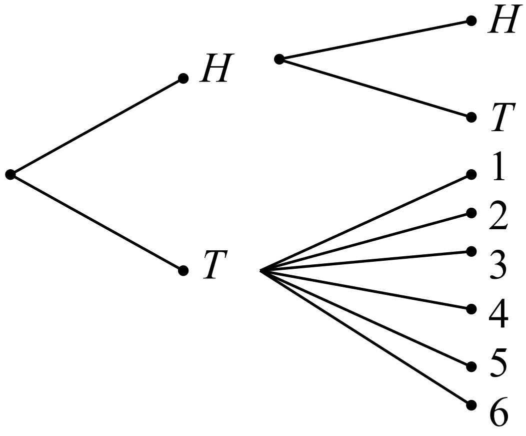
Example: If we toss a coin, the sample space is $S = \{H, T\}$ when H and T correspond to heads and tails.

Example: An experiment consists of flipping a coin and then flipping it a second time if a head occurs, if a tail occurs on the first flip, and then a die is tossed once. List the elements of the sample space.

Solution

We construct the following tree diagram

First Outcome **Second Outcome** **Sample Space**



HH

HT

$T1$

$T2$

$T3$

$T4$

$T5$

$T6$

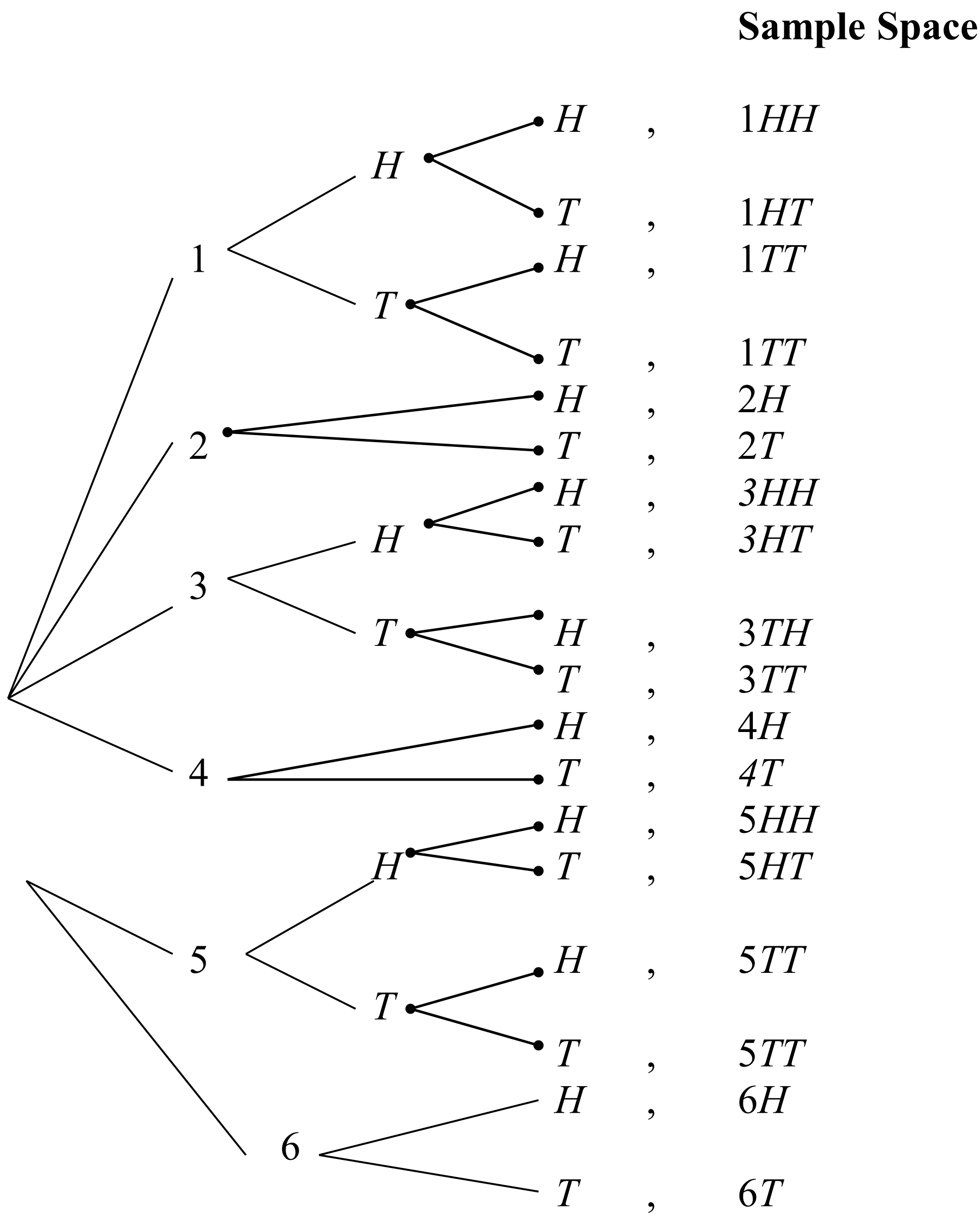
By proceeding along all paths, we see that the sample space is

$$S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$$

Example: An experiment consists of tossing a die and then flipping a coin once if the number on the die is even. If the number on the die is odd, the coin is flipped twice. Construct a tree diagram to show the elements of the sample space *S*.

Solution

First, we construct the following tree diagram



Remark:

Sample spaces with a large or infinite number of sample points are best described by a rule. For example if the possible out comes of an experiment are the set of all cities in the word with a population over one million

$S = \{ x \mid x \text{ is a city with a population over one million } \}$

Events: An event is a subset of a sample space.

Example: Toss a pair of dice, then describe the event *A* that the total number of points rolled with the pair of dice is τ .

Solution

$S=\{(x,y) \mid x=1,2,...,6; y=1,2,...,6\}$ or

$$
\begin{aligned}
S=\{ &(1,1),(1,2)....,(1,6)\\
&(2,1),(2,2),...,(2,6)\\
&(3,1),(3,2),...,(3,6)\\
&(4,1),(4,2),...,(4,6)\\
&(5,1),(6,2),...,(6,6)\\
&(6,1),(6,2),...,(6,6)
\end{aligned}
$$

then *B* can be written as

$$B=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,11)\}$$

Note that: the sample space *S* is the **sure event** and the empty set Φ is called the impossible event.

By using a set operations on events in *S*, we can obtain other events in *S*. Thus if *A* and *B* are events, then

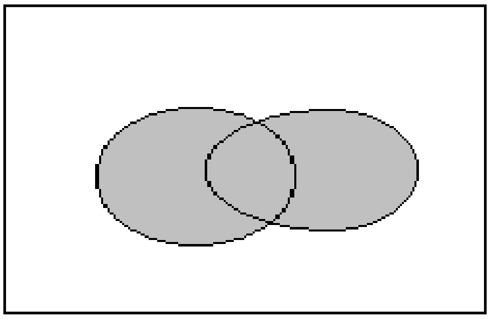
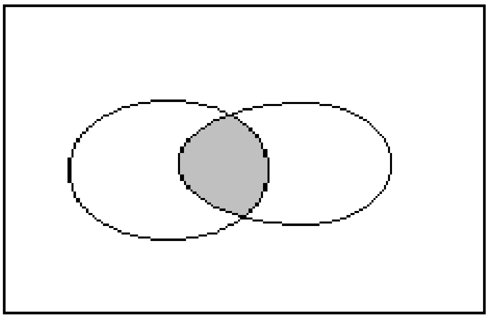
① $A \cup B$ is the event "either *A* or *B* or both".

② $A \cap B$ is the event "both *A* and *B*".

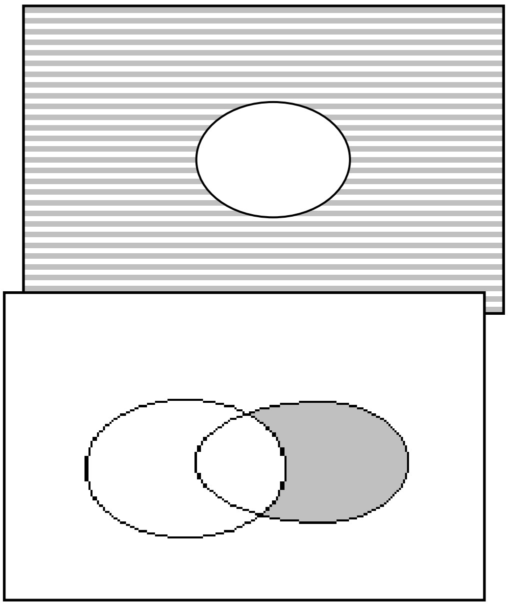
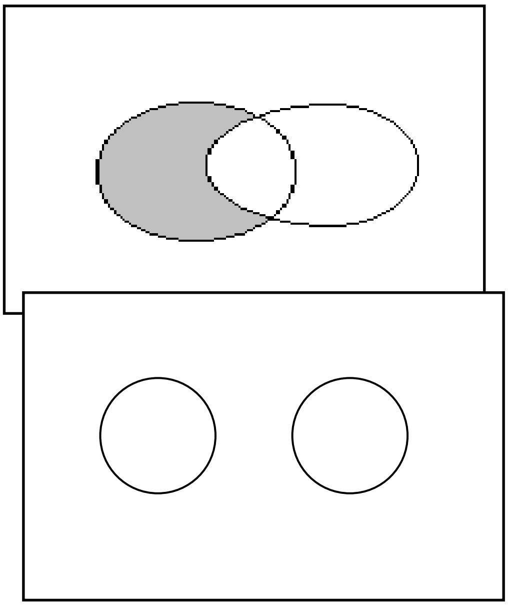
③ A' is the event "not *A*".

④ $A - B$ is the event "*A* but not *B*".

Definition: Two events *A* and *B* are **mutually exclusive** or **disjoint** if $A \cap B = \Phi$

$A \bigcup B$ $\qquad$ $A \bigcap B$

$B - A$ $\qquad$ $A \bigcap B = \Phi$

Example: Toss coin twice and let *A* be the event "at least one head occurs" and *B* the event "the second toss result is a tall", then $S = \{HH, TH, TT\}, A = \{HT, TH, HH\}$,
$B = \{HT, TT\}, A \bigcup B = \{HT, TH, HH, TT\}, A \bigcap B = \{HT\}$,
$A' = \{TT\}, A - B = \{TH, HH\}$

Axioms of probability:

Suppose we have a sample space *S*. To each event *A* in the class *C* of events, we associate a real number $P(A)$, then $P(A)$ is called the probability of the event *A* if the following axioms are satisfied

Axiom 1. $1 \ge P(A) \ge 0$
Axiom 2. $P(S) = 1$
Axiom 3. For any number of mutually exclusive events

$A_1, A_2, \ldots,$ in the class C

$$P(A_1 \cup A_2 \cup A_3 \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

exclusive events A_1, A_2

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Some important theorems on probability:

Theorem: $P(\Phi) = 0$
Proof: Since $A = A \cup \Phi$, then

$$P(A) = P(A \cup \Phi) = P(A) + P(\Phi)$$
$$\Rightarrow P(\Phi) = 0$$

Theorem: If A' is the complement of A then

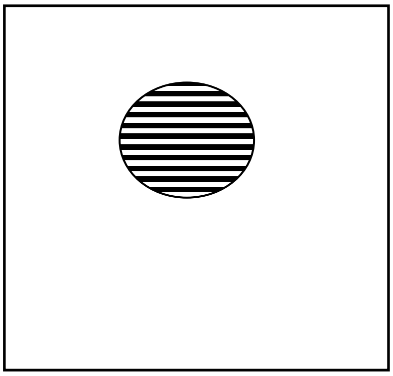
$$P(A')\, 1 - P(A)$$

Proof: Since $S = A \cup A^c$, then

$$P(S) = 1 = P\left[A \cup A^c\right]$$
$$= P(A) + P(A')$$

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

Theorem: If $A \subset B$, then $P(A) \le P(B)$
Proof:



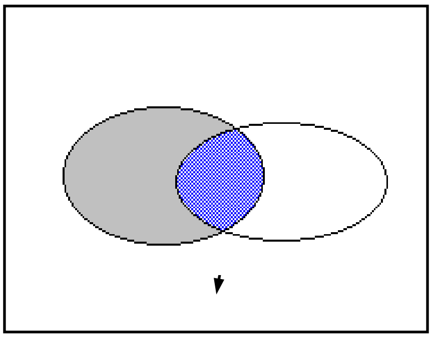
$$B = A \cup (B - A) \text{ , then}$$
$$P(B) = P\left[A \cup (B - A)\right]$$
$$P(B) = P(A) + P(B - A)$$

since $P(B - A) \ge 0$, so

$$P(A) \le P(B)$$

Theorem: For any events A and B

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



Proof:

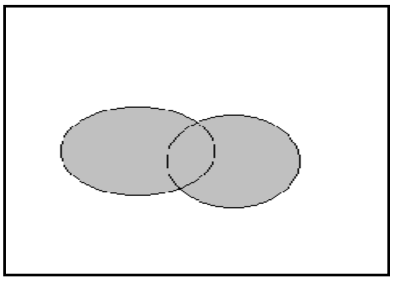
$$A=(A-B)\cup(A\cap B)$$

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

so, $\quad P(A-B)=P(A)-P(A\cap B)$

Theorem: For any event A and B

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



Proof: $\quad A\cup B=B\cup(A-B)$

$$P(A\cup B)=P[B\cup(A-B)]$$

$$=P(B)+P(A-B)$$

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

Then

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

Problem: For any three events A , B and C, show that

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

Proof: Let $A\cup B=D$

$$P[A\cup B\cup C]=P(A\cup D)=P(A)+P(D)-P(A\cap D)$$

But $P(D)=P(B\cup C)=P(B)+P(C)-P(B\cap C)$, and

$$P(A\cap D)=P[A\cap(B\cup C)]=P[(A\cap B)\cup(A\cap C)]$$

$$=P(A\cap B)+P(A\cap C)-P(A\cap B\cap C)$$

Then

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

Corollary:

If $A_1, A_2, ..., A_n$ is a partition of a sample space S, then

$$P(A_1\cup A_2\cup ...\cup A_n) = P(A_1)+P(A_2)+...+P(A_n)$$
$$= P(S) = 1$$

Theorem: If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}$$

Problems: Prove that

$$P(A_1\cup A_2 ...\cup A_n) \le P(A_1)+P(A_2)+...+P(A_n)$$

which is known Booles' inequality.

Examples

1. Given that $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ and $P(A\cap B) = \frac{1}{4}$, find

(i) $P(A\cup B)$ (ii) $P(A')$
(iii) $P(B')$ (iv) $P(A'\cap B')$
(v) $P(A'\cup B')$ (vi) $P(B\cap A')$
(vii) $P(B\cap A')$

Solution

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

$$= \frac{3}{8}+\frac{1}{2}-\frac{1}{4} = \frac{5}{8}$$

(ii) $P(A') = 1 - P(A) = 1 - \frac{3}{8} = \frac{5}{8}$

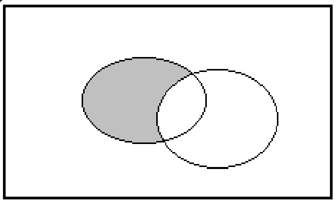
(iii) $P(B') = 1 - P(B) = 1 - \frac{1}{2} = \frac{1}{2}$

(iv) $P(A' \cap B') = P\left[(A \cup B)^c\right] = 1 - P(A \cup B)$

$$= 1 - \frac{5}{8} = \frac{3}{8}$$

(v) $P(A' \cup B') = P\left[(A \cap B)'\right] = 1 - P(A \cap B)$

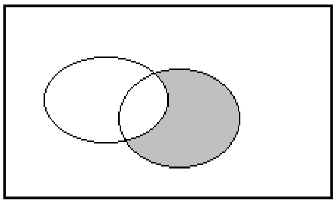
$$= 1 - \frac{1}{4} = \frac{3}{4}$$



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

$$= \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$$

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



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

$$= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

2. Given that $P(A \cup B) = \frac{3}{4}$, $P(A') = \frac{2}{3}$, $P(A \cap B) = \frac{1}{4}$, find $P(A)$, $P(B)$, $P(A \cap B)$

Solution

$$P(A) = 1 - P(A') = 1 - \frac{2}{3} = \frac{1}{3}$$

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

$$\frac{3}{4} = \frac{1}{3} + P(B) - \frac{1}{4}$$

$$P(B) = \frac{3}{4} - \frac{1}{3} + \frac{1}{4} = \frac{2}{3}$$

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

$$= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

3. Given that $P(A)=\frac{1}{2}, P(A\cup B)=\frac{3}{4}$ and $P(B')=\frac{5}{8}$, find
$P(A'\cap B'), P(A'\cup B')$ and $P(A'\cap B)$.

4. Let $P(A)=0.9$ and $P(B)=0.8$ show that
$P(A\cap B)\geq 0.7$

Solution

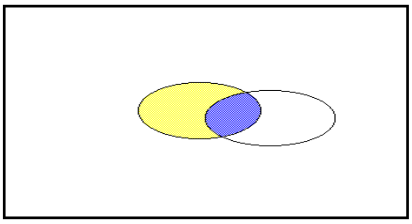
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

But $0\leq P(A\cup B)\leq 1$, so

$$P(A)+P(B)-P(A\cap B)\leq 1 \Rightarrow$$
$$P(A\cap B)\geq P(A)+P(B)-1=0.9+0.8-1$$
$$P(A\cap B)\geq 0.7$$

5. Show that $P(A) = P(A\cap B')+P(A\cap B)$

Solution



Since $A=(A\cap B')\cup(A\cap B)$
and $(A\cap B')\cap(A\cap B) = \Phi$
Then

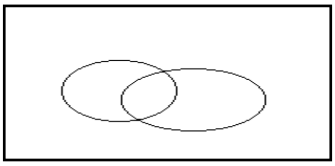
$$P(A)=P(A\cap B') + P(A\cap B)$$

6. Let A and B any events in S, Express the following events in terms of A and B

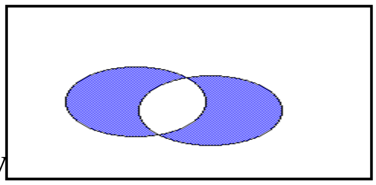
(i) At least one of the events occurs.
(ii) Exactly one of the two events occurs.

Solution

(i) $A\cup B$



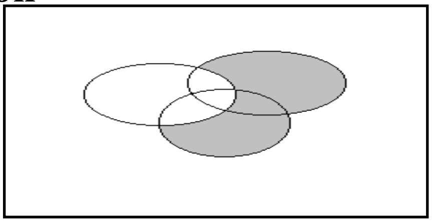
(ii) $(A\cap B')\cup(B\cap A')$
or $(A-B)\cup(B-A)$



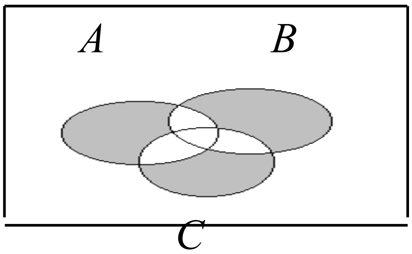
7. Let A, B and C be any three ev
Express the following events in terms of these events
(i) Either B or C occurs, but not A.
(ii) Exactly one of the events occurs.

(iii) Exactly two of the events occurs.

Solution

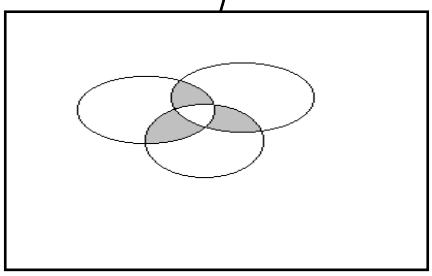


(i) $(B \cup C) \cap A'$



(ii) $(A \cap B' \cap C') \cup (B \cap A' \cap C') \cup (C \cap A' \cap B')$

$(A \cap B) \cap C'$



(iii) $((A \cap B) \cap C') \cup ((A \cap C) \cap B') \cup ((B \cap C) \cap A')$

8. Let A, B and C and three events in S.

If $P(A) = P(B) = \frac{1}{4}, P(C) = \frac{1}{3}, P(A \cap B) = \frac{1}{8}, P(A \cap C) = \frac{1}{6}$ and $P(B \cap C) = 0$, find $P(A \cup B \cup C)$?

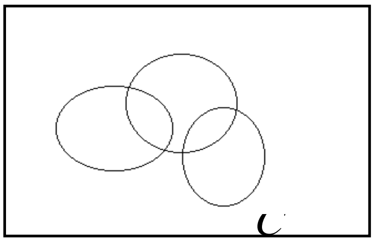
Solution

Since $P(B \cap C) = 0 \Rightarrow (B \cap C) = \Phi$

So,

$(A \cap B \cap C) = A \cap (B \cap C) = A \cap \Phi = \Phi \Rightarrow P(A \cap B \cap C) = 0$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)$$
$$-P(A\cap B)-P(A\cap C)-P(B\cap C)$$
$$+P(A\cap B\cap C)$$
$$=\frac{1}{4}+\frac{1}{4}+\frac{1}{3}-\frac{1}{8}-\frac{1}{6}-0+0=\frac{13}{24}$$



Example: A coin is weighted such that $P(H)=3P(T)$ find $P(H)$ and $P(T)$

Solution

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

$$P(S)=1=P(H)+P(T)$$

$$1=3P(T)+P(T)$$

$$4P(T)=1\Rightarrow\ P(T)=\frac{1}{4}$$

$$P(H)=\frac{3}{4}$$

Conditional Probability:

Definition: The conditional probability of *B*, given *A*, denoted by $P(B|A)$, is defined by

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

Example: The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$, the probability that it arrives on time is $P(A) = 0.83$, and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane

(a) Arrives on time given that it departed on time.
(b) Departed on time given that it has arrived on time.

Solution

$P(D) = 0.83, \ P(A) = 0.82, \quad P(D \cap A) = 0.78$

(a) The probability that a plane arrives on time given that it departed on time is

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.78}{0.83} = 0.94$$

(b) The probability that a plane departed on time given that it has arrived on time is

$$P(D|A) = \frac{P(D \cap A)}{P(A)} = \frac{0.78}{0.82} = 0.95$$

Independent Events:

Definition:
Two events *A* and *B* are independent if and only if

$$P(B|A) = P(B) \quad \text{and} \quad P(A|B) = P(A)$$

otherwise, *A* and *B* are dependent.

Theorem:

Two events A and B are independent if and only if

$$P(A \cap B) = P(A)\,P(B)$$

Example: The probability that A hits a target is $\frac{1}{4}$ and the probability that B hits it is $\frac{2}{5}$. what is the probability that the target will be hit if A and B each shoot at the target?

Solution

$P(A)=\frac{1}{4}$, $P(B)=\frac{2}{5}$, we seek $P(A \cup B)$.

The event that A hits the target is independent of the event that B hits the target:

$$P(A \cap B) = P(A)\,P(B)$$

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

$$= P(A) + P(B) - P(A)\,P(B)$$

$$= \frac{1}{4} + \frac{2}{5} - \frac{1}{4} . \frac{2}{5} = \frac{11}{20}$$

Theorem (1): If A and B are independent then A and B' are independent, and A' and B are independent.

Theorem (2) : If A and B are independent then A' and B' are independent.

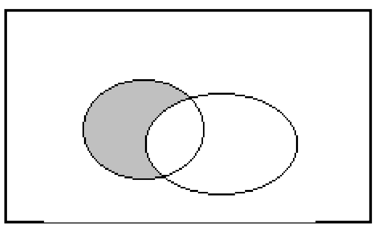
Proof of theorem (1):

Since A and B are independent then

$$P(A \cap B) = P(A)\,P(B)$$

we seek $P(A \cap B') = P(A)\,P(B')$

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



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

$$=P(A)-P(A)\,P(B)$$

$$=P(A)\left[1-P(B)\right] \qquad\qquad A\cap B'$$

$$=P(A)\,P(B')$$

Proof of theorem (2):

Since A and B are independent

$$P(A\cap B)=P(A)\,P(B)$$

We seek $P(A'\cap B')=P(A')P(B')$

$$P(A'\cap B')=P\left[(A\cup B)'\right]=1-P(A\cup B)$$

$$=1-\left[P(A)+P(B)-P(A\cap B)\right]$$

$$=1-P(A)-P(B)+P(A)\,P(B)$$

$$=(1-P(A))-P(B)\left[1-P(A)\right]$$

$$=(1-P(A))(1-P(B))$$

$$=P(A')P(B')$$

Theorem: Three events A, B and C are independent if :

(i) $P(A\cap B)=P(A)\,P(B)$

$P(A\cap C)=P(A)\,P(C)$

$P(B\cap C)=P(B)\,P(C)$

i.e., if events are pair wise independent, and

(ii) $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

Problem: Give an example to show that three events may be pair wise independent but not independent themselves.

Solution

Let a pair of fair coins be tossed, and consider the events

A = {heads on the first coin}

B = { " " "second"}

C = { " " "exactly one coin}

Then $S = \{HH, HT, TH, TT\}$

$$P(A) = P\{HH, HT\} = \frac{1}{2}$$

$$P(B) = P\{HH, TH\} = \frac{1}{2}$$

$$P(C) = P\{HT, TH\} = \frac{1}{2}$$

$$P(A \cap B) = P(HH) = \frac{1}{4}\ , \ P(A \cap C) = P(HT) = \frac{1}{4}$$

$$P(B \cap C) = P(TH) = \frac{1}{4}$$

Then the events are pair wise independent but

$P(A \cap B \cap C) = P(\Phi) = 0$, so

$$P(A \cap B \cap C) = P(\Phi) = 0 \neq P(A)\,P(B)\,P(C)$$

and so the three events are not independent.

Example: Let A and B be events with $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$,

$P(A \cap B) = \frac{1}{4}$, find

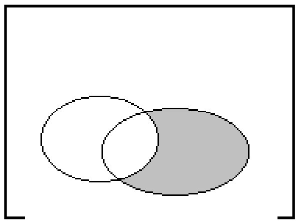
(i) $P(A|B)$ (ii) $P((A\cup B)|A')$

(iii) $P(A'|B'-A')$

Solution

(i) $P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{1/4}{1/3}=\dfrac{3}{4}$

(ii) $P[(A\cup B)|A']=\dfrac{P[(A\cup B)\cap A']}{P(A')}$

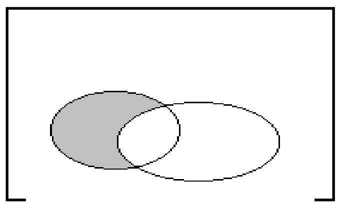


$$=\frac{P[(A'\cap A)\cup(A'\cap B)]}{P(A')}$$

$$=\frac{P[\Phi\cup A'\cap B]}{P(A')}=\frac{P(A'\cap B)}{1\ P(A)}=\frac{P(B-A)}{1-P(A)}$$

$$=\frac{P(B)-P(A\cap B)}{1-P(A)}=\frac{1/3-1/2}{1-1/2}=\frac{1/12}{1/2}=\frac{1}{6}$$

(ii) $P[A'|B'-A']=\dfrac{P[A'\cap(B'-A')]}{P(B'-A')}$



$$=\frac{P(\Phi)}{P(A\cap B')}=0$$

$A-B=B'-A'$

Example: Let A, B, C are independent events then A and $B\cup C$ are independent.

Solution

We seek $P[A\cap(B\cup C)]=P(A)\,P(B\cup C)$

But $P[A\cap(B\cup C)]=P[(A\cap B)\cup(A\cap C)]$

$=P(A\cap B)+P(A\cap C)-P(A\cap B\cap C)$

$$=P()\,P(B)+P(A)\,P(C)-P(A)\,P(B)\,P(C)$$
$$=P(A)\left[\,P(B)+P(C)-P(B)\,P(C)\,\right]$$
$$=P(A)\left[\,P(B)+P(C)-P(B\cap C)\right]$$
$$=P(A)\left[\,P(B\cup C)\right]$$

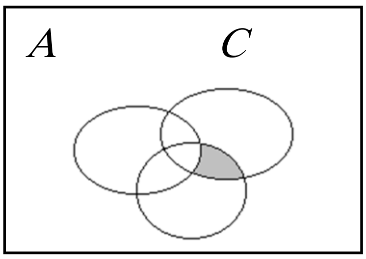
Theorem: Suppose *A*, *B* and *C* are independent events, then any of the $A',B,C;A,B',C,...,A',B',C,....,A',B',C'$ are also independent events.
We shall prove three cases.
If *A*, *B* and *C* are independent, then
(1) A',B,C are independent.
(2) A',B',C " " .
(3) A',B',C' " " .

Case (1):
Since *A*, *B* and *C* are independent, then

$$P(A\cap B)=P(A)\,P(B)$$
$$P(A\cap C)=P(A)\,P(C)$$
$$P(B\cap C)=P(B)\,P(C)$$
$$P(A\cap B\cap C)=P(A)\,P(B)\,P(C)$$

We seek
(i) $P(A'\cap B)=P(A')P(B)$
(ii) $P(A'\cap C)=P(A')P(C)$
(iii) $P(B'\cap C)=P(B')P(C)$
(iv) $P(A'\cap B\cap C)=P(A')P(B)\,P(C)$
(i) and (ii) are left to the reader.
(iii) are given
Now, we went to show that

$$P\left[\,A'\cap B\cap C\,\right]=P(A')P(B)\,P(C)?$$



$$P[\,A' \cap B \cap C\,] = P[\,(B \cap C) - A\,]$$

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

B

$$= P(B)\,P(C) - P(A)\,P(B)\,P(C)$$

$$= \;[1 - P(A)\,]P(B)\,P(C)$$

$(B \cap C) \cap A'$

Case (2): we seek

(i) $P(A' \cap B') = P(A')P(B')$

(ii) $P(A' \cap C) = P(A')P(C)$

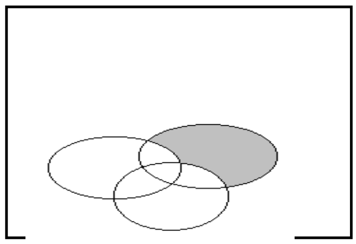
(iii) $P(B' \cap C) = P(B')P(C)$

(iv) $P(A' \cap B' \cap C) = P(A')P(B')P(C)$

(i), (ii) and (iii) are left to the reader

(iv) $P(A' \cap B' \cap C) = P\left[(A \cup B)' \cap C\,\right]$

A



$$= [\,C - A \cup B\,]$$

$$= P(C) - P[\,C \cap (A \cup B)]$$

B

$$= P(C) - P[\,(C \cap A) \cup (C \cap B)]$$

$$= P(C) - P(C \cap B) - P(C \cap B) + P(A \cap B \cap C)$$

$$= P(C) - P(A)\,P(C) - P(B)\,P(C) + P(A)\,P(B)\,P(C)$$

$$= P(C)[1 - P(A) - P(B) + P(B)\,P(C)\,]$$

$$= P(C)[\,(1 - P(A))\,(1 - P(B)\,)]$$

$$= P(C)\,P(A')P(B') = P(A')P(B')P(C)$$

Case (3) : We seek

(i) $P(A'\cap B') = P(A')P(B')$

(ii) $P(B'\cap C') = P(B')P(C')$

(iii) $P(B'\cap C') = P(B')P(C')$

(iv) $P(A'\cap B'\cap C') = P(A')P(B')P(C')$

(i), (ii) and (iii) are left to the reader.
we want to show that $P(A'\cap B'\cap C') = P(A')P(B')P(C')$?

$$P[A'\cap B'C'] = P\left|(A\cup B\cup C)'\right|$$

$$=1-P(A\cup B\cup C)$$

$$=1-[P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$$
$$-P(B\cap C)+P(A\cap B\cap C)]$$

$$=1-P(A)-P(B)-P(C)+P(A)P(B)+P(A)P(C)$$
$$+P(B)P(C)-P(A)P(B)P(C)$$

$$=(1-P(A))(1-P(B))(1-P(C))$$

$$=P(A')P(B')P(C')$$

Example: The probabilities that three men hit a target are respectively $\frac{1}{6},\frac{1}{4}$ and $\frac{1}{3}$. Each shoats once at the target.

(i) Find the probability that exactly one of than hits the target.
(ii) If only one hit the target, what is the probability that it was the first man?

Solution

Consider $A = \{\text{first man hits the target}\}$

$B = \{\text{second " " " " }\}$

and $C = \{\text{ third " " " " }\}$

then $P(A) = \frac{1}{6}, P(B) = \frac{1}{4}$ and $P(C) = \frac{1}{3}$ the three events are independent.

(i) Let $E = \{$ exactly one man hits the target $\}$ then

$$E = (A \cap B' \cap C') \cup (A' \cap B \cap C') \cup (A' \cap B' \cap C)$$

so

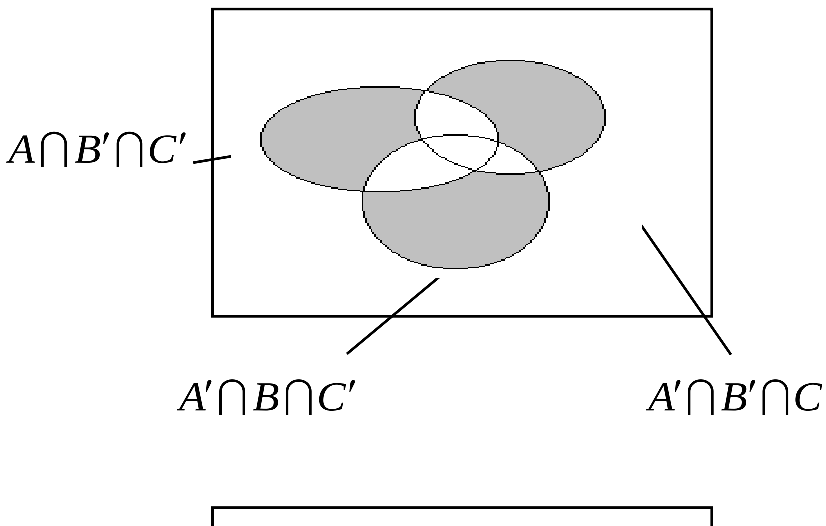
$$P(E) = P(A)P(B')P(C') + P(A')P(B)P(C') + P(A')P(B')P(C)$$

$$= \frac{1}{6}.\frac{3}{4}.\frac{2}{3} + \frac{5}{6}.\frac{1}{4}.\frac{2}{3} + \frac{5}{6}.\frac{3}{4}.\frac{1}{3} = \frac{31}{72}$$

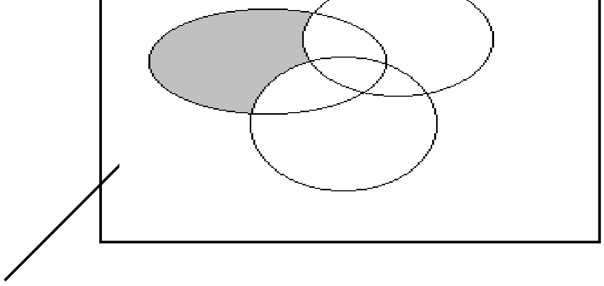
(ii) we see $P(A|E) = \dfrac{P(A \cap E)}{P(E)} = \dfrac{P(A \cap B' \cap C')}{P(E)}$

$$= \frac{1/12}{31/72} = \frac{6}{31}$$

(i)



(ii)



B

$A \cap E = A \cap B' \cap C'$

Problem: Let *A* and *B* be events defined in a sample space *S*. Show that if both $P(A)$ and $P(B)$ are non zero, then events *A* and *B* cannot be both mutually exclusive and independent.

Solution

Let *A* and *B* be mutually exclusive events and $P(A) \neq 0$, $P(B) \neq 0$, then $P(A \cap B) = P(\Phi) = 0$, but $P(A)\,P(B) \neq 0$.
Since $P(A \cap B) \neq P(A)\,P(B)$
A and *B* cannot be independent.

Multiplicative Rules:

Theorem (1): If in an experiment the events *A* and *B* can both occur, then

$$P(A \cap B) = P(A)\,P(B|A)$$

For the above theorem, we can also write

$$P(A \cap B) = P(B)\,P(A|B)$$

Theorem (2) : If in an experiment the events $A_1, A_2, ..., A_n$ can occur, then

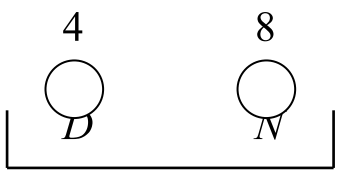
$$P(A_1 \cap A_2 \cap A_3 ... \cap A_n) =$$
$$P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)...P(A_n|A_1 \cap A_2 ... A_{n-1})$$

Example: A lot contains 12 items of which 4 are defective. Three items are drawn at random from the lot one after the other; find the probability that all three are non defective.

Solution

Let D : defective item

N : non defective item



$$P(N_1 \cap N_2 \cap N_3) = P(N_1)P(N_2 | N_1)P(N_3 | N_1 \cap N_2)$$

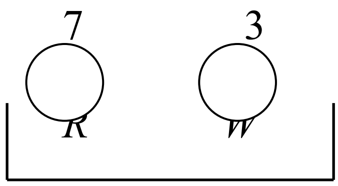
$$= \frac{8}{12} \cdot \frac{7}{11} \cdot \frac{6}{10} = \frac{14}{55}$$

Example: An urn contains 7 red marbles and 3 white marbles. Three marbles are drawn from the urn one after the other. Find the probability that the first two are red and the third is white.

Solution

Let R : read marble

W : white marble



We seek $P(R_1 \cap R_2 \cap W) = P(R_2)P(R_2 | R_1)P(W | R_1 \cap R_2)$

$$= \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{3}{8} = \frac{7}{40}$$

Example: The students in a class are selected at random, one after the other, for an examination. Find the probability that boys and girls in the class alternate, if the class contains 4 boys and 3 girls.

Solution

If the boys and girls are to alternate, then the first student examined must be a boy.

We seek

$$P(B_1 \cap G_2 \cap B_3 \cap G_4 \cap B_5 \cap G_6 \cap B_7)$$

$$= P(B_1)\, P(G_2 | B_1)P(B_3 | B_1 \cap G_2)P(G_4 | B_3)$$

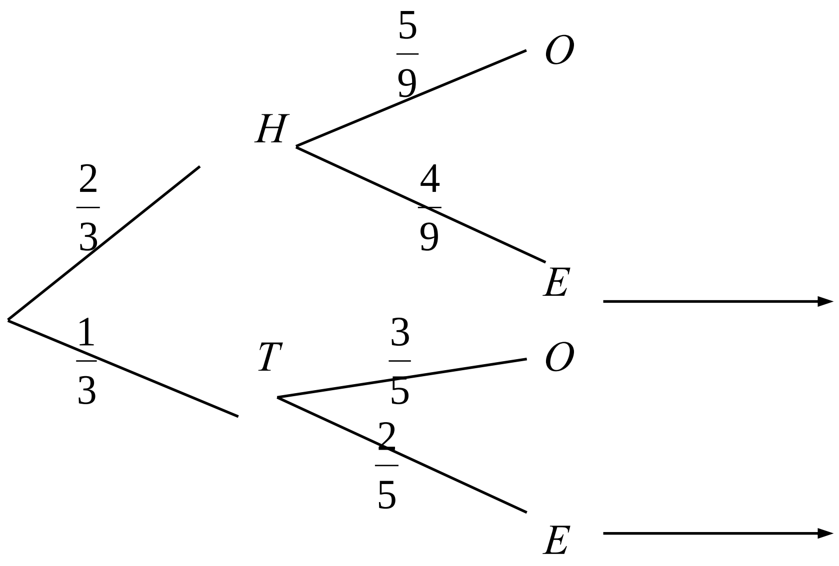
$$P(B_5 | B_1 \cap G_2 \cap B_3 \cap G_4)P(G_6 | \dots)P(B_7 | \dots)$$

$$=\frac{4}{7}.\frac{3}{6}.\frac{3}{5}.\frac{2}{4}.\frac{2}{3}.\frac{1}{2}.\frac{1}{1}=\frac{1}{35}$$

Example: A coin weighted so that $P(H)=\frac{2}{3}$ and $P(T)=\frac{1}{3}$, is tossed. If heads appears, then a number is selected at random from 1 through 9, if tails appears, then a number is selected at random from 1 through 5. Find the probability that an even is selected.

Solution

The three diagrams with respective probability is

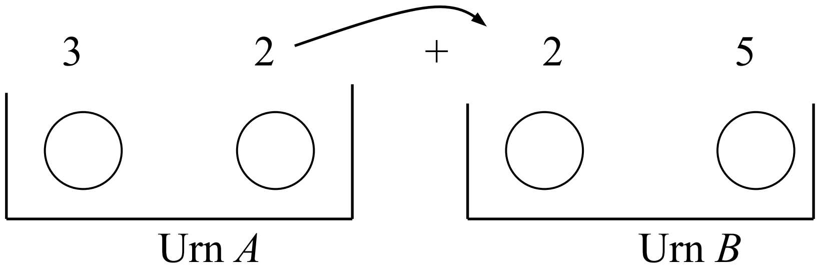


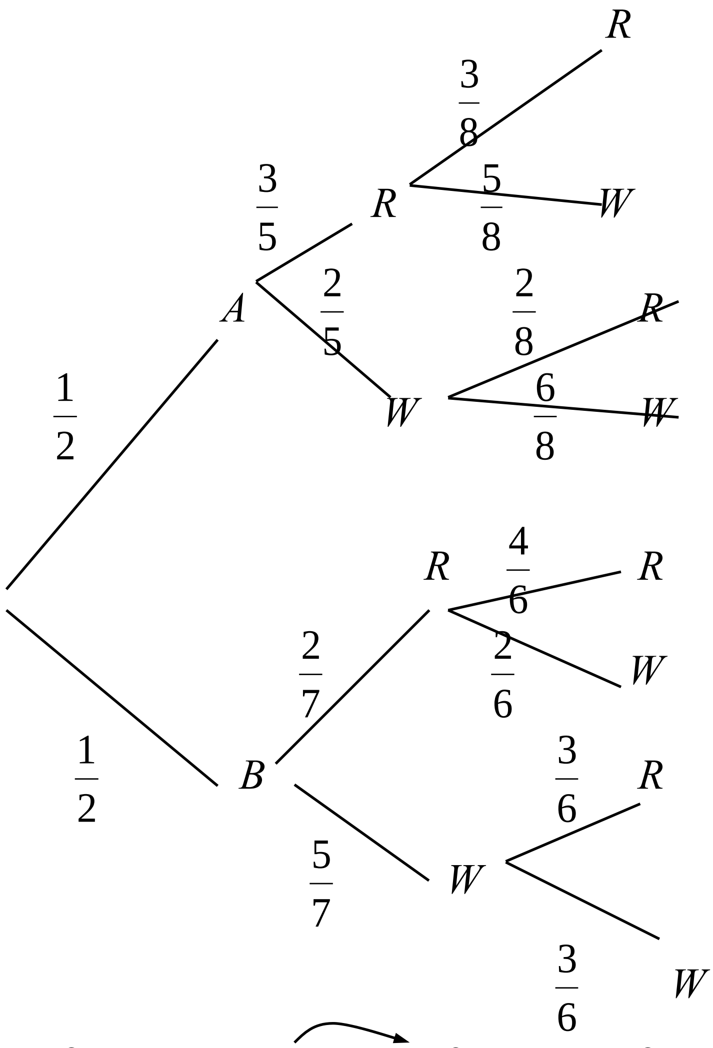
We seek $P[(H\cap E) \text{ or } (T\cap E)] = P(H\cap E)+P(T\cap E)$

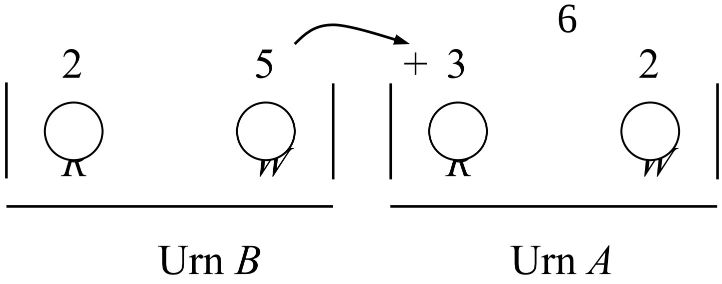
$$=P(H)\,P(E|H)+P(T)\,P(E|T)\ =\frac{2}{3}.\frac{4}{9}+\frac{1}{3}.\frac{2}{5}=\frac{58}{135}$$

Example: We are given two urns as follows:
Urn A contains 3 red and 2 white marbles.
Urn B " 2 " " 5 " " .

An urn is selected at random, a marble is drowning and put into the other urn; and then a marble is drawn from the second urn. Find the probability that both marbles drawn are of the same color.







We seek $A \cap R_1 \cap R_2$ or $A \cap W_1 \cap W_2$ or $B \cap R_1 \cap R_2$ or $B \cap W_1 \cap W_2$.

i.e., we want

$$P\left[(A \cap R_1 \cap R_2) \cup (A \cap W_1 \cap W_2) \cup (B \cap R_1 \cap R_2) \cup (B \cap W_1 \cap W_2)\right] =$$

$$P(A\cap R_1\cap R_2) + P(A\cap W_1\cap W_2) + P(B\cap R_1\cap R_2)$$
$$+ P(B\cap W_1\cap W_2)$$

$$= P(A_1)P(R_1 \mid A)P(R_1 \mid A\cap R_1) + P(A)P(W_1 \mid A)P(W_2 \mid A\cap W_1)$$

$$+ P(B)P(R_1 \mid B)P(R_2 \mid A\cap R_1) + P(B)P(W_1 \mid B)P(W_2 \mid B\cap W_1)$$

$$= \frac{1}{2}.\frac{3}{5}.\frac{3}{8} + \frac{1}{2}.\frac{2}{5}.\frac{6}{8} + \frac{1}{8}.\frac{2}{7}.\frac{4}{6} + \frac{1}{2}.\frac{5}{7}.\frac{3}{6} = \frac{901}{1680}$$

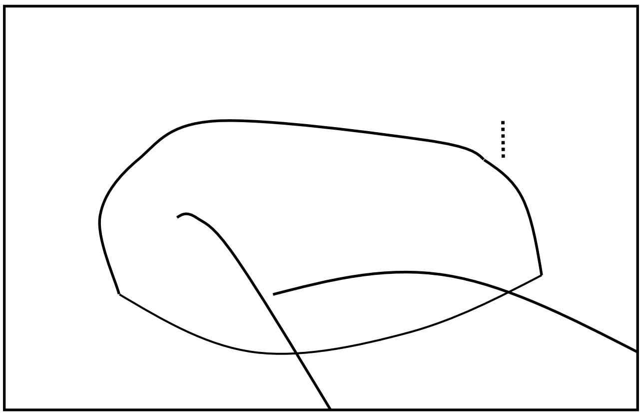
Problem (1):
Urn *A* contains *x* red marbles and *y* white marbles and Urn *B* contains *z* red marbles and *v* white marbles.
(i) If an Urn is selected at random and a marble drawn, what is the probability that the marble is red?
(ii) If a marble is drawn from Urn *A* and put into Urn *B* and then a marble is drawn from Urn *B*, what is probability that the second marble is red?

Problem (2):
Let *A* and *B* be events, with $P(A) = \frac{1}{2}, P(A\cap B) = \frac{1}{3}$ and $P(B \mid A^c) = \frac{1}{4}$, find $P(A \mid B^c)$.

Bayes Theorem:
Suppose $A_1, A_2, \ldots, A_n$ is a partition of *S* and *B* is any event. Then for any *i*

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)}$$

Proof:



Suppose the events $A_1, A_2, ..., A_n$ from a partition of a sample space *S*., Let *B* be any event, then

$$B = S \cap B = (A_1 \cup A_2 \cup ... \cup A_n) \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) ... \cup (A_n \cap B)$$

where $A_i \cap B$ are mutually exclusive, then

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_n \cap B)$$

Thus by multiplication theorem,

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ... + P(A_n)P(B|A_n)$$

But $\quad P(A_i|B) = \dfrac{P(A_i \cap B)}{P(B)}$

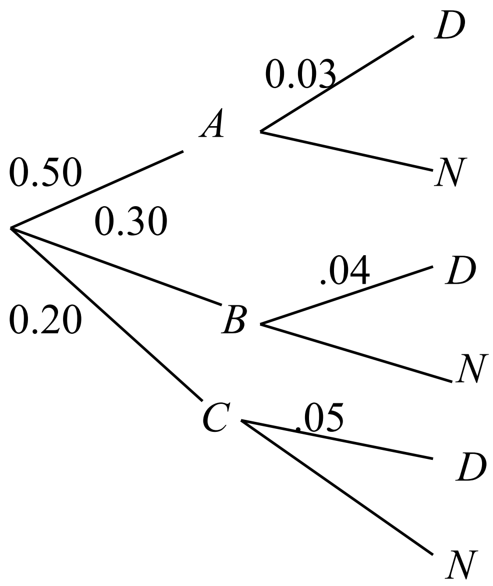
$$= \frac{P(A_i)P(B|A_i)}{P(B)}$$

$$= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Example : Three machines *A*, *B*, and *C* produce respectively 50%, 30% and 20% of the total number of items of a factory.

The percentages of defectives output of these machines are 3%, 4% and 5%. Suppose an item is selected at random is found to be defective. Find the probability that the item was produced by machine A.

Solution



We seek $P(A|D)$

By Bayes theorem

$$P(A|D)=\frac{P(A)\,P(D|A)}{P(A)\,P(D|\,D)+P(B)\,P(D|B)+P(C)\,P(D|C)}$$

$$=\frac{(0.50)\,(0.03)}{(0.50)\,(0.03)+(.30)\,(.04)+(.20)\,(.05)}=\frac{15}{37}$$

Problem (1):

In a factory four machines produce the same product. Machine A produce 10% of the output, machine B, 20%, machine C, 30% and machine D, 40%. The proportion of defective items produced by these follows

Machine A : .001

" B : .0005

" C : .005
" D : .002

An item selected at random is found to be defective. What is the probability that the item was produced by machine C?

Problem (2):
Suppose that a laboratory test to detect a certain disease has the following statistics. Let

A = event that the tested person has the disease
B = event that the tested result is positive

It is known that

$$P(B|A) = 0.99 \text{ and } P(B|\overline{A}) = 0.005$$

and 0.1 percent of the population actually has the disease. What is the probability that a person has the disease given that the result is positive?

Chapter 2

Random Variables

Definition: A **random variable** x an a sample space S is a function from S into the set R of real numbers such that the preimage of every interval of R in an event of S.

Note that: If X and Y are random variables on the same sample S, then $X+Y$, $X+k$, kX and XY (where is a real number) are the function on S defined by

$$(X+Y)(s)=X(s)+Y(s)$$

$$(kX)(s)=kX(s)$$

$$(X+k)(s)=X(s)+k$$

$$(XY)(s)=X(s)Y(s)$$, for every $s \in S$

It can be shown that these are also random variables.

Definition: A random variable is called **discrete random variable** if its set of possible outcomes is countable.
When a random variable can take on values on a continuous scale, it is called a **continuous random variable**.
Definition: The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if for each possible outcome x,

1. $f(x) \geq 0$
2. $\sum_{x} f(x) = 1$
3. $P(X = x) = f(x)$

Mean and Variance:

Let X be a random variable with the following probability distribution

x_1	x_2	$\dots$	x_n
$f(x_1)$	$f(x_2)$	$\dots$	$f(x_n)$

Then the mean or expectation of X, denoted by $E(X)$ or μ_X is defined by

$$E(X)=\sum_{i=1}^{n} x_i \; f(x_i)$$

and the variance of X, denoted by $Var(X)$, is defined by

$$Var(X)=\sum_{i=1}^{n} (\, x_i-\mu_X \,)\, f(x_i)=E(X-\mu_X)^2$$

The standard deviation of X, denoted by σ_X is

$$\sigma_X=\sqrt{Var\,(X)}$$

Theorem: $Var(X)=E\left(X^2\right)=\left(E(X)\right)^2$

Example: A pair of fair dice is tossed. Let x be assign to each point (a,b) in S the maximum of its numbers, i.e. $X(a,b)=\max\,(a,h)$. Find the distribution, expectation, variance and standard deviation of X.

Solution

$$S=\{\,(1,1),(1,2),\dots,(6,6)\,\}$$

Then X is a random variable with image set

$$X(s)=\{1,2,3,4,5,6\}$$

we compute the distribution f of X:

$$f(1)=P(X=1)=P(\,\{\,(1,1)\,\}\,)=\frac{1}{36}$$

$$f(2)=P(X=2)=P(\,\{(2,1),(2,2),(1,2)\,\}\,)=\frac{3}{36}$$

$$f(3)=P(X=3)=P\left(\{(3,1),(3,2),(3,3),(2,3),(1,3)\}\right)=\frac{5}{36}$$

$$f(4)=P(X=4)$$

$$=P\left(\{(4,1),(1,4),(4,2),(2,4),(4,3),(3,4),(4,4)\}\right)=\frac{7}{36}$$

$$f(5)=P(X=5)=\ldots=\frac{9}{36}$$

$$f(6)=P(X=6)=\ldots=\frac{11}{36}$$

Then the distribution of X follows

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

The mean or expectation of X is

$$E(X)=\sum x\,f(x)=1.\frac{1}{36}+2.\frac{3}{36}+\ldots+6.\frac{11}{36}=4.47$$

The variance of X is

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

But $E(X^2)=\sum x^2\,f(x)=1^2.\frac{1}{36}+2^2.\frac{3}{36}+\ldots+6^2.\frac{11}{36}$

$$=21.97$$

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

$$=21.97-(4.47)^2=1.99$$

So, $\sigma_X=\sqrt{Var(X)}=\sqrt{1.99}=1.4$

Example: A coin weighted so that $P(H)=\frac{2}{3}$ and $P(T)=\frac{1}{3}$ is tossed three times.

Let X be the random variable which assigns to each point in S the largest number of successive heads occurs. Find the distribution, expectation, and variance of X.

Solution

$$S=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$P(HHH)=\frac{2}{3}.\frac{2}{3}.\frac{2}{3}=\frac{8}{27} \quad , \quad P(THH)=\frac{1}{3}.\frac{2}{3}.\frac{2}{3}=\frac{4}{27}$$

$$P(HHT)=\frac{2}{3}.\frac{2}{3}.\frac{1}{3}=\frac{4}{27} \quad , \quad P(THT)=\frac{1}{3}.\frac{2}{3}.\frac{1}{3}=\frac{2}{27}$$

$$P(HTH)=\frac{2}{3}.\frac{1}{3}.\frac{2}{3}=\frac{4}{27} \quad , \quad P(TTH)=\frac{1}{3}.\frac{1}{3}.\frac{2}{3}=\frac{2}{27}$$

$$P(HTT)=\frac{2}{3}.\frac{1}{3}.\frac{1}{3}=\frac{2}{27} \quad , \quad P(TTT)=\frac{1}{3}.\frac{1}{3}.\frac{1}{3}=\frac{1}{27}$$

So, $X(TTT)=0$, $X(HTH)=1$, $X(HTT)=1$, $X(THT)=1$, $X(TTH)=1$, $X(HHT)=2$, $X(THH)=2$, $X(HHH)=3$

The image set of X is $X(s)=\{0,1,2,3\}$

$$f(1)=P(X=1)=P\{HTH, HTT, THT, TTH\}$$

$$=\frac{4}{27}+\frac{2}{27}+\frac{2}{27}+\frac{2}{27}=\frac{10}{27}$$

$$f(2)=P(X=2)=P\{HHT, THH\}$$

$$=\frac{4}{27}+\frac{4}{27}=\frac{8}{27}$$

$$f(3)=P(X=3)=P(HHH)=\frac{8}{27}$$

The distribution X is as follows

x	0	1	2	3
$f(x)$	$\frac{1}{27}$	$\frac{10}{27}$	$\frac{8}{27}$	$\frac{8}{27}$

So, $E(X)=\sum x\, f(x)=0.\frac{1}{27}+1.\frac{10}{27}+2.\frac{8}{27}+3.\frac{8}{27}$

$= 1.85$

and $Var(X)=E\left(X^2\right)=\left(E(X)\right)^2$

But $E\left(X^2\right)=\sum x^2\, f(x)=0^2.\frac{1}{27}+1^2.\frac{10}{27}+2^2.\frac{8}{27}+3^2.\frac{8}{27}$

$$= \frac{10+32+72}{27} = \frac{114}{27}$$

Then $Var(X) = 4.22 - (1.85)^2 = 0.7975$

Theorem (1): Let X a random variable and k a real number. Then

(i) $E(kX) = kE(X)$

(ii) $E(X+k) = E(X)+k$

Theorem (2): Let X and Y be random variable, an the same sample space S. Then

$$E(X+Y) = E(X)+E(Y)$$

Theorem (3): Let X be a random variable and k a real number. Then

(i) $Var(X+k) = Var(X)$

(ii) $Var(kX)^2 = k^2 Var(X)$

Problem: Prove that $Var(a+bX) = b^2 Var(X)$ when a and b are real numbers.

Example: A fair die is tossed. Let X denote twice the number appearing, and let Y denote 1 or 3 according as an odd or even number appears.
Find the distribution, expectation, variance of

(i) X
(ii) Y
(iii) $X+Y$
(iv) XY

Solution

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$ and each number appears with probability $\frac{1}{6}$.

(i) $X(1)=2$, $X(2)=4$, $X(3)=6$, $X(4)=8$, $X(5)=10$,
$X(6)=12$. Thus $X(s)=\{2,4,6,8,10,12\}$

and each number has probability $\frac{1}{6}$.

The distribution of X is as follows

x_i	2	4	6	8	10	12
$f(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Accordingly,

$$E(X) = \sum x\, f(x) = 2.\frac{1}{6}+4.\frac{1}{6}+...+12.\frac{1}{6}$$

$$E(X^2) = \sum x^2\, f(x) = 4.\frac{1}{6}+16.\frac{1}{6}+...+144.\frac{1}{6}=60.7$$

$$Var(X)=E(X^2)-(E(X))^2=60.7-(7)^2=11.7$$

$$\sigma_X=\sqrt{11.7}=3.4$$

(ii) $Y(1)=1$, $Y(2)=3$, $Y(3)=1$, $Y(4)=3$, $Y(5)=1$, $Y(6)=3$

Thus $Y(s)=\{1,3\}$, and $g(1) = P(Y=1)=P\{(1,3,5)\}=\frac{1}{2}$

$$g(3) = P(Y=3)=P\{(2,4,6)\}=\frac{1}{2}$$

Thus the distribution of Y is as follows

y	1	3
$g(y)$	$\frac{1}{2}$	$\frac{1}{2}$

According,

$$E(Y) = \sum y\, g(y) = 1.\frac{1}{2}+3.\frac{1}{2} = 2$$

$$E(Y^2) = \sum y^2\, g(y) = 1.\frac{1}{2}+9.\frac{1}{2}=5$$

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

$$= 5 - 4 = 1$$

(iii) Using $(X+Y)(s)=X(s)+Y(s)$, we obtain

$$(X+Y)(1)=X(1)+Y(1)=2+1=3$$

$$(X+Y)(2)=X(2)+Y(2)=4+3=7$$

$$(X+Y)(3)=X(3)+Y(3)=6+1=7$$

$$(X+Y)(4)=X(4)+Y(4)=8+3=11$$

$$(X+Y)(5)=X(5)+Y(5)=10+1=11$$

$$(X+Y)(6)=X(6)+Y(6)=12+3=15$$

Hence $(X+Y)(s)=\{3,7,11,15\}$

3 and 15 occurs with probability $\frac{1}{6}$.

7 and 11 occurs with probability $\frac{2}{6}$.

Thus the distribution of $X+Y$ is as follows

z	3	7	11	15
$P(z)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

Thus,

$$E(X+Y)=3.\frac{1}{6}+7.\frac{3}{6}+11.\frac{2}{6}+15.\frac{1}{6}=9$$

$$E\left((X+Y)^2\right)=9.\frac{1}{6}+49.\frac{3}{6}+121.\frac{2}{6}+225.\frac{1}{6}=95.7$$

$$Var(X+Y)=E\left((X+Y)^2\right)-\left(E(X+Y)\right)^2$$

$$= 95.7 - 81 = 14.7$$

$$\sigma_{X+Y}=\sqrt{14.7}=3.8$$

(iv) Using $(X+Y)(s)=X(s)Y(s)$, we obtain

$$(X\,Y)(1)=X(1)Y(1)=2\times1=2$$

$$(X\,Y)(2)=X(2)Y(2)=4\times3=12$$

$$(X\,Y)(3)=X(3)Y(3)=6\times 1=6$$

$$(X\,Y)(4)=X(4)Y(4)=8\times 3=24$$

$$(X\,Y)(5)=X(5)Y(5)=10\times 1=10$$

$$(X\,Y)(6)=X(6)Y(6)=12\times 3=36$$

Thus, the distribution of $X\,Y$ is as follows

w	2	6	10	12	24	36
$h(w)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Thus $E(X\,Y)=2.\frac{1}{6}+6.\frac{1}{6}+...+36.\frac{1}{6}=15$

$$E\left((X\,Y)^2\right)=4.\frac{1}{6}+36.\frac{1}{6}+...+(36)^2.\frac{1}{6}=359.3$$

$$Var\,(X\,Y)=E\left(X\,Y\right)^2-\left(E(X\,Y)\right)^2$$

$$= 359.3 - 225 = 134.3$$

$$\sigma_X=\sqrt{134.3}=11.6$$

Chapter 3

Some important probability distributions

Discrete Probability Distribution:

Binomial Distribution

A binomial experiment is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.
3. The outcomes of each trial must be independent of one another.
4. The probability of a success must remain the same for each trial.

A binomial experiment and its results give rise to a special probability distribution called the binomial distribution.

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a binomial distribution.

In binomial experiments, the outcomes are usually classified as successes or failures. For example, the correct answer to a multiple-choice item can be classified as a success, but any of the other choices would be incorrect and hence classified as a failure. The notation that is commonly used for binomial experiments and the binomial distribution is defined now.

In a binomial experiment, the probability of exactly X successes in n trials is

$$P(X = \mathrm{x}) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,\dots,n$$

An explanation of why the formula works is given following Example.

Example 1: A coin is tossed 3 times. Find the probability of getting exactly two heads.

Solution: This problem can be solved by looking at the sample space. There are three ways to get two heads.

$\mathrm{HHH}, \underline{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}, \mathrm{TTH}, \mathrm{THT}, \mathrm{HTT}, \mathrm{TTT}}$

The answer is $\frac{3}{8}$, or 0.375.

Looking at the problem in Example 1 from the standpoint of a binomial experiment, one can show that it meets the four requirements.

1 There are a fixed number of trials (three).

2 There are only two outcomes for each trial, heads or tails.

3 The outcomes are independent of one another (the outcome of one toss in no way affects the outcome of another toss).

4 The probability of a success (heads) is $\frac{1}{2}$ in each case.

In this case, $\mathrm{n}=3, \mathrm{X}=2, \mathrm{p}=\frac{1}{2}$, and $\mathrm{q}=\frac{1}{2}$. Hence, substituting in the formula gives

$$P(2 \text{ heads }) = \frac{3!}{(3-2)!\,2!} \cdot \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{3}{8} = 0.375,$$

which is the same answer obtained by using the sample space.

The same example can be used to explain the formula. First, note that there are three ways to get exactly two heads and one tail from a possible eight ways. They are HHT, HTH, and THH. In this case, then, the number of ways of obtaining two heads from three coin tosses is ${}_3\mathrm{C}_2$, or 3 . In general, the number of ways to get X successes from n trials without regard to order is

$$ {}_nC_X = \frac{n!}{(n-X)!\,X!} $$

This is the first part of the binomial formula. (Some calculators can be used for this.)

Next, each success has a probability of $\frac{1}{2}$ and can occur twice. Likewise, each failure has a probability of $\frac{1}{2}$ and can occur once, giving the $\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1$ part of the formula. To generalize, then, each success has a probability of p and can occur X times, and each failure has a probability of q and can occur $n-X$ times. Putting it all together yields the binomial probability formula.

Example 2: Survey on Doctor Visits

A survey found that one out of five Americans says he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

Source: Reader's Digest.

Solution: In this case, $n = 10, X = 3, p = \frac{1}{5}$, and $q = \frac{4}{5}$. Hence,

$$ P(3) = \frac{10!}{(10-3)!\,3!}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7 = 0.201 $$

Example 3: Survey on Employment

A survey from Teenage Research Unlimited (Northbrook, Illinois) found that 30% of teenage consumers receive their spending money from part-time jobs. If 5 teenagers are selected at random, find the probability that at least 3 of them will have part-time jobs.

Solution: To find the probability that at least 3 have part-time jobs, it is necessary to find the individual probabilities for 3 , or 4 , or 5 , and then add them to get the total probability.

$$P(3)=\frac{5!}{(5-3)!\,3!}(0.3)^3(0.7)^2=0.132$$
$$P(4)=\frac{5!}{(5-4)!\,4!}(0.3)^4(0.7)^1=0.028$$
$$P(5)=\frac{5!}{(5-5)!\,5!}(0.3)^5(0.7)^0=0.002$$

Hence,

P (at least three teenagers have part-time jobs)

$$=0.132+0.028+0.002=0.162$$

Computing probabilities by using the binomial probability formula can be quite tedious at times, so tables have been developed for selected values of n and p. Table B in Appendix C gives the probabilities for individual events. Example 4 shows how to use Table B to compute probabilities for binomial experiments.

Example 4: Tossing Coins

Solve the problem in Example 4 by using Table B.

Solution: Since $\mathrm{n}=3, \mathrm{X}=2$, and $\mathrm{p}=0.5$, the value 0.375 is found as shown in the following Figure.

Using Table B for **Example 4**

							p					
n	X	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0											
	1											
	2											
3	0						0.125					
	1						0.375					
	2						0.375					
	3						0.125					

$p = 0.5$

$n = 3$

$X = 2$

Example 5: Survey on Fear of Being Home Alone at Night
Public Opinion reported that 5% of Americans are afraid of being alone in a house at night. If a random sample of 20 Americans is selected, find these probabilities by using the binomial table.

a. There are exactly 5 people in the sample who are afraid of being alone at night.

b. There are at most 3 people in the sample who are afraid of being alone at night.

c. There are at least 3 people in the sample who are afraid of being alone at night.

Source: 100% American by Daniel Evan Weiss.

Solution:

a. $n = 20, p = 0.05$, and $X = 5$. From the table, we get 0.002.

b. n $= 20$ and p $= 0.05$. "At most 3 people" means 0 , or 1 , or 2 , or 3 . Hence, the solution is

$$P(0) + P(1) + P(2) + P(3)$$
$$= 0.358 + 0.377 + 0.189 + 0.060 = 0.984$$

c. $n = 20$ and $p = 0.05$. "At least 3 people" means $3,4,5,\ldots,20$. This problem can best be solved by finding $P(0) + P(1) + P(2)$ and subtracting from 1 .

$$P(0) + P(1) + P(2) = 0.358 + 0.377 + 0.189 = 0.924$$

$$1 - 0.924 = 0.076$$

Mean, Variance, and Standard Deviation for the Binomial Distribution

Find the mean, variance, and standard deviation for the variable that has the binomial distribution can be found by using the following formulas.

$$\text{Mean: } \mu = n \cdot p$$

$$\text{Variance: } \sigma^2 = n \cdot p \cdot q$$

$$\text{Standard deviation: } \sigma = \sqrt{n \cdot p \cdot q}$$

These formulas are algebraically equivalent to the formulas for the mean, variance, and standard deviation of the variables for probability distributions, but because they are for variables of the binomial distribution, they have been simplified by using algebra. The algebraic derivation is omitted here.

Example 6: Tossing a Coin

A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

Solution: With the formulas for the binomial distribution and $n = 4, p = \frac{1}{2}$, and $q = \frac{1}{2}$, the results are

$$\begin{aligned} \mu &= n \cdot p = 4 \cdot \frac{1}{2} = 2 \\ \sigma^2 &= n \cdot p \cdot q = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1 \\ \sigma &= \sqrt{1} = 1 \end{aligned}$$

From **Example 6,** when four coins are tossed many, many times, the average of the number of heads that appear is 2 , and the standard deviation of the number of heads is 1 . Note that these are theoretical values.

As stated previously, this problem can be solved by using the formulas for expected value. The distribution is shown.

No. of heads X	0	1	2	3	4
P(X = x)	1/16	4/16	6/16	4/16	1/16

$$\mu = E(X) = \sum X \cdot P(X)$$

$$= 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16}$$

$$= \frac{32}{16} = 2$$

$$\sigma^2 = \sum \ X^2 \cdot P(X) - \mu^2$$

$$\sigma^2 = 0^2 \cdot \frac{1}{16} + 1^2 \cdot \frac{4}{16} + 2^2 \cdot \frac{6}{16} + 3^2 \cdot \frac{4}{16} + 4^2 \cdot \frac{1}{16} - 2^2$$

$$= \frac{80}{16} - 4 = 1$$

$$\sigma = \sqrt{1} = 1$$

Example 7: Rolling a Die

A die is rolled 360 times. Find the mean, variance, and standard deviation of the number of 4 s that will be rolled.

Solution: This is a binomial experiment since getting a 4 is a success and not getting a 4 is considered a failure. Hence $n = 360, p = \frac{1}{6}$, and $q = \frac{5}{6}$.

$$\mu = n \cdot p = 360 \cdot \frac{1}{6} = 60$$

$$\sigma^2 = n \cdot p \cdot q = 360 \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = 50$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{50} = 7.07$$

On average, sixty 4 s will be rolled. The standard deviation is 7.07.

Continuous Probability Distribution :

Definition: The function *f*(*x*) is a **probability density function** for the continuous random variable *X*, defined over the set of real numbers *R*, if

1. $f(x) \geq 0$, for all $x \in R$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P(a < X < b) = \int_{a}^{b} f(x)\,dx$

The **Expectation** $E(X)$ is defined by

$$E(X) = \int_R X \, f(x)\,dx$$, when it exists.

The Variance $Var(X)$ is defined by

$$Var(X)\ E(X - E(X))^2 = \int_R (X - E(X))^2 \, f(x)\,dx$$

when it exists.

or $$Var(X) = E\left(X^2\right) - \left(E(X)\right)^2$$

$$= \int_R X^2 \, f(x)\,dx - \left(\int_R X \, f(x)\,dx \right)^2$$

Example: Let X be a continuous random variable with distribution

$$f(x) = \begin{cases} \dfrac{x}{6} + k & , \quad 0 \le x \le 3 \\ 0 & , \quad \text{o.w.} \end{cases}$$

(i) Evaluate k

(ii) Find $P\left(1 \le X \le 2\right)$

(iii) Find $P(-1 < X < 1)$

Solution

(i) $$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 \left(\frac{x}{6} + k\right) dx$$

$$= \left[\frac{x^2}{12} + kx \right]_0^3 = \left(\frac{9}{12} + 3k\right) - 0$$

$$\Rightarrow 1 - \frac{9}{12} = 3k \Rightarrow k = \frac{1}{12}$$

(ii) $$P\left(1 \le X \le 2\right) = \int_1^2 f(x)\,dx = \int_{x=1}^{x=2} \left(\frac{x}{6} + \frac{1}{12}\right) dx$$

$$= \left(\frac{x^2}{12} + \frac{x}{12}\right)_{x=1}^{2} = \frac{1}{12}\left[6 - 2\right] = \frac{4}{12} = \frac{1}{3}$$

(iii) $$P(-1 < X < 1) = \int_{-1}^{1} f(x)\,dx$$

0

$$=\cancel{\int_{-1}^{0} f(x)\,dx}+\int_{0}^{1} f(x)\,dx$$

$$=\int_{0}^{1}\left(\frac{x}{6}+\frac{1}{12}\right)dx=\frac{1}{12}\left(x^2+x\right)\Big|_{x=0}^{1}=\frac{1}{6}$$

Example: Consider the normal density function

$$f(x)=\frac{k}{a}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad,\ -\infty<x<\infty$$

(i) Evaluate k

(ii) Find $E(X)$ and $Var(X)$

Solution

(i) $$1=\frac{k}{a}\int_{-\sigma}^{\infty}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx\ \Rightarrow$$

let $\dfrac{x-\mu}{\sigma}=y\ \Rightarrow\ x=\mu+\sigma y\ ,\ dx=\sigma\,dy$

$$1=\frac{2k}{a}\int_{0}^{\infty}e^{-\frac{1}{2}y^2}\sigma\,dy$$

$$=2k\int_{0}^{\infty}e^{-\frac{1}{2}y^2}dy$$

let $y=\sqrt{2}\,z^{\frac{1}{2}}\ \ dy=\dfrac{\sqrt{2}}{2}z^{-\frac{1}{2}}dz$

$$\int_{0}^{\infty}e^{-z}z^{n-1}dz=\Gamma(n)$$

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

$$=2k\int_{0}^{\infty}e^{-z}\frac{\sqrt{2}}{2}z^{-\frac{1}{2}}dz$$

$$= \frac{2k\sqrt{2}}{\sqrt{2}} \int_0^\infty e^{-z} z^{-\frac{1}{2}} dz$$

$$= \sqrt{2}\, k \int_0^\infty e^{-z} z^{\frac{1}{2}-1}\, dz$$

$$1 = \sqrt{2}\, k\, \Gamma\left(\frac{1}{2}\right)$$

$$1 = \sqrt{2}\ \sqrt{\pi}\, k \ \Rightarrow k = \frac{1}{\sqrt{2\pi}}$$

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} x\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

let $\dfrac{x-\mu}{\sigma} = y$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\, \sigma \int_{-\infty}^{\infty} (\mu + \sigma y)\, e^{-\frac{1}{2}y^2}\, \sigma\, dy$$

$x = \sigma y + \mu$

$dx = \sigma\, dy$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma y)\, e^{-\frac{1}{2}y^2}\, \sigma\, dy$$

$$= \mu \overbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\, dy}^{1} \ + \ + \overbrace{\sigma \int_{-\infty}^{\infty} y\, e^{-\frac{1}{2}y^2}\, dy}^{0}$$

$$= \mu$$

(ii) $\quad Var(X) = E\left(X - E(X)\right)^2$

$$= E\left(X-\mu\right)^2 = \int_{-\infty}^{\infty} \left(X-\mu\right)^2\, f(x)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{\left(x-\mu\right)^2}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

let $\dfrac{x-\mu}{\sigma} = y$

$$= \frac{\sigma^2}{\sqrt{2\pi\sigma}} \int_{-\sigma}^{\infty} y^2 \; e^{-\frac{1}{2}y^2} \; \sigma \, d\, y \qquad \begin{aligned} x &= \sigma\, y + \mu \\ d\, x &= \sigma\, d\, y \end{aligned}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \; e^{-\frac{1}{2}y^2} \; d\, y$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} y^2 \; e^{-\frac{1}{2}y^2} \; d\, y$$

$$\text{let } \frac{1}{2} y^{-2} = z \;\Rightarrow\; y = \sqrt{2}\, z^{\frac{1}{2}}$$

$$d\, y = \frac{\sqrt{2}}{2} z^{-\frac{1}{2}} d\, z \;\;,\;\; d\, y = \frac{1}{\sqrt{2}} z^{-\frac{1}{2}} d\, z$$

$$Var(X) = \frac{2\sigma^2}{\sqrt{2}\sqrt{\pi}} \int_{0}^{\infty} 2\, z\, e^{-z} \frac{1}{\sqrt{2}} z^{-\frac{1}{2}} d\, z$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-z} z^{-\frac{1}{2}} d\, z = \frac{\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2, \quad Var(X) = \sigma^2$$

Gamma Function

$$\int_0^\infty x^{\alpha-1} e^{-x}\, dx = \Gamma(\alpha) \ , \ \alpha > 0$$

where $\Gamma(1)=1$

$$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$$

$$\Gamma(n)=(n-1)!$$

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

Example: The continuous random variable X has a Weibull distribution, with parameters α and β if its density function is given by

$$f(X)=\begin{cases} K\,X^{\beta-1} e^{-\alpha X^{\beta}} & , X>0 \\ 0 & , \text{o. w.} \end{cases}$$

1. Express K in terms of α and β .
2. Show that the mean and variance of the Weibull distribution are given by

$$E(X)=\alpha^{-\frac{1}{\beta}}\,\Gamma\left(1+\frac{1}{\beta}\right)$$

$$Var(X)=\alpha^{-\frac{2}{\beta}}\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\}$$

Solution

$$1 = K\int_0^\infty X^{\beta-1} e^{-\alpha X^{\beta}}\, dX$$

let $\alpha X^{\beta} = y$

$$=\frac{K}{\alpha\beta}\int_0^\infty e^{-y}\,dy=\frac{K}{\alpha\beta}\left(e^{-y}\right)_\infty^0$$

$$X^\beta=\frac{y}{\alpha}$$

$$\beta X^{\beta-1}\,dX=\frac{dy}{\alpha}$$

$$X^{\beta-1}\,dX=\frac{dy}{\alpha\beta}$$

$$1=\frac{K}{\alpha\beta}$$

$$\Rightarrow K=\alpha\beta$$

$$E(X)=\alpha\beta\int_0^\infty X\,X^{\beta-1}e^{-\alpha X^\beta}\,dX$$

$$\alpha X^\beta=y$$

$$X^\beta=\frac{y}{\alpha}$$

$$=\alpha\beta\int_0^\infty X^\beta e^{-\alpha X^\beta}\,dX$$

$$X=\left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}}y^{\frac{1}{\beta}}$$

$$=\alpha\beta\int_0^\infty \frac{y}{\alpha}e^{-y}\,\frac{\alpha^{-\frac{1}{\beta}}}{\beta}\,y^{\frac{1}{\beta}-1}\,dy$$

$$X=\alpha^{-\frac{1}{\beta}}y^{\frac{1}{\beta}}$$

$$dX=\alpha^{-\frac{1}{\beta}}\,\frac{1}{\beta}\,y^{\frac{1}{\beta}-1}\,dy$$

$$=\alpha^{-\frac{1}{\beta}}\int_0^\infty y^{\left(1+\frac{1}{\beta}\right)-1}e^{-y}\,dy$$

$$E(X)=\alpha^{-\frac{1}{\beta}}\,\Gamma\left(1+\frac{1}{\beta}\right)$$

$$Var(X)=E\left(X^2\right)-\left(E(X)\right)^2$$

But $E\left(X^2\right)=\alpha\beta\int_0^\infty X^2\,X^{\beta-1}e^{-\alpha X^\beta}\,dX$

$$\alpha X^\beta=y$$

$$X^\beta=\frac{y}{\alpha}$$

$$=\alpha\beta\int_0^\infty X^{\beta+1}e^{-\alpha X^\beta}\,dX$$

$$dX=\frac{\alpha^{-\frac{1}{\beta}}}{\beta}\,y^{\frac{1}{\beta}-1}\,dy$$

$$=\alpha\beta\int_0^\infty \alpha^{-1-\frac{1}{\beta}}\,y^{1+\frac{1}{\beta}}\,e^{-y}\,\frac{\alpha^{-\frac{1}{\beta}}}{\beta}\,y^{\frac{1}{\beta}-1}\,dy$$

$$X=\alpha^{-\frac{1}{\beta}}y^{\frac{1}{\beta}}$$

$$= \alpha^{-\frac{2}{\beta}} \int_0^\infty e^{-y}\, y^{\left(1+\frac{2}{\beta}\right)-1}\, dy \qquad\qquad dX = \frac{\alpha^{-\frac{1}{\beta}}}{\beta}\, y^{\frac{1}{\beta}-1}\, dy$$

$$= \alpha^{-\frac{2}{\beta}} \int_0^\infty e^{-y}\, y^{\left(1+\frac{2}{\beta}\right)-1}\, dy \qquad\qquad X^{\beta+1} = \alpha^{-\left(\frac{\beta+1}{\beta}\right)}\, y^{\frac{\beta+1}{\beta}}$$

$$= \alpha^{-\frac{2}{\beta}}\, \Gamma\left(1+\frac{2}{\beta}\right) \qquad\qquad = \alpha^{-\left(1+\frac{1}{\beta}\right)}\, y^{1+\frac{1}{\beta}}$$

$$Var(X) = E\left(X^2\right) - \left(E(X)\right)^2$$

$$= \alpha^{-\frac{2}{\beta}}\, \Gamma\left(1+\frac{2}{\beta}\right) - \alpha^{-\frac{2}{\beta}} \left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2$$

$$= \alpha^{-\frac{2}{\beta}} \left\{ \Gamma\left(1+\frac{2}{\beta}\right) - \left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2 \right\}$$

Beta Function

$$\int_0^1 X^{\alpha-1}(1-X)^{\beta-1}\,dX = B(\alpha,\beta)$$

$$= \frac{\Gamma(\alpha).\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Definition: A random variable X has a beta distribution and its is referred to as a beta random variable, if and only if its probability density is given by

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ +\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & ,0<x<1 \\ 0 & ,\text{o. w.} \end{cases}$$

where $\alpha>0$ and $\beta>0$

Problem (1): Show that mean and the variance of the beta distribution are given by

$$E(X) = \mu = \frac{\alpha}{\alpha+\beta} \qquad \text{and} \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Problem (2): Show that the parameters of the beta distribution can be expressed as follows in terms of the mean (μ) and the variance (σ^2) of this distribution

(i) $\alpha = \mu\left[\dfrac{\mu(1-\mu)}{\sigma^2} - 1\right]$

(ii) $\beta = (1-\mu)\left[\dfrac{\mu(1-\mu)}{\sigma^2} - 1\right]$

Problem (3): A random variable X has a Reyleigh distribution if and only if its probability density is given by

$$f(x) = \begin{cases} 2\alpha\ X e^{-\alpha X^2} & ,\ X>0 \\ 0 & ,\text{o. w.} \end{cases}$$

Where $\alpha > 0$. Show that for this distribution

(i) $E(X) = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$

(ii) $Var(X) = \frac{1}{\alpha}\left(1 - \frac{\pi}{4}\right)$

Problem (4): A random variable X has a Pareto distribution if and only if its probability density is given by

$$f(x) = \begin{cases} \frac{\alpha}{X^{\alpha+1}} & , X > 1 \\ 0 & , \text{o. w.} \end{cases}$$

Where $\alpha > 0$. Show that for Pareto distribution

$E(X) = \frac{\alpha}{\alpha - 1}$, provided $\alpha > 1$.

Normal Distribution

The normal probability density, also called the Gaussian density, might be the most commonly used probability density function in statistics.

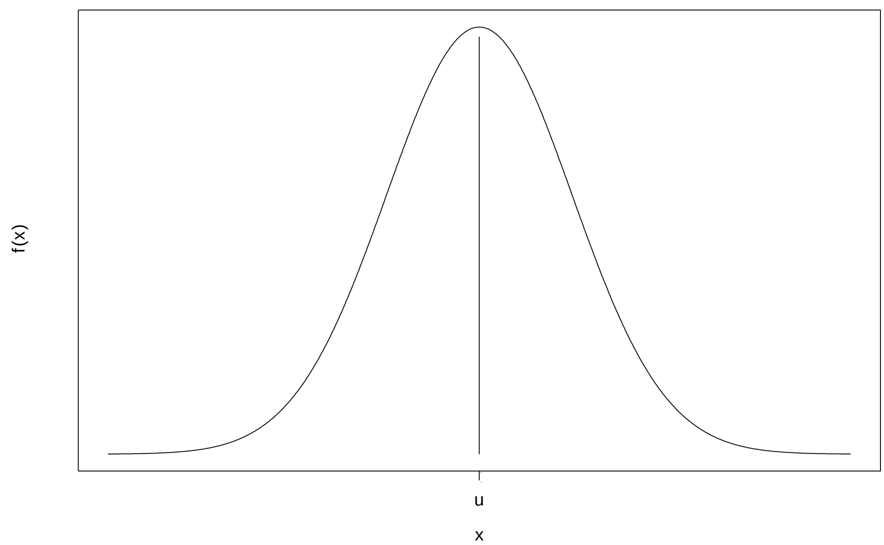
Normal Probability Density Function:
A random variable X taking values in $[-\infty,\infty]$ has the normal probability density function $f(x)$ if

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

where

$$\mu=E(X),\ \sigma^2=Var(X),\ \ \pi=3.14159$$

The graph of $f(x)$ is



Properties of Normal Density Function:

(a) $\mu=E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{\infty}x\cdot\frac{1}{\sqrt{2\pi\sigma}}\cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx$

and

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

(b)

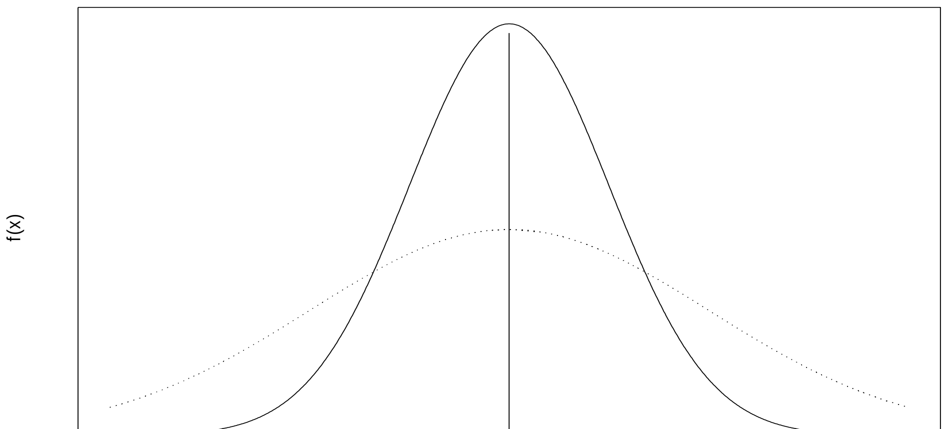
$\mu =$ the mean of the normal random variable X

$=$ the median of the normal random variable X $(P(X \le \mu) = P(X \ge u))$

$=$ the mode of the normal probability density $(f(u) > f(x),\ x \ne \mu)$

(c) X is a random variable with the normal density function. X is denoted by $X \sim N(\mu, \sigma^2)$

(d) The standard deviation determine the width of the curve. The normal density with larger standard deviation would be more dispersed than the one with smaller standard deviation. In the following graph, two normal density functions have the same means but different standard deviations, one is 1 (the solid line) and the other is 2 (the dotted line):



(e) The normal density is symmetric with respect to mean. That is, $f(u-c) = f(u+c)$, where c is any number

(f) The probability of a normal random variable follows the empirical rule introduce previously. That is,

$$P(u-\sigma \le X \le \mu+\sigma)=0.6826 \approx 68\%$$
$$P(\mu-2\sigma \le X \le \mu+2\sigma)=0.9544 \approx 95\%$$
$$P(\mu-3\sigma \le X \le \mu+3\sigma)=0.9973 \approx 100\%$$

(g) i.e., the probability of X taking values within one standard deviation is about 0.68, within two standard deviations about 0.95, and within three standard deviation about 1.

(h) Standard Normal Probability Density Function:

(i) A random variable Z, taking values in $[-\infty,\infty]$ has the standard normal probability density function $f(x)$ if

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \text{ , } -\infty \le x \le \infty \text{ ,}$$

where

$$\mu = E(Z)=0,\ \sigma^2 = Var(Z)=1 \text{.}$$

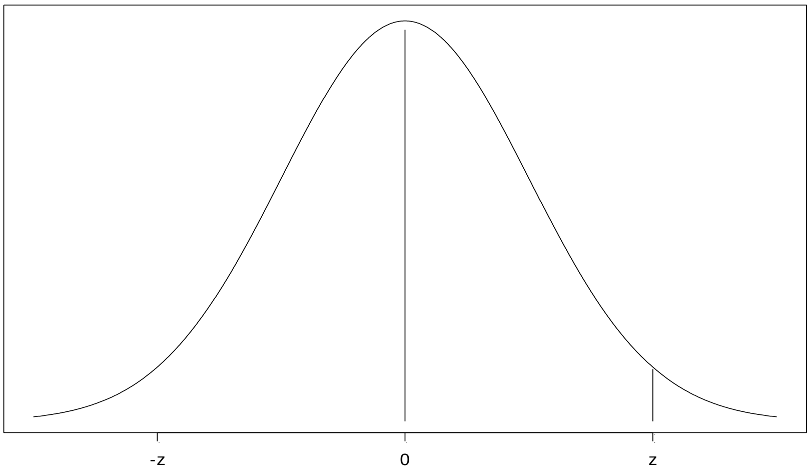
Note: we denote Z as $Z \sim N(0,1)$

The probability of Z taking values in some interval can be found by the normal table. The probability of Z taking values in [0,z], $z \ge 0$, can be obtained by the normal table. That is,

$P(0 \le Z \le z) =$ the area of the region between two vertical lines

$$= \int_0^z \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx$$

The graph is given below:



Example:

$$P(0 \le Z \le 1.0) = 0.3413$$
$$P(0 \le Z \le 1.03) = 0.3485$$

$$\begin{aligned} P(-1.0 \le Z \le 1.0) &= P(-1 \le Z \le 0) + P(0 \le Z \le 1) \\ &= 2P(0 \le Z \le 1) = 0.6826 \ \ (\text{symmetryof } Z) \end{aligned}$$

$$\begin{aligned} P(Z \ge 1.5) &= 1 - P(Z \le 1.5) = 1 - \left[P(Z \le 0) + P(0 \le Z \le 1.5)\right] \\ &= 1 - (\frac{1}{2} + 0.4332) = 0.0668 \end{aligned}$$

$$\begin{aligned} P(Z \ge -1.5) &= P(-1.5 \le Z \le 0) + P(Z \ge 0) \\ &= P(0 \le Z \le 1.5) + \frac{1}{2} \ \ (\text{symmetryof } Z) \\ &= 0.4332 + 0.5 = 0.9332 \end{aligned}$$

$$\begin{aligned} P(1 \le Z \ge 1.5) &= P(0 \le Z \le 1.5) - P(0 \le Z \le 1) \\ &= 0.4332 - 0.3413 = 0.0919 \end{aligned}$$

Example:

$P(Z \geq x) = 0.0099$. **What** is x?

Solution:

$$P(Z \leq x) = 1 - P(Z \geq x) = 1 - 0.0099 = 0.9901$$

$$\Rightarrow\ x \geq 0\ \ (\text{if}\ \ x \leq 0, \text{then}\ P(Z \leq x) \leq \frac{1}{2})$$

$$\Rightarrow\ P(Z \leq x) = P(Z \leq 0) + P(0 \leq Z \leq x) = \frac{1}{2} + P(0 \leq Z \leq x) = 0.9901$$

$$\Rightarrow\ P(0 \leq Z \leq x) = 0.9901 - \frac{1}{2} = 0.4901\ \Rightarrow\ x = 2.33$$

Computing Probabilities for any Normal Random Variable:
Once the probability of the standard normal random variable can be obtained, the probability of any normal random variable (not standard) can be found via the following important rescaling:

$$X \sim N(\mu, \sigma^2)\ \Rightarrow\ \frac{X - \mu}{\sigma} = Z \sim N(0,1)$$

Example:

Let $X \sim N(1,4)$. Please find $P(1 \leq X \leq 3)$.

Solution:

$\mu = 1$ and $\sigma = 2$. Then,

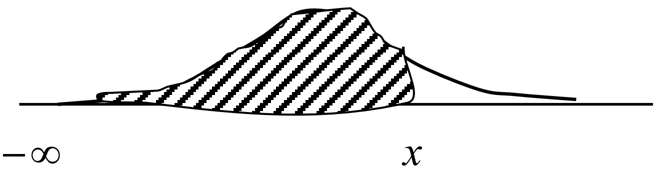
$$P(1 \leq X \leq 3) = P(1-1 \leq X - 1 \leq 3-1) = P(\frac{0}{2} \leq \frac{X-1}{2} \leq \frac{2}{2}) = P(0 \leq Z \leq 1) = 0.3413$$

Cumulative Distribution Function

Definition (1): The Cumulative distribution $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t),\ \text{for}\ \ -\infty < x < \infty$$

Definition (2) : The Cumulative distribution $F(x)$ of a continuous random variable X with density function $f(x)$ is

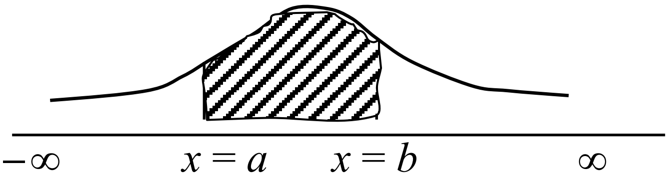


$$F(X) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt \quad , \ -\infty < x < \infty$$

Remark: Using the above definition one can write the results

(1) $P(a, X < b) = F(b) - F(a)$

(2) $F(x) = \dfrac{d\,F(x)}{d\,x}$, if derivative exists.



Example: Find the Cumulative distribution function *F*(*x*) for the random variable *X* of the distribution given by

x	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

and graph this Cumulative distribution function.

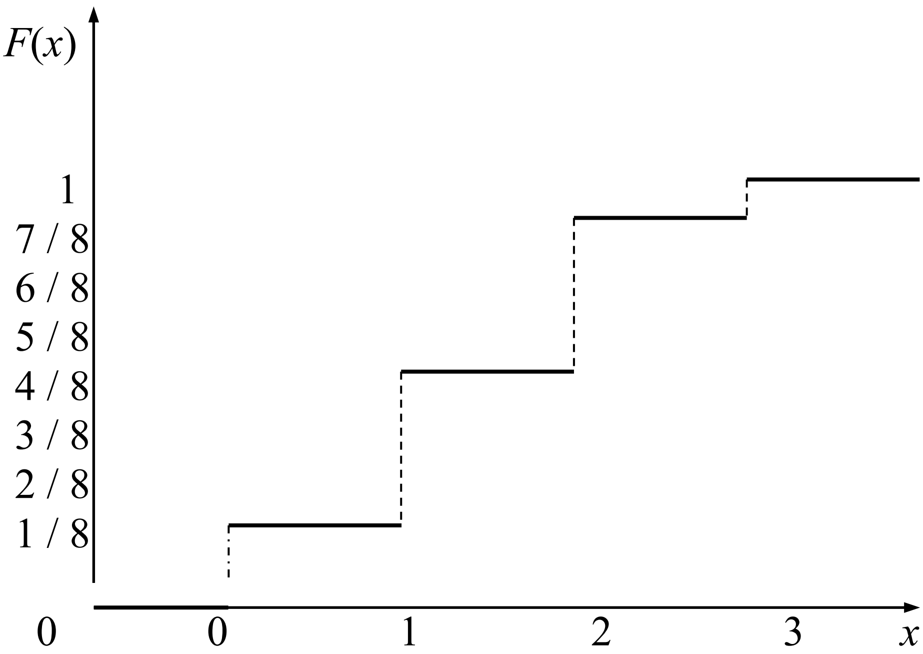
Solution

$$F(x) = P(X \le x) = \sum_{t \le x} f(t)$$

Using the above table, we obtain

$$F(X)=\begin{cases} 0 & , \quad -\infty<x<0 \\ \frac{1}{8} & , \quad 0\le x<1 \\ \frac{1}{2} & , \quad 1\le x<2 \\ \frac{7}{8} & , \quad 2\le x<3 \\ 1 & , \quad 3\le x<\infty \end{cases}$$

The graph of the Cumulative distribution function $F(x)$ is shown in the following Fig.



Example: A random variable X has the density function

$$f(x)=\frac{C}{x^2+1} \quad , \quad \text{where} \quad -\infty<x<\infty$$

(a) Find the value of the constant C.

(b) Find $P\left(\frac{1}{3}<X^2<1\right)$

(c) Find the Cumulative distribution

Solution

(a) We must have $\int_{-\infty}^{\infty} f(x)\,dx = 1,$

i.e., $\int_{-\infty}^{\infty} \frac{C\,dx}{x^2+1} = C\tan^{-1} x\Big|_{-\infty}^{\infty} = C\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = C\pi\ 1$

so that $C = \frac{1}{\pi}$

Remark: If $\alpha^2 < x^2 < \beta^2$ then $\alpha < x < \beta$ or $-\beta < x < -\alpha$

(b) If $\frac{1}{3} < x^2 < 1$, then either $\frac{\sqrt{3}}{3} < x < 1$ or $-1 < x < -\frac{\sqrt{3}}{3}$

Thus the required probability is

$$P\left(\frac{1}{3} < X^2 < 1\right) = P\left(\frac{\sqrt{3}}{3} < X < 1\right) + P\left(1 < X < -\frac{\sqrt{3}}{3}\right)$$

$$= \frac{1}{\pi}\int_{-1}^{-\sqrt{3/3}} \frac{x}{x^2+1} + \frac{1}{\pi}\int_{\sqrt{3}/3}^{1} \frac{x}{x^2+1} = \frac{2}{\pi}\int_{\sqrt{3}/3} \frac{dx}{x^2+1}$$

$$= \frac{2}{\pi}\left[\tan^{-1}(1) - \tan^{-1}\left(\frac{\sqrt{3}}{3}\right)\right] = \frac{2}{\pi}\left(\frac{\pi}{4} - \frac{\pi}{6}\right)$$

$$= \frac{2}{\pi} \times \frac{\pi}{12} = \frac{1}{6}$$

(c) $F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\pi}\int_{-\infty}^{x} \frac{dt}{t^2+1} = \frac{1}{\pi}\left[\tan^{-1}\Big|_{-\infty}^{x}\right]$

$$= \frac{1}{\pi}\left[\tan^{-1} x - \tan^{-1}(-\infty)\right]$$

$$= \frac{1}{\pi}\left[\tan^{-1} x + \frac{\pi}{2}\right] = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x$$

Example: The Cumulative distribution function for a random variable X is

$$F(x)=\begin{cases} 1-e^{-2x} & ,x\geq 0 \\ 0 & ,x<0 \end{cases}$$

(a) Find the density function.

(b) Find $P(X>2)$

(c) Find $P(-3<X\leq 4)$

Solution

(a) $$f(x)=\frac{d}{dx}F(x)=\begin{cases} 2e^{-2x} & ,x\geq 0 \\ 0 & ,x<0 \end{cases}$$

(b) $P(X\leq 2)=F(2)=1-e^{-4}$, Hence

$$P(X>2)=1-\left(1-e^{-4}\right)=e^{-4}$$

(c) $P(-3<X\leq 4)=F(4)-F(-3)$

$$=\left(1-e^{-8}\right)-(0)=1-e^{-8}$$

Problems

1. A random variable is said to have a Cauchy distribution if its density given by

$$f(x) = \frac{\frac{\beta}{\pi}}{(x-\alpha)^2+\beta^2} \quad \text{for } -\infty < x < \infty$$

(a) Find the Cumulative distribution function $F(x)$

(b) Find $P\left(\left|4X^2-3\right|<1\right)$ in the case of $\alpha=0$, and $\beta=1$.

2. A random variable X is said to have a Laplace distribution if its density is given by

$$f(x)=\frac{1}{2\beta}\exp\left\{-\frac{|x-\alpha|}{\beta}\right\}$$

$-\infty<x<\infty$, $x-\infty<\alpha<\infty$, $\beta>0$

(a) Show that $E(X)=\alpha$ and $Var(X)=2\beta^2$

(b) Show that

$$F(x)=\begin{cases}\frac{1}{2}\exp\left(\frac{x-\alpha}{\beta}\right) & ,x\le\alpha \\ 1-\frac{1}{2}\exp\left(\frac{-x+\alpha}{\beta}\right) & ,x>\alpha\end{cases}$$

3. Consider the extreme value density function

$$f(x)=\frac{1}{\beta}\exp\left\{-e^{-(x-\alpha)/\beta}-\left(\frac{x-\alpha}{\beta}\right)\right\},$$

$-\infty<x<\infty$, $x-\infty<\alpha<\infty$, $\beta>0$

(a) Find $F(x)$

(b) Find $P\left(0<x^2+6x<16\right)$ in the case of $\alpha=0$ and $\beta=1$.

Chapter 4

Moments and Moment Generating Functions

Definition: The r[th] moment about the origin of the random variable X is given by

$$\mu_r' = E\left(X^r\right) = \begin{cases} \sum_x x^r \ f(x) & \text{,if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} x^r \ f(x)\, d\,x & \text{, if } x \text{ is continuou} \end{cases}$$

Remark: We can write the mean and the variance of a random variable as

$$\text{Mean} = E(X) = \mu = \mu_1' \text{ , Variance } = Var(X) = \mu_2' - \mu^2$$

Definition: The moment generating function of the random variable X is given by $E\left(e^{t\,x}\right)$ and is denoted by $M_X(t)$. Hence

$$M_X(t) = E\left(e^{t\,x}\right) = \begin{cases} \sum_x e^{t\,x} \ f(x) & \text{,if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} e \quad f(x)\, d\,x & \text{, if } x \text{ is continuou} \end{cases}$$

Theorem: Let X be a random variable with moment generating function $M_X(t)$. Then

$$\left.\frac{d^r\, M_X(t)}{d\,t^r}\right|_{t=0} = \mu_r'$$

Example: Find the generating function of the binomial random variable X

$$f(x)=\binom{n}{x} p^x q^{n-x} \quad , x=0,1,2,\ldots,n \quad , \; q=1-p$$

and then use it to verify that

Mean = $E(X) = n\,p$ and Variance = $Var(X) = n\,p\,q$

Solution

$$M_X(t) = \sum_{x=0}^{n} e^{t\,x} \binom{n}{x} p^x\, q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(p\,e^t\right)^x q^{n-x}$$

$$=\left(p\,e^t+q\right)^n$$

Now $\dfrac{d}{d\,t}(M_X(t)) = n\left(p\,e^t+q\right)^{n-1} p\,e^t$, and

$$= n\,p\left(p\,e^t+q\right)^{n-1} e^t$$

$$\frac{d^2}{d\,t}(M_X(t)) = n\,p\left(p\,e^t+q\right)^{n-1} e^t +$$

$$n\,p\,(n-1)\left(p\,e^t+q\right)^{n-2} p\,e^{2t}$$

Setting $t = 0$, we get

$$\mu_1' = \frac{d}{d\,t}\left(M_X(t)\right)_{t=0} = n\,p$$

$$\mu_2' = \frac{d^2}{d\,t^2}\left(M_X(t)\right)_{t=0} = n\,p + n\,(n-1)\,p^2$$

So, $Var(X) = \mu_2' - \mu^2 = n\,p + n(n-1)\,p^2 - n^2\,p^2$

$$= n\,p - n\,p^2 = n\,p\,(1-p)$$

$$= n\,p\,q$$

Example: Show that the moment general function of the random variable X, hearing a gamma density function

$$f(x) = \frac{e^{-x\theta}\, x^{\alpha-1}}{\Gamma(\alpha)}\theta^{\alpha} \;,\; x>0,\; \theta>0 \;,\; \alpha>0$$

is given by $M_X(t)=\left(\dfrac{\theta}{\theta-t}\right)^{\alpha}$

Solution

$$M_X(t)=\int_0^{\infty} e^{tx}\,\frac{e^{-x\theta}\,x^{\alpha-1}}{\Gamma(\alpha)}\theta^{\alpha}\,dx$$

$$=\frac{\theta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty} e^{-(\theta-t)x}\,x^{\alpha-1}\,dx$$

let $\quad (\theta-t)x=y$

$$x=\frac{y}{\theta-t}$$

$$dx=\frac{1}{\theta-t}\,dy$$

then $\quad M_X(t)=\dfrac{\theta^{\alpha}}{\Gamma(\alpha)}\displaystyle\int_0^{\infty} e^{-y}\left(\frac{y}{\theta-t}\right)^{\alpha-1}\frac{dy}{\theta-t}$

$$=\frac{\left(\dfrac{\theta}{\theta-t}\right)^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty} e^{-y}\,y^{\alpha-1}\,dy$$

$$=\left(\frac{\theta}{\theta-t}\right)^{\alpha}\frac{\Gamma(\alpha)}{\Gamma(\alpha)}$$

$$=\left(\frac{\theta}{\theta-t}\right)^{\alpha}$$

Problems

1. Show that the r[th] moment about the origin of the gamma distribution

$$f(x) = \frac{e^{-x\theta}\, x^{\alpha-1}}{F(\alpha)} t^{\alpha} \;,\; x > 0$$

   is given by $\mu_r' = \dfrac{\theta^{-r}\,\Gamma(\alpha+r)}{\Gamma(\alpha)}$

2. A random variable X has the geometric distribution
   $f(x) = p\,q^{x-1}, x = 1, 2, 3, \ldots$ where $q = 1 - p$
   show that the moment generating function of X is

$$M_X(t) = \frac{p\,e^t}{1 - q\,e^t}$$

   and then use $M_X(t)$ to find the mean and variance of the geometric distribution.

3. A random variable X has the Poisson distribution

$$f(x) = \frac{e^{-\mu}\,\mu^x}{x!},\; x = 0, 1, 2, \ldots$$

   Show that the moment generating function of X is

$$M_X(t) = e^{\mu\left(e^t - 1\right)}$$

   Using $M_X(t)$, find the mean and variance of the Poisson distribution.

Chapter 5
Sampling Distribution

The probability distribution of a statistic is called a sampling distribution, i.e. the distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called sampling distribution of that statistic.

We commonly use the terms population and sample in our study.

A **population** is the complete collection of all elements to be studied.

A **sample** : is a subcollection of elements drawn from a population.

Definition : Any function of the random variables constituting a random sample is called **a statistic**.

Now, let $X_1, X_2, \ldots, X_n$ be n independent random variables, each having the same probability distribution f(x), we then define $X_1, X_2, \ldots, X_n$, to be a **random sample** of size n from the population f(x) and write its joint probability distribution as

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \ldots f(x_n)$$

Sample Mean : If $X_1, X_2, \ldots, X_n$ constitute a random sample, then

$$\overline{X} = \sum_{i=1}^{n} X_i \Big/ n$$

is called the sample mean.

Sample Variance : If $X_1, X_2, \ldots, X_n$ constitute a random sample, then

$$S^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \Big/ (n-1)$$

is called a sample variance.

If we use the values of the random variables, we might calculate

$$\overline{x} = \sum_{i=1}^{n} x_i \Big/ n \quad \text{and} \quad S^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})}{(n-1)}$$

The Distribution of the Mean :

Theorem : If $X_1, X_2, \ldots, X_n$ constitute a random sample from an infinite population with mean μ and variance σ^2, then

$$E(\overline{X}) = \mu \quad \text{and} \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

Proof : We use the following theorem

$$E(aX + b) = a\,E(X) + b$$

$$Var(aX + b) = a^2\, Var(X)$$

Then $$E(\overline{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu$$

$$= \frac{1}{n}.n\,\mu = \mu$$

$$Var(\overline{X}) = Var\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\left(\sum \sigma^2\right)$$

$$Var(\overline{X}) = \frac{1}{n^2}.n\,\sigma^2 = \frac{\sigma^2}{n}$$

It is customary to write $E(\overline{X})$ as $\mu_{\overline{X}}$ and $Var(\overline{X})$ as $\sigma_{\overline{X}}$, and refer to $\sigma_{\overline{X}}$ as the standard error of the mean.

Theorem : If $X_1, X_2, \ldots, X_n$ constitute a random sample from $N(\mu, \sigma^2)$, then $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$.

Proof : If $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\mu t + t^2 \frac{\sigma^2}{2}}$

So, $$M_{\overline{X}}(t) = E\left(e^{t\overline{X}}\right) = E\left(e^{\frac{t}{n}\sum X_i}\right)$$

$$= \prod_{i=1}^{n} E\left(e^{\frac{tX}{n}}\right) = \left[E\left(e^{\frac{tX}{n}}\right)\right]^n$$

But $$E\left(e^{\frac{tX}{n}}\right)=e^{\frac{\mu t}{n}+\frac{t^2\sigma^2}{2n^2}}$$

$$M_{\overline{X}}(t)=\left(e^{\frac{\mu t}{n}+\frac{t^2\sigma^2}{2n^2}}\right)^n$$

$$=e^{\mu t+\frac{t^2\sigma^2}{2n}}$$

So, we can write $\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$

Central Limit Theorem :

If $X_1, X_2, \ldots, X_n$ constitute a random sample from an infinite population with the mean μ , the variance σ^2 , and the moment generating function $M_X(t)$, then the limiting distribution of

$Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ as $n\rightarrow\infty$ is the standard normal distribution.

Proof : $M(t)=E\left(e^{tz}\right)=E\left(e^{t\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)}\right)$

$$=e^{-\frac{\mu t}{\mu/\sqrt{n}}}E\left(e^{\frac{t\overline{X}}{\sigma/\sqrt{n}}}\right)$$

$$=e^{-\frac{\mu t}{\mu/\sqrt{n}}}E\left(e^{\frac{t}{\sigma/\sqrt{n}}\sum X_i}\right)$$

$$=e^{-\frac{\mu t}{\mu/\sqrt{n}}}\prod_{i=1}^{n}E\left(e^{\frac{tX_i}{\sigma\sqrt{n}}}\right)$$

$$=e^{-\frac{\mu t}{\mu/\sqrt{n}}}\left[E\left(e^{\frac{tX}{\sigma\sqrt{n}}}\right)\right]^n$$

and hence that

$$\ln\left(M_Z(t)\right)=\frac{-\mu t}{\sigma/\sqrt{n}}+n\ln E\left(e^{\frac{tX}{\sigma\sqrt{n}}}\right)$$

$$=\frac{-\mu t}{\sigma/\sqrt{n}}+n\ln E\left[1+\frac{tX}{\sigma\sqrt{n}}+\frac{t^2X^2}{2\sigma^2 n}+\ldots\right]$$

$$= \frac{-\mu\, t}{\sigma / \sqrt{n}} + n \ln \left[1 + \frac{t\, \mu}{\sigma \sqrt{n}} + t^2 \frac{E\left(X^2\right)}{2\sigma^{2n}} + \ldots \right]$$

But we know that

$$\ln (1+y) = y - \frac{1}{2} y^2 + \frac{1}{3} y^3 - \frac{1}{4} y^4 + \ldots$$

$$\text{So, } \ln \left(M_z(t)\right) = \frac{-\mu\, t}{\sigma / \sqrt{n}} + n \left[\left(\frac{\mu\, t}{\sigma \sqrt{n}} + t^2 \frac{E\left(X^2\right)}{2\sigma^{2n}} + \ldots \right) \right.$$

$$\left. - \frac{1}{2} \left(\frac{\mu\, t}{\sigma \sqrt{n}} + t^2 \frac{E\left(X^2\right)}{2\sigma^{2n}} + \ldots \right) + \frac{1}{3} (\ldots)^3 \ldots \right]$$

$$= \frac{-\mu\, t}{\sigma / \sqrt{n}} + \frac{\mu\, t}{\sigma / \sqrt{n}} t^2 \frac{E\left(X^2\right)}{2\sigma^2} - \frac{n}{2} \left(\frac{\mu\, t}{\sigma \sqrt{n}} + t^2 \frac{E\left(X^2\right)}{2\sigma^{2n}} + \ldots \right)^2 + \ldots$$

$$= \frac{t^2}{2\sigma^2} E\left(X^2\right) - \frac{1}{2} t^2 \frac{\mu^2}{\sigma^2} + O\left(\frac{1}{\sqrt{n}} \right)$$

$$= \frac{1}{2} \frac{t^2}{2\sigma^2} \left(E\left(X^2\right) - \mu^2 \right) + O\left(\frac{1}{\sqrt{n}} \right) = \frac{1}{2} \frac{t^2}{\sigma^2} . \sigma^2 + O\left(\frac{1}{\sqrt{n}} \right)$$

$$= \frac{t^2}{2} + O\left(\frac{1}{\sqrt{n}} \right) \Rightarrow \lim_{n \to \infty} \ln M_Z(t) = \frac{1}{2} t^2$$

and hence $\lim\limits_{n \to \infty} M_Z(t) = e^{\frac{t^2}{2}} \Rightarrow Z \sim N(0,1)$

<u>Sampling Distribution of the Difference Between Two Averages:</u>

Theorem : If $X_{11}, X_{12}, X_{13}, \ldots, X_{1n_1}$ ‘ $X_{21}, X_{22}, \ldots, X_{2n_2}$ are independent random variables the first n_1 constituting a random sample from an infinite population with the mean μ_1 and the variance σ_1^2 and the other n_2 constituting a random sample from an infinite population with the mean μ_2 and the variance σ_2^2 , then

(a) $E\left(\overline{X}_1 - \overline{X}_2\right) = \mu_1 - \mu_2$

(b) $Var\left(\overline{X}_1 - \overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Proof : using the following theorems

$$E(a X + b) = a E(X) + b$$
$$Var (a X + b) = a^2 Var (X)$$
$$E(\overline{X}) = \mu$$
$$Var (\overline{X}) = \frac{\sigma^2}{n}$$

we have,

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

$$Var(\overline{X}_1 - \overline{X}_2) = Var(\overline{X}_1) + Var(-\overline{X}_2)$$
$$= Var(\overline{X}_1) + Var(\overline{X}_2)$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2}{n_2}$$

Theorem : Consider the two independent random variables

$$(X_{11}, X_{12}, ..., X_{1n_1}) \sim N(\mu_1, \sigma_1^2) \text{ and}$$
$$(X_{21}, X_{22}, ..., X_{2n_1}) \sim N(\mu_2, \sigma_2^2)$$

Show that $Z = (\overline{X}_1 - \overline{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Proof : $M_Z(t) = M_{\overline{X}_1 - \overline{X}_2}(t) = E\left(e^{t(\overline{X}_1 - \overline{X}_2)}\right)$

$$= E\left(e^{t\overline{X}_1}\right) . E\left(e^{-t\overline{X}_2}\right)$$

$$M_Z(t) = E\left(e^{\frac{t}{n}\Sigma X_1}\right) . E\left(e^{-\frac{t}{n}\Sigma X_2}\right)$$

$$= \prod_{i=1}^{n_1} E\left(e^{t\frac{X_1}{n_1}}\right) . \prod_{i=1}^{n_2} E\left(e^{-t\frac{X_2}{n_2}}\right)$$

$$= \left(E\left(e^{t\frac{X_1}{n_1}}\right)\right)^{n_1} . \left(E\left(e^{-t\frac{X_2}{n_2}}\right)\right)^{n_2}$$

$$= \left(e^{\frac{\mu_1 t}{n_1} + t^2\frac{\sigma_1^2}{2n_1^2}}\right)^{n_1} . \left(e^{-\frac{\mu_2 t}{n_2} + t^2\frac{\sigma_2^2}{2n_2^2}}\right)^{n_1}$$

$$= e^{\mu_1 t + t^2 \frac{\sigma_1^2}{2n_1}} . e^{-\mu_2 t + t^2 \frac{\sigma_2^2}{2n_2}}$$

$$= e^{(\mu_1 - \mu_2) t + t^2 \left(\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2} \right)}$$

So, we can write

$$Z = \left(\overline{X}_1 - \overline{X}_2\right) \sim N\left(\mu_1 - \mu_2 , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)$$

Exercises

(1) If $X_1, X_2, ..., X_n$ are independent random variables having identical Bernoulli distribution with parameter θ , then $\overline{X}$ is the proportion of successes in n trials, which we denote by $\hat{\theta}$, verify that $E(\hat{\theta}) = \theta$ and $Var(\hat{\theta}) = \frac{\theta (1-\theta)}{n}$.

(2) Consider $X_{11}, X_{12}, ..., X_{1n_1}$ and $X_{21}, X_{22}, ..., X_{2n_2}$ are independent random variables, with the first n_1 random variables have Bernoulli distribution with the parameter θ_1 and the other n_2 random variables have Bernoulli distributions with the parameter θ_2 , show that in the notation of the above exercise :

(i) $E(\hat{\theta}_1 - \hat{\theta}_2) = \theta_1 - \theta_2$

(ii) $Var(\hat{\theta}_1 - \hat{\theta}_2) = \frac{\theta_1 (1-\theta_1)}{n_1} + \frac{\theta_2 (1-\theta_2)}{n_2}$

The Chi-Square Distribution :

A random variable X has the chi-square distribution with ν degrees if its probability density is given by

$$f(x)=\begin{cases} \dfrac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{\frac{\nu-2}{2}}\, e^{-x/2} & , \quad x>0 \\ 0 & , \text{ o.w.} \end{cases}$$

The mean and the variance of the chi-square distribution with ν degrees of freedom ν and 2ν and the moment generating function is given by

$$M_X(t)=(1-2t)^{-\nu/2}$$

The chi-square distribution has several important mathematical properties, which are given throught the following theorems.

Theorem : If the random variable X is $N(\mu,\sigma^2)$ then the random variable $W=\dfrac{X-\mu}{\sigma}$ is N(0, 1).

Proof : The distribution function $G(\omega)$ of W is

$$G(\omega)=\Pr\left(\frac{X-\mu}{\sigma}\le\omega\right)=\Pr\left(X\le\omega\,\sigma+\mu\right)$$

that is, $G(\omega)=\int_{-\infty}^{\omega\,\sigma+\mu}\dfrac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$

we change the variable of integration by writing $y=\dfrac{(x-\mu)}{\sigma}$, then

$$G(\omega)=\int_{-\infty}^{\omega}\frac{1}{\sqrt{2\pi}}\,e^{-y^2/2}\,dy$$

The p.d.f. $g(\omega)=\dfrac{d}{d\omega}\left[G(\omega)\right]$ of the random variable W is

$$g(\omega)=\frac{1}{\sqrt{2\pi}}\,e^{-\omega^2/2} \quad , \; -\infty<\omega<\infty$$

Thus W is N(0, 1).

Theorem : If the random variable X is $N(\mu,\sigma^2)$ then the random variable $V=\dfrac{(X-\mu)^2}{\sigma^2}$ is $\chi^2(1)$.

Proof : Because $V = W^2$, where $W = \frac{X - \mu}{\sigma}$ is N(0, 1)

the distribution function $G(\nu)$ of V is

$$G(\nu) = \Pr\left(W^2 \le \nu\right) = \Pr\left(-\sqrt{\nu} \le W \le \sqrt{\nu}\right)$$

That is,

$$G(\nu) = \begin{cases} 2\int_0^{\sqrt{\nu}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2}\, d\omega , & \nu \ge 0 \\ 0 & , \text{o.w.} \end{cases}$$

we change the variable of integration by writing $\omega = \sqrt{y}$, then

$$G(\nu) = \int_0^{\nu} \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2}\, dy ,$$

The p.d.f. $g(\nu) = G'(\nu)$, so,

$$g(\nu) = \begin{cases} \frac{1}{\sqrt{\pi}\sqrt{2}\sqrt{\nu}} e^{-\nu/2} , & \nu > 0 \\ 0 & , \text{o.w.} \end{cases}$$

$$= \begin{cases} \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{2}} \nu^{-\frac{1}{2}-1} e^{-\nu/2} , & \nu > 0 \\ 0 & , \text{o.w.} \end{cases}$$

so, V is $\chi^2(1)$.

Theorem : Let $X_1, X_2, \ldots, X_n$ be independent random variables having, respectively, the normal distributions $N\left(\mu_1, \sigma_1^2\right)$, $N\left(\mu_2, \sigma_2^2\right)$, …, $N\left(\mu_n, \sigma_n^2\right)$, the random variable $Y = \sum_{i=1}^{n} K_i X_i$, where K_i , i = 1, …, n are real constants, is normal with mean $\sum_{i=1}^{n} K_i \mu_i$ and variance $\sum_{i=1}^{n} K_i^2 \sigma_i^2$, that is Y is

$$N\left(\sum_{i=1}^{n} K_i \mu_i , \sum_{i=1}^{n} K_i^2 \sigma_i^2\right) .$$

Proof : $M_Y(t) = E\{\exp\left[t\left(K_1X_1 + K_2X_2 + \ldots + K_nX_n\right)\right]\}$

$$= E\left(e^{tK_1X_1}\right)E\left(e^{tK_2X_2}\right)\ldots E\left(e^{tK_nX_n}\right)$$

Hence $E\left(e^{tK_iX_i}\right) = \exp\left[\mu_i\,(K_i\,t) + \sigma_i^2\,\frac{(K_i\,t)^2}{2}\right]$

That is $\quad M_Y(t) = \prod_{i=1}^{n} \exp\left[(K_iX_i)\,t + \frac{\left(K_i^2\sigma_i^2\right)t^2}{2}\right]$

$$= \exp\left[\left(\sum_{i=1}^{n} K_i\,\mu_i\right)t + \frac{\left(\sum_{i=1}^{n} K_i^2\,\sigma_i^2\right)t^2}{2}\right]$$

$$\Rightarrow\ Y \sim N\left(\sum_{i=1}^{n} K_i\,\mu_i\,,\ \sum_{i=1}^{n} K_i^2\,\sigma_i^2\right)$$

Theorem : Let $X_1, X_2, \ldots, X_n$ be independent variables that have, the chi-square distributions $\chi^2(r_1), \ldots, \chi^2(r_n)$, then $Y = \sum_{i=1}^{n} X_i$ is $\chi^2(r_1 + \ldots + r_n)$.

Proof : $\quad M_Y(t) = E\left(e^{t(X_1 + X_2 + \ldots + X_n)}\right)$

$$= E\left(e^{tX_1}\right)E\left(e^{tX_2}\right)\ldots E\left(e^{tX_n}\right)$$

Since $\quad E\left(e^{tX_i}\right) = (1 - 2t)^{-r_i/2}\ ,\ t < \frac{1}{2},\ i = 1, 2, \ldots, n$

We have $\quad M_Y(t) = (1 - 2t)^{-(r_1 + r_2 + \ldots + r_n)/2}$

That is $\quad Y = \sum_{i=1}^{n} X_i \sim \chi^2(r_1 + r_2 + \ldots + r_n)$

Theorem : Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that is $N\left(\mu, \sigma^2\right)$, then $Y = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$ is $\chi^2(n)$.

Proof : If $X \sim N\left(\mu, \sigma^2\right)$, then $\left(\frac{X - \mu}{\sigma}\right)^2 \sim \chi^2(1)$ using the above theorem, we have $\quad Y = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$

The chi-square distribution has many important applications, foremost, there are those based directly or indirectly, on the following theorem.

Theorem : If $\overline{X}$ and S^2 are the mean and the variance of the random sample if size n from a normal population with the mean μ and the variance σ^2, then

1. $\overline{X}$ and S^2 are independent.
2. The random variable $\frac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with (n-1) degrees of freedom.

Proof : The proof of part 1. is left to the reader .

To prove part 2, let us begin with the identity

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

Now, if we divide each term by σ^2 and substitute $(n-1)S^2$ for $\sum (X_i - \overline{X})^2$, we have

$$\sum \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)}{\sigma^2} S^2 + \left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2$$

since $$\sum \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

$$\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2 \sim \chi^2(1)$$

Now, since $\overline{X}$ and S^2 are independent, it follows that $\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2$ and $\frac{(n-1)}{\sigma^2} S^2$ are independent, and, hence $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

<u>The t distribution :</u>

We showed that for random samples from a normal population with the mean μ and the variance σ^2, $\overline{X}$ has a normal distribution with the mean μ and the variance $\frac{\sigma^2}{n}$, that is, $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ has the standard normal distribution. The major difficulty in applying this result is that in actual practice σ is usually unknown, which makes it necessary to replace it with a value of the sample standard deviation S. the theory which follows leads to the exact distribution of $\frac{\overline{X} - \mu}{S / \sqrt{n}}$ for random samples from normal population.

To derive this sampling distribution, let us first study the more general problem stated in the following theorem.

Theorem : If y and z are independent random variables, y has a chi-square distribution with ν degrees of freedom, and z has the standard normal distribution, then the distribution of

$$t = \frac{z}{\sqrt{y/\nu}}$$

is given by :

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} . \left(1 + \frac{t^2}{\nu}\right) \quad \text{for } -\infty < t < \infty \ .$$

and it is called the t distribution with ν degrees of freedom.

Proof: Since y and z are independent , their joint density is given by

$$f(y,z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} . \frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}$$

for y > 0 and $-\infty < z < \infty$, and f(y, z) = 0 elsewhere. Then, to use the change of variable technique, we solve $t = \frac{z}{\sqrt{y/\nu}}$ for z getting

$z = t\sqrt{y/\nu}$ and, hence, $\frac{\partial z}{\partial t} = \sqrt{y/\nu}$. Thus, the joint density of y and t is given by

$$g(y,t)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}}\, y^{\frac{\nu-1}{2}}\, e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)} & \text{for } y>0 \text{ and } -\infty<t<\infty\\ 0 & \text{elsewhere}\end{cases}$$

and, integrating out y with the aid of the substitution $w=\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)$, we finally get

$$f(t)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\cdot\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \text{for } -\infty<t<\infty$$

The t distribution was first obtained by W. S. Gosset, who published his research under the pen name "Student"; hence, the distribution is also known as the **Student-t distribution**, or **Student's-t distribution**.

Theorem : If $\bar{x}$ and s^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the variance σ^2, then

$$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

has the t distribution with n - 1 degrees of freedom.

Proof : The random variables

$$y=\frac{(n-1)s^2}{\sigma^2} \quad \text{and} \quad z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

have, respectively, a chi-square distribution with n - 1 degrees of freedom and the standard normal distribution. Since they are, independent, substitution into the formula for the above theorem yields

$$t=\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}}=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

and this complete the proof.

The F Distribution :

Another distribution which plays an important role in connection with sampling from normal population is the **F distribution**. We shall define this distribution as the sampling distribution of the ratio of two independent chi-square random variables, each divided by its respective degrees of freedom.

Theorem : If u and v are independent random variables having chi-square distributions with ν_1 and ν_2 degrees of freedom, then the distribution of

$$F = \frac{u/\nu_1}{v/\nu_2}$$

is given by

$$g(F) = \begin{cases} \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} . F^{\frac{\nu_1}{2}-1} \left(1+\dfrac{\nu_1}{\nu_2}.F\right)^{-\frac{1}{2}(\nu_1+\nu_2)} & \text{for } F > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and it is called the **F distribution** with ν_1 and ν_2 degrees of freedom.

Proof : The joint density of u and v is given by

$$f(u, v) = \frac{1}{2^{\nu_1/2}\,\Gamma\left(\dfrac{\nu_1}{2}\right)} . u^{\frac{\nu_1}{2}-1} e^{-\frac{u}{2}} . \frac{1}{2^{\nu_2/2}\,\Gamma\left(\dfrac{\nu_2}{2}\right)} . v^{\frac{\nu_2}{2}-1} e^{-\frac{v}{2}}$$

$$= \frac{1}{2^{(\nu_1+\nu_2)/2}\,\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} . u^{\frac{\nu_1}{2}-1} v^{\frac{\nu_2}{2}-1} e^{-\frac{u+v}{2}}$$

for u > 0 and v > 0, and f(u, v) = 0 elsewhere. Then, to use the change of variable technique, we solve $F = \dfrac{u/\nu_1}{v/\nu_2}$ for u getting $u = \dfrac{\nu_1}{\nu_2}. vF$ and , hence, $\dfrac{\partial u}{\partial F} = \dfrac{\nu_1}{\nu_2}. v$. Thus, the joint density F and v is given by

$$g(F, v) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{2^{(\nu_1+\nu_2)/2}\,\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} . F^{\frac{\nu_1}{2}-1} v^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{v}{2}\left(\frac{\nu_1 F}{\nu_2}+1\right)}$$

for F > 0 and v > 0, and g(F, v) = 0 elsewhere. Now, integrating out v by making the substitution $w = \frac{\nu}{2}\left(\frac{\nu_1 F}{\nu_2}+1\right)$, we finally get

$$g(F) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} . F^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}F\right)^{-\frac{1}{2}(\nu_1+\nu_2)} & \text{for } F > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Theorem : If s_1^2 and s_2^2 are the variances of independent random samples of size n_1 and n_2 from normal populations with the variances σ_1^2 and σ_2^2 , then

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} = \frac{\sigma_2^2\, s_1^2}{\sigma_1^2\, s_2^2}$$

has an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

Proof : Left to the reader.

Exercises

(1) Let $X_1, X_2, \ldots, X_{25}$ and $Y_1, Y_2, \ldots, Y_{25}$ be two random samples from two independent normal distributions N(0, 16), and N(1, 9), respectively, $\overline{X}$ and $\overline{Y}$ denote the corresponding sample means. Find $\Pr\left(\overline{X} > \overline{Y}\right)$.

(2) Find the mean and variance of $S^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n}$ where $X_1, X_2, \ldots, X_n$ is a random sample from $N\left(\mu, \sigma^2\right)$.

(3) Let X_1 and X_2 be two independent random variables so that the variances of X_1 and X_2 are $\sigma_1^2 = k$ and $\sigma_2^2 = 2$, given that the variance of $Y = 3X_2 - X_1$ is 25, find k.

(4) Consider that $X \sim F(r_1, r_2)$, show that $Y = \frac{1}{X} \sim F(r_2, r_1)$.

(5) Consider that $T = \frac{X}{\sqrt{Y/n}}$, where

$$X \sim N(0, 1)$$

$$Y \sim \chi^2(n)$$

Show that $T^2 \sim F(1, n)$

(6) Consider that X ~ F(m, n). Find E(X).

(7) Let X_1, X_2 be a random independent sample from N(0, 1).

(i) what is the distribution of $\frac{X_2 - X_1}{\sqrt{2}}$?

(ii) " " " " " $\frac{(X_1 + X_2)^2}{(X_2 - X_1)^2}$?

(iii) " " " " " $\frac{(X_2 + X_1)}{\sqrt{(X_2 - X_1)^2}}$?

" " " " " $\frac{1}{Z}$ if

TABLE IV

THE NORMAL DISTRIBUTION

$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\,dw$$

$$[\Phi(-z) = 1 - \Phi(z)]$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

z	1.282	1.645	1.960	2.326	2.576
$1 - \Phi(z)$	0.100	0.050	0.025	0.010	0.005

References

1- Barnett, V. 1975, Elements of Sampling theory, EUP, London
2- Cochran, W. G. 1980 Statistical methods Iowa state U. Press
3- Cooke, D. and others, Basic Statistical computing, Arnold, London.
4- Durran, J. H.,1970 Statistics and Probability Cambridge U. Press, Cambridge
5- Thomas, G. H. 1970, Probability Wesley reading, mass .
6- Birnbaum, Z. W 1962, Introduction to probability and Mathematical Statistics.
7- Freund, J. E, Walpole, R. E 1980 Mathematical Statistics.
8- Kupur, J. N, Saxena, H. C Mathematical Statistics.
9- Meyer, P. L. 1970, Introductory Probability and Statistical Applications.
10- Smernove, N. V, Dynin, I. V. 1965, Course of Probability Theory and Mathematical Statistics.
11- Douglas C. Montgomery , George C. Runger 2003 Applied Statistics and Probability for Engineering