**ADVENTIST UNIVERSITY OF CENTRAL AFRIQUE**

**FACULTY OF BUSINESS ADMINISTRATION**

**ACADEMIC YEAR: 2018-2019**

**COURSE: DESCRIPTIVE STATISTICS**

**COURSE CODE: STAT 122**

**Prepared by Mr. Pascal Kayisire**

**Masoro, January, 2018**

**Course objectives**

As the collection and analysis of data are fundamental to business and economics; when managers are well informed about a company’s internal operations (e.g. sales, production, inventory levels, time to market) and competitive position (e.g. market share, customer satisfaction, repeat sales) can take appropriate actions to improve their business. In this case manager needs to correct data and analyze the gathering data for making a reasonable decision. It is why in this course the students learn how to correct and analyze the quantitative data, and then this course builds on the material that was learn in the initial collection, analysis, and interpretation of uncertain data. Students get to learn advance statistical methods that will use designing valid experiments and drawing reliable conclusions from the data they produce. This course is a gentle introduction to statistics and its simple applications.

**Learning outcomes**

Upon the completion of this course, the student should be able to:

• Understand and explain some statistics concepts as tools of analysis of real life situations, specifically in business.

• Use techniques developed in statistics to model and solve real life problems related to business.

**Indicative content**

**Part 1. Descriptive statistics**

1.1. Overview of statistics

1.2. Numerical summaries

**1.3.** Simple linear regression.

**Part 2. probability theory**

2.1. Introduction

2.2. Some probability distribution

**References**

1. Lind, A. D., Marchal, G. W. and Mason, D. R. (2002) Statistical Techniques in Business & Economics, 11th Ed., by McGraw-Hill.
2. Thomas H., Ronald J. (1977). Introductory Statistics Third edition, New York.
3. Robert V. Hogg.(2006). Probability and Statistical Inference. 7e edition. New Jersey.
4. Murray R., Larry J.. (1998).Schaum’s outlines Statistics Third Edition. New York.

**Part I: DESCRIPTIVE STATISTICS**

1. **INTRODUCTION**

**Question: What is statistics?**

Basically, statistics is the “science of data.” There are three main tasks in statistics: (A) collection and organization, (B) analysis, and (C) interpretation of data.

1. **Collection and organization of data**: We will see several methods of organizing data: graphically (through the use of charts and graphs) and numerically (through the use of tables of data). The type of organization we do depends on the type of analysis we wish to perform.

**Quick Example**: Let us collect the status (freshman, sophomore, junior, senior) of a group of 20 students in this class. We could then organize the data in any of the above ways.

1. **Analysis of data:** Once the data is organized, we can go ahead and compute various quantities (called *statistics* or *parameters*) associated with the data.

**Quick Example** Assign 0 to freshmen, 1 to sophomores etc. and compute the mean.

1. **Interpretation of data:** Once we have performed the analysis, we can use the information to make assertions about the real world (e.g. the average student in this class has completed x years of college).

**Definition: Statistics** is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

**1.1. TYPES OF STATISTICS**

The study of statistics is usually divided into two categories; descriptive and inferential statistics.

1. **Descriptive statistics** is the methods of organizing, summarizing, and presenting data in an informative ways.

The data will refer to a given set of “objects” called a population.

2. **Inferential statistics** is the methods used to determine something about a population, based on sample.

Inferential statistics is also called statistical inference and inductive statistics.

**DEFINITION**: - population is a collection of all possible individuals, objects, or measurements of interest.

- Sample is a portion, or part of population of interest.

* 1. **TYPES OF VARIBLES**

They are two basic types of data: (1) those obtained from a qualitative population and

(2) Those obtained from quantitative population.

When the characteristics or variable being studied is non numeric, it is called a **qualitative variable** or an **attribute**.

Examples of qualitative variables are gender, religious affiliation, type of automobile owned, state of birth, and eye color.

When the data being studied are qualitative, we are usually interested in how many or what proportion fall in each categories.

For example, what percent of population has blue eyes? How many Catholics and how many Protestants are there in the United States? Qualitative data are often summarized in charts and bar graphs.

When the variables studied can be reported numerically, the variable is called quantitative variable, and population is called a quantitative population. Examples of quantitative variables are balance in your checking account, the ages of company presidents, the number of children in a family, the life of a battery (such as 42 moths) , and so on.

**Quantitative variables** are either **discrete** or **continuous**.

Discrete variables can assume only certain values, and there are usually “gaps” between the values. Example: number of bedrooms in a house (1,2,3,4,etc.) ,number of students in each section of statistics course (25 in section A ,42 in section B and 18 in section C).

Notice that a home can have 3 or 4 bedrooms, but it cannot have 3.56 bedrooms. Thus there is a “gap” between possible values. Typically, discrete variables result from counting. Observations of continuous variable can assume any values within a specific range. Typically continuous variables result from measuring something.

**Examples**: -Heights of the children in a school,

-The weights (in kg) of students in the class.

- Temperature, pressure, typically, continuous variables result from measuring something.

**1.3. LEVELS OF MEASUREMENT**

Data can be classified according to levels of measurements. The levels of measurements of the data often dictate the calculations that can be done to summarize and present the data. It will also determine the statistical tests that should be performed.

There are actually four levels of measurement: Nominal, ordinal, interval, and ratio. The “lowest”, or the most primitive, measurements are the nominal level. The highest, or the level that gives us the most information about the observation, is the ratio level of measurement.

**1. Nominal level data**

In the nominal level of measurement, the observations can only be classified and counted. There is no particular order to the labels. It means that if data are classified into categories and the order of those categories is not important, it is the nominal level of measurement.

**Example:** - Gender (male, female) and political affiliation (republican, democrat, independent, all others). If it makes no difference whether male or female is listed first. They are nominal level.

To summarize, the nominal level data have the following properties:

1. Data categories are **mutually exclusive** and **exhaustive.**
2. Data categories have no logical order.

**Definition**: -**Mutually exclusive** is a property of a set of categories such that an individual, object, or measurement is included in only one category.

-**Exhaustive** is a property of a set of categories such that each individual, object, or measurement must appear in a category.

**2. Ordinal level data**

The next higher level of data is the ordinal level. Data that can be logically ranked are referred to as ordinal measures. **Example:** Consumer response to the sound of a new speaker might be excellent, very good, fair, or poor.

In summary, the properties of ordinal level data are:

* The data classifications are mutually exclusive and exhaustive.
* Data classifications are ranked or ordered according to the particular trait they possess.

**3. Interval level data**

The interval level of measurement is the next higher level. It includes all the characteristics of the ordinal level, but in addition, the difference between values is a constant size. (If one observation is greater than another by a certain amount and the zero point is arbitrary, the measurement is the one an interval scale). **Example**: - The difference between temperatures of 70 degrees and 80 degrees is 10 degrees. Likewise a temperature of 90 degrees is 10 degrees more than a temperature of 80 degrees.

The properties of interval level data are:

* The data classifications are mutually exclusive and exhaustive.
* Data classifications are scaled according to the amount of the characteristic they possess.
* Equal differences in the characteristic are represented by equal differences in the measurements.

**4. Ratio level data**

Practically all quantitative data are the ratio level of measurement. The ratio level is the “highest” level of measurement. It has all the characteristic of the interval level but in addition the 0 point is meaningful and the ratio between two numbers is meaningful. Example of ratio scale of measurement include: wages, units of production, weight, changes in stock prices, distance between branch offices, and height. Money is a good illustration. If you have 0 dollars, then you have no money. The ratio of two numbers is also meaningful. If Jim earns$300for selling insurance and Rob earns $ 600 for year selling cars, then Rob earns twice as such as Jim. EXAMPLE 2: the difference between $ 300 and $200 is $100.

The properties of ratio level data are:

* The data classifications are mutually exclusive and exhaustive.
* Data classifications are scaled according to the amount of the characteristics they possess.
* Equal differences in the characteristic are represented by equal differences in the number assigned to the classifications.
* The zero point is the absence of the characteristic.

1. **Descriptive data**

**2.1.1. Constructing a Frequency Distribution**

Frequency distribution is a grouping of data exclusive classes showing the number of observation in each. How do we develop a frequency distribution? The first step is to tally the data into a table that shows the classes (categories) and the number of observation in each category. The steps in constructing a frequency distribution are best described using an example. Remember our goal is to make a table that will quickly reveal the shape of the data.

Consider the following data table which summarize the selling prices of the vehicles sold last month at whitener Pontiac

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| $20,197 | $20,372 | $17,454 | $20,591 | $23,651 | $24,453 | $14,266 | $15,021 | $25,683 | $27,872 |
| 16,587 | 20,169 | 32,851 | 16,251 | 17,047 | 21,285 | 21,324 | 21,609 | 25,670 | 12,546 |
| 12,935 | 16,873 | 22,251 | 22,277 | 25,034 | 21,533 | 24,443 | 16,889 | 17,004 | 14,357 |
| 17,155 | 16,688 | 20,657 | 23,613 | 17,895 | 17,203 | 20,765 | 22,783 | 23,661 | 29,277 |
| 17,642 | 18,981 | 21,052 | 22,799 | 12,794 | 15,263 | 32,925 | 14,399 | 14,968 | 17,346 |
| 18,442 | 18,722 | 16,331 | 19,817 | 16,766 | 17,633 | 17,962 | 19,845 | 23,285 | 24,896 |
| 26,076 | 29,492 | 15,890 | 18,740 | 22,449 | 19,374 | 21,571 | 25,337 | 17,642 | 20,613 |
| 21,220 | 27,655 | 19,442 | 14,891 | 17,818 | 23,237 | 17,445 | 18,556 | 18,639 | 21,296 |

We refer to the unorganized information in table as **raw data** or ungrouped data. The raw data are more easily interpreted if organized into a frequency distribution.

**Step 1**: **Decide on the number of classes**. The goal is to use just enough groupings or classes to reveal the shape of distribution.

A useful guidelines to determine the number of classes “2 to the k rule” this guide suggest, you select the smallest number (k) for the number of classes such that  is a greater than the number of observations (n).

In the whiter Pontiac example, there were 80 vehicles sold so n=80 if we try k=6, then .somewhat less than 80. Hence 6, is not enough class. If we let k=7, then, which is greater than 80. So the recommended number of classes is 7.

**Step 2**: **Determine the class interval or width**. Generally the class interval or width should be the same for all classes. The classes all taken together must cover at least the distance from the lowest value in the raw data up to the highest value. Expressing these words in a formula: ****

Where  is the class interval,  is the highest observed value,  is the lowest observed value, and  is the number of classes.

In whitener Pontiac case the lowest value is $12,546 and the highest value is $32,925. Then we need 7 classes, the interval should be at least



**In practice this interval size is usually rounded up to some convenient number, such as a multiple of 10 or 100**. The value of $3,000 might readily be used in this case.

In frequency distributions, equal class intervals are preferred. However, unequal class intervals may be necessary in certain situations to avoid a large number of empty, or almost empty, classes.

**Step 3:** **Set individual class limits.** State clear class limits so you can put each observation into only one category this means you must avoid unclear class limits. Class such as $1,300-$1,400 and $1,400 -$1,500 should not be used because it is not clear whether the value $ 1,400 is in 1st or 2nd class. In this text we will generally use the format $1,300 up to $1,400 and $1,400 up to $1,500 etc. or [1,300-1,400[with this format. it is clear that $1,399 goes into the 1st class and $1,400 in the 2nd class. Now in the case of whiter Pontiac the class would be defined as

|  |
| --- |
| $12,000 up to 15,000  15,000 up to 18,000  18,000 up to 21,000  21,000 up to 24,000  24,000 up to 27,000  27,000 up to 30,000  30,000 up to 33,000 |

Or

[12,000-15,000[

[15,000-18,000[

[18,000-21,000[

[21,000-24,000[

[24,000-27,000[

[27,000-30,000[

[30,000-33,000[

It means Lower limit of the 1st class is price interval<15,000 the lower limit of the 2nd class.

**Step 4**: **Tally the vehicle selling prices into the classes**. To begin, the selling price of the first vehicle in table is $20,197. It is tallied in the $18,000 up to $21,000 class. Therefore, when all the selling prices are tallied, the table would appear as:

|  |
| --- |
| Class tallied |
| [12,000-15,000[ ~~||||~~ |||  [15,000-18,000[ ~~||||~~ ~~||||~~  ~~||||~~ ~~||||~~ |||  [18,000-21,000[ ~~||||~~ ~~||||~~ ~~||||~~ ||  [21,000-24,000[ ~~||||~~ ~~||||~~ ~~||||~~ |||  [24,000-27,000[ ~~||||~~ |||  [27,000-30,000[ ||||  [30,000-33,000[ || |

**Step 5:** **count the number of item on each class**. The number of observation in each class is called the frequency. In the $ 12,000 up to $15,000 class there are 8 observations and in the $15,000 up to $ 18,000 class there are 23 observations. Therefore, the class frequency in the first class is 8 and the class frequency in second class is 23. There are a total of 80 observations or frequencies in the entire set of data.

The following table summarizes, Frequency Distribution of Selling Prices at Pontiac Last month

|  |
| --- |
| Selling prices Frequency |
| [12,000-15,000[ 8  [15,000-18,000[ 23  [18,000-21,000[ 17  [21,000-24,000[ 18  [24,000-27,000[ 8  [27,000-30,000[ 4  [30,000-33,000[ 2  **Total 80** |

**Exercise:** 1.The commissions earned for the first quarter of last year by the eleven members of sales staff at Master Chemical Company are:

|  |
| --- |
| $1,650 $1,475 $1,510 $1,670 $1,595 $1,760 $1,540 $1,495 $1,590 $1,625 $1,510 |

1. What are the values such as $1,650 and $1,475 called?
2. Using $1,400 up to $1,500 as the first class, $1,500 up to $1,600 as second class, and so forth, organize the quarterly commissions into a frequency distribution.
3. What are the numbers in the right column of your frequency distribution called?
4. Describe the distribution of monthly commissions, based on the frequency distribution. What is the largest amount of commission earned? What is the smallest? What is the typical amount earned?
   * 1. **Class intervals and Class Midpoints**

We will use two other terms frequently: **class midpoint** and **class interval**. The midpoint, also called the **class mark**, is halfway between the lower limits of the consecutive classes. It can be computed by adding the lower class limit to the upper class limit and dividing by 2. Referring to the table of Frequency Distribution of Selling Prices at Pontiac Last month, for the first class the lower class limit is$12,000 and the next limit is $15,000. The class midpoint is $13,500, found by ($12,000+$15,000)/2.

To determine the class interval, subtract the lower limit of the class from the lower limit of the next class. The class interval of the vehicle selling price data is $3,000, which we found by subtracting the lower of the first class, $12,000, from the lower limit of the next class; that is, $15,00 - $412,000=$3,000.You can also determine the class interval by finding the difference between consecutive midpoints. The midpoint of the first class is 413,500 and the midpoint of the second class is $16,500. The difference is $3,000.

**2.1.3. Relative Frequency Distribution**

It may be desirable to convert class frequencies to relative class frequencies to show the fraction of the total number of observations in each class. To convert a frequency distribution to relative frequency distribution, each of the class frequencies is divided by total number of observations.

The Relative frequency Distribution of the Prices of Vehicles sold Last Month at Whitner Pontiac.

|  |  |  |  |
| --- | --- | --- | --- |
| Selling Price in $ | Frequency | Relative Frequency | Found by |
| 12000 up to 15000 | 8 | 0.1 | 8/80 |
| 15000 up to 18000 | 23 | 0.2875 | 23/80 |
| 18000 up to 21000 | 17 | 0.2125 | 17/80 |
| 21000 up to 24000 | 18 | 0.225 | 18/80 |
| 24000 up to 27000 | 8 | 0.1 | 8/80 |
| 27000 up to 30000 | 4 | 0.05 | 4/80 |
| 30000 up to 33000 | 2 | 0.025 | 2/80 |
| Total | 80 | 1 |  |

Exercise: 1. Refer to above table, which shows the Relative frequency Distribution of the Prices of Vehicles sold Last Month at Whitner Pontiac.

1. How many vehicles sold for $15,000 up to $18,000?
2. What percent of the vehicles sold for a price between $15,000 and $18,000?
3. What percent of the vehicles sold for $27,000 or more?
4. Assume the table below describes the number of patients received by Gisenyi Hospital within 30 days of June.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 83 | 64 | 84 | 76 | 84 | 54 | 75 | 59 | 70 | 61 |
| 63 | 80 | 84 | 76 | 68 | 52 | 65 | 90 | 52 | 77 |
| 95 | 36 | 78 | 61 | 59 | 84 | 95 | 47 | 87 | 60 |

1. How many classes and class interval would you recommend?
2. Organize the data into a frequency distribution and determine the relative frequency?

**Stem-and- Leaf Displays**

In the previous section, we showed how to organize data into a frequency distribution so we could summarize the raw data into a meaningful form. The major advantage to organizing data into a frequency distribution is that we get a quick visual picture of the shape of the distribution without doing any further calculation. That is, we can see where the data are concentrated and also determine whether there are any extremely large or small values. There are two disadvantages, however to organizing the data into a frequency distribution: (1) we lose the exact identity of each value and (2) we are not sure how the values within each class are distributed.

One technique that is used to display quantitative information in a condensed form is the **Stem-and- leaf display**. An advantage of the stem-and-leaf display over a frequency distribution is that we do not lose the identity of each observation. A stem-and-leaf display is similar to frequency distribution with more information, i.e. data values instead of tallies.

**Definition**: **STEM-AND-LEAF DISPLAY** is a statistical technique to present a set of data. Each numerical value is divided into two parts. The leading digit(s) becomes the stem and the trailing digit the leaf. The stems are located along the vertical axis, and the leaf values are stacked against each other along the horizontal axis.

Example: Listed in table bellow is the number of 30-second radio advertising sports purchased by each of 45 members of the Greater Buffalo Automobile Dealers Association last year. Organize data into a stem-and-leaf display. Around what values do the numbers of advertising sports tend to cluster? What is the fewest number of sports purchased by a dealer? The largest number purchased?

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 96 | 93 | 88 | 117 | 127 | 95 | 113 | 96 | 108 | 94 | 148 | 156 |
| 139 | 142 | 94 | 107 | 125 | 155 | 155 | 103 | 112 | 127 | 117 | 120 |
| 112 | 135 | 132 | 111 | 125 | 104 | 106 | 139 | 134 | 119 | 97 | 89 |
| 118 | 136 | 125 | 143 | 120 | 103 | 113 | 124 | 138 |  |  |  |

From the data in table we note that the smallest number of spots purchased is 88. So we will make the first stem value 8. The largest number is 156, so we will have a stem values begin at 8 and continue to 15. The first number in table is 96, which will have a stem value of 9 and a leaf value of 6. Moving across the top row, the second value is 93 and the third is 88.

Organizing all the data, the stem-and-leaf chart looks as follows.

|  |  |
| --- | --- |
| Stem | Leaf |
| 8 | 8 9 |
| 9 | 6 3 5 6 4 4 7 |
| 10 | 8 7 3 4 6 3 |
| 11 | 7 3 2 7 2 1 9 8 3 |
| 12 | 7 5 7 0 5 5 0 4 |
| 13 | 9 5 2 9 4 6 8 |
| 14 | 8 2 3 |
| 15 | 6 5 5 |

The usual procedure is to sort the leaf values from the smallest to largest. The last line the row referring to the values in the 150s, would appear as:

|  |  |
| --- | --- |
| 15 | 5 5 6 |

The Final table would appear as follows, where we have sorted all of the leaf values.

|  |  |
| --- | --- |
| Stem | Leaf |
| 8 | 8 9 |
| 9 | 3 4 4 5 6 6 7 |
| 10 | 3 3 4 6 7 8 |
| 11 | 1 2 2 3 3 7 7 8 9 |
| 12 | 0 0 4 5 5 5 7 7 |
| 13 | 2 4 5 6 8 9 9 |
| 14 | 2 3 8 |
| 15 | 5 5 6 |

You can draw several conclusions from the stem-and- leaf display. First the lowest number of spots purchased is 88 and the largest is 156. Two dealers purchased less than 90 sports, and three purchased 150 or more. You can observe, for example, that the three dealers who purchased more than 150 spots actually purchased 155, 155, and 156 spots. The concentration of the number of spots is between 110 and 130. There were 9 dealers who purchased between 110 and 119 spots and 8 who purchased between 120 and 129 spots.

**Exercise:** 1) The price-earnings ratios for 21 stocks in the retail trade category are:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8.3 | 9.6 | 9.5 | 9.1 | 8.8 | 11.2 | 7.7 | 10.1 | 9.9 | 10.8 |  |
| 10.2 | 8 | 8.4 | 8.1 | 11.6 | 9.6 | 8.8 | 8 | 10.4 | 9.8 | 9.2 |

Organize this information into a stem-and-leaf display.

1. How many values are less than 9.0?
2. List the values in the 10.0 up to 11.0 category.
3. What is the middle value?
4. What are the largest and the smallest price-earnings ratios?
   1. **Graphic Presentation of Frequency Distribution**
      1. Graphic Presentation of Qualitative Data

The most common graphic form to present a qualitative variable is **a bar chart**.

**BAR CHART** A graph that shows qualitative classes on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

Example: The table below lists the number of workers who belong to different categories based on the amount of stress they experience in their jobs.

|  |  |
| --- | --- |
| Category | Number of workers |
| High stress | 23 |
| Moderate stress | 42 |
| Minimal stress | 25 |
| No stress | 10 |

The Bar chart bellow shows the number of workers experienced stress in their jobs

Another useful type of chart for depicting qualitative information is a **pie chart.**

**PIE CHART** A chart that shows the proportion or percentage that each class represents of the total number of frequencies.

We can still use the same data from above example of number of workers who experienced stress in their jobs to represent pie chart graphically.

|  |  |
| --- | --- |
| Category | Number of works |
| High stress | 23 |
| Moderate stress | 42 |
| Minimal stress | 25 |
| No stress | 10 |

Note that, Pie charts and bar charts serve much the same function. What are the criteria for selecting one over the other? In most cases, pie charts are the most informative

* + 1. Graphic Presentation of Quantitative Data

Sales managers, stock analysts, hospital administrators, and other busy executives often need a quick picture of the trends in sales, stock prices, or hospital costs. These trends can often be depicted by the use of charts and graphs. Three charts that will help portray a frequency distribution graphically are the histogram, the frequency polygon, and the cumulative frequency polygon.

1. **Histogram**

One of the most common ways to portrays to frequency distribution is a **histogram**

**Histogram** is a graph in which the classes are marked on horizontal axis and the class frequencies on the vertical axis. The class frequencies are drawn adjacent to each other.

For instance consider the following table of the Relative frequency Distribution of the Prices of Vehicles sold Last Month at Whitner Pontiac

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frequency Distribution - Quantitative | | | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  | ***Data*** |  |  |  |  |  | *cumulative* | |
|  | *lower* |  | *upper* | *midpoint* | *width* | *frequency* | *Relative frequency* | *frequency* | *percent* |
|  |  |  |  |  |  |  |  |  |  |
|  | 12,000 | < | 15,000 | 13,500 | 3,000 |  | 0.10 | 8 | 10.0 |
|  | 15,000 | < | 18,000 | 16,500 | 3,000 | 23 | 0.288 | 31 | 38.8 |
|  | 18,000 | < | 21,000 | 19,500 | 3,000 | 17 | 0.213 | 48 | 60.0 |
|  | 21,000 | < | 24,000 | 22,500 | 3,000 | 18 | 0.225 | 66 | 82.5 |
|  | 24,000 | < | 27,000 | 25,500 | 3,000 | 8 | 0.100 | 74 | 92.5 |
|  | 27,000 | < | 30,000 | 28,500 | 3,000 | 4 | 0.050 | 78 | 97.5 |
|  | 30,000 | < | 33,000 | 31,500 | 3,000 | 2 | 0.025 | 80 | 100.0 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | 80 | 100.0 |  |  |

The Histogram of the Prices of Vehicles sold Last Month at Whitner Pontiac.

1. **Frequency Polygon**

A **frequency polygon** is similar to histogram. It consists of line segments connecting the points formed by the intersections of the class midpoints and the class frequencies.

The Frequency polygon of the Prices of Vehicles sold Last Month at Whitner Pontiac.

1. **Cumulative Frequency Distributions**

Consider once again the distribution of selling prices of vehicles at Whitner Pontiac. Suppose we were interested in the number of vehicles that sold for less than $18,000 or the value below which 40 percent of the vehicles sold. These numbers can be approximated by developing a **cumulative frequency distribution** and portraying it graphically in a **cumulative frequency polygon.**

The cumulative frequency distribution of the Prices of Vehicles sold Last Month at Whitner Pontiac

**Exercise:** The following represents the marks obtained by 30 students in a month class in final exam 66,61,65,59,87,61,58,70,77,94,45,80,58,78,49,72,75,97,84,82,79,75,68,75,90,78,85,67,57,52

i) Construct a stem and leaf display for this data.

ii) Construct a frequency distribution for this data.

iii) Construct for this data: - a histogram, - a frequency polygon

1. **MEASURES OF CENTRAL TENDENCY**

Measure of central tendency is a single value that summarizes a set of data. It locates the center of the values. There is not just one measurement of central tendency; in fact, there are many. We will consider five: the arithmetic mean, the weight mean, the geometric mean, median and the mode. The mean, the mode, and the median are three of the most commonly used.

* 1. **Mean**

1. **The populations mean** is the sum of all the values in the population divided by the number of values in population.



Using mathematical symbols

|  |
| --- |
|  |

Where: : Represents the population mean. Greek lowercase letter

: The number of items in the population

: Represent any particular value

: The Greek capital letter “sigma” and indicate the operation of addition.

 : The sum of the x values.

Any measurable characteristic of a population is called a **parameter**. The mean of a population is an example of a parameter.  **PARAMETER** is a characteristic of a population**.**

**Example**: Find the mean of the number: 24, 26, 27, 29, 32, 42

=

 Where stands for the sample mean and the lower case is the number in the sample

The mean of a sample, or any other measure based on sample data, is called a **statistic.**

**STATISTIC** is a characteristic of a sample**.**

**Exercises**

1. The annual incomes of a sample of several middle management employees at wasting house are: $62,900, $69,100, $58,300 and $76,500.

a) Give the formula for the sample mean

b) Find the sample mean

c) Is the mean you computed in (b) a statistic or a parameter

d) What is your best estimate of the population mean?

2. All Students in advanced computer science 411 are considering the population. Their course grades are 92, 96, 61, 86, 79 and 84.

a) Give the formula for the population mean

b) Compute the mean course grade

c) Is the mean you computed in (b) a statistic or a parameter why?

3. Springers sold 95 Antonelli men’s suits for the regular price of $400. For the spring sale the suits were reduced to $200 and 126 were sold. At the final clearance, the price was reduced to $100 and the remaining 79 suits were sold.

a) What was the weighted mean price of an Antonelli suit?

b) Springer paid $200 a suit for the 300 suits. Comment on the store’s profit for suit if a salesperson receives a $25 commission for each one sold.

**Definition**: The Arithmetic mean of n observation isor 

Properties of the arithmetic mean

The arithmetic mean is used as a measure of central tendency. It has several important properties:

1. E very set of interval level data has a mean.
2. All the values are included in computing the mean.
3. A set of data has only one mean.
4. The mean is useful measure for comparing two or more populations.
5. The arithmetic mean is the only measure of central tendency where the sum of the deviation of each value from the mean will always be zero.

Expressed symbolically: .

**PROOF**: Let denoted the Arithmetic mean of  

Let  where d: deviation



1. **The Weighted mean**

The weighted mean is a special case of the arithmetic mean.

It occurs when there are several observations of the same value which might occur, if the data have been grouped into a frequency distribution.

In general the weighted mean of a set of numbers designated by  with the corresponding weights is computed by

Weighted mean: 

**Example 1** Find the mean from the following frequency distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable: | 10 | 11 | 12 | 13 | 14 |
| Frequency | 5 | 10 | 11 | 8 | 6 |

**Answer**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 10 | 5 | 50 |
| 11 | 10 | 110 |
| 12 | 11 | 132 |
| 13 | 8 | 104 |
| 14 | 6 | 84 |
|  | 40 | 480 |

Therefore the mean is 

**Example** 2: The carter construction company pays its hourly employees $6.50, $7.50 and $8.50 per hour. There are 26 hourly employees. 14 are paid at the $6.50 rate, 10 at the $7.50 rate and 2 at the $8.50 rate. What is the mean hourly rate paid the 26 employees?

Answer: 

The weighted mean hourly wage is rounded to $7.04

Example: Suppose the nearly Wendy’s restaurant sold median large, and biggie-sized soft drinks for $0.90, $1.25, and $1.50 respectively of the last ten drinks sold, 3 were medium, 4 were large, and 3 were biggie-sized.

Solution: To find the mean amount of the last ten drinks sold, we could use formula.



1. **Geometric mean**

The geometric mean is useful in finding the average of percentages, ratios, indexes, or growth rates.

It has wide application in business and economics because we are often interested in finding the percentage changes in sallies or economic figure, such as the gross national product, which compound or building on each others.

The geometric mean of a set of n positive number is defined the nth root of the product of n values. The formula for the geometric mean is written:

Geometric mean 

Example: The profits earned by Atkins Construction Company on four recent projects were 3 percent, 2 percent, 4 percent and 6 percent. What is the geometric mean profit?

**Answer**

The geometric mean is 3.46 percent, found by



The geometric mean is the fourth root of 144 or 3.46

The geometric mean profit is 3.46 percent

The geometric mean will always be less than or equal to arithmetic mean.

Note also that all the data value must be positive to determine the geometric mean

**Question**: Show that 

Proof: let  be two quantities then







 This is always true.

**Exercise:** The annual dividends, in percent, of four oil stocks are: 4.91, 5.75, 8.12, and 21.60. Find the geometric mean dividend. Find the Arithmetic mean dividend. Is the Arithmetic mean equal to or greater than the geometric mean?

A second application of the geometric mean is to find an average percent increase over a period of time.

|  |
| --- |
| AVERAGE PERCENT INCREASE OVER TIME  n is the number of period. |

Example: The population Harlan, Alaska, in 1990 was 2 persons, by 2000 it was 22. What is the average annual rate of percentage increase during the period?

Answer: There are 10 years between 1990 and 2000 so n=10. The formula for the geometric mean as applied to this type of problem is:

The final value is 0.271. So the annual rate of increase is 27.1percent. This means that the rate of population growth in Harlan is 27.1 percent per year.

Exercise: 1) In 1990 there were 9.19 million cable TV subscribers. By 2000 the number of subscribers increased to 54.87 million. What is the geometric mean annual increase for the period?

2) Production of cables stocks increased from 23,000 units in 1980 to 120,520 units in 2000.

Find the geometric mean annual percent increase.

1. **The mean of grouped data**

The approximate the arithmetic of data organized into a frequency distribution, we begin by assuming the observation in each class are represented by the midpoint of class.

ARITHMETIC MEAN OF GROUPED DATA:

Example

|  |  |  |  |
| --- | --- | --- | --- |
| Selling price ($thousands) | Frequency(f) | Midpoint(x) | f\*x |
| [12-15[ | 8 | $13.5 | $108 |
| [15-18[ | 23 | 16.5 | 379.5 |
| [18-21[ | 17 | 19.5 | 331.5 |
| [21-24[ | 18 | 22.5 | 405 |
| [24-27[ | 8 | 25.5 | 204 |
| [27-30[ | 4 | 28.5 | 114 |
| [30-33[ | 2 | 31.5 | 63 |
| total | 80 |  | $1,605 |



So we conclude that the mean vehicle selling place is about $20.1.

Exercises: Determine the estimated mean of the following frequency distribution.

1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| class | [0-5[ | [5-10[ | [10-15[ | [15-20[ | [20-25[ |
| frequency | 2 | 7 | 12 | 6 | 3 |

2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| class | [20-30[ | [30-40[ | [40-50[ | [50-60[ | [60-70[ |
| frequency | 7 | 12 | 21 | 18 | 12 |

* 1. **The Median**

1. Median of ungrouped data is the midpoint of the values after they have been ordered from the smallest to the largest, or largest to the smallest.

Fifty percent of the observation is above the median and fifty percent below the median.

The data must be at least ordinal level of measurement.

To find the median, arrange the scores in ascending order. If n is odd, median is the middle number, otherwise, it is the average of the two middle numbers.

If n (n: number of observations) is odd, then the median (me) is  where 

If n is even, then the median (me) is.

Or 

Example: Find the median of the numbers

1. 12, 9, 17, 16, 10, 10, 13.
2. 56, 46, 61, 57, 48, 50, 47, 44.

Solution: (a) In ascending order the numbers are 9, 10, 10, 12, 13, 16, and 17.

There are 7 numbers and so the median is in position 

Therefore the median is 

(b) In ascending order the numbers are 44, 46, 47, 48, 50, 56, 57, and 61.

They are 8 numbers and so the two middle numbers are in 

Therefore the median is 

Exercises: Calculate the median of {10, 15, 12, 11, 14, 18, 19, 17} and {40, 19, 32, 12, 35, 27, 33}

1. **The median of grouped data**

The median of grouped data, can be estimated, however, by (1) locating the class in which the median lies and the (2) interpolating within that class to arrive the median.

The rational for this approach is that the members of the median class are assumed to be evenly spaced throughout the class.

The formula is 

Where is the lower limit of the class containing the median.

 is the total number of frequencies.

 is the frequency in the median class

 is the cumulative number of frequencies in all the classes preceding the class containing the median.

h is the width of the class in which the median lies.

Example: compute the median from the following data

C.I is class interval

C.F is cumulative frequency.

C.I 10-15 15-20 20-25 25-30 **30-35** 35-40 40-45 45-50

F 10 20 125 270 **303** 197 65 10

C.F 10 30 155 425 **728** 925 990 1000

Underline interval (30-35) indicate the median class

 , L=30 , h=5 , CF=425 , f=303

Example: prices of 80 new vehicles sold last month at written Pontiac (prices in thousands)

C.I= [12-15[ [15-18[ [18-21[ [21-24[ [24-27[ [27-30[ [30-33[

F 8 23 17 18 8 4 2

CF 8 31 48 66 74 78 80



* 1. **The mode**

The mode is another measure of central tendency.

1. MODE (ungrouped data) is the value of observation that appears most frequently.

Example: Find the mode in {2, 3, 3, 5, 4, 3, 6, 5, 4, 3, 4, 2}

Solution: The most appears observation is 3 which has frequency equal to 4(i.e. it appears 4 times) the mode is especially useful in describing nominal and ordinal levels of measurement.

1. The mode of grouped data

For grouped data, the modal class is the class with large frequency. And the modal can be approximated by midpoint of the class containing the largest number of class frequency.

The mode in an empirical formula: can be calculated as

|  |
| --- |
|  |

Where: calculate the mode for the following data:

C.I: [0-10[ [10-20[ [20-30[ [30-40[ [40-50[ [50-60[

F: 14 17 22 26 23 18

Those underline are the class modal.



Where is the lower limit of the class containing the mode.

 is the frequency in the class preceding that of mode.

 is the frequency in the mode class.

 is the frequency in the class next of that containing the mode.

1. **MEASURE OF DISPERSION**

In this section also we continue to develop measures to describe a set of data, concentrating on measures that describe the dispersion or variability of the data.

We will consider several measures of dispersion. The range is based on location of the largest and the smallest values in the data set. The mean deviation, the variance, and the standard deviation are all based on deviations from the mean.

* 1. **Range**

The simplest measure of dispersion is the range.

It is the difference between the highest and the lowest values in a data set. In the form of an equation:

|  |
| --- |
|  |

Example: consider the following data sets.

1. 8, 9, 10, 11, 12. B)-45, -8, 2, 42, 59.

Each has a mean of 10, but the second set is more spread out than the first. In fact the range of the first is 12-8=4 but the range of second is 59+45=104.

* 1. **Mean Deviation**

It measures the mean amount by which the values in a population or sample vary from their mean.

In terms of definition

Mean deviation the arithmetic mean of the absolute values of the deviation from the arithmetic mean.

Where: is the value of each observation?

 Is the arithmetic mean of the values?

 Is the number of observations in the sample?

is Indicates the absolute value.

Example: The number of patients seen in the emergency room at St. Luke’s memorial hospital for a sample of 5 days last year was: 103, 97, 101, 106 and 103. Determine the mean deviation and interpret.

Answer:

The mean number of patients is 102, found by (103+97+101+106+103)/5=102.

|  |
| --- |
| Number of cases () absolute deviation |
| 103 103-102=1 1  97 97-102=-5 5  101 101-102=-1 1  106 106-102=4 4  103 103-102=1 1  Sum is 0 total=12 |



The mean deviation is 2.4 patients per day. The number of patients deviates, on average, by 2.4 patients from the mean of 102 patients per day.

Exercises: 1) The weights of a group of crates being shipped to Ireland are (in pounds): 95 103 105 110 104 105 112 90

1. What is the range of the weights?
2. Compute the arithmetic mean weight
3. Compute the mean deviationof weights
4. Calculate the (a) range, (b) arithmetic mean, and (c) mean deviation and (d) interpret the range and the mean deviation.
5. The department of statistics at western state university offers eight sections basic statistic. Following are the number of studies involved in these sections 34 46 52 29 41 38 36 and 28.
6. Dave’s automatic door installs automatic garage door opens the following list indicates the number of minutes needed to install a sample of 10 doors: 28 32 24 46 42 44 40 54 38 32 and 42.
   1. **Variance and standard deviation**

The variance and standard deviation are also based on the deviations from the mean.

The variance: The arithmetic mean of the squared deviations from the mean.

Note that the variance is no-negative, and it is zero only if all observations are the same.

Standard deviation: The square root of the variance

1. **Population variance:** The formulas for the population variance and the sample variance are slightly different. The population variance for ungrouped data, that is data not tabulated into a frequency distribution, is found by:

Population variance  

Where: is the symbol for the population variance (is the lower-case Greek letter sigma)

: The value of an observation in the population

 : The arithmetic mean of the population

: The number of observation in the population

Example: The ages of all patients in the solution ward of Yellowstone hospital are 38, 26, 13, 41 and 22 years. What is the population variance? 

|  |
| --- |
| Age() |
| 38 +10 100  26 -2 4  13 -15 225  41 +13 169  22 -6 36 |
| 140 0 534  Sum of deviations from men must equal zero. |

Population standard deviation

The variance is difficult to interpret for a single set of observations. The variance (from previous example) of 106.8 for the ages of the patients in solution is not in terms of years, but rather” years squared.”

There is a ways out of this dilemma. By taking the square root of the population variance, we can transform it to the same unit of measurement used for the original data. The square root of 106.8 years squared is 10.3years. The square root of the population variance is called population standard deviation.

In terms of a formula for ungrouped data

Population standard deviation ****

Exercises: 1) Consider these five values a population: 8, 3, 7,3and 4.

1. Determine the mean of the population
2. Determine the variance
3. Determine the standard deviation.
4. Consider these six values a population: 13, 3, 8, 10, 8, and 6.
5. Determine the mean of the population
6. Determine the variance and standard deviation
7. The annual reports of Dennis industries cited these primary earnings per common share for the past five years: $2.68, $1.03, $2.26, $4.30, and $3.58. if we assume these are population values, what is:
8. The arithmetic mean primary earnings per share of common stock?
9. The variance?

1. **Sample variance**

The difference between the population variance and the sample variance it requires a change in the denominator. Instead of substituting n (number in the sample) for N (number in the population), the denominator is n-1.

Thus the formula for the sample variance is  where S2 is the sample variance

 : The value of sample

: The mean of the sample

 : The number of observation in the sample.

Sample variance, direct formula:

|  |
| --- |
|  |

To show the formula of sample variance, direct formula. First we can show that. 









 But 





Therefore 

Hence, we recommend direct formula for calculating a sample variance.

**Example**

The hourly wages for a sample of part-time employees at fruit packers, inc. are= $2, $10,$6,$8and $9. What is the sample variance?

Answer: 

Using squared deviations from the mean

|  |  |  |
| --- | --- | --- |
| Hourly wage (X) |  |  |
| 2 | -5 | 25 |
| 10 | 3 | 9 |
| 6 | -1 | 1 |
| 8 | 1 | 1 |
| 9 | 2 | 4 |
| SUMM=35 | SUMM=0 | SUMM=40 |

 S2=in dollars squared

Using the direct formula

|  |  |
| --- | --- |
| Hourly wage (X) | X2 |
| $2 | 4 |
| 10 | 100 |
| 6 | 36 |
| 8 | 64 |
| 9 | 81 |
| SUM=$35 | SUM=285 |



Sample standard deviation the sample standard is used as an estimator of the population standard deviation.

The sample standard deviation is the square root of the sample variance.

The sample standard deviation for ungrouped data is most easily determined by

S: Standard deviation of sample,  direct formula

**Example:** The sample variance in the previous example involving hourly wage was computed to be 10. What is the sample standard deviation?

**Answer:** The sample standard deviation is $3.16, found by.

1. **Measures of dispersion for data grouped into a frequency distribution**
   1. **Range:** Recall that the range is the difference between the highest and lowest values.

To estimate the range from data already grouped into a frequency distribution, subtract the lower limit of the lowest class from the upper limit of the highest class.

**Example:** Suppose a sample of 47 hourly wages was grouped into this frequency distribution:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Hourly Earnings | [$5-$10[ | [$10-$15[ | [$15-$20[ | [$20-$25[ | [$25-$30[ |
| Frequency | 6 | 12 | 19 | 7 | 3 |

The range is $25 found by $30-$5=$25

* 1. **Standard Deviation of grouped data**

Recall that for ungrouped data, one formula for the sample standard deviation is : 

If the data of interest are in grouped form (in a frequency distribution), the sample standard deviation can be approximated by substituting for and  for 

Standard deviation, grouped data 

Where s is the symbol for the sample observation

 is the midpoint of a class

is the class frequency

 is the total number of sample observation

Example: Consider the following sample frequency distribution of profit-sharing plan by employees. What the standard deviation of data? What is the sample variance?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Amount invested | [$30-35[ | [$35-40[ | [$40-45[ | [$45-50[ | [$50-55[ | [$55-60[ | [$60-65[ | [$65-70[ |
| Number of Employees | 3 | 7 | 11 | 22 | 40 | 24 | 9 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Amount Invested | Frequency f | Midpoint x | f\*x | f\*x2 |
| [$30-$35[ | 3 | $32.50 | $97.50 | 3168.75 |
| [$35-$40[ | 7 | 37.5 | 262.5 | 9843.75 |
| [$40-$45[ | 11 | 42.5 | 467.5 | 19868.75 |
| [$45-$50[ | 22 | 47.5 | 1045 | 49637.5 |
| [$50$-55[ | 40 | 52.5 | 2100 | 110250 |
| [$55-$60[ | 24 | 57.5 | 1380 | 79350 |
| [$60-$65[ | 9 | 62.5 | 562.5 | 35156.25 |
| [$65-$70[ | 4 | 67.5 | 270 | 18225 |
| Total | 120 |  | $6,185 | $325,500 |

To find the standard deviation of these grouped data into a frequency distribution.

**Step 1:** each frequency is multiplied by its class midpoint. That is multiply f times x.

Thus, for the first class 3\*$32.50=$97.50

**Step 2:** Calculate. This could be written. For the first class it would be

$97.50\*$32.50=$3168.75 and so on

**Step 3:** Sum theand  columns. The totals are $6185 and 325500 respectively



The sample standard deviation is $ 7.50. The variance is ($7.51)2 or about 56.40(in dollars squared)

**Exercises** Compute the range, the standard deviation, and variance

1. Refer to the following frequency distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | [0-5[ | [5-10[ | [10-15[ | [15-20[ | [20-25[ |
| Frequency | 2 | 7 | 12 | 6 | 3 |

1. Refer to following frequency distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | [20-30[ | [30-40[ | [40-50[ | [50-60[ | [60-70[ |
| Frequency | 7 | 12 | 21 | 18 | 12 |

* 1. **Interpretation and uses of the Standard Deviation**

The standard deviation is commonly used as a measure to compare the spread in two or more sets of observations.

**5.3.1. Chebyshev’s Theorem**

We have stressed that a small standard deviation for a set of values indicates that these values are located close to the mean. Conversely, a large standard deviation reveals that the observation are widely scattered about the mean.

The Russian mathematician P.L. chebyshev (1821-1894) develop a theorem that allows determining the minimum proportion of the values that lie within a specified number of standard deviation of the mean. For example, based on chebyshev’s theorem, at least three of four values, or 75 percent, must lie between the mean and two standard deviations and the mean minus two standard deviation. This relationship applies regardless of the shape of the distribution. Further, at least eight of nine values, or 88.9 percent, will lie between plus three standard deviation and minus three standard deviations of the mean. At least 24 of 25 values, or 96 percent, will lie between plus and minus five standard deviations of the mean.

CHEBYSHEV’S THEOREM: For any set of observations (sample or population) the proportion of the values that lie within K standard deviation of the mean is at least , where k is any constant greater than 1.

**Example:** 1) In the previous example and solution, the arithmetic mean was $51.54, and the standard deviation was computed to be $7.51. At least what percent of the contributions lie within plus 3.5 standard deviations and minus 3.5 standard deviations of the mean?

**Answer**: About 92 percent, found by 

1. According to chebyshev’s theorem, at least what percent of any set of observations will be within 1.8 standard deviation of the mean?

**Answer:** 

**Exercise**: Consider these five values of a population: 8, 3, 7, 3, and 4.Determine the mean of the population. Determine the standard deviation. Using chebyshev’s rule, in what interval can we find at least 96% of the population.

* + 1. **The Empirical rule**

Chebyshev’s theorem is concerned with set of value; that is the distribution of values can have any shape. However, for a symmetrical, bell-shaped distribution, we can be more precise in explaining the distribution about the mean these relationships involving the standard deviation and the mean are the **empirical rule**, sometimes called the normal rule.

**Empirical Rule**

For a symmetrical ,bell-shaped frequency distribution approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean ;about 95 percent of the observations will lie within plus and minus two standard deviation of the mean; and practically all (99.7percent ) will lie within plus and minus three standard deviations of the mean.

Thus, if  and S=10, practically all the observations lie between 100+3(10) and 100-3(10), or 70 and 130.the range is therefore 60, found by 130-70

***Example 1*** I)A sample of the monthly amount spent for food by a senior citizen living alone approximates a symmetrical, bell-shaped frequency distribution.

The sample mean is $150; the standard deviation is $20. Using the empirical rule:

1. About 68 percent of the monthly food expenditures are between what two amounts?
2. About 95 percent of the monthly food expenditures are between what two amounts?
3. Almost all the monthly expenditures are between what two amounts?

***Answer***

1. About 68 percent are between $130 and $170,found by 
2. About 95 percent are between $110 and $190,found by 
3. Almost all (99.7 percent ) are between $90 and $210, found 
4. The Pitney Pipe Company is one of several domestic manufactures of pvc pipe. The quality control department sampled 600 10-foot lengths. At a point 1 foot from the end of the pipe they measured the outside diameter. The mean was 14.0 inches and standard deviation 0.1inches.
5. If the shape of the distribution is not known, at least what percent of the observations will be between 13.85 inches and 14.15 inches?
6. If we assume that the distribution of diameters is symmetrical and bell-shaped, about 95 percent of the observations will be between what two values?

Answer (a) 



(b) 13.8 and14.2 found by14.0-2(0.1) and 14.0+2(0.1)

* 1. **Relative dispersion**

Karl Pearson (1857-1936), who contributed significantly to the science of statistics, developed a relative measure called the coefficient of variation (CV). It is a very useful measure when:

1. The data are in different unit (such as dollars and days absent)
2. The data are in the same units, but the means are far apart.

Coefficient of variation is the ratio of the standard deviation to the arithmetic mean, expressed as a percent.

Example: A study of the amount of bonus paid and the years of employee resulted in these statistics: the mean bonus paid was $200; the standard deviation was $40. The mean number of year of service was 20 years; the standard deviation was 2 years. Compare the relative dispersion in the two distributions using the coefficient of variation.

Answer: The distribution is in different units (dollars and years of service). Therefore, they are converted to coefficients of variation.

For the bonus paid for years of service  

Interpreting, there is dispersion relative to the mean in the distribution of bonus paid compared with the distribution of years of service (because 20 percent>10 percent)

The same procedure is used when the data are in the same unit but the means are for a part.

**Example**: The variation in the annual incomes of executives is to be compared with the variation in incomes of unskilled employees. For a sample of executives,  and . For a sample of unskilled employees,  and. We are tempted to say that there is more dispersion in the annual incomes of the executives because $50,000>$3,200. The means are also for a part; however, that we need to convert the statistics to coefficients of variation to make a meaningful comparison of the variations is annual incomes.

**Answer**: For the executives for the unskilled employees:

There is no difference in the relative dispersion of two groups.

Exercises:

1. For the sample of student in the college of business administration at mid –Atlantic university, the mean grade point average is 3.10 with a standard deviation of 0.25. Compute the coefficient of variation.
2. A large group of Air force inductees was given two experimental tests-a mechanical aptitude test and a finger dexterity test. The arithmetic mean score on the mechanical aptitude test was 200, with a standard deviation of 10. The mean and standard deviation for the finger dexterity test where: . Compare the relative dispersion in the two groups.
   1. **SKEWNESS**

Another characteristic of a set of data is shape. There are four shapes commonly observed: symmetric, positive skewed, negatively skewed and bimodal.

* In a symmetric set of observations the mean and the median are equal and the data values are evenly spread around these values. i.e. mean=median or sk=0.
* A set of values is skewed to the right or positive skewed if there is a single peak and the values extend much further to the right of peak than to the left of the peak. i.e. mean>median or sk>0.
* In a negatively skewed distribution there is a single peak but the observations extend further to the left, in the negative direction, than to the right. i.e. (mean<median) or SK<0.
* A binomial distribution will have two or more peaks. This is often the case when the values are from two or more populations.

Note that positive skewed distributions are more common. Salaries often follow this pattern. There are several formulas in statistical literature used to calculate skewness. The simplest, developed by Professor Karl Pearson, is based on difference between the mean and the median.

PEARSON’S COEFFICIENT OF SKEWNESS

|  |
| --- |
|  |

Using the relationship the coefficient of skewness can range from -3 up to 3.

Note: when mean and median are equal sk =0.

The coefficient of skewness to represent output from the statistical software package (Excel and MINITAB) is based on the cubes deviation from the mean.

The formula is: SOFTWARE COEFFICIENT OF SKEWNESS



**Example**: Following are the earning per share for a sample of 15 software companies for year 2000. The earnings per share are arranged from smallest to largest.

$0.09 $0.13 $0.41 $0.51 $1.12 $1.20 $1.49 $3.18 $3.50 $6.36 $7.83 $8.92 $10.13 $12.99 $16.40.

Compute the mean, median, and standard deviation.

Find the coefficient of skewness using Pearson’s estimation and the software methods. What is your conclusion regarding the shape of distribution?

**Answer**: Firstly determine the mean using ungrouped sample data formula.



The median of ungrouped sample data: if n is odd 

The sample standard deviation

Pearson’s coefficient of skewness is 1.017. Found by 

This indicates there is moderate positive skewness in the earnings per share data.

Note that we obtain a similar, but not exactly the same from the software method.

Now, the details of calculation are shown below for the coefficient of skewness using the software method.

|  |  |  |
| --- | --- | --- |
| Earnings per share |  |  |
| 0.09  0.13  0.41  0.51  1.12  1.20  1.49  3.18  3.50  6.36  7.83  8.92  10.13  12.99  16.40 | -09310  -0.9234  -0.8697  -0.8506  -0.7337  -0.7184  -0.6628  -0.3391  -0.2778  0.2701  0.5517  0.7605  0.9923  1.5401  2.1935 | -0.8070  -0.7873  -0.6579  -0.6154  -0.3950  -0.3708  -0.2912  -0.0390  -0.0214  0.0197  0.1679  0.4399  0.9772  0.6539  1O.5537  TOTAL=11.8274 |



**EXERCISE** for exercises 1-3, do the following:

1. Determine the mean, median, and the standard deviation.
2. Determine the coefficient of skewness using Pearson’s method.
3. Determine the coefficient of skewness using the software method
4. The following values are the starting salaries, in $000, for a sample of five accounting graduates who accepted position in public accounting last year. 36.0 26.0 33.0 28.0 31.0
5. Listed below are the salaries, in $000, for a sample of 15 executives in the electronics industry. $516.0 $548.0 $534.0 $586.0 $566.0 $529.0 $523.0 $546.0 $538.0 $523.0 $551.0 $552.0 $486.0 $558.0 $574.0
6. Listed below are the commission earned ($000) last year by the sales representatives at the furniture patch.

$3.9 $5.7 $10.6 $13.0 $13.6 $15.1 $15.8 $17.1 $7.3 $17.4 $17.6 $22.3 $38.6 $43.2 $87.7

* 1. **Other measures of dispersion**

The standard deviation is the most widely used measure of dispersion. However, there are other way of describing the variation or spread in a set of data. One method is to determine the location of values that divide a set of observation into equal parts. These measures include **quartiles**, **deciles**, and **percentiles.**

**Quartiles** divide a set of observation into four equal parts. The first quartile, usually labeled Q1, is the value below 25percent of the observations occur and the third quartile usually labeled Q3, is the value below which 75percent of observation occur. Logically, Q2 is the middle value of a set of data arranged from smallest to the largest.

In a similar fashion **deciles** divide a set of data into 10 equal parts and **percentiles** into 100 equal parts.

To formalize the computational procedure, let Lp refer to the location of a desired percentile. So if we wanted to find the 33rd percentile we would use L33 and if we wanted the median, the 50th percentile, then L50. The number of observations is n, so if we want to locate the middle observation, its position is at (), or we would write this as, where p is the desired percentile.

Location of percentile 

I.e. location of Q1 is 

Location of Q2  is 

Location of Q3 is 

Example: 1) If n+1 is divided by 4. Calculate Q1, Q2, Q3 for the following data marks: 20 24 40 12 30 15 50.

Answer: First we order the values from smallest to largest.

12 15 20 24 30 40 50 in this case n=7 which is odd number.

Q1is locate at 

i.e. Q1 locate at 2nd place or is the 2nd value which is 15. Thus Q1=15

Q2 is locate at  thus Q2=24

Q3is locate at  thus Q3=40

2) If n+1is not divided by four. Suppose a data set contained six values: 91, 75, 61, 101, 43, and 104. We want to locate the first and the third quartile.

Answer: we order the value from the smallest to the largest: 43, 61, 75, 91, 101, and 104.

The first quartile is locate at 

The position formula tells us that the first quartile is locate between the first and the second value and that it is 0.75 of distance between the first and the second value. The first value is 43 and the 2nd is 61 so the distance between these two values is 18. To locate the Q1we need to move 0.75 of the distance between the 1st and the 2nd value, so 0.75 \*18=13.5. To complete procedure, we add 13.5 to the first value and report that the first quartile is located at 56.5. (found by 43+13.5= 56.5).

**Quartile, deciles, percentile of grouped data**

The quartile for grouped data is calculated as follows:

 h:width of the class

 l: lower limit of the class corresponding to(either ) quartile.

f: Frequency corresponding to the class of quartile.

CF: Cumulative frequency of the class preceding corresponding to quartile

**Example:** Calculate Q1, Q2, and Q3 for the following data:

C.I: [0-10[ [10-20[ [20-30[ [30-40[ [40-50[

F; 8 15 23 16 9

CF: 8 23 46 62 71









To calculate deciles, percentiles for grouped data we use the following formulas; D: deciles

P: percentile.









**Quartile deviation** (Q.D)

**Inter quartile range.** The inter quartile range is the distance between the 1st and 3rd quartile.

Now, quartile deviation is one-half (a half) of inter quartile

Its formula

Inter quartile=Q3-Q1

Quartile deviation = (Q.D is the measure of dispersion)

Coefficient of quartile deviation = (is the relative measure)

**Example:** 1) Compute QD and coefficient of QD from the following data

|  |  |
| --- | --- |
| Mark | 10 20 30 40 50 80 |
| No student | 4 7 15 8 7 2 |
| C.F | 4 11 26 34 41 43 |

 Term in this case.Q1=20

 Term in this case Q3=40

Q.D=; coefficient C.D=

**Exercise**: compute Q.D and coefficient of QD from the following data

|  |  |
| --- | --- |
| X | [10-20[ [20-30[ [30-40[ [40-50[ [50-60[ [60-70[ |
| F | 12 19 5 10 9 6 |
| cf | 12 31 36 46 55 61 |

**5.7. Box Plots and Outliers**

5.7.1. **Box plot**

A box plot is a graphical display, based on quartiles, that helps us picture a set of data. To construct a box plot, we need only five statistics: the minimum value, Q1 (the first quartile), the median, Q3(third quartile), and the maximum value.

**5.7.2. Outliers**

An outlier is a value that is consistent with the rest of data. The standard definition of an outlier is a value that is more than 1.5 times the inter quartile range smaller than Q1 or larger than Q3.

Outlier >Q3+1.5 (Q3-Q1)

Outlier <Q1-1.5 (Q3-Q1).

Example: The following represents the marks obtained by 30 students in a month class in final exam 66,61,65,59,87,61,58,70,77,94,45,80,58,78,49,72,75,97,84,82,79,75,68,75,90,78,85,67,57,52

Answer: The following summary in the output and the box plot are from Excel (Megastat). (Megastat can be downloaded and installed, then it will be in excel window automatically)

|  |  |
| --- | --- |
| Descriptive statistics |  |
|  | *# 1* |
| count | 30 |
| mean | 71.47 |
| sample variance | 176.40 |
| sample standard deviation | 13.28 |
| minimum | 45 |
| maximum | 97 |
| range | 52 |
|  |  |
| 1st quartile | 61.00 |
| median | 73.50 |
| 3rd quartile | 79.75 |
| interquartile range | 18.75 |
| mode | 75.00 |
|  |  |
| low outliers | 0 |
| high outliers | 0 |

**Exercise** : I) A police officer checking the speed of cars on Kanombe road highway recorded the following speeds for 30 vehicles the data is given below 58,60,57,70,67,63,63,68, 66,67,52,48,48,60, 56,64,63,60,59,63, 54,50,54,62,58,56,61,65,52,46

1. Arrange the data in ascending order.
2. Compute the mean, mode, and median.
3. Find the range.
4. Construct a frequency distribution.
5. Calculate variance and standard deviation.
6. Compute the coefficient of skewness.
7. Are these data skewed? if so how ?
8. The following is the number of minutes students commute from home to school
9. 25 48 37 17 32 26 16 23

23 29 36 31 26 21 32 25 31

43 35 42 38 33 28 41

i) How many classes and class interval would you recommend?

ii) By taking lower value minus one as a lower limit of the first class, organise the data into a frequency distribution and determine the relative frequency.

* 1. Draw a histogram and frequency polygon of minutes commuted.

iv) Determine first, second, third quartile and comment on result?

1. A sample of 50 student’s marks was taken for survey of performance in mathematics. The following frequency distribution describes the result obtained at school.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class intervals of mathematics marks | [0 - 10[ | [10 - 20[ | [20 - 30[ | [30 - 40[ | [40 - 50[ |
| Number of students (frequency) | 7 | 6 | 15 | 12 | 10 |

1. Calculate Pearson’s coefficient of skewness and interpret it.
2. Calculate interquartile range.

6. SIMPLE LINEAR REGRESSION MODEL

6.1.1 SIMPLE REGRESSION MODEL

A regression model is a mathematical equation that describes the relationship between two or more variables.

A simple regression model includes only two variables: one independent variable and one dependent. The dependent variable (Y) is the one being explained, and the independent variable (X) is the one used to explain the variation in the dependent variable.

For example:  are equations of regression lines.

6.1.2 LINEAR REGRESSION

A simple regression model that gives a straight line relationship between two variables is called a linear regression model.

The equation of a straight line is of the form: .

 is the intercept, i.e. the value of  for 

 is the slope, i.e. the change of one unit of  corresponds to a change of  units for .

In the examples above, only  represents a simple linear model.

Intercept:  ; slope: 

6.2. SIMPLE LINEAR REGRESSION ANALYSIS

Model  is called a deterministic model since it gives an exact relationship between and. Real life model are not deterministic, but contains an error term, denoted . The error term is assumed to be normally distributed with constant variance , and mean 0,

i.e. , with 

6.2.1 Scatter Diagram.

A plot of paired observations of the form  is called a scatter diagram.

EXAMPLE 6.1

Suppose we take a sample of 7 households from a low-to moderate- income neighborhood and collect information on their income and food expenditures for the past month. The information obtained in hundreds of dollars is given in the following table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Income | 35 | 49 | 21 | 39 | 15 | 28 | 25 |
| food expenditure | 9 | 15 | 7 | 11 | 5 | 8 | 9 |

Note that this scatter diagram is approximately linear. The method of least squares gives an estimate of the line which approaches as closely as possible each data point

6.2.2. LEAST SQUARES LINE

Suppose that  is an estimate of model 

 is called the predicted value of . It can be shown that 

To find and , we have to minimize the error sum of squares,

  and  are called least squares estimate of  and  respectively. The line Ŷ=a+bx is called regression line or least square lines. The values of  and are:





Example 6.2

Find the least squares line for the example 6.1

Solution

|  |  |  |  |
| --- | --- | --- | --- |
| Income | Food expenditure |  |  |
| 35 | 9 | 315 | 81 |
| 49 | 15 | 735 | 225 |
| 21 | 7 | 147 | 49 |
| 39 | 11 | 429 | 121 |
| 15 | 5 | 75 | 25 |
| 28 | 8 | 224 | 64 |
| 25 | 9 | 225 | 81 |
| 212 | 64 | 2150 | 646 |





b=

a=

The regression line is 

Since the slope is positive, food expenditure increases with income.

6.3. STANDARD DEVIATION OF RANDOM ERRORS.

The standard deviation of errors is given by:

, or

, where: 



The degrees of freedom for a simple line of regression is  since we lose

One degree of freedom to calculate  and one for 

EXAMPLE 6.3

Compute the standard deviation of errors in Example 10.1

Solution

|  |  |  |
| --- | --- | --- |
| Income x | Food expenditure Y |  |
| 35 | 9 | 1225 |
| 49 | 12 | 2401 |
| 21 | 7 | 441 |
| 39 | 11 | 1521 |
| 15 | 5 | 225 |
| 28 | 8 | 784 |
| 25 | 9 | 625 |
| 212 | 61 | 7222 |









**6.4. Coefficient of determination**

The total sum of squares is given by: 

In the above example 

The error sum of Squares is: 

The regression sum of Squares is: 

The coefficient of determination is the ratio: , or , and 

represents the proportion of  that is explained by the use of the regression model. A value closed to 1 or 100%, indicates a model that fits very well the data.

**Example 6.4**

Find the coefficient of determination for the example 6.1. Interpret the results.

Solution

. Hence,



Thus, 92% of the total variation (SST) is explained by the regression model.

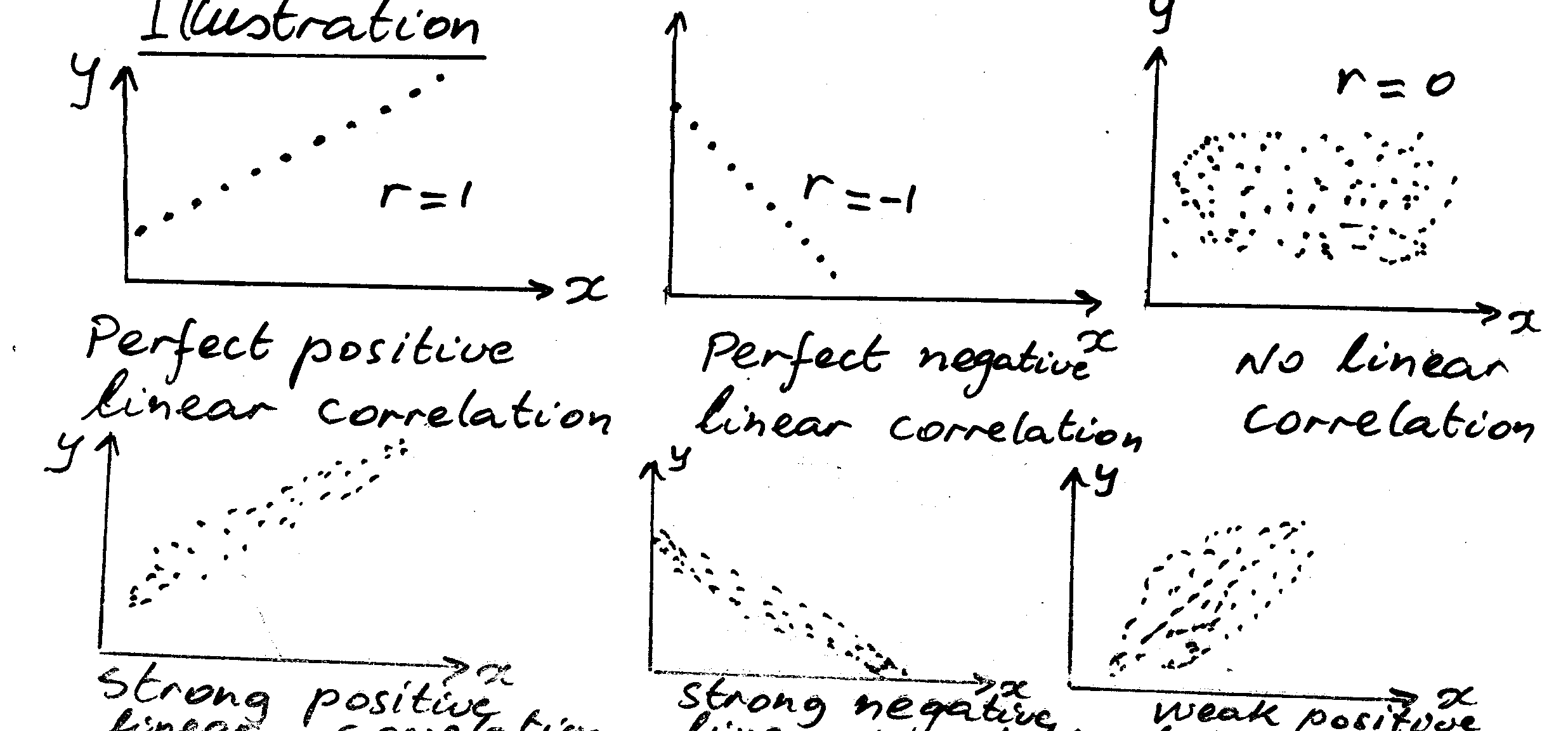
**6.5. LINEAR CORRELATION**

**6.5.1. LINEAR CORRELATION COEFFICIENT**

The coefficient of correlation measures the strength of the linear relationship between two samples or two populations. It is denoted by for a population, and by for a sample, and 

**Characteristics of the correlation coefficient**

* + 1. It shows the direction and strength of the linear relationship between two interval- or ratio-scale variables.
    2. It ranges from -1 up to and including +1.
    3. A value near 0 indicates there is little relationship between the variables.
    4. A value near +1 indicates a direct or positive relationship between the variables.
    5. A value near -1 indicates inverse or negative relationship between the variables.



The value of the linear correlation is given by: where are as defined above.

**Example 6.5**

Calculate the correlation coefficient for the example 6.1.

Solution





For statistical software, the results from linear regression report analysis of variance (ANOVA) as in multiple regressions as well as in the Analysis of variance (ANOVA).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of variation | Degree of freedom (d.f) | Sum of square (S.S) | Mean square (M.S) | Value of the test statistics F |
| Regression |  |  |  |  |
| Residual or Errors |  |  |  |  |
| Total |  |  |  |  |

 is the sample size and  is the number of variables. For a simple linear regression .

Exercises

1. The following sample observations were randomly selected.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x: | 5 | 3 | 6 | 3 | 4 | 4 | 6 | 8 |
| y: | 13 | 15 | 7 | 12 | 13 | 11 | 9 | 5 |

1. Draw a scatter diagram.
2. Determine the coefficient of correlation.
3. Determine the coefficient of determination. Interpret
4. Determine the regression and interpret the values of coefficient.
5. Estimate y when x=7.
6. The following regression equation was computed from a sample of 20 observations:. SSE was found to be 100 and SS total 400.
7. Determine the standard error of estimate.
8. Determine the coefficient of determination.
9. Determine the coefficient of correlation. (Caution: watch the sign!)
10. An ANOVA table is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SOURCE | DF | SS | MS | F |
| Regression | 1 | 50 |  |  |
| Error |  |  |  |  |
| Total | 24 | 500 |  |  |

1. Complete the ANOVA table
2. How large was the sample?
3. Determine the standard error.
4. Determine the coefficient of determination.
5. The following is a partial ANOVA table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SOURCE | DF | SS | MS | F |
| Regression | 2 |  |  |  |
| Error |  |  | 20 |  |
| Total | 11 | 500 |  |  |

1. Complete the ANOVA table
2. How large was the sample?
3. Determine the standard error.
4. Determine the coefficient of determination.

Part 2: CHAPTER 7: INTRODUCTION TO THE PROBABILITY THEORY

7.1. PERMUTATIONS AND COMBINATIONS

7.1.1. Introduction

This sections deal with concepts required for the study of probability and statistics. Statistics is a branch of science that is an outgrowth of the theory of probability. Combinations and permutations are used in both statistics and probability, and they in turn, involve operations with factorial notation. Therefore combinations, permutations, and factorial notation are discussed in this section.

This section aims at teaching a method of approach to certain problems involving arrangements and selections. This section helps us to count all elements of finite set; for counting all elements of finite set the injection, surjection and Cartesian product play important role.

**7.1.2. Cartesian product**

Consider  and  as the finite sets.

A Cartesian product between  and is defined as:



In general 





**Example 7.1:**

How many words of three alphabet can you form using the alphabet letters such that the first and third are consonants and the second is vowel.

Answer:

Consider  and 

Words

**7.1.3. The union of 2 sets**

Consider  and  as the finite sets.





 if 







**7.1.4. Arrangement with repetition**

Definition: If we take p elements in n elements of E one after another with repetition means that before taking the second object in E we must put the first taken into E; we continue this process until we take p elements is called arrangement with repetition.

Notation:



Example7.2: How many numbers can you write composed by 3 digits numbers different to zero?

Answer:  ,

We obtain  numbers

**7.1.5. Arrangement without repetition**

Definition: If we take r elements in n elements of E one after another without repetition means that taking one after the other in E; we continue this process until we take p elements is called arrangement without repetition.

Notation: 

Example2.3: How many permutations of six objects taken two a time can be made?

Answer: 

How many four-digit can be formed from the digits 2, 3, 4, 5, 6, and 7 without repetition?

Answer: 

**7.1.6. Particular cases.**

Permutation (bijection)

We call permutation of finite set E of #E= n, all arrangement at n elements of E is 

Example7.4: How many possibilities can you arrange 4 cars into a parking of 4 places?

Answer: 4!= 24 possibilities

Permutation with repetition

The number of arrangements of n items, where there are k groups of like items of size

, respectively, is given by 

Example7.5: How many different arrangements of the letters in the word ROOM can be made?

Answer: n=4, 

 Arrangements

**7.1.6. Combinations**

A combination is defined as a possible selection of a certain number of objects taken from a group without regard to the order.

Consider E non empty finite set of cardinal n and r is a natural number such that 

We call combination at r elements of E all part of r elements of E.



Example7.6: How many possible 14 students can form a committee of 6 students? Each committee is a party of 6 students of the set of 14 students

Answer:   


**7.2. Probability notion**

**7.2.1. Introduction and terminology**

Probability is a branch of mathematics that deals with measuring or determining quantitatively the likelihood that an event or experiment will have a particular outcome. It studies random experiment i.e. possible outcomes are more or equal to 2 and are due to chance. Probability is based on the study of permutations and combinations and is the necessary foundation for statistics.

The foundation of probability is usually ascribed to the 17th-century French mathematicians Blaise Pascal and Pierre de Fermat, but mathematicians as early as Gerolamo Cardano had made important contributions to its development. Mathematical probability began in an attempt to answer certain questions arising in games of chance, such as how many times a pair of dice must be thrown before the chance that a six will appear is 50-50. Or, in another example, if two players of equal ability, in a match to be won by the first to win ten games, are obliged to suspend play when one player has won five games, and the other seven, how should the stakes be divided?

* + 1. **What is a probability?**

In general, it is a number that describes the chance that something will happen.

**Definition: Probability** is a value between 0 and 1, inclusive, describing the relative possibility (chance or likelihood) an event will occur.

There are three key words that are used in study of probability: (1) Experiment, (2) outcome, and (3) event. These terms are used in our everyday language, but in statistics they have specific meanings.

**Experiment** is an action or process that leads to the occurrence of one and only one of several possible results (observations). Experiment can be also defined as an occurrence whose result or outcome is uncertain.

**An Outcome** is a particular result (observation) of an experiment.

**A sample space** of a random experiment is a set (list) of all possible outcomes of an experiment. The outcomes must be exhaustive and mutually exclusive.

In set notation, the sample space and its outcomes are represented as: where  is the number of outcomes.

**Examples7.7**

|  |  |  |
| --- | --- | --- |
| EXPERIMENT | OUTCOMES | SAMPLE SPACE |
| 1. Inspection of a part | Good (G), Defective (D) |  |
| 1. Toss of a coin once | Head (H), Tail (T) |  |
| 1. Roll of a die once | 1, 2, 3, 4, 5, 6 |  |
| 1. Toss of a coin twice | HH, HT, TH, TT |  |
| 1. Play a lottery | Win (W), Lose (L) |  |
| 1. Take a stats Test | Pass (P), Fail (F) |  |
| 1. Select a person | Male (M), Female (F) |  |

**Event** is a collection of one or more outcomes of an experiment. Given a sample space S, an event E is a subset of S (including the empty set and sample space S).

Event E is a **Simple event** if it contains only one element and a **compound event** if it contains more than one element.

**Examples 7.8:**

1. Experiment: Roll a die and observe the number facing up 

Event E: The number observed is odd



A: The number observed is even



B: the number observed is prime



C: the number is neither prime nor composite



1. Experiment: Roll two distinguishable dice and observe the numbers facing up.



Event F: the dice show the same number



Event G: the sum of the numbers is 1.



1. Experiment: Draw a hand of two cards from a deck of 52.

Event H: both cards are hearts.

* + 1. **Requirements of probabilities**

Performing the same experiment repeatedly, may result in different outcomes. Therefore, the best we can do is to consider the probability or the likelihood of occurrence of a certain outcome.

Given a sample space, the probabilities assigned to the outcomes must satisfy two requirements:

1. The probability of any outcome must lie between 0 and 1. That is, for each 
2. The sum of the probabilities of all the outcomes in a sample space must be 1. That is, 

**Probability of events**

The probability of an event is the sum of the probabilities of the simple events that constitute the event.

Example: Consider tossing a coin once. If head and tail are equally likely, then since 

If we toss a coin twice, the sample space is . If the simple events ,,, are equally likely each one has a probability of .

Let’s find the probability of the event A=”get at least one head”.



If it was an experiment of a die tossed once, the event “get an event number” is and has probability 

* + 1. **Three approaches to assigning probabilities**

1. **Classical approach**

If all outcomes of an experiment are equally likely to occur, the probability of an event  is the number of outcomes favourable to  divided by the total number of outcomes. That is; 

**Example 7.9:** consider an experiment of rolling a die. What is the probability of the event “ the number observed is even”.





Probability of an even number 

Mutually exclusive is the occurrence of one event means that none of the others can occur at the same time. Example: in the die tossing experiment, the event “an even number” and the event “an odd number” are mutually exclusive. If an even number occurred, it could not also be an odd number.

Collective exhaustive: at least of the events must occur when an experiment is conducted. For instance consider the above case for the die tossing experiment, every outcome will be either even or odd. So the set is collectively exhaustive.

If the set of events is collectively exhaustive and the events are mutually exclusive, the sum of the probabilities equals 1.

1. **Relative frequency approach (empirical concept)**

Another way to determine probability is based on relative frequencies. The probability of an event happening is determined by observing what fraction of time similar events happened in the past. That is, if an experiment is repeated  times and an event is observed times, then according to the relative frequency approach of probability: 

**Example 7.10:** suppose that last year 1000 students took the statistics course and that 200 obtained grade A. The relative frequency is . This is an estimate of the probability of obtaining grade A in the course. In other words relative frequencies are not probabilities, but approximate probabilities when the experiment is performed over and over. This is known as law of large numbers.

**Definition**: (Law of large numbers) If an experiment is repeated again and again, the probability of en event obtained from the relative frequency tends to the actual or theoretical probability.

1. **Subjective approach**

If there is little or no past experience on which to base a probability, it may be arrived at subjectively. Essentially, this means evaluating the opinions and other information and then estimating or assigning the probability.

Subjective concept of probability they likelihood (probability) of a particular event happening that is assigned by an individual based on whatever information is available.

Subjective probability is often influenced by biased preferences, and experience of the person assigning the probability.

**Examples 7.11:** (1) an investor would like to know the probability that a particular stock will increase in value. Using the subjective approach he may state that this probability is 65%, i.e. 0.65. The reasoning may be an analysis of factors associated with the stock and the stock market in general. (2) Estimating the likelihood you will earn an A in this course. (3) Estimating the likelihood the England Patriots will play in the super Bowl next year.

* 1. **JOINT, MARGINAL, AND CONDITIONAL PROBABLITY**

7.3.1. INTERSECTION **OF EVENT AND JOINT PROBABILITY**

The intersection of events A and B is the event that occurs when both A and B occur. It is denoted as , or as in the book by: A and B. The probability of the intersection is called the joint probability.

**Example 7.11:** Why are some mutual fund managers more successful than other? One possible factor is where the manager earned his or her MBA. Suppose that a potential investor examined the relationship between how well the mutual fund performs and where the fund manager earned his or her MBA. The following table gives the different joint probabilities. Analyse these probabilities and interpret.

|  |  |  |
| --- | --- | --- |
|  | Mutual fund outperforms market B1 | Mutual fund not outperform market B2 |
| Top-20 MBA program A1 | 0.11 | 0.29 |
| Not top -20MBA program A2 | 0.06 | 0.54 |

Let A1: Fund manager graduated from a top-20 MBA program

A2: Fund manager did not graduate from a top-20 MBA program

B1: Fund outperforms the market

B2: fund does not outperform the market

We have, 







These mean that;

1. 11% of all mutual funds outperform the market and their managers graduated from a 20-top MBA program.
2. The other three joint probabilities are interpreted similarly.

**7.3.2. Marginal probability**

Marginal probability is the probability of a single event without consideration of any other event.

Marginal probabilities, computed by adding across rows and down columns, are so named because they are calculated in the margins of the table.

**Example 7.12:** Calculate the marginal probabilities for the example 7.11

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mutual fund outperforms market B1 | Mutual fund not outperform market B2 | Marginal probability |
| Top-20 MBA program A1 | 0.11 | 0.29 | 0.40 |
| Not top -20MBA program A2 | 0.06 | 0.54 | 0.60 |
| Marginal probability | 0.17 | 0.83 | 1 |

The four marginal probabilities are the following:









**7.3.3. CONDITIONAL PROBABILITY**

Conditional probability is the probability that an event will occur given that another event has already occurred. If A and B are two events, the conditional probability of A given B is written as  and is defined by  with . Similarly,

 with 

**Example 7.13:** Suppose that in example 7.12 we select one mutual fund at random and discover that it did not perform the market. What is the probability that a top- 20 MBA program graduate manages it?

**Solution**

We need to calculate . Using the conditional probability formula, we find:



Thus, 34.94% of all mutual funds that do not outperform the market are managed by top-20 MBA program graduates.

**7.3.4. INDEPENDENCE**

One of the objectives of calculating conditional probability is to determine whether two events are related. Two events are independent if the probability of one event is not affected by the occurrence of the other event. This is expressed mathematically, for two events A and B, by or . Two events that are not independent are said to be dependent.

**Example 7.14:** Determine whether the fund outperforming the market and whether the manager graduated from a top-20 MBA program are independent events

**Solution**

There are four combinations of event, so four ways of solving the problem. Let’s use one of them. That is compared and 





Since , the event  and  are depend.

Similarly, it is easy to show that and , and , and are dependent.

**7.3.5. UNION**

The union of events A and B is the event that occurs when either A or B or both occur. It is denoted as A or B.

**Example 7.15:** determine the probability that a randomly selected fund outperforms the market or the manager graduated from top-20 MBA program.

**Solution**

We want to calculate  the union or occurs when any of the following joint events occur:

1. Fund outperforms the market and the manager graduated from a top-20 MBA program: and 
2. Fund outperforms the market and the manager did not graduate from a top-20 MBA program: and 
3. Fund does not outperform the market and the manager graduated from a top-20 MBA program: and 

Thus,



**Note:**  As 



* 1. **PROBABILITY RULES AND TREES**

**7.4.1. COMPLEMENT RULE**

The complement of event , denoted as or , is the events that occurs when does not occur.

Diagram: S



The probability of the complement event is calculated by: 

This implies that 

**7.4.2. MULTIPLICATION RULE**

The multiplication rule is based on the formula for conditional probability, and is used to calcutate the joint probability is used to calculate the joint probability of two events. The joint probability of two events  and is; , since 

Altering the notation, we also have: . When  and are independent, then 

**Example 7.16**: A graduate statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the **two students chosen are female?**

**Solution**

Let be the event “the first student chosen is female”

 be the event “the second student chosen is also a female”.

We need to calculate



**Example 7.17:** Refer to example 1.16. The professor who teaches the course is suffering from the classes. The professor’s replacement will teach the next two classes. His style is to select one student at random and pick on him or her to answer questions during that class. What is the probability that the two students chosen are female?

**Solution**

Here the professor chooses one student in each of the two classes. The choice in the second class, so that the events  and  are independent. Using the multiplication rule for independent events, we get: 

**7.4.3. ADDITION RULE**

The addition rule provides a straight forward way of computing the probability of the union of two events, namely 

**Example 7.18:** Recall the example 1.8. Calculate 

**Solution:** According to the addition rule,



Note: **Addition rule for mutually exclusive events**

When two events are mutually exclusive (which means that the two events cannot occur together), their joint probability is 0. Therefore, the probability of the union of two mutually exclusive events  and  is 

**Example 7.19:** In large city two newspapers are published, the **sun** and the **post**. The circulation departments report that 22% of the city’s households have a subscription to the **sun** and 35% Subscribe to the **post**. A survey reveals that 6% of all households subscribe to both newspapers. What proportion of the city’s household subscribe to **at least one** newspaper?

**Solution**

Let A: be the event “the household subscribes to the sun”

B: be the event “the household subscribes to the post”.

We need to calculate: . According to the addition rule, 

That is 51% of the city’s households subscribe to at least one newspaper.

**7.4.4. PROBABILITY TREES**

A probability tree is a useful device for calculating probabilities or using probability rules.

**Example 7.20:** Recall example 1.16. of selecting two female students (without replacement), if there are 3 female students in a class of 10. The tree diagram is drawn as follows:

First choice second choice joint probability

2/9 

F|F

3/10 F M|F 7/9 

M F|M 3/9 

7/10

M|M 6/9 

Thus, as found in Example 1:16.

Similarly, the tree diagram for the Example 1.17 is

First choice second choice joint probability

3/10 

F|F

3/10 F M|F 7/10 

M M|M 7/10 

7/10

F|M 3/10 

Thus,as found previously.

Note: One can extend the questions and use the tree diagram. For example “what is the Probability that one student is female and the other is male”. The solution is simple; for the 1st diagram (mutually exclusive events). Likewise, for the second diagram, we have: 

**Example 7.21:** Students who graduate from law schools must still pass a bar exam before becoming lawyers. Suppose that in a particular jurisdiction the pass rate for the first- time test takers is 72%. Indicates who fail the first exam may take it again several months later. Of those who fail their first test, 88% pass their second attempt. Find the probability that a randomly selected law school graduate becomes a lawyer. Assume that candidates cannot take the exam more than twice.

**Solution:**

Use of a tree diagram

First Exam Second Exam Joint probability

Pass 0.72 

Fail 0.28 Pass|Fail 0.88

Fail|Fail 0.12



Thus, using the addition rule for mutually Exclusive events, we have:



* 1. **BAYES’ LAW**

Conditional probability is used to find probability of an event given that one of its possible causes has occurred. We used Bayes’ law to find the probability of the possible given an event has occurred.

**Example7.22**

Medical tests can produce false-positive and false –negative results. A particular test is found to perform as follows:

* Correctly diagnose “positive” 94% of the time
* Correctly diagnose “negative” 98% of the time.

It is known that 4% of man in the general population suffer from an illness. What is the probability that a man is suffering from the illness, if the test result is positive?

**Solution**

Let:  has a disease

Does not have the disease

 Positive test result

 Negative test result

The above information can be summarized in the following tree diagram.

Disease Test result Joint probability

(Prior Probabilities) (likelihood probability) (Posterior probability)

0.94 

PT|D

D 0.04 NT|D 0.06 



0.96  0.02 

 0.98 

Recall that we need to calculate 

We know that , using the above tree diagram, we get: 

Thus, 

There is a 66.2% probability that the man has the disease. The probability is high, but considerably lower than the man would fear.

Note: similarly 





In general, Bayes’ theorem is written as



**CHAP 8. DISCRETE RANDOM VARIABLE AND THEIR PROBABILIT DISTRIBUTIONS**

* 1. **RANDOM VARIABLES**

**Definition**

A Random variable is a variable whose value is determined by the outcome of a random experiment.

EXAMPLE 8.1 Number of vehicles owned. (p.194)

The table below gives the number of vehicles owned, the associated frequencies and relative frequencies for all 2000 families living in a small town.

|  |  |  |
| --- | --- | --- |
| Number of vehicles (x) | Frequency (n) | Rel. Frequency (f) |
| 0 | 30 | 30/2000= 0.0015 |
| 1 | 470 | 470/2000= 0.235 |
| 2 | 850 | 850/2000= 0.425 |
| 3 | 490 | 490/2000= 0.245 |
| 4 | 160 | 160/2000= 0.080 |
| Total | N=2000 |  |

The random experiment is selecting randomly a family in the population.

The random variable, denoted X with value denoted x, is “the number of vehicles owned” Note that the values of X are 0, 1, 2, 3, 4.

* + 1. DISCRETE RANDOM VARIABLE

Definition

A random variable that assumes countable values is called a discrete random variable.

Example 8.2

The following are examples of discrete random variables.

1. The number of vehicles owned by a family in example 8.1 is a random variable X TAKING VALUES (COUNTABLE) 0, 1, 2,3, and 4.
2. The number of flats in a university is a discrete random variable.
3. The number of customers who visit a bank during any hour is a discrete random variable.
   * 1. CONTINUOUS RANDOM VARIABLE

Definition

A random variable that can assume any value contained in one or more intervals is called a continuous random variable.

EXAMPLE 8.3

The following are some examples of continuous random variables:

1. The height of a person.
2. The time taken to complete an exam.
3. The weight of a fish.
4. The price of a house.
   1. PROBABLITY DISTRIBUTION OF A DISCRETE RANDOM VARIABLE

Definition

The probability distribution of a discrete random variable lists all possible values that the random variable can assume and their corresponding probabilities.

Note:

1. 0 for each value x of X
2.  all x

EXAMPLE 8.4

Using the relative frequencies of example 8.1, the probability distribution for this example is:

|  |  |
| --- | --- |
| Number of vehicles owned x | Probability p (x) |
| 0 | 0.015 |
| 1 | 0.235 |
| 2 | 0.425 |
| 3 | 0.245 |
| 4 | 0.080 |
| Total |  |

This distribution can be represented graphically as follows:

EXAMPLE 8.5

Which of the following tables represents a probability distribution?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A) | x | p(x) | B) | x | P(x) | C) | x | P(x) |
|  | 0 | 0.08 |  | 2 | 0.25 |  | 7 | 0.7 |
|  | 1 | 0.11 |  | 3 | 0.34 |  | 8 | 0.5 |
|  | 2 | 0.39 |  | 4 | 0.28 |  | 9 | -0.2 |
|  | 3 | 0.27 |  | 5 | 0.13 |  |  |  |

1. Is not a probability distribution since 0.85 1.
2. Is a probability distribution since for every x, we have 0 p(x)1 and

.

1. Is not a probability distribution -0.20<0 even if .

EXAMPLE 8.6

The following table lists the probability distribution of the number of breakdowns per week for a machine based on past data.

|  |  |
| --- | --- |
| Breakdown per week | 0 1 2 3 |
| Probability | 0.15 0.20 0.35 0.30 |

1. Present the probability distribution graphically.
2. Find the probability that the number of breakdowns for this machine during a given week is:
3. Exactly 2; (ii) 0 t0 2; (iii) more than 1; (iv) at most 1.

**Solution**

1. Probability distribution

|  |  |
| --- | --- |
| x | P(x) |
| 0 | 0.15 |
| 1 | 0.20 |
| 2 | 0.35 |
| 3 | 0.30 |
|  |  |

1. (i) p(x=2)=0.35
2. P(0x2)= p(x=0)+p(x=1)+p(x=2)

=0.15+0.20+0.35

=0.70

1. P(x>1)=p(x=2)+p(x=3)

=0.35+0.30

=0.65

Or, p(x>1)=1-p(x1)

=1-[p(x=0)+p(x=1)]

=1-(0.15+0.20)

=1-0.35=0.65

1. P(x1)=p(x=0)+p(x=1)

=0.15+0.20=0.35

EXAMPLE 8.7 (EXAMPLE 1.5)

According to a survey, 60% of all students at a large university suffer maths anxiety. Two students are randomly selected from this university. Let x denote the number of students in this sample who suffer from maths anxiety. Develop the probability distribution of x.

**Solution**

Let us define the following two events

N= the student selected does not suffer from maths anxiety. M = the student selected suffers from maths anxiety.

If two students are selected randomly, we have the following tree diagram:

First students, second student final outcomes and probabilities

N 0.4 

N 0.4 M 0.6 

M

0.6 N 0.4 

M

0.6 

In a sample of two students, let x be the number of students who suffer from maths anxiety. Then x can take the value 0 for the event , 1 for the events and 2 for the event .

Hence,





This, the probability distribution is:

|  |  |  |  |
| --- | --- | --- | --- |
| x | 0 | 1 | 2 |
| P(x) | 0.16 | 0.48 | 0.36 |

* 1. **MEAN OF A DISCRETE RANDOM VARIABLE**

The mean, also called the expected value, of a discrete random variable , denoted by is defined by: 

In other words, **** is a weighted average of the possible values of  where the weights are the probabilities.

**Example8.8:**

Recall the Example on the number of cars owned by family, and calculated the mean number of cars owned.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 0 | 0.015 | 0 |
| 1 | 0.235 | 0.235 |
| 2 | 0.425 | 0.85 |
| 3 | 0.245 | 0.735 |
| 4 | 0.08 | 0.32 |
|  |  |  |

The mean is . That is, on average, every family has 2.14 cars.

* 1. **VARIANCE (VAR X) AND STANDARD DEVIATION** **OF A DISCRETE RANDOM VARIABLE**

Let X be a r.v having the pdfand the  where 

is constant. The variance of X, written Var (X) is given by: .Alternatively,



The variance is defined as the average of the sum of squared deviationsfrom different values of gains (losses) around the expected value (=mean) of gain (or loss).

The standard deviation is a parameter that indicates the spread of the gains or the losses.

Properties



Example 8.9:

An r.v X has the following distribution:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 |
| P(X=x) | 0.1 | 0.3 | 0.2 | 0.3 | 0.1 |

Find E(X), Var( X ) and 

Solution







Exercise

Find the Mean, Variance, and Standard Deviation for the following probability distribution.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | sum |
| p(x) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 6/6=1 |

Hint:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | sum |
| p(x) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 6/6 = 1 |
| x p(x) | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 6/6 | 21/6 = 3.5 |
| x^2 p(x) | 1/6 | 4/6 | 9/6 | 16/6 | 25/6 | 36/6 | 91/6 = 15.1667 |

Exercises

1. Three girls Aileen, Barbara and Cathy pack biscuits in a factory. From the batch allotted to them Aileen packs 55%, Barbara 30% and Cathy 15%. The probability that Aileen breaks some biscuits in a pack is 0.7 and the respective probability for Barbara and Cathy are 0.2 and 0.1.

What is the probability that a packet with broken biscuits found by the checker was packed by Aileen.

2. Calculate the expected value, the variance and the standard deviation for the following pdf:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | -5 | -4 | 1 | 2 |
|  |  |  |  |  |

3. A perfect coin is tossed until one obtains either Head or 5 Tails. Compute



* 1. **SPECIAL DISCRETE PROBABILITY DISTRIBUTIONS**
     1. **BINOMIAL DISTRIBUTION**

Binomial experiment: an experiment with a fixed number of independent trials. Each trial can only have two outcomes, or outcomes which can be reduced to two outcomes. The probability of each outcome must remain constant from trial to trial.

Binomial distribution: the outcomes of a binomial experiment with their corresponding probabilities.

Let consider an experiment with n independent trials and only two possible outcomes. We call one of the outcomes successful outcome and the probability of its occurring is P(s) =p. The other outcome is called fail outcome and the probability of its occurring is P (f) =q=1-p.

The probability for getting exactly k successes in n trials is .

Then the pdf of X is given by 

We write  and read “X is a random variable following the binomial distribution with parameter n and p”.

Example 8.10:

The probability that a person supports party A is 0.6.Find the probability that a randomly selected sample of 8 voters, there are:

Exactly 3 who support party A.

More than 5 who support party A

Solution

Consider “support of party A to be success”. Then p=0.6 and q=1-0.6=0.4.

Let X be the r.v.” the number of party A supporters”

~

~and its pdf is given by 

Thus 

k=3

P(X=3)=

P(X>5)=P(X=6)+P(X=7)+P(X=8)

=

=

=0.315

**Expectation value and variance**

If the r.v X is such that ~, then 

Example: If the probability that it will be a fine day is 0.4. Find the expected number of fine days in a week end and the standard deviation.

Solution

Let “fine day” be “success”. Then, p=0.4 and q=0.6.

Let X be an r.v. “number of fine days in a week “.Then ~, where n=7 and p=0.4. Hence ~







Mode

Let the r.v. X be such that . To compute the mode of the probability distribution of X, we only consider the values of X close to E(X).

Example:

If ~, find the mode of the probability distribution of X.

Solution

Here X has pdf as follow:



So that 

But E(x) =np=10\*0.45=4.5

Now 







Here X with the highest probability is 4. Therefore, the mode of X is 4.

* 1. **POISSON DISTRIBUTION**

Poisson distribution: a probability distribution used when a density of items is distributed over a period of time. The sample size needs to be large and the probability of success to be small.

Let X be a discrete r.v. It is said to follow the Poisson distribution if its pdf is of the form:

Then the Poisson distribution is defined by This distribution is an infinitely countable distribution.

Example 8.11

If, find: a)  ; b); c)

Solution









**Expectation value and variance**

If X is a discrete r.v such that, then 

Example 8.12

If with standard deviation 3. Find

E(X)

P(X<4)

Solution

Since standard deviation , we have 

Consequently, since  ,

Thus the pdf of X is given by 





Uses of Poisson distribution

There are two main practical uses of Poisson distribution:

When we consider the distribution of random events

When an event is randomly scattered in time (or in space) and has mean number of occurrences  in a given interval and if X is a random variable “the number of occurrences in a given interval, then.

Examples: Car accidents on a particular street of road in one day

-Accidents in a factory per week

-Telephone calls made to a switchboard in a given minute.

Exercise 8.13: The mean number of bacteria per millimetre of a liquid is known to be 4. Assuming that the number of bacteria follows a Poisson distribution, find the probability that 1 ml of liquid there will be:

No bacteria

4 bacteria

Less than 3 bacteria

Solution





Where X is the r.v “the number of bacteria in 1 ml of liquid”











When we approximate the binomial distribution.

Let X be a random variable such that . Now X can be approximated by a Poisson distribution with parameter  if n is large (> 50 ) and p is small (< 0.1).

The approximation gets better as.

Example 8.14: A factory packs bolts in boxes of 500. The probability that a bolt is defective is 0.002.

Find the probability that a box contains 2 defective bolts.

Solution: Let X be the random variable. The number of defective bolts in a box:

~since n=500, p=0.002, then ~

Thus the pdf of X is given by:



Since n=500, p=0.002, we use the Poisson approximation with 

~With pdf

We require computing:



Mode: Let~. Assume that ,thus the pdf of X is given by:



…………………………………

………………………………….

We notice that the modes are 0 and 1.

In general, if is an integer, then there are two modes and these occur when 

Let 

~

The mode is 1.

In general if is nor an integer then the mode is an integer such that 

* 1. **Continuous random variables**
     1. Introduction

A continuous random variable is a theoretical representation of a continuous variable such as height, mass or time.

Let X a random variable such that

Y



y=f(x)

0 a x b B

One says that X is a continuous random variable and f(x) is the probability density function.

This pdf satisfies the following conditions:



If f(x) is the probability density function (pdf) on, then:



Example8.15: The continuous random variable X has pdf



Find the expectation, the variance and the standard deviation of X.

Solution: 

Mode

The mode is the value of X for which f(x) is greatest in the given range of X.

For some pdf, it is possible to determine the mode by finding the maximum point on the curve y=f(x).

Example 8.16

The continuous random variable X has pdf



Find the mode.

Solution

Y

0.3

X

Mode

 Mode=1

8.7.2. THE NORMAL DISTIBUTION



The normal distribution is important in statistics for four main reasons:

Numerous phenomena measured on continuous scales have been shown to follow ( or can be approximated by ) the normal distribution.

We can use the normal distribution to approximate various discrete probability distributions, such as the binomial and the Poisson.

It provides the basis for the statistical process-control

It provides the basis for classical statistical inference.

The Standard Normal distribution

In general, we have: 

Such that the density distribution is given by this mathematical expression:



By standardizing a normally distributed random variable, we will need only one table. By using the transformation formula for , the standard normal score 

In this case we have: 

The distribution of a random variable which follows the normal law is symmetric according to its expected.









Properties:

1.

2. 

3. 

4. 

Examples 8.17:

1.. Calculate  ?

Solution : 

2. . Calculate  ?



EXERCISES

1. Consider a variable follows a standard normal distribution.

Calculate: a)  b)  c)  d)  e)  f) 

2. Consider  a variable follows a standard normal distribution. Determine  in each follows cases: a)  b)  c) 

d)  e) 

3. Number of cars that pass through a car wash between 16:00 and 17:00 has the following probability distribution the table below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 4 | 5 | 6 | 7 | 8 | 9 |
|  | 1/12 | 1/12 | 1/6 | 1/6 | 1/4 | 1/4 |

What is the expected number of cars washed?

4. Find the expected number of boys on a student committee of 3 selected at random from 4 boys and 3 girls?

5. For the standard normal distribution, find the area:

a. between -1.27 and 1.86 b. below 1.7 c. above 1.18 d. between -0.47 and -0.35

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Positive Z (Normal distribution) | | | |  |  |  |  |  |  |  |
| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
| 0 | 0.5 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.5279 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.5438 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.6293 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.6591 | 0.66276 | 0.6664 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.7054 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.7224 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.7549 |
| 0.7 | 0.75804 | 0.76115 | 0.76424 | 0.7673 | 0.77035 | 0.77337 | 0.77637 | 0.77935 | 0.7823 | 0.78524 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79955 | 0.80234 | 0.80511 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 | 0.82894 | 0.83147 | 0.83398 | 0.83646 | 0.83891 |
| 1 | 0.84134 | 0.84375 | 0.84614 | 0.84849 | 0.85083 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.86433 | 0.8665 | 0.86864 | 0.87076 | 0.87286 | 0.87493 | 0.87698 | 0.879 | 0.881 | 0.88298 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 | 0.89435 | 0.89617 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.9032 | 0.9049 | 0.90658 | 0.90824 | 0.90988 | 0.91149 | 0.91309 | 0.91466 | 0.91621 | 0.91774 |
| 1.4 | 0.91924 | 0.92073 | 0.9222 | 0.92364 | 0.92507 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |
| 1.5 | 0.93319 | 0.93448 | 0.93574 | 0.93699 | 0.93822 | 0.93943 | 0.94062 | 0.94179 | 0.94295 | 0.94408 |
| 1.6 | 0.9452 | 0.9463 | 0.94738 | 0.94845 | 0.9495 | 0.95053 | 0.95154 | 0.95254 | 0.95352 | 0.95449 |
| 1.7 | 0.95543 | 0.95637 | 0.95728 | 0.95818 | 0.95907 | 0.95994 | 0.9608 | 0.96164 | 0.96246 | 0.96327 |
| 1.8 | 0.96407 | 0.96485 | 0.96562 | 0.96638 | 0.96712 | 0.96784 | 0.96856 | 0.96926 | 0.96995 | 0.97062 |
| 1.9 | 0.97128 | 0.97193 | 0.97257 | 0.9732 | 0.97381 | 0.97441 | 0.975 | 0.97558 | 0.97615 | 0.9767 |
| 2 | 0.97725 | 0.97778 | 0.97831 | 0.97882 | 0.97932 | 0.97982 | 0.9803 | 0.98077 | 0.98124 | 0.98169 |
| 2.1 | 0.98214 | 0.98257 | 0.983 | 0.98341 | 0.98382 | 0.98422 | 0.98461 | 0.985 | 0.98537 | 0.98574 |
| 2.2 | 0.9861 | 0.98645 | 0.98679 | 0.98713 | 0.98745 | 0.98778 | 0.98809 | 0.9884 | 0.9887 | 0.98899 |
| 2.3 | 0.98928 | 0.98956 | 0.98983 | 0.9901 | 0.99036 | 0.99061 | 0.99086 | 0.99111 | 0.99134 | 0.99158 |
| 2.4 | 0.9918 | 0.99202 | 0.99224 | 0.99245 | 0.99266 | 0.99286 | 0.99305 | 0.99324 | 0.99343 | 0.99361 |
| 2.5 | 0.99379 | 0.99396 | 0.99413 | 0.9943 | 0.99446 | 0.99461 | 0.99477 | 0.99492 | 0.99506 | 0.9952 |
| 2.6 | 0.99534 | 0.99547 | 0.9956 | 0.99573 | 0.99585 | 0.99598 | 0.99609 | 0.99621 | 0.99632 | 0.99643 |
| 2.7 | 0.99653 | 0.99664 | 0.99674 | 0.99683 | 0.99693 | 0.99702 | 0.99711 | 0.9972 | 0.99728 | 0.99736 |
| 2.8 | 0.99744 | 0.99752 | 0.9976 | 0.99767 | 0.99774 | 0.99781 | 0.99788 | 0.99795 | 0.99801 | 0.99807 |
| 2.9 | 0.99813 | 0.99819 | 0.99825 | 0.99831 | 0.99836 | 0.99841 | 0.99846 | 0.99851 | 0.99856 | 0.99861 |
| 3 | 0.99865 | 0.99869 | 0.99874 | 0.99878 | 0.99882 | 0.99886 | 0.99889 | 0.99893 | 0.99896 | 0.999 |
| 3.1 | 0.99903 | 0.99906 | 0.9991 | 0.99913 | 0.99916 | 0.99918 | 0.99921 | 0.99924 | 0.99926 | 0.99929 |
| 3.2 | 0.99931 | 0.99934 | 0.99936 | 0.99938 | 0.9994 | 0.99942 | 0.99944 | 0.99946 | 0.99948 | 0.9995 |
| 3.3 | 0.99952 | 0.99953 | 0.99955 | 0.99957 | 0.99958 | 0.9996 | 0.99961 | 0.99962 | 0.99964 | 0.99965 |
| 3.4 | 0.99966 | 0.99968 | 0.99969 | 0.9997 | 0.99971 | 0.99972 | 0.99973 | 0.99974 | 0.99975 | 0.99976 |
| 3.5 | 0.99977 | 0.99978 | 0.99978 | 0.99979 | 0.9998 | 0.99981 | 0.99981 | 0.99982 | 0.99983 | 0.99983 |
| 3.6 | 0.99984 | 0.99985 | 0.99985 | 0.99986 | 0.99986 | 0.99987 | 0.99987 | 0.99988 | 0.99988 | 0.99989 |
| 3.7 | 0.99989 | 0.9999 | 0.9999 | 0.9999 | 0.99991 | 0.99991 | 0.99992 | 0.99992 | 0.99992 | 0.99992 |
| 3.8 | 0.99993 | 0.99993 | 0.99993 | 0.99994 | 0.99994 | 0.99994 | 0.99994 | 0.99995 | 0.99995 | 0.99995 |
| 3.9 | 0.99995 | 0.99995 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99997 | 0.99997 |