MARMARA ÜNİVERSİTESİ YAYINEVİ

T.C.
MARMARA ÜNİVERSİTESİ
İSTATİSTİK ve KANTİTATİF ARAŞTIRMALAR MERKEZİ
İSKAR Yayın No.96/2

STATISTICAL PROBLEMS AND SOLUTIONS FOR ECONOMICS AND BUSINESS ADMINISTRATION

Prof. Dr. Ateş VURAN

Dr. Şule ÖZMEN

Marmara University
Faculty of Economic and Administrative Sciences
Department of Business Administration

İSTANBUL, 1996

MARMARA ÜNİVERSİTESİ YAYINEVİ

T.C.
MARMARA ÜNİVERSİTESİ
İSTATİSTİK ve KANTİTATİF ARAŞTIRMALAR MERKEZİ
İSKAR Yayın No.96/2

STATISTICAL PROBLEMS AND SOLUTIONS FOR ECONOMICS AND BUSINESS ADMINISTRATION

Prof. Dr. Ateş VURAN

Dr. Şule ÖZMEN

Marmara University
Faculty of Economic and Administrative Sciences
Department of Business Administration

İSTANBUL, 1996

MARMARA ÜNİVERSİTESİ YAYINEVİ

PREFACE

This book is especially written for the students of the Faculty of Economic and Administrative Sciences studying Business Administration, Economics, International Relations in English, as a supplementary book of the 'Statistics Courses'. The problems and solutions are related with the topics mostly covered in first semester of a two semester Statistics course. The related topics required to solve the problems are explained briefly in the solutions of the problems. Therefore the student may easily recognize the contents of the subjects necessary for solving the problem.

The problems are formulated in a way that fits to the situations and conditions of Turkey unlike the books that are written by foreign authors in English. Though many problems that the managers and decision makers face in economics and business environment are mostly similar, there may be differences in countries like different economic and business environments, different money units and measurement units as length, weight, area, etc. Therefore, the problems are appropriate for the needs of graduates in their professional life in Turkey.

The text has been written in a somewhat conversational style as if giving lecture in a class-room to increase the understanding of the students. The origins of the problems are mainly the researches that we have conducted or the examples that are prepared for the lectures and the questions of the statistics exams for years. Therefore we had to opportunity to see the mistakes that the students had made while solving these problems and tried to reveal the students about the potential mistakes by giving explanations under the headings "***Attention***". Interpretations of the results are also made to facilitate the students to use the statistical techniques as tool in decision making process in their professional life..

Though, there may be the effect of the books both written in English or Turkish while formulating some of these problems, none of them are exactly taken from those books. However there are some typical problems that can be faced in any statistics books like deciding to accept or reject a lot on the basis of the probability of observing specific number of defective items and the like.

MARMARA ÜNİVERSİTESİ YAYINEVİ

The topics that are covered in this problems book are mainly; "Data Classification, Data Summarisation, Measures of Central Tendency and Measures of Dispersion" in Part I, "Probability" in Part II and "Probability Distributions" in Part III.

In the problems of the first part, analysis of both the raw data and classified data is made. The steps that should be followed in classification procedure is explained and in addition to the frequency distribution tables, classified data are presented by graphical displays and mainly Excel and SPSS software packages are used for displaying these graphics. Arithmetic mean, weighted mean, geometric mean, mode, median, quartiles, range, interquartile range, mean absolute deviation, variance, standard deviation, coefficient of variation are covered and some of these are computed both for ungrouped and grouped data.

In the problems of the second part, concept of probability is covered and the types of the events like mutually exclusive, complementary, collectively exhaustive dependent and independent events are explained within the problems. Based on the characteristics of the events, application of addition and the multiplication rules, in combining the events are explained and the probabilities like marginal, joint and conditional probabilities are computed in the problems. Problems related with "Bayes Rule" and problems that are formulated as contingency tables are also solved.

The problems of the third part is related with the discrete and continous probability distributions. Characteristics of binomial probability distributions, Poisson process and normal probability distribution are discussed in the problems. For each of the problems related with normal distribution, by using Excel program, graphs are prepared to show the areas corresponding to the required probabilities.

The problems are mostly solved by using statistical software packages. These are mainly Microsta, NCSS, SPSS. "Data Analysis" and "Functions" option of Excel program is also used to solve the problems.

İstanbul 1996

Prof. Dr. Ateş VURAN

MARMARA ÜNİVERSİTESİ YAYINEVİ

TABLE OF CONTENTS

PART I Page No.

Data Classificaton, Data Summarization, Measures of Central Tendency and Measures of Dispersion 1

PART II

Probability 42

PART III

Probability Distributions 55

MARMARA ÜNİVERSİTESİ

MARMARA ÜNİVERSİTESİ 1883

MARMARA ÜNİVERSİTESİ YAYINEVİ

PART I

Data Classification, Data Summarization, Measures of Central Tendency and Measures of Dispersion:

The problems in this section are mainly related with summarizing mass of data by constructing frequency distributions and describing data by measures of central tendency and dispersion. Presentation of data by graphical displays like histogram, frequency polygon and ogives are also shown in the solutions.

1- Weekly payment of randomly selected 50 textile workers are given in the following table.

Table 1.1 Weekly Payment of 50 Workers (in 1,000,000 TL)

2.32,	2.63,	1.55,	1.14,	1.95,	**3.28,**	2.92,	1.11,	1.70,
1.83,	1.97,	1.34,	1.42,	2.89,	2.63,	1.60,	1.70,	2.26,
2.38,	2.36,	1.41,	1.67,	1.18,	1.27,	1.31,	1.67,	**0.91,**
1.48,	1.53,	1.62,	1.75,	2.18,	1.82	2.87,	1.30,	1.78,
2.07,	1.55,	2.10,	1.85,	2.00,	2.43,	1.80,	1.78,	1.79,
1.80,	2.45,	2.00,	1.50,	2.20				

a - Classify the data set, construct the frequency and relative frequency distribution tables.

b - Construct cumulative frequency distribution table and cumulative relative frequency distribution tables.

c - Draw histogram.

d - Draw relative frequency histogram.

e - Draw frequency polygon.

f - Draw ogives.

SOLUTION

a - First step in classification procedure for the purpose of constructing a frequency distribution is to find out the range of the data set.

Range = Highest value - Lowest value. (These values are given as bold in Table 1.1)
Range = 3.28 - 0.91 = 2.37.

Second step is to decide about the number of classes and class widths. The number of the classes used for grouping data is selected arbitrarily based on the nature of the data and the purpose of the user. However, a reasonable number of classes and class width should be selected. What is important is; to avoid the loss of information by selecting few classes. On the other hand, it should be noticed that if too many classes are used, the purpose of classification and summarization would not be achieved. For example; if empty classes or classes with frequency of 1 or 2 occur, it may indicate that number of classes are too large.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Alternative number of classes, class widths can be used and final choice can be made after evaluating the alternatives.

Following rule may help to determine the class width.

$$\text{Class width} = \frac{\text{Range}}{\text{No of classes}}$$

Alternative class widths based on the number of classes chosen may be as follows:

6 classes:

2.38 / 6 = 0.396 Instead of 0.396, it is better to round it and use 0.4 as class width. Because grouping the data into classes and computations by grouped data can be simplified. Representation of data and class limits are also better if rounded class widths and multiples of 5, or 0.5 are used.

or

5 classes:

2.38 / 5 = 0.476 Here again the width can be assumed as 0.5.

As seen above, number of classes and class width should be determined together as they depend each other.

Attention: Raw data can be classified by the help of statistics software programs. In such statistics software packages, you have the chance of trying different class widths. This provides opportunity to select the most appropriate class width and number of classes in accordance with the purpose.

In the following table, weekly payments are grouped into 5 classes and the class width is 0.5. Notice that lower class limit of the first class is 0.90, rather than the minimum value (0.91) in the raw data set. That is; you do not necessarily start with the minimum value as the lower limit of the first class. Both for making the computations easier and good representation of data, the following classification alternative is better to be considered rather than the classes like $0.91 < 1.41$; $1.41 < 1.91$ etc.

Table 1.2 Frequency distribution and relative frequency distribution table.

Weekly Payment (in million TL)	No. of Workers (frequency)	Relative (as percent)	Frequency (as proportion)
.90 less than 1.40	8	16.00 %	0.16
1.40 less than 1.90	22	44.00 %	0.44
1.90 less than 2.40	12	24.00 %	0.24
2.40 less than 2.90	6	12.00 %	0.12
2.90 less than 3.40	2	4.00 %	0.04
Total	50	100 %	1.00

MARMARA ÜNİVERSİTESİ YAYINEVİ

Following descriptions that are given as examples can be made about weekly payments from the above frequency distribution table (Table 1.2).

- Out of 50 workers, 22 workers' weekly payment is between 1.4 and 1.9 million TL. 1.4 is included as the classes are defined as including the lower class limit. 1.90 is excluded as the values up to that value is included in considered class.
- 24 % of the workers have weekly payment between 1.9 and 2.4 million TL.
- Proportion of workers who have weekly payment between 900.000 and 1.4 million TL is 0.16.

Attention: "less than" quotations are used between the class boundaries. Note that, value(s) equal to the lower class limit of the classes are included in that class but the data equal to the upper class limit of the classes are not included in that class but included in the succeeding class. This approach is more convenient than constructing classes like 0.90 to 1.39, 1.40 to 1.89, both for simplifying the computations of summary measures like mean and for covering every possible data like 1.395 or 1.896.

Attention: For simplifying the presentation in frequency distribution tables, "less than" quotations will not be repeated in each of the tables in the following problems and solutions. It should be accepted that each class includes the values equal or more than the lower class limit up to the upper class limit excluding the values that are exactly equal to the upper class limit.

b - Cumulative and relative cumulative frequency distributions can be constructed either in ascending order (Table 1.3) or in descending order (Table 1.4).

Table 1.3 Cumulative frequency and cumulative relative frequency distribution tables in ascending order.

Weekly Payment (in million TL)	Cumulative Frequency (ascending)	Cumulative Relative Frequency (ascending)	
		(as percent)	(as proportion)
.90 - 1.40	8	16.00 %	0.16
1.40 - 1.90	30	60.00 %	0.60
1.90 - 2.40	42	84.00 %	0.84
2.40 - 2.90	48	96.00 %	0.96
2.90 - 3.40	50	100.00 %	1.00

Following descriptions can be made about weekly payments from the above cumulative frequency distribution table in ascending order (Table 1.3).

- 48 workers' weekly payment is less than 2.9 million
- 84 percent of the workers' weekly payment is less than 2.4
- Proportion of the workers who have weekly payment less than 1.4 million is 0.16
- Proportion of the workers who have weekly payment less than 1.9 million is 0.60

MARMARA ÜNİVERSİTESİ YAYINEVİ

Table 1.4 Cumulative frequency and cumulative relative frequency distribution tables in descending order.

Weekly Payment (in million TL)	Cumulative Frequency (descending)	Cumulative Relative Frequency (descending) (as percent)	(as proportion)
.90 - 1.40	50	100.00 %	1.00
1.40 - 1.90	42	84.00 %	0.84
1.90 - 2.40	20	40.00 %	0.40
2.40 - 2.90	8	16.00 %	0.16
2.90 - 3.40	2	4.00 %	0.04

Following descriptions can be made about weekly payments from the above cumulative frequency distribution table in descending order (Table 1.4).

- 42 workers' weekly payment is equal or more than 1.4 million.
- Proportion of the workers who have weekly payment equal or more than 1.9 million is 0.40.
- 16 percent of the workers' weekly payment is equal or more than 2.4.
- Proportion of the workers who have weekly payment equal or more than 2.9 million is 0.04.

Attention: Ascending order cumulative frequencies corresponding to a class, represent total number of workers who receives *less than* the amount stated as the *upper class limit* of the considered class. Descending order cumulative frequencies corresponding to a class, represent total number of workers who receives *more than or equal* to the amount stated as the *lower class limit* of the considered class.

c - Histogram is a graphical presentation of data set grouped into classes. Cartesian coordinates are used for drawing histograms. Classes are recorded on the apsis and the frequencies (absolute or relative) are recorded on ordinate. However frequencies are represented by the areas of the rectangles whose bases are the widths of the classes. That is heights of the rectangles are not necessarily correspond to the frequencies of the classes in case of unequal class widths. For example; if the frequency distribution has a class that has a width equal to 2 units where the other classes' widths are 1 unit (ie. twice of the other classes' width), then the height of the rectangle representing that unequal class should be half of the class frequency. Because the product of the base (2 units) and the height (half of the frequency) should be equal to the frequency. As a result the areas of the rectangles represent the frequencies.

Refer Problem 10 on page 28 as an example of a histogram with unequal class widths.

MARMARA ÜNİVERSİTESİ YAYINEVİ

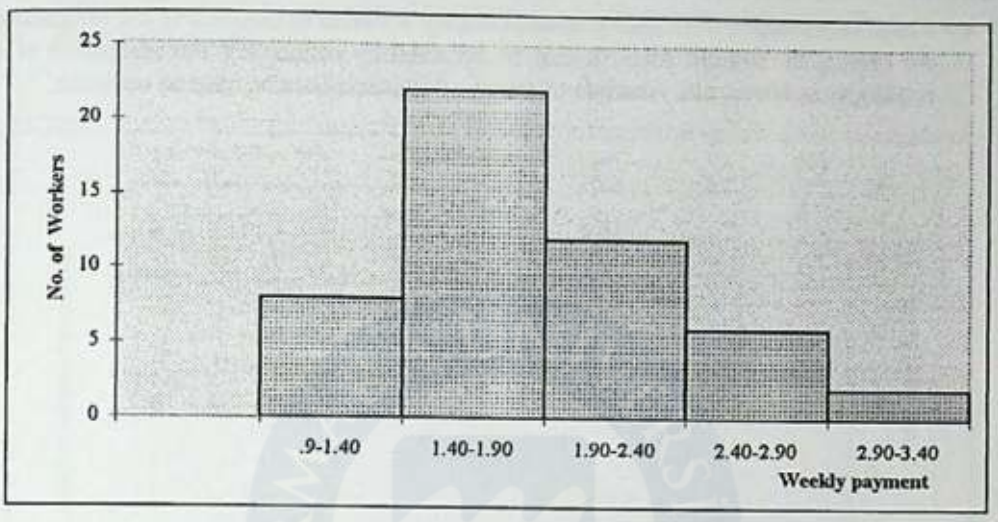


Fig.1.1. Histogram of Weekly Payments of 50 Textile Workers.

d - The only difference between the histogram shown in fig 1.1. and the relative frequency histogram is that the vertical axis shows relative frequencies instead of absolute frequencies.

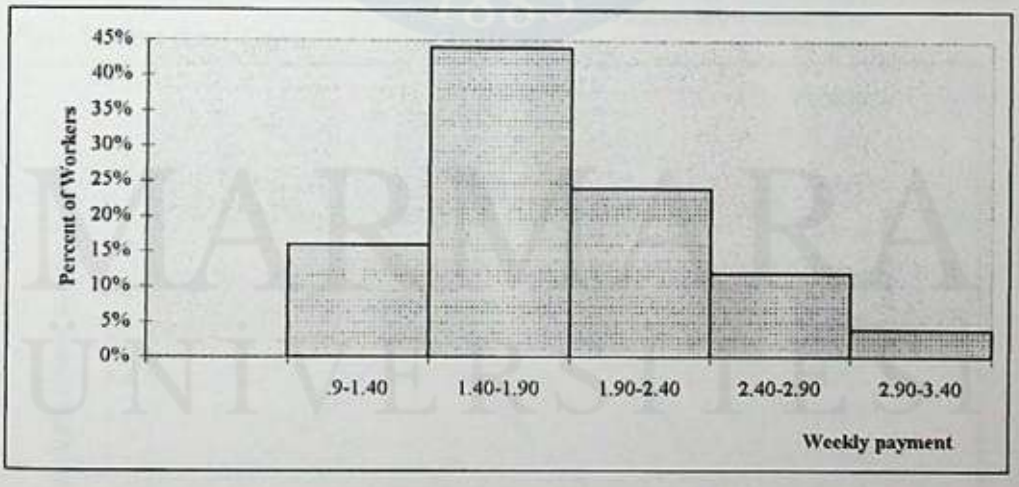


Fig.1.2. Relative Frequency Histogram of Weekly Payments of 50 Textile Workers

MARMARA ÜNİVERSİTESİ YAYINEVİ

e - Frequency polygon is another graphical display obtained by connecting the midpoints of the classes by straight lines. It can be obtained by connecting the midpoints of the rectangles in histogram. Absolute or relative frequencies can be used on ordinate.

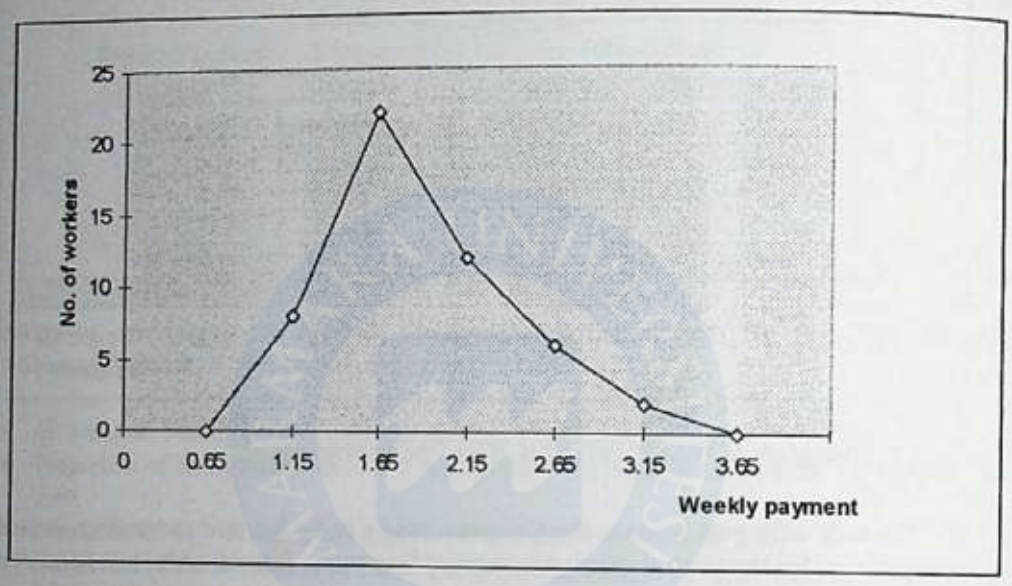


Fig. 1.3. Frequency Polygon of Weekly Payments of 50 Textile Workers (Scale of orninate is absolute frequencies)

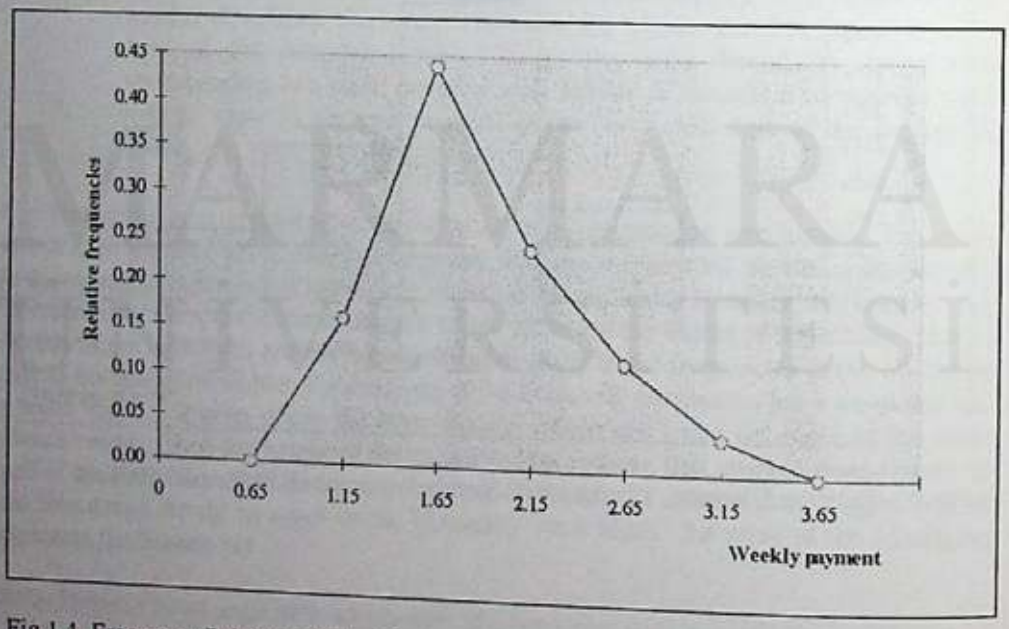


Fig.1.4. Frequency Polygon of Weekly Payments of 50 Textile Workere (Scales of ordinate is relative frequencies)

MARMARA ÜNİVERSİTESİ YAYINEVİ

f - Ogives are drawn on the basis of the cumulative or relative cumulative frequencies. In the following figures "less than ogive" based on ascending order relative cumulative frequencies and "or more ogive" based on descending order cumulative frequencies are given respectively.

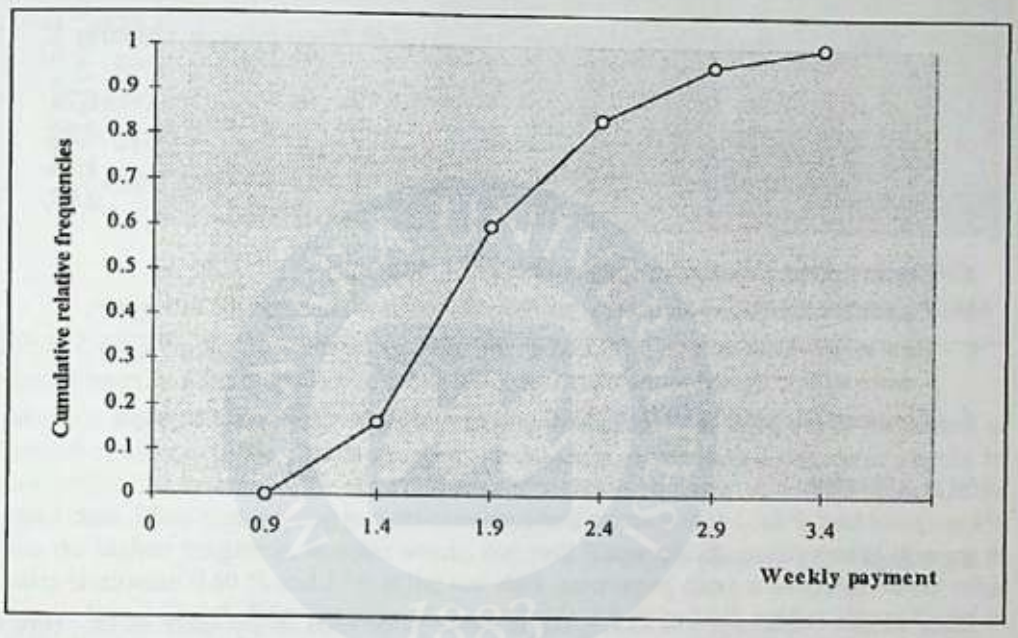


Fig.1.5. "Less than" Ogive.

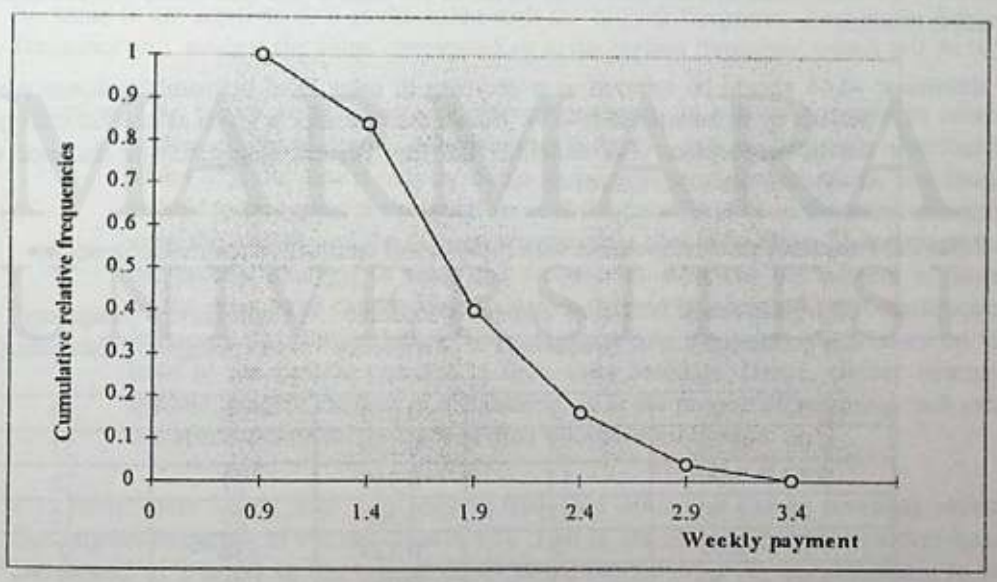


Fig.1.6. "Or more" Ogive.

MARMARA ÜNİVERSİTESİ YAYINEVİ

2 - Product manager of a new product - that has been recently introduced to the market - wants to evaluate the pattern of the demand of this product. For this purpose, percent change in sales amount in recent month with respect to the previous month is computed by using the data collected from 40 stores located in different territories. In the following table increases and decreases in sales amount of those stores are recorded as percent changes.

Table 2.1. Percent Changes in Sales Amount of 40 Stores between 2 Months.

-0.45,	-0.30,	-0.25,	0.05,	0.10,	0.15,	0.40,	0.60,	0.70,	0.70,
0.70,	0.70,	0.75,	0.80,	0.80,	0.80,	0.85,	0.90,	0.90,	0.95,
0.95,	1.00,	1.00,	1.10,	1.10,	1.10,	1.20,	1.20,	1.25,	1.30,
1.30,	1.35,	1.40,	1.50,	1.50,	1.60,	1.80,	1.90,	2.00,	2.10,

a - Classify percent change in amount of sales of 40 stores.

b - Construct the relative frequency histogram and show mode on the histogram.

c - How many stores will be classified as the most successful stores if only top 5 % of the stores will be treated as the most successful stores according to sales increase criteria?

d - Construct "or more" and "less than" ogives and show median on the graph.

SOLUTION

a - Range = 2.1 - (-0.45) = 2.55

Class width = 2.55 / 6 = 0.425 0.5 is better to be used as the class width, both for simplifying computations and for good representation. It is also more reasonable to start with -0.50 rather than to start with -0.45 as the lower limit of the first class, because of the same reasons.

Attention: -0.45 should be inferred as a decrease in sales level between the consecutive periods by an amount of 0.45 %. Notice that the data is given as percent changes not as proportions. (For example, the first percent change is the decrease of 0.45% but not a decrease of 45 %).

Table 2.2.Frequency distribution table with relative and cumulative relative frequencies.

Change in Sales (as percentage)	No. of Stores (frequency)	Relative Frequency	Cumulative (ascending)	Frequency (descending)
- 0.50 - 0.00	3	0.075	3	40
0.00 - 0.50	4	0.100	7	37
0.50 - 1.00	14	0.350	21	33
1.00 - 1.50	12	0.300	33	19
1.50 - 2.00	5	0.125	38	7
2.00 - 2.50	2	0.050	40	2

b - Relative frequency histogram is drawn by the using the relative frequencies either expressed as proportion or percentage.

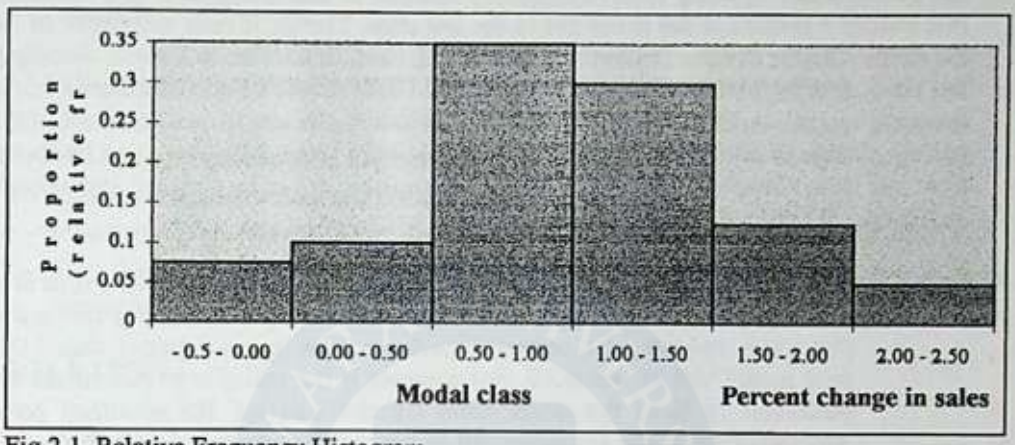


Fig 2.1. Relative Frequency Histogram

Mode is the most frequently observed value. Like median and mean, it is a measure of central tendency. If the data is grouped, modal class rather than mode value should be considered. Modal class is the class that has the highest frequency. In this example, it is the third class. Class representing percent changes equal or more than 0.50 % and less than 1% has the highest frequency. In other words; the most frequently observed percent changes in sales is between 0.50 % and 1 %. If the raw data (ungrouped data) is available, mode value can be found directly from the data set. From Table 2.1 on page 8, mode value is found as 0.70. Because in 4 of the stores -that is the highest frequency in raw data-, percent change in sales is 0.70 %. Pay attention that mode value is **not equal to** the highest frequency, that is, mode is not equal to 4, it is the value with the highest frequency. Four is the highest frequency and mode is the value corresponding to the highest frequency, which is 0.70 %.

Attention: One of the purposes in descriptive statistics is to present the data in tabular and/or graphical forms. After summarizing the data as frequency distribution tables or if the data is only available as frequency distribution tables, you should be able to recognize the original form of the data. That is, in the above example, what the classes and the frequencies represent should be clear. Classes represent the percent changes in sales and frequencies represent the number of stores whose sales have changed by the values defined by classes. While constructing frequency distribution tables, you are classifying the stores into classes on the basis of the percent changes in their sales amounts. Hence, classes represent specific percent changes and classification is the process of assigning each store to a specific class on the basis of their sales amount change.

c - Refer Table 2.2 on page 8. In relative frequency column, it can be obviously noticed that relative frequency of the last class is 5 %. That is, the most successful 5 percent stores correspond to 2 stores in the sample of 40 stores (40 x 0.05 = 2) and according to the grouped data, percent increase in sales of those 2 stores is between 2.00 percent and less than 2.50 percent. When the raw data is available, exact values can be found after organizing them as an ordered array. In this example, from Table 2.1, exact percent change

MARMARA ÜNİVERSİTESİ YAYINEVİ

in sales amount of most successful 2 stores can be obtained. They are 2.00 % and 2.10 %. However, if only grouped data is available as given in Table 2.2, answer for such a question can be computed by using relative frequency column. In this example, it is a coincidence that exactly 5 percent of the stores fall in the last class. That is, if most successful 10 % of the stores' change in sales amount was questioned, then, in addition to 2 stores classified in last class, another 2 stores classified in class 1.50 - 2.00 should be also considered, because the most successful 10 % stores correspond to 4 stores (40 x 0.10 = 4). As a result, the percent change in sales amount, -that 10 % of successful stores fall above,- can be obtained from raw data. However if only grouped data is available, it can be obtained approximately by the help of ogives shown in part **(d)**.

Attention: The policy of the producer of this product may be to reward the top 5 % of all the stores (population of stores). Depending on this sample result, they can quantify this policy and state that, stores achieving a sales increase greater than 2.00 % next month will be rewarded. For example if the policy is to reward the most successful 10 % of the stores, from the raw data set, the minimum percent increase above which, the stores will be rewarded can be found as 1.80.

d - "Less than" and "or more" ogives are shown in the following graph. Median is the intersection point of two ogives. Notice that "less than" and "or more" ogives intersect at a point that is very close to 1 % and that percent change in sales amount correspond to 50th percentile (that is to 0.50 on ordinate). Hence 50 % of the stores' sales amount is changed by less than the median value and 50 % of stores sales amount is changed by more than the median value.

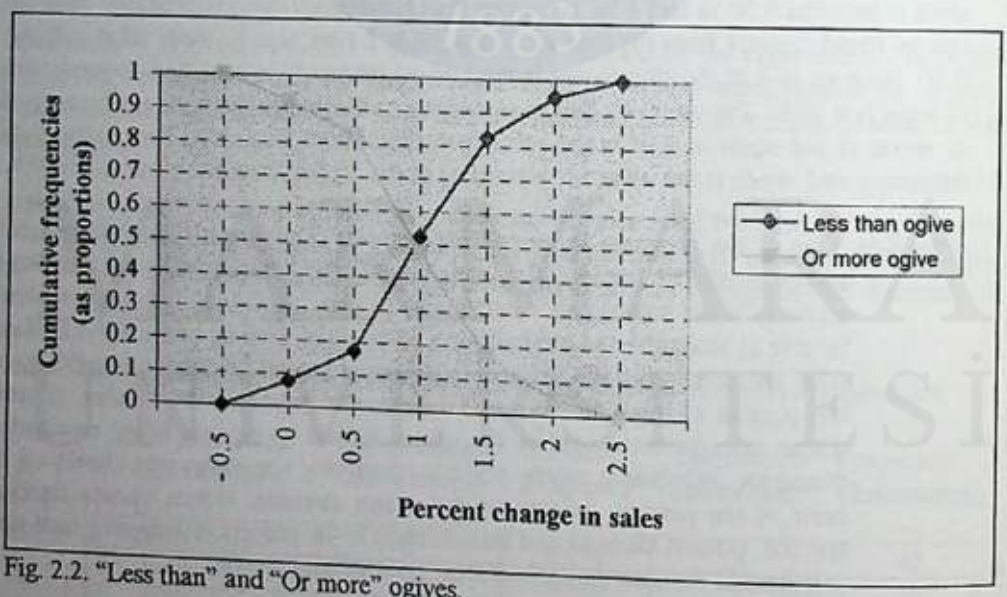


Fig. 2.2. "Less than" and "Or more" ogives.

Median is the middlemost value which divides the data values into half. Like mode it is a measure of central tendency. If the raw data (ungrouped data) is available, median value can be found directly from the data set. If data set contains odd number of data values there will be a single middlemost value. Otherwise (in case of even number of data values), median is the average of the middlemost two values. In this example, as seen in Table 2.1, median value is the 0.95 %. That is; average of 20th and 21st values. (0.95 + 0.95)/2 = 0.95.

MARMARA ÜNİVERSİTESİ YAYINEVİ

3 - Refer to the weekly payment data given in 1st Problem to answer the following questions.

a - Calculate mean (arithmetic mean) and standard deviation from raw data.
b - Calculate mean, mean absolute deviation, standard deviation from the grouped data.
c - Find median from raw data and compute from grouped data.
d - Find mode from raw data and compute from grouped data.
e - What is the proportion of the workers whose weekly payment is less than 2.4 million?
f - What is the proportion of the workers whose weekly payment is less than the average payment?

SOLUTION

a - Arithmetic mean of weekly payments of 50 workers is computed by dividing the sum of 50 workers' weekly payments to the total number of workers which is 50. (Refer Table 1.1)

$$\text{Arithmetic mean} = \frac{2.32 + 2.63 + \ldots\ldots + 1.50 + 2.20}{50} = 1.881$$

$\mu \Rightarrow$ Symbol for arithmetic mean for population.

N $\Rightarrow$ Symbol for total number of observations if the set of data belongs to a population.

$\bar{X} \Rightarrow$ Symbol for arithmetic mean for sample

n $\Rightarrow$ Symbol for total number of observations if the set of data belongs to a sample.

Attention: **a)** If data is ungrouped, arithmetic mean is computed by dividing the sum of the observations (Σxi) to the total number of observations (N or n). **b)** If data values are only classified without grouping, that is; if frequency of each data value is available, sum of the observations can be computed by multiplying each data value with its frequency ($\Sigma xi.fi$) and to obtain mean, this sum is divided by the sum of the frequencies (Σfi). Sum of the frequencies is equal to total number of observations. **c)** If data values are classified with grouping, in other words grouped into classes, then, the sum of the observations can be *approximately* computed by using the midpoints of the classes ($\Sigma mi.fi$). Similarly, to obtain mean, this sum is divided by the sum of the frequencies (Σfi).

Standard deviation: For computing standard deviation; first of all, deviation of data values from their mean is found by subtracting each data point from the mean. Each deviation is squared and then all squared deviations are added. Finally, mean of the sum of the squared deviations are computed by dividing the sum of the squared deviation to the total number of observations. Note that standard deviation is a descriptive measure of dispersion and shows the average deviation of each data value from its mean.

MARMARA ÜNİVERSİTESİ YAYINEVİ

$\sigma \Rightarrow$ Symbol for standard deviation for population.

$S \Rightarrow$ Symbol for standard deviation for sample.

$$S = \sqrt{\frac{(2.32-1.881)^2+(2.63-1.881)^2+.....+(1.50-1.881)^2+(2.20-1.881)^2}{50-1}} = 0.520374$$

Attention: If a set of data belongs to a population, sum of the squared deviations is divided by N (population size). If it belongs to a sample, sum of squared deviations is divided by (n - 1). In this example, a randomly selected sample of 50 textile workers are considered. Therefore denominator is (n - 1).

b - Computations of mean, mean absolute deviation and standard deviation of the grouped data values are given in the following table. Descriptive measures like mean, median, mode, standard deviation that are computed from grouped data are approximate values. However, if the raw data is not available, they are useful approximations.

Table 3.1 - Computation of average from grouped data.

Weekly Payment (in million)	Class midpoints m_i	Frequency f_i	Classmidpts. x frequencies $m_i \times f_i$
.90 - 1.40	1.15	8	1.15 x 8 = 9.2
1.40 - 1.90	1.65	22	1.65 x 22 = 36.3
1.90 - 2.40	2.15	12	2.15 x 12 = 25.8
2.40 - 2.90	2.65	6	2.65 x 6 = 15.9
2.90 - 3.40	3.15	2	3.15 x 2 = 6.3
Total		50	$\Sigma (m_i f_i)$ = 93.5

Mean = $\Sigma m_i f_i / (\Sigma f_i)$ = 93.5 / 50 = 1.87

Mean computed from the **raw data** is found as **1.881.**
Mean computed from **grouped data** is found as **1.87.**
As you see, mean computed from the grouped data is a good approximation.

Table 3.2 - Computation of mean absolute deviation from grouped data.

Weekly Payment (in million)	Class midpoints m_i	f_i	Deviations = $m_i - \bar{x}$	Absolute deviations x frequencies
.90 - 1.40	1.15	8	1.15 - 1.87 = - 0.72	0.72 x 8 = 5.76
1.40 - 1.90	1.65	22	1.65 - 1.87 = - 0.22	0.22 x 22 = 4.84
1.90 - 2.40	2.15	12	2.15 - 1.87 = 0.28	0.28 x 12 = 3.36
2.40 - 2.90	2.65	6	2.65 - 1.87 = 0.78	0.78 x 6 = 4.68
2.90 - 3.40	3.15	2	3.15 - 1.87 = 1.28	1.28 x 2 = 2.56
Total		50		$\Sigma \lvert m_i - \bar{x} \rvert f_i$ = 21.2

MARMARA ÜNİVERSİTESİ YAYINEVİ

Mean Absolute Deviation = $(\Sigma \mid m_i - \bar{x} \mid f_i) / n - 1 = 21.2 / 49 = 0.4326$

Table 3.3 - Computation of standard deviation from grouped data.

Weekly Payment (in million)	m_i	f_i	Deviations = $m_i - \bar{x}$	Squared deviations x frequencies
.90 - 1.40	1.15	8	1.15 - 1.87 = - 0.72	0.518 x 8 = 4.144
1.40 - 1.90	1.65	22	1.65 - 1.87 = - 0.22	0.048 x 22 = 1.065
1.90 - 2.40	2.15	12	2.15 - 1.87 = 0.28	0.078 x 12 = 0.936
2.40 - 2.90	2.65	6	2.65 - 1.87 = 0.78	0.608 x 6 = 3.648
2.90 - 3.40	3.15	2	3.15 - 1.87 = 1.28	1.638 x 2 = 3.276
Total		50		Total = 13.069

$\Sigma f_i (m_i - \bar{x}_i)^2 = 13.069 \quad S = \sqrt{(\Sigma f_i (m_i - \bar{x}_i)^2) / n - 1} = \sqrt{13.069 / 49} = \sqrt{0.267}$

$S = \sqrt{0.267} = 0.517$

Sum of the squared deviations in the above table is 13.069 due to rounding. If the product of the frequencies and squared deviations would not be rounded, sum of the squared deviations is found as 13.08.

c - Median is the middlemost value. In order to find the middlemost value from the raw data, it should be sorted. Sorting can be either in ascending order or in descending order Sorted data in ascending order is given below.

Table 3.4. Sorted data about the weekly payments of 50 textile workers.

0.91,	1.11,	1.14,	1.18,	1.27,	1.30,	1.31,	1.34,	1.41,	1.42,
1.48,	1.50,	1.53,	1.55,	1.55,	1.60,	1.62,	1.67,	1.67,	1.70,
1.70,	1.75,	1.78,	1.78,	*1.79,*	*1.80,*	1.80,	1.82,	1.83,	1.85,
1.95,	1.97	2.00	2.00,	2.07,	2.10,	2.18,	2.20,	2.26,	2.32,
2.36,	2.38,	2.43,	2.45,	*2.63,*	*2.63,*	2.87,	2.89,	2.92,	3.28.

As the set of data contains even number of data values, median is the average of two middlemost values (Refer Table 3.4)

Median = (1.79 + 1.80) /2 = 1.795

Median is interpreted as follows; half of the textile workers weekly payment is below 1,795,000 TL. and the other half receives above 1,795,000 TL.

Median is computed approximately from grouped data in the following table. (Table 3.5). For computing median from grouped data, first step is to find out the class that the middlemost observation falls. Cumulative frequencies in ascending order can be used for

MARMARA ÜNİVERSİTESİ YAYINEVİ

this purpose. When these frequencies are examined, it can be deduced that second class includes the median value because 25th observation (the middlemost observation) falls in this class. 8 of the workers receive less than 1.40 and 30 receives less than 1.90. Therefore 25th worker's weekly payment should be within the limits of the second class.

Table 3.5. Computation of median from grouped data.

	Weekly Payment (in million TL)	f_i	Cumulative Frequency (ascending)
	.90 - 1.40	8	8
Median class ⇒	1.40 - 1.90	22	30
	1.90 - 2.40	12	42
	2.40 - 2.90	6	48
	2.90 - 3.40	2	50

$$Md = l + c\,\frac{(N/2) - F_p}{F_m} \qquad Md = 1.40 + 0.50\,\frac{(50/2) - 8}{22} = 1.786$$

l = Lower class limit of median class.
c = Class width of the median class.
N = Total number of observations.
F_p = Cumulative frequency of the class preceding the median class.
F_m = Frequency of the median class.

d - Mode value is the most frequently observed value. However in this data set, 5 data values are repeated two times. These are 1.55, 1.67, 1.78, 2.00 and 2.63. (Refer Table 3.4.) Hence, there may be more than one mode value in the data set. When the grouped data is examined, class with the highest frequency is the second class. That is; 22 of the workers or 44 % of the workers' weekly payment is between 1.40 and less than 1.90. (Refer Table 3.6) In grouped data set, it is also possible to observe more than one modal class. For example, if there are two mode values, then the distribution is said to be a bimodal distribution.

Table 3.6 Computation of mode from grouped data.

	Weekly Payment (in million TL)	Frequency
	.90 - 1.40	8
Modal class ⇒	1.40 - 1.90	22
	1.90 - 2.40	12
	2.40 - 2.90	6
	2.90 - 3.40	2

$$Mo = l + c\,\frac{d_1}{d_1 + d_2} \qquad Mo = 1.40 + 0.50\,\frac{14}{14 + 10} = 1.69167$$

MARMARA ÜNİVERSİTESİ YAYINEVİ

l = Lower class limit of median class.
c = Class width of the median class.
d_1 = Difference between the frequency of the modal class and the preceding class.
d_2 = Difference between the frequency of the modal class and the succeeding class.

e - Proportion of the workers whose weekly payment less than 2,400,000 million TL is 0.84 or 84 % of the workers' weekly payment is less than 2,400,000 million TL. Note that workers, whose payment is between 0.90 and less than 2.4 million (first, second and third class frequencies) are considered.

Table 3.7 Computation of proportion of the workers receiving less than 2,400,000 TL

Weekly Payment (in million TL)	No. of Workers (frequency)	Relative Frequency	Cumulative Relative Frequency
.90 - 1.40	8	16.00 %	16.00 %
1.40 - 1.90	22	44.00 %	60.00 %
1.90 - 2.40	12	24.00 %	84.00 %
2.40 - 2.90	6	12.00 %	96.00 %
2.90 - 3.40	2	4.00 %	100.00 %

f - Mean computed from the raw data set is equal to 1.881. From the raw data set, it can be counted that 30 data values are less than the mean. That is; **(30 / 50) x 100 = 60 %** of the workers' weekly payment is less than the average weekly payment.

The mean computed from grouped data is equal to 1.87. Computation of approximate proportion of the workers receiving less than mean, from the grouped data is as follows;

The mean falls in the second class, therefore 8 workers in the first class + some amount of the workers in second class receives less than 1.87. To find out that 'some amount', ratio of the 'class width of the class containing mean' to 'range between mean and the lower limit of that class is considered.

0.50 Width of the class containing mean.
1.87 - 1.40 = 0.47 The difference between the mean and the lower limit of that class

0.50 22 If there are 22 workers within the range of 0.50, how many workers
0.47 ? correspond to a range of 0.47 approximately?

(0.47 x 22) / 0.50 = 20.68 ≈ 21 workers

0.90 - 1.40 8 workers receives less than 1.40.
1.40 - 1.87 21 workers approximately receives between 1.40 and 1.87.

0.90 - 1.87 29 workers (58 %) approximately receives less than mean.

As a result based on the raw data, 60 %; based on grouped data, 58 percent of the workers' weekly payment is less than the average weekly payment.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Table 3.8. Summary of the measures of central tendency and dispersion for ungrouped and grouped data for question 3.

Measures of central tendency and dispersion	Ungrouped data	Grouped data
Arithmetic mean	1.881	1.87
Median	1.795	1.786
Mode	3 mode values	1.692
Standard deviation	0.520374	0.517

4 - An investor is evaluating the alternative of purchasing a small fast food restaurant. Among other things, he is also interested in the average number of customers per day and the extent of variation in number of customers from day to day. He has collected the number of customers who have visited this restaurant for past 50 days to obtain this information. After classifying the data, frequency distribution given in the following table is obtained.

Number of Customers	No of Days (frequency)
0 - 20	10
20 - 40	20
40 - 60	15
60 - 80	5

a - Draw histogram.

b - Draw frequency polygon.

c - Draw ogives.

d -Compute an average which provides you information about the measure of central tendency

e - Compute standard deviation as a measure of dispersion

f - Compute median, first quartile, second quartile and third quartile.

g - In what percent of the days, number of customers are i) above 40, ii) below 40, iii) above 50, iv) below 50?

h - According to the distribution, on a given day, what is the probability that i) at least 20 customer visit? ii) at most 39 customer visit?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

a - Histogram

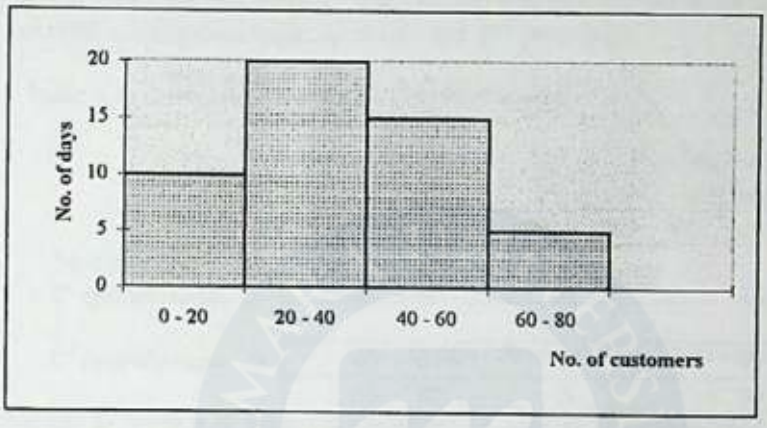


Fig. 4.1. Histogram of the Number of Customers .

b - Frequency polygon

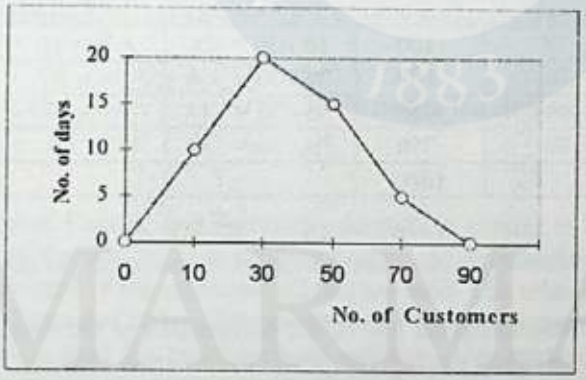


Fig. 4.2. Frequency Polygon of No. of Customers Per Day

c - For drawing ogives cumulative distributions are used.

Table 4.1 - Cumulative frequency distributions table.

No. of Customers	No. of Days	Cum. Freq. (ascending)	Cum. Freq. (descending)
0 - 20	10	10	50
20 - 40	20	30	40
40 - 60	15	45	20
60 - 80	5	50	5

MARMARA ÜNİVERSİTESİ YAYINEVİ

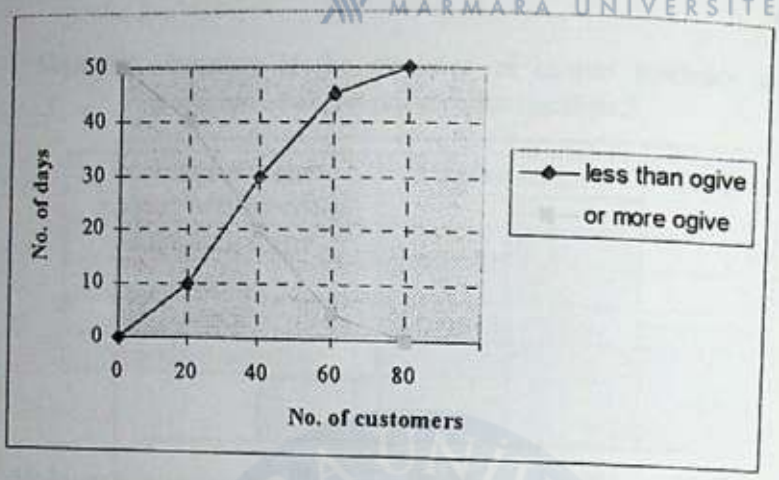


Fig. 4.3. "Less than" and "or more" ogives

d - Table 4.2 Computation of average and standard deviation of number of customers per day

No. of Customers	No. of Days (f_i)	Class Midpoints (m_i)	(f_i x m_i)	Deviations (m_i - *mean*)	Squared Deviation x Frequencies
0 - 20	10	10	100	10 - 36 = -26	676 x 10 = 6760
20 - 40	20	30	600	30 - 36 = - 6	36 x 20 = 720
40 - 60	15	50	750	50 - 36 = 14	96 x 15 = 2940
60 - 80	5	70	350	70 - 36 = 34	1156 x 5 = 5780
Total	**50**		**1800**		**16200**

Mean = $\bar{X} = \Sigma(m_i f_i)/(\Sigma f_i) = 1800/50 = 36$. On the average 36 customers per day.

Attention: You should pay attention to the unit of the denominator of the term used to compute mean. In other words; 'per what' the mean is required. In this example, denominator 50 is the number of days. Therefore; 36 is the average number of the ***customers*** visiting restaurant ***per day***.

e - For computing standard deviation, first; sum of the product of squared deviations shown in last column of Table 4.2 will be divided by 49 (n - 1) and the squareroot of the result will be taken.

$$S = \sqrt{16200/(50-1)} = \sqrt{330.61224} = 18.182745$$

f - For computing median, first quartile and third quartile from grouped data set, first step is to find out the corresponding classes from a cumulative frequency distribution table. (See Table 4.3) Remember that median is the value which divides data values into 2 equal pieces. Quartiles are the values which divide the data values into 4 equal pieces. Deciles are the values which divides the data values into 10 equal pieces.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Percentiles are the values which divide the data values into 100 qual pieces That is ¼ of the data values are below the 1st quartile and below the 25th percentile. 1st quartile is equal to the 25th percentile. ¾ of the data values are above the first quartile and above the 25th percentile. Second quartile and median is equal, because half of the observations fall below the 2nd quartile and half of them fall above the 2nd quartile. 2nd quartile is also equal to the 5th decile and 50th percentile.

Table 4.3. Cumulative frequency distribution table.

	No. of Customers	No. of Days	Cumulative Frequency (ascending)
	0 - 20	10	10
Median class ⇒ 1st quartile class ⇒	20 - 40	20	30
3rd quartile class ⇒	40 - 60	15	45
	60 - 80	5	50

Attention: Median class and/or first quartile and/or third quartile may fall in the same class.

Median is approximately 25th (N/2 = 50/2=25) observation. When cumulative frequencies are examined, it is observed that it falls in the second class. (Between 20 and 40).

$$Md = 20 + 20\ \frac{(50/2) - 10}{20} = 35$$ In half of the days no. of customers per day is below 35.

Quartiles, Deciles, and Percentiles formula is similar to the median formula. For *l*, lower limit of the classes corresponding to the required quartiles, deciles or percentiles and for c class widths of the respective classes are used. The other difference is the numerator of the ratio. Instead of N/2 used for median, N/4 is used for quartiles, N/10 is used for Deciles and N/100 is used for percentiles. In addition, this term is multiplied by the number that correspond to the quartile, the decile or the percentile to be computed.

$$Q_1 = l + c\ \frac{(1(N/4)) - F_p}{F_{Q1}} \qquad Q_2 = l + c\ \frac{(2(N/4)) - F_p}{F_{Q2}} \qquad Q_3 = l + c\ \frac{(3(N/4)) - F_p}{F_{Q3}}$$

l = Lower class limit of 1st (or 2nd or 3rd) quartile class.
c = Class width of the 1st (or 2nd or 3rd) quartile class.
N = Total number of observations.
F_p = Cumulative frequency of the class preceding the 1st (or 2nd 3rd) quartile class.
F_{Q1}= Frequency of the class containing the 1st quartile
F_{Q2}= Frequency of the class containing the 2nd quartile
F_{Q3}= Frequency of the class containing the 3rd quartile

MARMARA ÜNİVERSİTESİ YAYINEVİ

$$P_{25} = l + c\,\frac{(25(N/100)) - F_p}{F_{P25}} \qquad P_{75} = l + c\,\frac{(75(N/100)) - F_p}{F_{P75}}$$

Explaination of the terms for percentiles will not be repeated as they are similar to the terms explained for quartiles with minor differences that can be easily recognized.

Computation of 1st and 3rd quartiles are as follows. 2nd quartile is equal to median which has been already computed.

$$Q_1 = 20 + 20\,\frac{(50/4) - 10}{20} = 22.5 \qquad Q_3 = 40 + 20\,\frac{(3(50/4)) - 30}{15} = 50$$

$Q_2 = \text{Md} = 35.$

In ¼ of the days (or 25 % of the days) no. of customers is less than 22.5. In ¾ of the days (or 75 % of the days) no. of customers is above 22.5.

In ¾ of the days (or 75 % of the days) no. of customers is less than 50. In ¼ of the days (or 25 % of the days) no. of customers is above 50.

g - Percentage of the days that the number of the customers are below or above a specific value can be obtained from the following table. (Table 4.4)

Table 4.4. Cumulative relative frequency distribution of number of customers.

Number of Customers	No. of Days..	Cum. Relative freq. (ascending)	Cum. Relative freq. (descending)
0 - 20	10	20 %	100 %
20 - 40	20	60 %	80 %
40 - 60	15	90 %	40 %
60 - 80	5	100 %	10 %

i) In 40 percent of the days, number of customers per day are above 40.
ii) In 60 percent of the days number of customers per day are below 40.
iii) 50 is the midpoint of the third class. In 15 days no. of customers is between 40 and 60. Hence; approximately in half of the days (7.5 days) number of customer per day can be assumed as between 40 and 50 and in the other half of the days between 50 and 60. Therefore;

	50 - 60	7.5 days	approximately
	60 - 80	5 days	
Total	50 - 80	12.5 days	approximately

$$(12.5/50) \times 100 = 25\ \%$$

Approximately in 25 percent of the days number of customers are above 50.

MARMARA ÜNİVERSİTESİ YAYINEVİ

iv)

	0 - 20	10 days	
	20 - 40	20 days	
	40 - 50	7.5 days	approximately
Total	0 - 50	37.5 days	approximately

(37.5 / 50) x 100 = 75 %

Approximately in 75 percent of the days number of customers per day are below 50.

h - i)The proportion of the days that at least 20 customers visit the restaurant is 0.80. In other words in 80 % of the days number of customers visiting restaurant are equal or more than 20. On the basis of this information, it can be stated that; on a given day, the probability that at least 20 customer will visit the restaurant is 0.80.

ii) The proportion of the days that at most 39 customers visit the restaurant is 0.60. In other words in 60 % of the days number of customers visiting restaurant are less than 40. Note that in second class, number of days in which 20 to less than 40 customer visit restaurant is grouped. Therefore 40 customers are not included in that class. As no. of customers is countable, less than 40 customers can be inferred as 39 customers.

5 - For the purpose of evaluating their finished goods inventory policy, DUR Company, which produces durable goods gathered sales data from 40 stores which sell their products. Sales data for a group of durables showing similar demand patterns are collected from 40 stores and classified as given in table below. Each store has sent the previous years' annual sales quantities.

Table 5.1 Annual Demand of Durable goods

Annual Demand (No. of durable goods)	No of stores
500 - 600	1
600 - 700	11
700 - 800	15
800 - 900	10
900 - 1000	3
Total	40

a - Draw relative frequency histogram and relative frequency polygon for the previous years' annual demand of stores?

b - i)What is the average no of DUR products sold by a store? ii)What is the total number of products sold by 40 stores approximately?

c - What percent of stores sell above the average sales level?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

a - Relative frequency histogram and relative frequency polygon .

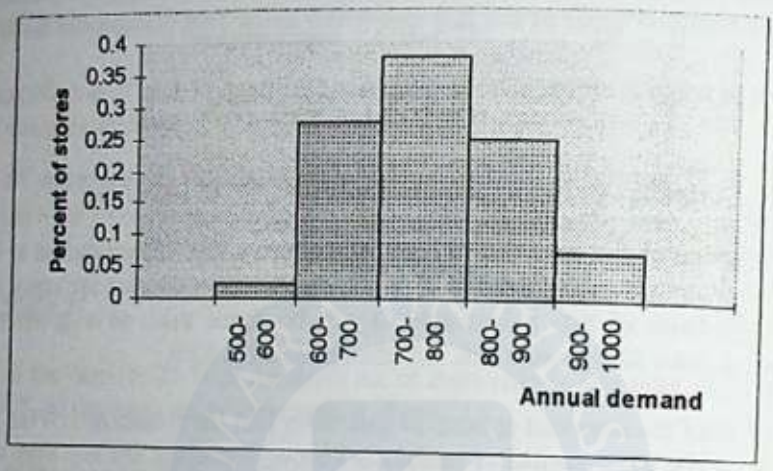


Fig 5.1. Relative Frequency Histogram of Annual Demand of Durable Goods

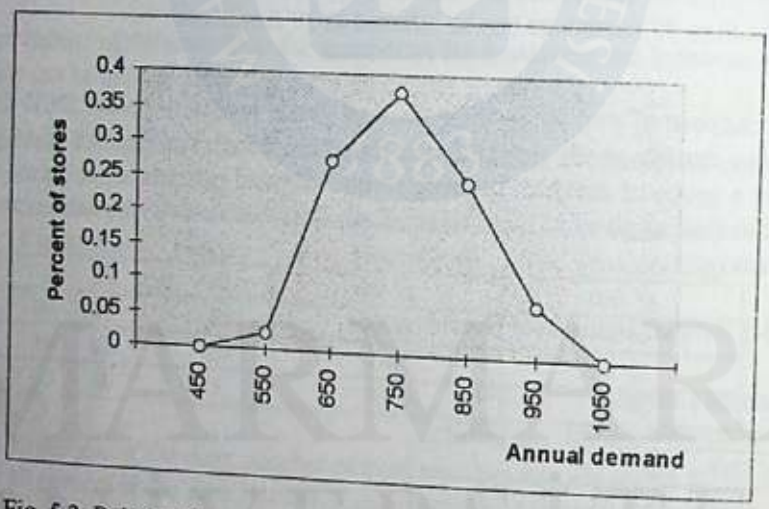


Fig. 5.2. Relative Frequency Polygon of Annual Demand of Durable Goods.

b - Table 5.1 Frequency and cumulative relative frequency distributions of annual demand.

Annual demand	No. of stores (f_i)	Class midpoints (m_i)	Cum. relative freq. (descending)	$m_i \times f_i$
500 - 600	1	550	100 %	550
600 - 700	11	650	97.5 %	7150
700 - 800	15	750	70 %	11250
800 - 900	10	850	32.5 %	8500
900 - 1000	3	950	7.5 %	2850
Total	**40**			**30300**

i) Mean = $\bar{X} = \Sigma(mi\,fi) / (\Sigma fi) = 30300 / 40 = 757.5$

ii)Total number of Dur products sold by 40 dealers ($\Sigma(mi\,fi)$) is approximately 30300. As the *product* of the class midpoints and the corresponding frequencies are *added*, total number of products is an approximation. In other words, class midpoints are treated as an average of the data values observed in that class and to find out the sum of the values in each class they are multiplied by the corresponding frequencies.

Compute standard deviation by yourself. (Answer; 97.1055)

c - Average falls in the third class (700 - 800) Class width is 100

	100	Width of the class containing mean.
800 - 757.5	= 42.5	The difference between the mean and the the upper limit of that class

100	15	If there are 15 stores which sell between 700 and 800 (or within a range of 100)
42.5	?	approximately how many stores sell between 757.5 and 800 (or within a range of 42.5)

(42.5 x 15) / 100 = 6.375 stores approximately. $\Rightarrow$ 6 stores

757.5 - 800	6	approximately in 6 stores, annual demand is between 757.5 and 800,
800 - 900	10	in 10 stores, annual demand is between 800 and 900,
900 - 1000	3	in 3 stores, annual demand is between 900 and 1000.
757.5 - 1000	19	stores approximately. (47.5 %)

47.5 percent of the stores annual sales is above the average sales level.
52.5 (100 - 47.5) percent of the stores annual sales is below the average sales level.

6 - Following data is the age distribution of 360 respondents questioned for a survey.

Ages	No. of respondents
15 - 25	40
25 - 35	70
35 - 45	110
45 - 55	80
55 - 65	60
Total	360

a - What is the mean age of the respondents?

b - What is the standard deviation?

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - Average age is computed below.

Table 6.1 - Computation of mean age and standard deviation of age.

Ages	f_i	m_i	$m_i \times f_i$	$f_i (m_i - \bar{x})^2$
15 - 25	40	20	800	40 (20 - 41.4) = 18318.4
25 - 35	70	30	2100	70 (30 - 41.4) = 9097.2
35 - 45	110	40	4400	110 (40 - 41.4) = 215.6
45 - 55	80	50	4000	80 (50 - 41.4) = 5916.8
55 - 65	60	60	3600	60 (60 - 41.4) = 20757.6
Total	360		14900	54305.6

$\bar{x} = 14900 / 360 = 41.3889$ rounded to **41.4.**

b - $s = \sqrt{54305.6 / (360 - 1)} = \sqrt{151.2691} = 12.29915$ rounded to **12.3**

Standard deviation computed without rounding mean is 12.4563.

7 - The following information is about the households income level of a town which is one of the alternative places for an investor planning to establish a supermarket. Compute and compare mean and median of the household income distribution

Household Income (in 1000000)	Relative frequencies
5 - 15	0.20
15 - 25	0.18
25 - 35	0.14
35 - 45	0.12
45 - 55	0.14
55 - 65	0.14
65 - 75	0.08

SOLUTION

In this example, instead of frequencies, relative frequencies are given. For computing mean and median from grouped data, relative frequencies can be used, because whether they are relative or not they can be treated as weights. As a result, arithmetic mean of a set of data values computed approximately from grouped data is, in a way, a weighted average where the weights are the frequencies or relative frequencies.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Table 7.1 - Computation of mean and median.

	Household Income	Relative Freq.	m_i	Cum.Rel. Freq.	m_i x rel freq.
	5 - 15	0.19	10	0.19	1.9
	15 - 25	0.25	20	0.44	5
Median class ⇒	25 - 35	0.17	30	0.61	5.1
	35 - 45	0.15	40	0.76	6
	45 - 55	0.12	50	0.88	6
	55 - 65	0.08	60	0.96	4.8
	65 - 75	0.04	70	1.00	2.8
	Total	**1.00**			**31.6**

$\bar{x} = \Sigma m_i f_i / (\Sigma f_i) = 31.6 / 1 = \mathbf{31.6}$ million TL ⇒ 31,600,000 TL

Attention: Note that the sum of the relative frequencies is equal to 1 if expressed as proportions or equal to 100 if expressed as percentage. Therefore if the relative frequencies expressed as proportions are used for computing mean, sum of the product of class midpoints and the relative frequencies will be equal to mean, if relative frequencies are expressed as percentages, then the sum of the product of class midpoints and the relative frequencies should be divided by 100 to find out mean.

$$Md = 25 + 10\ \frac{(1/2) - 0.44}{0.17} = 25 + 3.52941 = 28.52941 \text{ million TL.} \Rightarrow 28{,}529{,}410 \text{ TL.}$$

Approximately mean is 31,600,000 TL and median is 28,529,100 TL. 50 % of the households' income are less than 28,529,100 TL. As the mean is greater than median, more than 50 % of the housholds' income is below the mean and less than 50 % of the households' income is above the mean. This indicates that more households fall in lower income levels and less households falls in higher income levels.

8 - The following table shows monthly rental costs for a sample of 25 stores in a large shopping center APITOL, obtained from their rent contracts.

Rental cost in million TL	No. of stores
5 - 15	2
15 - 25	3
25 - 35	8
35 - 45	7
45 - 55	4
55 - 65	1

a - Compute mean rental cost.

b - Compute standard deviation.

c - Compute median

d - Compute first quartile.

e - Compute third quartile.

f - Compute Interquartile range and semi-interquartile range.

MARMARA ÜNİVERSİTESİ YAYINEVİ

. **a** - Table 8.1 Computation of mean and standard deviation.

Rental cost in million TL	No. of stores	*mi*	Cum. Freq.	*mi* x *fi*	$(mi - \bar{X})^2 \; fi$
5 - 15	2	10	2	20	595.36 x 2 = 1190.72
15 - 25	3	20	5	60	207.36 x 3 = 622.08
25 - 35	8	30	13	240	19.36 x 8 = 154.88
35 - 45	7	40	20	280	31.36 x 7 = 219.52
45 - 55	4	50	24	200	243.36 x 4 = 973.44
55 - 65	1	60	25	60	655.36 x 1 = 655.36
Total				**860**	**3816**

1st Q. and Md. class ⇒
3rd Q.class ⇒

Mean = $\bar{X} = \Sigma mi\,fi / (\Sigma fi) = 860 / 25 = 34.4$ million TL. 34,400,000 TL.

b -

$$S = \sqrt{(\Sigma fi(m_i - \bar{X}_i)^2) / n - 1} = \sqrt{3816/24} = \sqrt{159}$$

$$S = \sqrt{159} = 12.60952$$

Mean rental cost of the stores is 34,400,000 TL. and standard deviation is 12,609,520 million.

c - The median class is found as follows;

N/2 = 25/2= 12.5 Median is the value between 12th and 13th observation. In other words when the stores are sorted with respect to their rents in ascending order, the rent which is the average of the rent of the 12th and the 13th store is the median value.

$$Md = l + c\,\frac{(N/2) - F_p}{F_m} \qquad Md = 25 + 10\,\frac{(25/2) - 5}{8} = 25 + 9.375 = 34.375$$

Attention: The median value is found as closer to the upper class limit of the median class. Remember that the median value correspond to the average of the 12th and 13th observations Based on cumulative frequencies, it is inferred that 13 of the stores' rents are less than 35 million TL. and 5 of them are less than 25 million TL. Therefore the median value corresponding to the average of the 12th and 13th observations is expected to be between 25 and 35 million TL. and closer to 35 million. The number of stores with referred rental costs are 8. That is the average of 12th and 13th observation out of 25 observation correspond to the average of 7th and 8th observation in the third class. For that reason it is closer to the upper limit of the median class.

MARMARA ÜNİVERSİTESİ YAYINEVİ

$$Q_1 = l + c\,\frac{(1(N/4)) - F_p}{F_{Q1}} = 25 + 10\,\frac{(1(25/4)) - 5}{8} = 25 + 10\,\frac{6.25 - 5}{8} = 26.5625$$

Q_1 = 26.5625 million TL. $\Rightarrow$ 26,562,500 TL.

$$Q_3 = l + c\,\frac{(3(N/4)) - F_p}{F_{Q3}} = 35 + 10\,\frac{(3(25/4)) - 13}{7} = 35 + 10\,\frac{18.75 - 13}{7} = 43.21429$$

Q_3 = 43.21429 million TL. $\Rightarrow$ 43,214,290 TL.

Interquartile range is a measure of dispersion and equal to the difference between 3rd quartile and 1st quartile. $Q_3 - Q_1$.

IQR = 43,214,290 - 26,562,500 = 16,651,790 TL.

Semi - interquartile range = $SIQR = IQR / 2$ = 16,651,790 / 2 = 8,325,895

9 - 25 common stocks from textile sector is selected and percentage increases in their prices with respect to previous week is examined. The following table shows the percentage increases.

Increase (%)	1 - 3	3 - 5	5 - 7	7 - 9	9 - 11	11-13
No of stocks	1	3	6	8	5	2

a - What is the average percent increase in prices of 25 stocks

b - Compute standard deviation

c - Compute coefficient of variation

d - Show at which class the median value falls and compute median.

SOLUTION

a - Mean = $\Sigma m_i f_i / (\Sigma f_i) = \dfrac{(2x1) + (4x3) + (6x6) + (8x8) + (10x5) + (12x2)}{25} = 7.519$

Average increase per stock is approximately 7.5 %.

Attention: Notice that the mean is obtained by dividing the numerator by 25. 25 is the amount of the stocks and percent increases in their prices are computed between two consecutive weeks. As a result, the average is inferred as the percent increase per stock. If the average rate of increase of a stock per a time period is required, the mean that should be used is ***geometric mean*** but not the arithmetic mean. That is; if the rate of changes, -either decreases or increases- are considered on hourly, weekly, monthly, yearly etc. basis, then to find out the average rate of change per hour, per week, per month, per year, etc. geometric mean should be computed.

MARMARA ÜNİVERSİTESİ YAYINEVİ

b - Standard deviation of the sample of 25 common stocks.

$$s = \frac{1(2-7.5)^2 + 3(4-7.5)^2 + 6(6-7.5)^2 + 8(8-7.5)^2 + 5(10-7.5)^2 + 2(12-7.5)^2}{25-1} =$$

$$= \sqrt{154.25/2} = \sqrt{6.427083} = 2.535169$$

c - ***Coefficient of variation*** = (Standard deviation / mean) 100

$$= (2.535169 / 7.519) = 33.72\ \%$$

Standard deviation is 33.72 % of the mean.

Coefficient of variation expresses the standard deviation as a percent of the mean. It is a unitless measure and can be used to compare the variation between the set of data values measured by different units.

d - Median value falls in the 4th class, between 7 % and 9 %.

$$Md = 7 + 2\ \frac{(25/2) - 10}{8} = 7 + 0.625 = 7.625\ \%.$$

10 - The following relative frequency histogram represent the income distribution of 1000 households of a specific geographical area.

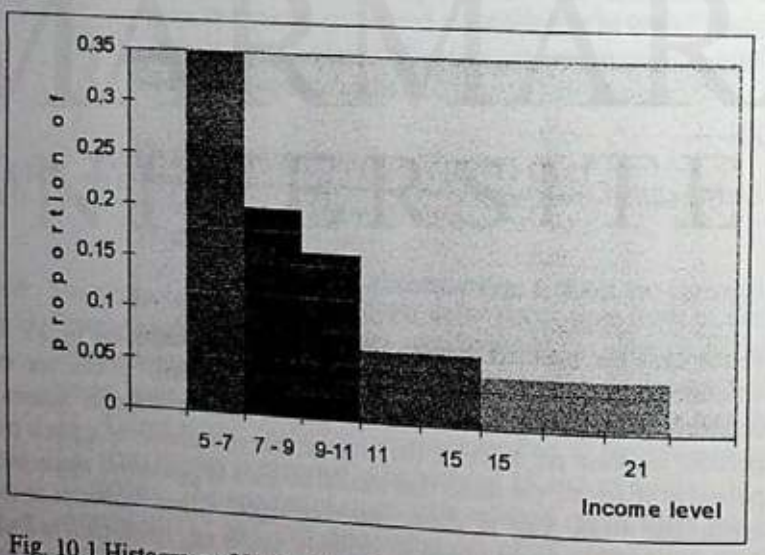


Fig. 10.1 Histogram of Income Level of thousand Households.

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - Construct the frequency distribution table. (Hint: heights of the rectangles is given because of the difficulty in reading them on graph, they are respectively 0.35, 0.20, 0.16, 0.07 and 0.05)

b - i)What percent of the household have an income level between 5 and 9 million? ii)How many household have an income level between 9 and 13 million?

c - Compute mean and standard deviation?

d- Compute first and third quartile and find interquartile range?

SOLUTION

a - Histogram shown in fig. 10.1 has unequal classes. Remember that in histograms frequencies of the classes are represented by areas not by heights. Therefore if the width of a class is different than the other classes, frequencies of those classes should be carefully interpreted.

Frequency (or relative frequency) = Base of the rectangle x Height of the rectangle.

On the basis of this rule, relative frequencies of the classes are as follows;

Table 10.1. Relative frequencies obtained from histogram (fig.10.1)

Households Incomes (in million TL)	Proportion of Households (p_i)	Number of households 1000x(p_i)	Cumulative frequency	Cumulative relative frequency
5 - 7	0.35	350	350	0.35
7 - 9	0.20	200	550	0.55
9 - 11	0.16	160	710	0.71
11 - 15	**0.14**	140	850	0.85
15 - 21	**0.15**	150	1000	1.00
Total	**1.00**	**1000**		

Note that base of the 1st, 2nd, 3rd classes are **2**
Base of the 4th class is **4** (that is twice of the previous classes) and
Base of the last class is **6** (that is three times the first three classes).

Therefore:

The height of the rectangle representing class "11 to less than 15" is multiplied by 2 to find out the relative frequency. (2 x 0.07 = 0.14)

The height of the rectangle representing class "15 to less than 21" is multiplied by 3 to find out the relative frequency. (3 x 0.05 = 0.15)

Total number of households that are surveyed are 1000. Therefore to find out the frequencies, relative frequency of each class is multiplied by 1000.

İSTANBUL TİCARET ÜNİVERSİTESİ
KÜTÜPHANESİ

MARMARA ÜNİVERSİTESİ YAYINEVİ

Attention: If the corresponding height of the unequal classes were directly assumed as the relative frequencies without making the above adjustment, it would be a mistake and the sum of the relative frequencies would not be equal to 1.

b - i) 55 % of the households' income are between 5 million and less than 9 million.

ii)	9 - 11		20 %	household. Out of 1000 it correspond	200
	11 - 13	14 % / 2 =	7 %	household. Out of 1000 it correspond	70
Total	9 - 13		27 %	household. Out of 1000 it correspond	**270**

c - $\bar{x} = \dfrac{(6x0.35)+(8x0.20)+(10x0.16)+(13x0.14)+(18x0.15)}{1} = 9.82$ million TL.

s = 4.151 million TL. (If you round the values during the computations the result that you obtain may have minor differences)

11 - 30 beverage producers' advertising expenditure as a percentage of gross sales revenue is given in the following table.

Table 11.1 Advertising expenditure as a percentage of gross sales revenue.

9, 10, 10, 11, 12, 12, 13, 13, 14, 15, 15, 15, 15, 17, 18,
19, 20, 20, 21, 22, 23, 24, 24, 25, 27, 27, 28, 29, 30, 31

a - Describe data set by a measure of central tendency and by a measure of dispersion.
b - Draw histogram.

SOLUTION

a - Arithmetic mean $= \dfrac{9+10+10+11+12+\ldots\ldots 27+28+29+30+31}{30} = 18.9667$

Advertising expenditure of 30 beverage producers is 18.97 % of gross sales revenue on the average.

$$S = \sqrt{\frac{(9-18.9667)^2+(10-18.9667)^2+\ldots\ldots+(30-18.9667)^2+(31-1.8.8667)^2}{30-1}}$$

$= 6.646147$

MARMARA ÜNİVERSİTESİ YAYINEVİ

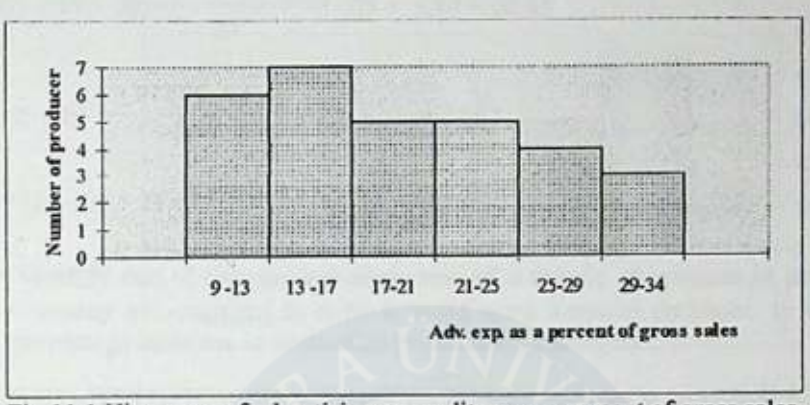


Fig 11.1 Histogram of advertising expenditure as a percent of gross sales.

12 - The following table shows the starting salaries of 50 recent graduates of business administration graduates. management

Salaries (in million TL)	No of graduates
55 - 65	23
65 - 75	15
75 - 85	6
85 - 95	4
95 - 105	2

a - Draw relative frequency histogram and show modal class

b - Draw "less than" and "or more" ogives. Show median on graph.

c - Compute mean and standard deviation

d - Compute interquartile range

SOLUTION

a - Table 12.1 - Relative and Cumulative Frequency Distributions Table

Salaries (in million TL.)	No. of graduates	Relative frequency	Cum. freq. (ascending)	Cum.freq. (descending)
55 - 65	23	0.46	23	50
65 - 75	15	0.30	38	27
75 - 85	6	0.12	44	12
85 - 95	4	0.08	48	6
95 - 105	2	0.04	50	2

MARMARA ÜNİVERSİTESİ YAYINEVİ

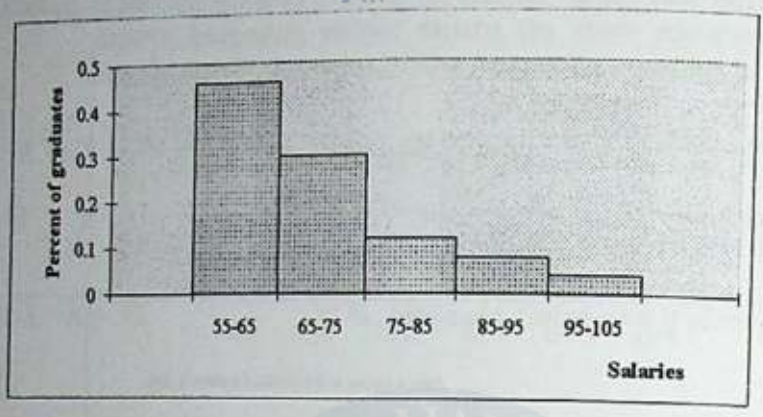


Fig. 12.1 Histogram of salaries of recent graduates

b -

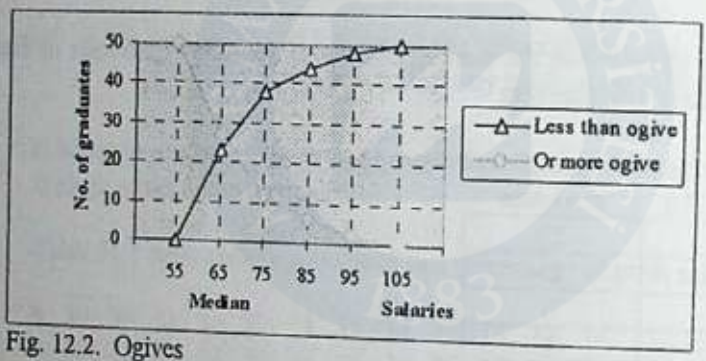


Fig. 12.2. Ogives

c -

Salaries (in million TL.)	f_i	m_i	$m_i \times f_i$	$(m_i - \bar{X})^2 f_i$
55 - 65	23	60	1380	88.36 x 23 = 2032.28
65 - 75	15	70	105	0.36 x 15 = 5.4
75 - 85	6	80	480	112.36 x 6 = 674.16
85 - 95	4	90	360	424.36 x 4 = 1697.44
95 - 105	2	100	200	936.36 x 2 = 1872.72
Total			**3470**	**6282**

1st Q. class ⇒ Median ⇒ and 3rd Q.class

Mean = $\bar{X} = \Sigma m_i f_i / (\Sigma f_i) = 3470 / 50 = 69.4$ million TL.

$$S = \sqrt{(\Sigma f_i (m_i - \bar{X}_i)^2) / n - 1} = \sqrt{6282 / 49} = \sqrt{128.2041}$$

$$S = \sqrt{128.2041} = 11.32272$$

MARMARA ÜNİVERSİTESİ YAYINEVİ

d - $Q_1 = 55 + 10\,\dfrac{(1(50/4)) - 0}{23} = 55 + 5.43 = 60.43$

$Q_3 = 65 + 10\,\dfrac{(3(50/4)) - 23}{15} = 65 + 9.67 = 74.67$

$IQR = Q_3 - Q_1 = 74.67 - 60.43 = 14.24.$

13 - Monthly rate of change in product costs of a sample of products in product mix of ABC Company are examined in order to make some financial decisions. In the following table, percentage increases in costs of 25 products are classified

Increase (%)	5 - 6	6 - 7	7 - 8	8 - 9	9 -10	10 -11
No of products	1	3	6	8	5	2

a - What is the average percent increase in costs of 25 products?

b - Compute and comment on standard deviation?

c - One of the products among these product mix is "bisçok". Rates of cost change of bisçok for three consecutive months are observed as 7 %, 15 % and 14 %. What is the monthly average rate of change in cost of bisçok?

SOLUTION

a - Arithmetic mean $= \dfrac{(5.5x1) + (6.5x3) + (7.5x6) + (8.5x8) + (9.5x5) + (10.5x2)}{25} = 8.26$

This is the average percent increase in cost of 25 products. That is; the average is "***per product***".

b - Standard deviation = 0.863752

c - The average rate of change in cost of product per month should be computed by geometric mean.

Geometric mean $= \sqrt[n]{x_1.x_2.x_3.\ldots\ldots x_n}$

Firstly, the rate of change at consecutive time periods should be expressed as a proportion and then 1 must added.

That is 7 % $\Rightarrow$ 107/100 = 0.07 $\Rightarrow$ 0.07 + 1 = 1.07
15 % $\Rightarrow$ 109/100 = 0.09 $\Rightarrow$ 0.09 + 1 = 1.09
14 % $\Rightarrow$ 114/100 = 0.14 $\Rightarrow$ 0.14 + 1 = 1.14

7 % change in cost between time periods t and t+1 can be inferred as follows cost of product in time t+1 is 1.07 times the cost of the product in time t. Similarly, in time t+2, cost of the product is 1.09 times the cost of the product in time t+1, in time t+3, cost of the product is 1.14 times the cost of the product in t+2.

Geometric mean of rate of change in cost of product is computed as follows.

$$GM = \sqrt[3]{1.07x1.09x1.14} = \sqrt[3]{1.40277} = 1.119426$$

Average percent change in cost of product is;

$1.119426 - 1 = 0.119426 \quad \Rightarrow 11.94\ \%$

Within 3 months, on the average, cost of bisçok is increased by 11.94 % per month. That is; the average is "***per month***".

14 - Human Resources Department of the ABC Company designs training programs to upgrade the supervisory skills of production line supervisors. As the programs are self administered, supervisors require different numbers of hours to complete these programs. Department manager has applied these programs in company's two different factories located in different places. From each factory 30 participants have attended. The number of the hours to complete the recently designed program applied to supervisors of factory A and factory B is given in the following tables respectively.

Table 14.1 Program completion time of 30 participants in factory A.

91, 120, 130, 136, 140, 147, 154, 160, 160, 166,
172, 176, 180, 180, 180, 187, 193, 195, 200, 206,
209, 210, 219, 233, 245, 249, 253, 277, 292, 328.

Table 14.2 Program completion time of 30 participants in factory B.

118, 154, 172, 179, 180, 192, 192, 200, 207, 208,
216, 225, 234, 242, 248, 251, 261, 267, 269, 270,
275, 280, 285, 290, 295, 295, 300, 310, 320, 355.

a - Which of the frequency polygons given in fig.14.1 and 14.2 correspond to the completion time of the supervisors in factory A and B?
b - For each of the factory; i) compute average completion time, ii) the completion time, above which 50 % of the observation lie.
c - In which factory, the proportion of the participants, who have completed the program less than the average value, is higher?
d - In which factory, the variation of completion time is higher?
e - If a participant in plant A completed the recent program in 180 hours and another participants in plant B in 234 hours, whose relative standing is higher?

MARMARA ÜNİVERSİTESİ YAYINEVİ

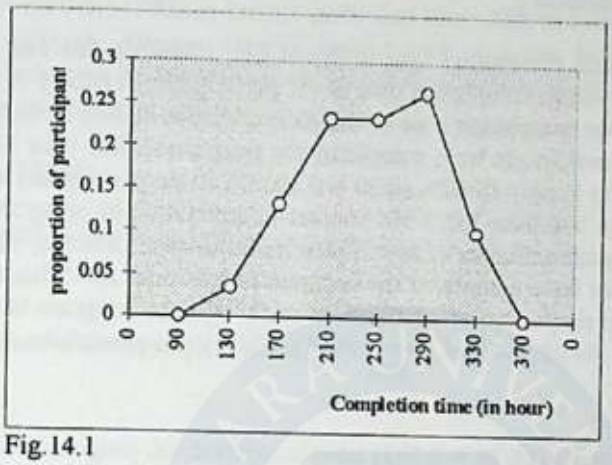


Fig.14.1

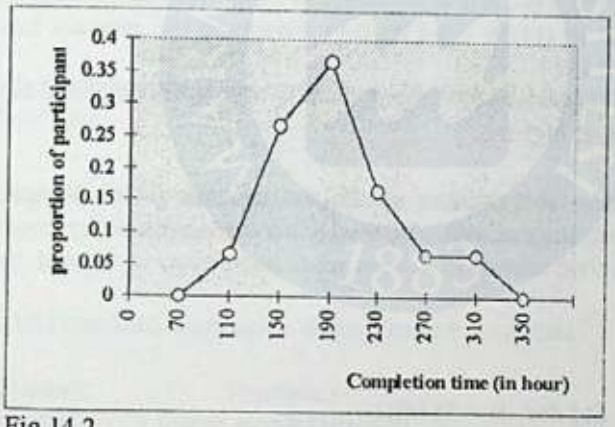


Fig.14.2

SOLUTION

a - Fig.14.1 belongs to data set given in Table 14.2. Fig.14.2 belongs to data set given in Table 14.1.

b - i) Mean completion time of the training program in factory A and B are as follows:

Mean $_A$ = 192.93 Mean $_B$ = 243

ii) The completion time above which 50 % of the observations lie is the median time.

Median $_A$ = (180 + 187) / 2 = 183.5 Median $_B$ = (248 + 251) / 2 = 249.5

In factory A, 50 % of the participants' completion time is less than 183.5.
In factory B, 50 % of the participants' completion time is less than 249.5.

Completion time of the program by the participants from two factories can be compared by their means and by their medians.

Attention: Note that shorter completion time is favourable. Based on this fact it can be inferred that average completion time of the participants in factory A is shorter than the average completion time of the participants in factory B. In factory A, 50 % of the participants have completed the program in less than 183.5 hour while in factory B only 16.67% (4/30 = 0.16667) of the participants completed the program in less than 183.5 %. Another interpretation by using the median value of completion time in factory B may be as follows. In factory B, 50 % of the participants have completed the program in less than 249.5 hour while in factory A, the proportion of participants completed the program in less than 249.5 hour is 86.67 %. (26 / 30 = 0.8667). That is this proportion is higher in factory A.

c - In factory A, 16 out of 30 (16/30 = 0.533) have completed the program less than the average completion time (192.93)

In factory B, 14 out of 30 (14/30 = 0.466) have completed the program less than the average completion time (243)

Hence; in plant A, the proportion of the participants, who have completed the program less than the average value, is higher.

d - Coefficient of variation is a measure of relative variation and it is expressing the standard deviation as a percentage of mean. That is; it is a unitless measure.

$$C.V. = \frac{\sigma}{\mu} \times 100$$

Std. dev. A = 52.8961 Std dev. B = 55.0041
Mean of A = 192.93 Mean of B = 243

$$\text{Coefficient of variation}_A = \frac{52.8961}{192.93} \times 100 = 27.4172\ \%$$

$$\text{Coefficient of variation}_B = \frac{55.0041}{243} \times 100 = 22.6354\ \%$$

Based on the coefficient of variations, it can be inferred that variation in completion time of the participants in factory A is more then the variation in B. Note that standard deviation of factory A is less than the standard deviation of factory B, but when their variations are compared with each other, relatively, variation in A is more than the variation in B.

MARMARA ÜNİVERSİTESİ YAYINEVİ

15 - A machine is filling tomato paste into three different cans. 250 gr. cans 500 gr. cans and 1 kg cans. Average amount poured into cans are 245 gr., 495 gr. and 995 gr. respectively and the standard deviation of the machine is 10 gr. In which of the cans the machine shows more variation relatively.

SOLUTION

$$\text{Coefficient of variation (250 gr. cans)} = \frac{10}{245} \times 100 = 4.1\ \%$$

$$\text{Coefficient of variation (500 gr. cans)} = \frac{10}{495} \times 100 = 2\ \%$$

$$\text{Coefficient of variation (1000 gr. cans)} = \frac{10}{995} \times 100 = 1\ \%$$

Relatively machine shows less variation while filling 1 kg cans when compared with 500 gr. and 250 gr cans.

16 - Demand distribution of a product during the reorder period is given in the following table. Reorder period is the period between the placement of the order and the receipt of the shipment. Frequency correspond to the number of reorder periods examined.

Table 16.1 Demand distribution during the reorder period.

Demand (in units)	Frequency (No of reorder periods)
150 - 180	20
180 - 210	15
210 - 240	30
240 - 270	20
270 - 300	10
300 - 330	5

a - If the inventory level is 180 at the beginning of the reorder period, in what proportion of the reorder periods the company will be unable to meet the total demand?

b - In what percent of reorder periods, demand is above 240

c - If the company wants to set the customer service level at 95 %. What should be the maximum amount of the inventory to be kept during the reorder period

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

a - If the inventory level is 180, demand up to 180 could be met. If the demand occurs more than 180 during the reorder period, inventory level would not be sufficient to meet total demand. The proportion of the reorder periods in which the demand is equal or greater than 180 can be obtained from relative cumulative frequencies in descending order.

Table 16.2 Cumulative relative distribution table.

Demand (in units)	Freq.	Cumulative (ascending)	relative freq. (descending)
150 - 180	20	0.20	1.00
180 - 210	15	0.35	0.80
210 - 240	30	0.65	0.65
240 - 270	20	0.85	0.35
270 - 300	10	0.95	0.15
300 - 330	5	1.00	0.05

That is; $(15 + 30 + 20 + 10 + 5) / 100 = 80 / 100 = 0.80$.

If the inventory level is 180 during the reorder period, and if demand occurs as more than 180 during 80 % of the reorder periods then, in 80 % of the times, demand of the customer in excess of 180 could not be met.

b - Similar to part a, percent of the reorder periods during which demand occurs more than 240 can be obtained from descending order cumulative relative frequencies.

That is $(20 + 10 + 15)/100 = 35/100 = 0.35$

In 35 % of the reorder periods demand is equal to 240 or more.

c -Customer service level is the percentage which shows the amount of the customers whose demand has been met. 95 % customer service level means that 95 % of customers' demand has been met by the safety stock kept. When ascending order cumulative frequencies are considered, it is observed that demand is up to 300 in 95 % of the times. That is; if the safety stock is kept as 300, in 95 % of the reorder periods customer demand will be met but in 5 % of the periods as demand will exceed 300 and as the safety stock will not be sufficient to meet the exceeding amount the demand, they will be in a shortage position. Thus, the maximum amount of the inventory to be kept should be 300 to set 95 % customer service level.

MARMARA ÜNİVERSİTESİ YAYINEVİ

17 - In the following table, annual data on number of Visa and EC/MC credit card for five years are given.

Table 17.1 Number of Visa and EC/MC credit cards

Years	No. of Visa credit cards	No. of EC/MC credit cards
1991	556,621	115,273
1992	802,573	97,462
1993	1,178,326	142,826
1994	1,301,020	166,112
1995	1,806,909	343,617

Source: BKM

a - Compute average rate of change in number Visa credit cards per year.
b - Compute average rate of change in number of EC/MC credit cards.

c - Compute average rate of change in total Visa and EC/MC credit cards.
d - Compare the average rate of change of each credit cards with each other and with the average rate of change of total cards.

SOLUTION

a - Geometric mean will be used here.

First step is to find out the rate of change between successive years. If 'r' shows the ratio of change, then r +1 is as follows.

Table 17.1 Computation of average rate of change in no. of Visa credit cards.

Years	No. of Visa C.	(r +1)
1991	556,621	---
1992	802,573	802,573 / 556,621 = 1.4419
1993	1,178,326	1,178,326 / 802,573 = 1.4682
1994	1,301,020	1,301,020 / 1,178,326 = 1.1041
1995	1,806,909	1,806,909 / 1,301,020 = 1.3888

⇒ No of credit cards in 1992 is 1.44 times the no. of credit cards in 1991

$GM = \sqrt[4]{1.44 x 1.47 x 1.10 x 1.39} = 1.3423$

Average rate of change = 1.3423 -1 = 0.3423 ⇒ 34.23 %

Between the years 1991 and 1995 number of Visa credit cards increased by 34.23 % on the average per year.

MARMARA ÜNİVERSİTESİ YAYINEVİ

b -

Table 17.2 Computation of average rate of change in no. of EC/MC credit cards.

Years	No. of EC/MC	(r +1)
1991	115,273	---
1992	97,462	97,462 / 115,273 = 0.8455
1993	142,826	142,826 / 97,462 = 1.4655
1994	166,112	166,112 / 142,826 = 1.1630
1995	343,617	343,617 / 166,112 = 2.0686

$GM = \sqrt[4]{0.85x1.47x1.16x2.07} = 1.3140$

Average rate of change = 1.3140 -1 = 0.3140 $\Rightarrow$ 31.40 %

Between the years 1991 and 1995 number of EC/MC credit cards increased by 31.40 % on the average per year.

c -

Table 17.4 Computation of average rate of change in total no. of Visa and EC/MC credit cards.

Years	VISA + EC/MC	(r +1)
1991	671,894	---
1992	900,035	900,035 / 671,894 = 1.34
1993	1,321,152	1,321,152 / 900,035 = 1.47
1994	1,467,132	1,467,132 / 1,321,152 = 1.11
1995	2,150,526	2,150,526 / 1,467,132 = 1.47

$GM = \sqrt[4]{1.34x1.47x1.11x1.47} = 1.326$

Average rate of change = 1.326 -1 = 0.326 $\Rightarrow$ 32.60 %

Between the years 1991 and 1995 number of Visa and EC/MC credit cards increased by 32.60 % on the average per year.

d - Comparison of the average rate of changes of Visa, EC/MC and sum of them between the years 1991 and 1995.

Visa	EC/MC	Total
34.23 %	31.40 %	32.60 %

Average rate of increase of Visa credit cards is more than the average rate of increase between 1991 and 1995.

MARMARA ÜNİVERSİTESİ YAYINEVİ

18 - Refer problem 17 and compute the average rate of changes of Visa, EC/MC and sum of Visa and EC/MC credit cards between 1992 and 1995.

SOLUTION

Table 18.1 Comparison of average rate of changes of Visa, EC/MC and sum of Visa and EC/MC.

Years	Growth rates of Visa	Growth rates of EC/MC	Growth rates of Visa and EC/MC
1993	1.4682	1.4655	1.4679
1994	1.1041	1.1630	1.1105
1995	1.3888	2.0686	1.4658
Geometric Mean	**1.3106**	**1.5220**	**1.3369**
Average rate of change	**31.06 %**	**52.20 %**	**33.69 %**

Average rate of increase in EC/MC is more than the average rate of change of Visa per year when the years 1992 to 1995 is considered. Examine the rate of change of EC/MC between the years 1991 and 1992 given in table 16.2. A decrease of 15.44 % (1 - 0.8455) between the years 1991 and 1992 caused a decrease in the average rate of change between the years 1991 and 1995. However when the years 1992 to 1995 is considered the decrease in 1992 with respect to 1991 would not be included and the average rate of change will come out as higher value when compared with Visa.

MARMARA ÜNİVERSİTESİ YAYINEVİ

PART II. Probability

The problems in this section are related with probability and the important concepts of probability. Types of the events like mutually exclusive, complementary, dependent and independent events are defined in the solutions and addition and multiplication rule are explained by the examples. Bayes rule is also covered and application of this rule to managerial and economics problems are considered.

19 - In the MBA program of a university, 200 students have completed the graduate courses and studying for their thesis. Cumulative grades of these students are evaluated and based on these grades they are rated as high honor, honor, satisfactory and unsatisfactory. It is reported that 20 of these students are high honor students, 50 are honor students, 90 are satisfactory and the rest is unsatisfactory. Besides this, the scores of the admission test which has been applied to these students before they were accepted to this program are examined and based on these scores, each student is included in one of the three categories which are classified as outstanding, good and moderate. The result is as follows, out of 20 high honor students, 7 of them had received an outstanding score from the admission test. Number of the students who had received outstanding score among 50 honor students is 10, among 90 satisfactory students, it is 12 and among 40 unsatisfactory students it is 5. Out of 20 high honor students , 8 of them had received a good score from the admission test. Number of the students who had received good score among 50 honor students is 28, among 90 satisfactory students, this number is 33 and among 40 unsatisfactory students it is 6.

a - Define all simple events and find the probabilities of these events.

b - Define all compound events and compute the probabilities of these events.

SOLUTION

Depending on the information provided in the question, the following table can be prepared. Total number of high honor, honor and satisfactory students are given as 20, 50 and 90 respectively. Number of unsatisfactory students can be obtained by subtracting these numbers from the total number (200) of students (200 - (20 + 50 + 90) = 40). Besides this, the result of admission test is given for each group of students classified as high honor, honor, satisfactory and unsatisfactory. The number of students in the third category (who have received moderate scores from the admission test) are obtained by subtracting the total number of students who have received outstanding and good scores from the total number of each of the high honor, honor, satisfactory and unsatisfactory groups of students. The number of students that are computed from the given values are given as bold italic in the following table.

Table 19.1 Contingency table that show the results of the admission test and MBA program graduation cumulative scores.

	High Honor	Honor	Satisfactory	Unsatisfactory	TOTAL
Outstanding	7	10	12	5	***34***
Good	8	28	33	6	***75***
Moderate	***5***	***12***	***45***	***29***	***91***
TOTAL	20	50	90	***40***	200

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - SIMPLE EVENTS and the corresponding probabilities in this example are as follows;

Simple events are the events that favor one outcome.

Being a high honor student	-	Event K.	P (K) = 20 / 200 = 0.10
Being a honor student	-	Event H	P (H) = 50 / 200 = 0.25
Being a satisfactory student	-	Event S	P (S) = 90 / 200 = 0.45
Being an unsatisfactory student	-	Event U	P (U) = 40 / 200 = 0.20

P (K) is read as "Probability that event K occurs"

Other simple events are;

Receiving an outstanding score	-	Event O.	P (O) = 34 / 200 = 0.17
Receiving a good score	-	Event G.	P (G) = 75 / 200 = 0.375
Receiving a moderate score	-	Event M.	P (M) = 91 / 200 = 0.455

P (K); P (H); P (S); P (U); P (O); P (G); P (M) are marginal probabilities

Probability that any one single event will occur is called **marginal probability**

Notice that; whatever the student's graduation score is, probability of selecting a student who had received an outstanding score from the admission test is a marginal probability. Or whatever the student's admission test score is, probability of selecting a student who had graduated as a honor student is a marginal probability. That is; probabilities at the margins of the contingency table are marginal probabilities.

Attention: Events can be classified as **mutually exclusive, collectively exhaustive and complementary** events.

Events K, H, S, U are ***mutually exclusive*** and ***collectively exhaustive*** events.

They are ***mutually exclusive*** because they don't have any common outcome, any two or more than two of those events can not occur together. In other words, a single student can not be a high honor, honor, satisfactory and unsatisfactory student at the same time.

They are ***collectively exhaustive*** because those events include every possible outcome. In other words, each of the 200 student in this program can be classified into one of these groups. It is not possible to observe a student who will not fall in any of these groups.

Events O, G and M are also ***mutually exclusive*** and ***collectively exhaustive events***.

However, events K and O are ***not mutually exclusive*** because both events can occur together. In other words a student can be a high honor student and at the same time she/he might have received an outstanding score on the admission test. Similarly events K and G; K and M; H and O; H and G; H and M; S and O; S and G; S and M; U and O; U and G; U and M are not mutually exclusive.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Complement of an event O is $\overline{O}$ (read as "not O")

Event $\overline{O}$ is observing students who have not received outstanding score. This event includes students whose scores are other than outstanding scores.

$P(\overline{O}) = 1 - P(O) = 1 - (34/200) = 176 / 200 = 0.83$

b) COMPOUND EVENTS that can be considered in this example, and the corresponding probabilities are as follows.

Compound events are formed by combining two or more events

Combining may be in two different ways which are called ***Intersection*** and ***Union***. Intersection ($\cap$)is the joint occurrence of two or more events. Union ($\cup$) is occurrence of any one of two or more simple events in addition to their joint occurrence.

Intersection combines events by **"and"**
Union combines events by **"or"**

i) Some examples of compound events composed by intersection (joint occurrence of two events) based on question 19 are as follows;

- Event O and K : Receiving an outstanding score in the admission test and being a high honor student. Seven students out of 200 students have received outstanding scores and they have become high honor students. P(O and K) = 7 / 200 = 0.035

- Event O and U : Receiving an outstanding score in the admission test and being an unsatisfactory student. Five students out of 200 students have received outstanding scores and then become unsatisfactory students. P(O and U) = 5 / 200 = 0.025

- Event M and U : Receiving a moderate score in the admission test and being an unsatisfactory student. Twentynine students out of 200 students have received moderate scores and then become unsatisfactory students. P(M and U) = 29 / 200 = 0.145

Probability of the occurrence of joint events like the events that are stated above are **joint probability.**

Attention: Mutually exclusive events can not be combined with an intersection as those events have no common outcome. (eg. Events O and G are mutually exclusive events. P(O and G) = 0 . Because a student receives either outstanding or good score but can not receive both scores from the same test at the same time. However as you see in the above examples about joint events, a student might have received an outstanding score in the admission test (event O) and then become a high honor student (event O and K) or an unsatisfactory student (event O and U).

MARMARA ÜNİVERSİTESİ YAYINEVİ

ii) Some examples of compound events composed by unions in question 19 are as follows;

- **If combined events are not mutually exclusive:**

Event K or O : Either being a high honor student or receiving an outstanding score in the admission test.

P(K or O) = (20 + 34 -7) / 200 = 0.235

Addition Rule

(Event K and event O are not mutually exclusive events)

$$P(K \cup O) = P(K) + P(O) - P(K \text{ and } O)$$

Event K considers 20 students, event O considers 34 students, however 7 of the students are considered both in event K and event O as they possess both of the attributes. When the total outcomes of event K and event O are added, those students who are high honor student and received outstanding score from the admission test (K and O), are counted two times. Therefore one of the counts are excluded by subtracting 7, to avoid double counting.

Event K or H or O ; P(K or H or O) = (20 + 50 + 34 - 7 - 10) / 200.
Either being a high honor student or honor student or receiving an outstanding score.

$$P(K \cup H \cup O) = P(K) + P(H) + P(O) - P(K \cap O) - P(H \cap O)$$

- **If the combined events are mutually exclusive:**

Event K or H : Either being a high honor or honor student.

P (K or H) = (20 + 50) / 200 = 0.35

According to addition rule;

$$P(K \cup H) = P(K) + P(H) - P(K \cap H)$$

Event K and event H are ***mutually exclusive events*** $P(K \cap H) = 0$ Therefore;

Addition Rule

(Event K and event O are mutually exclusive)

$$P(K \cup H) = P(K) + P(H)$$

Event K or S; Either being a high honor or satisfactory - P(K or S)= (20 + 90)/200 = 0.55

Event K or U; Either being a high honor or unsatisfactory- P (K or U)= (20 + 40)/200= 0.30

Applicant who received either good or moderate score - P (G or M) = (75 + 91)/200 = 0.83

Either being a high honor or honor or satisfactory student - Event K or H or S;
P (K or H or S) = (20 + 50 + 90) / 200 = 0.80

Attention: When mutually exclusive events are combined by union, there is no need to subtract the probability of joint occurrence as those events have no common outcome.

20 - Refer problem 19.

a - What proportion of high honor students have received outstanding scores?

b -What proportion of students who have received moderate scores are honor students?

SOLUTION

a - P (O / K) Read as probability of occurrence of event O given that event K has occurred

P (O / K) = 7 / 20 Proportion of students who have received outstanding scores among high honor students.

There are 20 students which favors event K (high honor students). Among 20 high honor students, 7 of them had also received an outstanding score. This probability can be interpreted as follows. Suppose that a student is selected among 200 students and it is observed that she/he is a high honor student (one of the 20 high honor students), and now we are interested whether this specific student had received an outstanding score or not from the admission test. As 7 of the high honor students had received an outstanding score from the admission test, the probability is 7 over 20.

If we are interested in the probability that a selected high honor student had not received and outstanding score, then it is computed as follows;

$P(\overline{O} / K) = (20 - 7) / 20 = 13 / 20.$

Attention: First of all, the total number of outcomes should be considered in P (O / K). Here the total number of outcomes is 20 not 200. Because it is given that the selected student is a high honor student. Secondly we are interested mainly in the proportion of the student who had received an outstanding score given that the student is a high honor student - that is among the high honor students. (It is 7 in 20).

Proportion of the students who had received outstanding score from the admission test is different among the high honor, honor, satisfactory and unsatisfactory students.

MARMARA ÜNİVERSİTESİ YAYINEVİ

It is 7 / 20 among high honor students. P (O / K) = 7 / 20
It is 10 / 50 among honor students. P (O / H) = 10 / 50
It is 12 / 90 among satisfactory students. P (O / S) = 12 / 90
It is 5 / 40 among unsatisfactory students. P (O / U) = 5 / 40

Probabilities of event "O" conditional to the occurrences of events K, H, S, and U are called **conditional probabilities.**

Attention: P (O / K) is not equal to P (O). That is; probability of observing an outstanding score is different from the probability of observing an outstanding score given that the student is high honor student. However if we are interested in the probability of observing a high honor student who had also received an outstanding score among all the students (event O and K), that is the joint probability and it is equal to 7 /200.

As a result, following relation exist between **conditional, joint** and **marginal probabilities**

$$P(O / K) = \frac{P(O \text{ and } K)}{P(K)} = \frac{7 / 200}{20 / 200} = \frac{7}{20}$$

b - P (H / M) = 12 / 91 Proportion of honor students who have received moderate scores from admission test .

$$P(H / M) = \frac{P(H \text{ and } M)}{P(M)} = \frac{12 / 200}{91 / 200} = \frac{12}{91}$$

21 - A production line operates effectively in 90 % of the time (in 10 % of the time it operates ineffectively). When it is operating effectively, it produces 95 % acceptable parts and 5 % defective parts. When it is operating ineffectively, it produces 30 % acceptable parts and 70 % defective parts. A part is selected from this production line and observed that it was a defective part.

a - What is the probability that, production line is operating effectively?
b - What is the probability that, production line is operating ineffectively?

SOLUTION

First step in solving probability problem is the identification of the events and the probabilities that are given corresponding to the events.

Event E is production line operating effectively - Event E; P (E) = 0.90
Event I is production line operating ineffectively - Event I; P (I) = 0.10

MARMARA ÜNİVERSİTESİ YAYINEVİ

Event A is producing acceptable parts - Event A
Event D is producing defective parts - Event D

DEPENDENT EVENTS and **conditional probabilities** considered in this example are as follows;

Events of producing acceptable or defective parts are dependent on whether production line operates effectively or not.

> Two events are said to be **dependent**;
> if the probability of occurrence of one of the event is affected by the occurrence of the other event.
> Otherwise they are said to be **independent**

Events A and E are dependent
Events A and I are dependent
Events D and E are dependent
Events D and I are dependent

Pay attention that, as the probability of the production line producing acceptable parts and defective parts are effected by the state of the machine operating effectively or ineffectively, those probabilities are given as conditional probabilities.

- producing acceptable parts given that the machine operates effectively; $P(A / E) = 0.95$
- producing defective parts given that the machine operates effectively; $P(D / E) = 0.05$
- producing acceptable parts given that the machine operates ineffectively; $P(A / I) = 0.30$
- producing defective parts given that the machine operates ineffectively; $P(D / I) = 0.70$

Attention :

i) $P (A / E)$ is basically the probability of event A.(After event E has occurred) It should not be confused with P (A and E). P (A and E) is the probability of joint occurrence of event A and event E.

ii) Probabilities of the events which add up to 1 are both the first pair and second pair of events stated above. $P(A / E) + P (D / E) = 1$; and $P(A / I) + P(D / I) = 1$. In conditional probabilities, the probabilities of all the possible outcomes occurring after an occurrence of an event adds up to 1. That is production line either operates effectively or ineffectively. After occurrence of any of these events, there are two possible outcomes, they produce either acceptable or defective products.

BAYES THEOREM is applied in solving this problem. The importance of by Bayes rule is that the probability of an event named as prior probability is revised when an additional information is available. The following formula of Bayes rule is stated according to this problem.

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - $$P(E/D) = \frac{P(D/E)P(E)}{P(D/E)P(E) + P(D/I)P(I)}$$

$$P(E/D) = \frac{(0.05)(0.90)}{(0.05)(0.90)+(0.70)(0.10)} = \frac{0.045}{0.045+0.07} = 0.3913$$

b - $$P(I/D) = \frac{P(D/I)P(I)}{P(D/I)P(I) + P(D/E)P(E)}$$

$$P(I/D) = \frac{(0.70)\ (0.10)}{(0.70)\ (0.10)+(0.05)\ (0.90)} = \frac{0.07}{0.07+0.045} = 0.6087$$

OR

$$P(I/D) = 1 - P(E/D) = 1 - 0.3913 = 0.6087$$

***Interpretation*:** Suppose that; the machine, whether operates effectively or ineffectively, will be controlled by checking the items produced by that machine. A part is selected among the produced items and controlled whether defective or not. The result of the control is an additional information which leads to revise the probability whether machine operates effectively or ineffectively. In this problem, selected part is defective and depending on this information, probability that the machine operates ineffectively is revised as 0.6087 which was 0.10 prior to the information obtained.

22 - Refer problem 21. If the selected part from the production line is good:

a - What is the probability that, production line operates effectively?

b - What is the probability that, production line operates ineffectively?

SOLUTION

a - $$P(E/A) = \frac{(0.95)\ (0.90)}{(0.95)\ (0.90)+(0.30)\ (0.10)} = \frac{0.855}{0.855+0.03} = 0.9661$$

b - $$P(I/A) = \frac{(0.30)\ (0.10)}{(\ 0.30)\ (0.10)\ +\ (0.95)\ (0.90)} = \frac{0.03}{0.03+0.855} = 0.0339$$

23 - In a multinational company, 66 % of the executives are Business Administration graduates. There are total of 400 employee in this company, 100 of them are executives and 150 of them are Business Administration graduates.

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - What is the probability that, an employee is an executive given that he/she is a Business Administration graduate?

b - What is the probability that a personnel is not an executive given that he/she is a Business Administration graduate?

SOLUTION

Events and the probabilities of those events are follows;

Event B is Business Administration graduates $P(B) = 150 / 400 = 0.375$

Event E is Executives $P(E) = 100 / 400 = 0.250$

Business Administration graduates among executives $P(B/E) = 0.66$

Executives among Business Administration graduates $P(E/B)$

$$\text{a - } P(E/B) = \frac{P(E \text{ and } B)}{P(B)} = \frac{P(B/E)\,P(B)}{P(B)} = \frac{(0.66)(0.25)}{0.375} = 0.44$$

b - Probability that a personnel is not an executive given that he/she is a Business Administration graduate is asked. ie. $P(\overline{E} / B)$. In part (a), it is found that among Business Administration graduates 44 % are executives. Then the rest 56 % are not executives. That is:

$$P(\overline{E} / B) = 1 - 0.44 = 0.56$$

24 - Refer problem 23. What is the probability that a personnel is not an executive and not a Business Administration graduate ?

SOLUTION

The required probability that a personnel is not an executive and not a Business Administration graduate is one of the four possible outcomes in the sample space is:

$\overline{E}$ and $\overline{B}$; not executive and not Business administration graduate. The others are;

$\overline{E}$ and B ; not executive but Business Administration graduate.
E and $\overline{B}$; executive but not Business Administration graduate.
E and B ; both executive and Business Administration graduate.

MARMARA ÜNİVERSİTESİ YAYINEVİ

From the probabilities given in the problem, the following probabilities can be computed.

$P(E\,B) = (0.66)(0.25) = 0.165 \Rightarrow P(E\,B) = 0.165$

$P(E) = P(E\,B) + P(E\,\overline{B})$

$P(E\,\overline{B}) = P(E) - P(E\,B) = 0.250 - 0.165 = 0.085$

$P(B) = P(B\,E) + P(B\,\overline{E})$

$P(\overline{E}\,B) = P(B) - P(B\,E) = 0.375 - 0.165 = 0.210$

$P(\overline{B}) = P(E\,\overline{B}) + P(\overline{E}\,\overline{B})$

$P(\overline{E}\,\overline{B}) = P(\overline{B}) - P(E\,\overline{B}) = 0.625 - 0.085 = 0.540$

Alternative solution is preparing a table which shows all possible outcomes and the number of occurrences in sample space.

Table 24.1 Contingency table of executives and Business Administration graduates

	Bus. Adm. Graduates	Not Bus.Adm Graduates	Total
Executives	**0.66**x100= 66	**100** - 66 34	100
Not Executives	**150** - 66 = 84	300-84 = 216	300
Total	**150**	250	**400**

	Event B	Event $\overline{B}$
Event E	EB	E$\overline{B}$
Event $\overline{E}$	$\overline{E}$B	$\overline{E}\,\overline{B}$

25 - Three similar machines are producing the same item and all the items enter the last process together. For that reason it becomes impossible to understand from which machine the defective item is originated. 40 % of all the items are produced by machine A, 45 % by machine B and the rest is produced by machine C. It is known from previous productions that, 2 % of the items produced by machine A is defective. Defective proportions for machines B and C are 3 % and 4 % respectively. If a defective item is observed, What are the probabilities that it is produced by **a)** machine A, **b)** machine B, **c)** machine C.

SOLUTION

Event D is occurrence of defective item
Event A, B and C are the productions of the items by machines A, B and C respectively.

$P(A) = 0.40$ $\quad P(D/A) = 0.02$
$P(B) = 0.45$ $\quad P(D/B) = 0.03$
$P(C) = 0.15$ $\quad P(D/C) = 0.04$

İSTANBUL TİCARET ÜNİVERSİTESİ KÜTÜPHANESİ

MARMARA ÜNİVERSİTESİ YAYINEVİ

a - $$P(A/D) = \frac{P(D/A)\ P(A)}{P(D/A)\ P(A) + P(D/B)\ P(B) + P(D/C)\ P(C)}$$

$$= \frac{(0.02)(0.40)}{(0.02)(0.40)+(0.03)(0.45)+(0.04)(0.15)} = \frac{0.008}{0.008+0.0135+0.006} = 0.29091$$

b - $$P(B/D) = \frac{P(D/B)\ P(B)}{P(D/A)\ P(A) + P(D/B)\ P(B) + P(D/C)\ P(C)}$$

$$= \frac{0.0135}{0.008+0.0135+0.006} = 0.49091$$

c - $$P(C/D) = \frac{P(D/C)\ P(C)}{P(D/A)\ P(A) + P(D/B)\ P(B) + P(D/C)\ P(C)}$$

$$= \frac{0.006}{0.008+0.0135+0.006} = 0.0275$$

26 - In credit card payments, card holders could either pay back the total amount of their credit card debts (referred as payment in full) or they can just pay the stated minimum amount on each month and interest is charged for the rest of the amount owed (referred as instalment payment). In a bank, product manager of credit cards intends to assess the proportion of the cardholders who will prefer payment in full, among the group of card holders classified as highest income group. It is already known that among all the cardholders 40 percent prefer payment in full, while the rest prefers instalment payment. Application forms of the cardholders are further examined in detail and it is obtained that, 90 percent of cardholders who prefer payment in full are in the highest income group. Proportion of this group (the highest income group) is only 5 percent among the cardholders who makes instalment payment. What is the probability which the product manager wants to assess?

SOLUTION

Events and the probabilities corresponding to those are as follows:

Event F ; payment in full. $P(F) = 0.40$
Event P ; instalment payment. $P(P) = 0.60$
Event H ; the highest income group. $P(H)$ is not given directly but proportions are stated dependent on payment preferences as follows:

MARMARA ÜNİVERSİTESİ YAYINEVİ

$P(H/F) = 0.90$ 90 percent of cardholders who prefer payment in full are in the highest income group.

$P(\overline{H}/F) = 0.10$

$P(H/P) = 0.05$ Proportion of the highest income group is 5 percent among the cardholders who make instalment payment.

$P(\overline{H}/P) = 0.95$

Event $\overline{H}$ (not "H"); cardholders other than the highest income group.

The manager wants to assess the probability of the cardholder who prefer payment in full, among the group of card holders classified as highest income group.

$$P(F/H) = \frac{P(H/F)P(F)}{P(H/F)P(F) + P(H/P)P(P)}$$

$$P(F/H) = \frac{(0.90)\ (0.40)}{(0.90)\ (0.40) + (0.05)\ (0.60)} = \frac{0.36}{0.36+0.03} = 0.92307$$

Attention : This problem is a typical application of Bayes rule to a management problem. The manager of the bank is interested in the proportion of the cardholders who make payment in full, among the groups of card holders classified according to their income level. The reason of this interest is as follows; one of the sources of banks' revenue is the interest charged to the debts of the cardholders who make instalment payment. However in case of full payment in time, no interest is charged to debts of the cardholders and this will not contribute to the interest revenue of the bank.

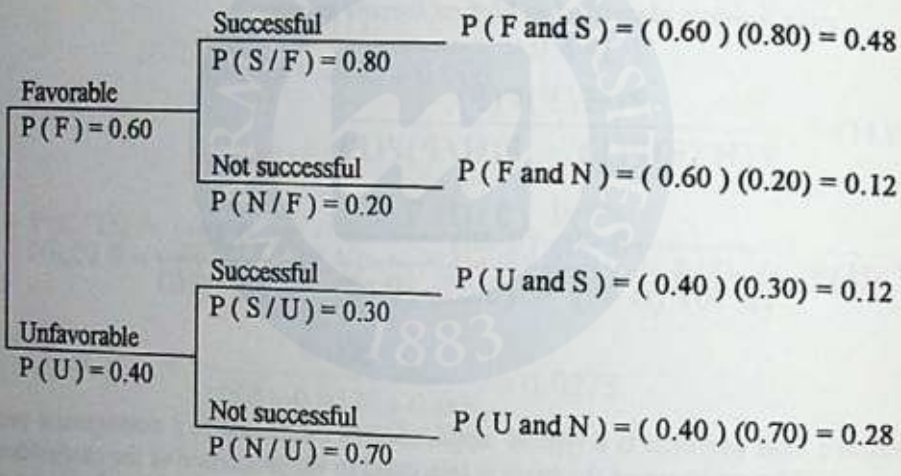
In this problem only the proportion of the high income group cardholders who make payment in full is questioned. In addition to this, proportion of high income group of cardholders who make instalment payment or proportion of the cardholders other than the high income group who make payment in full or instalment payment can be considered.

From the manager's point of view, what is important is to obtain that piece of information (that is; $P(F/H)$) in advance to reduce the uncertainty about the future revenue of the bank that will originate from the interest rates of the debts of the cardholders. They achieve this purpose by examining the records of the cardholders who make payment in full and examine the income level of those cardholders. In this example, cardholders who make payment in full are examined and it is observed that proportion of cardholders who are in high income group among the cardholders preferring payment in full is 90 %. $P(H/F)$ Besides this, proportion of high income group cardholders ($P(H)$) is also available.

MARMARA ÜNİVERSİTESİ YAYINEVİ

As a result by using bayes rule, From the probabilities P (H / F), P (H), P (H / P), and P (P); proportion of cardholders who will make payment in full among the high income group can be determined which will help the manager throughout the budgeting process.

27 - A product manager of Liver, manufacturing wide range of products in food and chemical industry wishes to determine whether or not to market a new detergent. Success criteria is the predetermined sales level. Before introducing the product, they intend to carry on a customer testing. Result of the test may be favorable or unfavorable. Probability assessed for the favorable test result is 60 %. If the result of the test is favorable, probability that a successful sales level will be achieved is 80 %. But for an unfavorable test result, the probability of success is assessed to be only 30%. Construct the probability tree diagram and show the probabilities.



P (F) = 0.60 Probability that the test result is favorable (Marginal Probability)
P (U) = 0.40 Probability that the test result is un favorable (Marginal Probability)

P (S / F) = 0.80 Probability of achieving a successful sales level given that the test result is favorable. (Conditional Probability)

P (N / F) = 0.20 Probability of unsuccessful sales level given that the test result is favorable. (Conditional Probability)

P (S / U) = 0.30 Probability of achieving a successful sales level given that the test result is unfavorable. (Conditional Probability)

P (N / U) = 0.70 Probability of unsuccessful sales level given that the test result is unfavorable. (Conditional Probability)

P (F and S) = 0.48 Probability of favorable test result and successful sales level (Joint Probability)

P (F and N) =0.12 Probability of favorable test result and unsuccessful sales level (Joint Probability)

P (U and S) = 0.12 Probability of unfavorable test result and successful sales level (Joint Probability)

P (U and N) = 0.28 Probability of unfavorable test result and unsuccessful sales level (Joint Probability)

MARMARA ÜNİVERSİTESİ YAYINEVİ

PART III. Probability Distributions

The problems in this section are related with the probability distributions. Discrete and continuous random variables are defined in the solutions and the discrete and continuous probability distributions that are binomial, Poisson, hypergeometric and normal distributions are discussed. Problems related Poisson and normal approximation to the binomial are also solved.

28 - Probability distribution of number of traffic accidents at an intersection during a week period is given on the following table.

Table 28.1 Number of accidents and probabilities

No. of accidents (X)	Prob. P(X)
0	0.35
1	0.43
2	0.12
3	0.06
4	0.04

a - Define random variable considered in this example.

b - What is the probability of occurrence of 2 accidents?

c - What is the probability of occurrence of less than or equal to 2 accidents?

d - What is the probability of occurrence of more than 2 accidents?

e - What is the expected number of accidents, variance and standard deviation of accidents?

SOLUTION

a - Random variable in this problem is the "number of accidents". It is a discrete random variable. It assumes values like 0, 1, 2, 3, 4.

> Random variables that take only distinct values are **discrete random variables**

> Random variables that take on any value on a continuum are **continuous random variables**

b - Probability of occurrence of 2 accidents is 0.12. $P(X = 2) = 0.12$

> **Probability distributions** provide probabilities corresponding to a value or range of values which the random variable assumes, by an algebraic formula, a table or a graph.

MARMARA ÜNİVERSİTESİ YAYINEVİ

c - $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.35 + 0.43 + 0.12 = 0.90$

Random variable assuming values less than or equal to 2 means that it takes values of 0, 1, and 2 as the random variable is discrete. Notice that, to find out the probability of events that 0 or 1 or 2 accidents, **addition rule** of the probability is applied. As three of the events are **mutually exclusive**, individual probabilities of occurrence of 0, 1 and 2 accidents are added to find out the required probability.

d - $P(X>2) = P(X=3) + P(X=4) = 0.06 + 0.04 = 0.10$

Alternative solution is ;

$P(X>2) = 1 - P(X \leq 2) = 1 - 0.90$ Events of greater than 2 accidents and less than or equal to 2 accidents are complementary events.

e - Table 28.2 Computation of expected number, variance and standard deviation of accidents

No. of accidents (x)	$P(x)$	$x \cdot P(x)$	$[x - E(x)]^2$	$[x - E(x)]^2 P(x)$
0	0.35	0	$(0 - 1.01)^2 = 1.0201$	$(1.0201)(0.35) = 0.357035$
1	0.43	0.43	$(1 - 1.01)^2 = 0.0001$	$(0.0001)(0.43) = 0.000043$
2	0.12	0.24	$(2 - 1.01)^2 = 0.9801$	$(0.9801)(0.12) = 0.117612$
3	0.06	0.18	$(3 - 1.01)^2 = 3.9601$	$(3.9601)(0.06) = 0.237606$
4	0.04	0.16	$(4 - 1.01)^2 = 8.9401$	$(8.9401)(0.04) = 0.357604$

$E(x) = \Sigma x \cdot P(x) = 1.01$ $V(x) = \Sigma P(x)[x - E(x)]^2 = 1.0699$

Standard Deviation $= \sqrt{1.0699} = 1.0343597$

29 - It is known that 10 % of the items produced by a machine is defective. A random sample of 10 items is selected among the products produced by this machine.

a - What is the probability of observing one defective item?

b - What is the probability observing two defective items?

c - What is the probability that none of the selected products will be defective?

d - What is the probability that less than 3 of the selected products will be defective?

e -What is the probability that less than or equal to 3 of the selected products will be defective?

f - What is the probability that greater than 4 of the selected products will be defective?

g -What is the probability that greater than or equal to 3 of the selected products will be defective?

MARMARA ÜNİVERSİTESİ YAYINEVİ

Probability distribution which will be used to find out the required probabilities in this question is binomial probability distribution. **Binomial probability distribution** is a probability distribution for discrete random variables that has the following formula.

$$P(X=r)=\binom{n}{r}\ p^r\ q^{n-r}\ =\left(\frac{n!}{n!\ (n-r)!}\right)\ p^r\ q^{n-r}$$

Above algebraic representation of binomial probability distribution is applied to find out the probability of specific number of occurrences (**r** occurrences) in fixed number of trials (**n** trials) given a constant probability of success (**p**) for each trial. Trials are named as Bernouilli trials. Characteristics of bernouilli trials are as follows:

1. There are n independent trials.
2. Each trial has **two** basic outcomes named as ***success*** or ***failure*** (success does not necessarily represent the favorable event, it is just to distinguish the outcomes in a trial.
3. Probability of success (p) is constant from trial to trial.
4. The main interest is the probability of observing specific number of success (n) or failure (n - r) in n trials.

In this example characteristics of bernouilli trials exist.

1. There are 10 trials. (n = 10) in this example. Here, trial is selecting and checking whether the item is defective or not. These trials are independent, that is; probability of the second item being defective is not affected from the first item being either defective or good. This assumption is valid if the selection is performed with replacement or can be assumed that it is so.

2. In each trial, the items checked are either defective or good. In other words there are two basic outcomes. Observing a defective item is one of two basic outcomes and could be named as success. Observing a good item is the other basic outcome and could be named as failure.

3. Probability of success - observing a defective item in each trial in this example - is 10 %. (p = 0.10) It is known that 10 % of the items produced by this machine is defective. Therefore each item selected has the same chance of being a defective item and that chance is 10 %. Selection is assumed to be made with replacement.

4. We are interested in the probability of observing specific number of successes (zero, one, two, less than three defective items) in 10 trials.

On the basis of above explanations probabilities of specific number of occurrences are found by both using the binomial probability formula and binomial cumulative probability table.

a - $P(X=1)\ =\binom{10}{1}\ 0.10^1\ \ 0.90^9\ \ =0.3874$

MARMARA ÜNİVERSİTESİ YAYINEVİ

From binomial cumulative probability table;

$P(X=1) = P(X \leq 1) - P(X=0) = 0.7361 - 0.3487 = 0.3874$

b - $P(X=2) = \binom{10}{2} 0.10^2 \ 0.90^8 = 0.1937$

From binomial cumulative probability table;

$P(X=2) = P(X \leq 2) - P(X \leq 1) = 0.9298 - 0.7361 = 0.1937$

c - $P(X=0) = \binom{10}{0} 0.10^0 \ 0.90^{10} = 0.3487$

From binomial cumulative probability table; $P(X=0) = 0.3487$

d - $P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$$= \binom{10}{0} 0.10^0 \ 0.90^{10} + \binom{10}{1} 0.10^1 \ 0.90^9 + \binom{10}{2} 0.10^2 \ 0.90^8$$

$$= 0.3487 + 0.3874 + 0.1937 = 0.9298$$

From binomial cumulative probability table;

$P(X<3) = P(X \leq 2) = 0.9298$

e - $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$ or

$= P(X<3) + P(X=3)$ In part c $P(X<3)$ is calculated as 0.9298.

$= 0.9298 + 0.0574 = 0.9872$

From binomial cumulative probability table; $P(X \leq 3) = 0.9872$

f - $P(X>4) = P(X=5) + P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$

or simply $= 1 - P(X \leq 3) = 1 - 0.9872 = 0.0128$

g - $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) + P(X=6) + P(X=7) +$
$P(X=8) + P(X=9) + P(X=10)$

or simply $= 1 - P(X<3) = 1 - P(X \leq 2) = 1 - 0.9298 = 0.0702$

MARMARA ÜNİVERSİTESİ YAYINEVİ

Attention: Binomial probability distribution is a discrete probability distribution. Random variable assumes discrete values. If x being less than 3 is considered, it means x assumes values 2, 1, or 0 or stated in other words, x assumes values less than or equal to 2.

30 - If the success probability is 50 % , in 8 trials what is the expected number and variance of success?

SOLUTION

Table 30.1 Computation of expected number and variance in binomial probability distributions.

No. of success (x)	$P(x)$*	$x \cdot P(x)$	$[x - E(x)]^2$	$[x - E(x)]^2 P(x)$ *
0	0.0039	0	$(0 - 4)^2 = 16$	(16) (0.0039) = 0.0624
1	0.0312	0.0312	$(1 - 4)^2 = 9$	(9) (0.0312) = 0.2808
2	0.1094	0.2188	$(2 - 4)^2 = 4$	(4) (0.1094) = 0.4376
3	0.2188	0.6564	$(3 - 4)^2 = 1$	(1) (0.2188) = 0.2188
4	0.2734	1.0936	$(4 - 4)^2 = 0$	(0) (0.2734) = 0
5	0.2188	1.094	$(5 - 4)^2 = 1$	(1) (0.2188) = 0.2188
6	0.1094	0.6564	$(6 - 4)^2 = 4$	(4) (0.1094) = 0.4376
7	0.0312	0.2184	$(7 - 4)^2 = 9$	(9) (0.0312) = 0.2808
8	0.0039	0.0312	$(8 - 4)^2 = 16$	(16) (0.0039) = 0.0624

*2nd column of the table is rounded up to 4 digits after decimal point. For that reason sum of the last column comes out to be 1.9992 instead of 2.

$$E(x) = \sum x \cdot P(x) = 4 \qquad V(x) = \sum [x - E(x)]^2 P(x) = 2$$

$$\text{Standard Deviation} = \sqrt{2} = 1.4142135$$

Expected value and variance in binomial probability distribution are as follows:

$$E(x) = n \cdot p \qquad V(x) = n \cdot p \cdot q$$

$$E(x) = 8 \times 0.50 = 4 \qquad V(x) = 8 \times 0.5 \times 0.5 = 2$$

31 - Market share of a toothpaste brand in toothpaste market is estimated as 32 %.

a - If '**n**' consumers are questioned about the brand of the toothpaste they prefer, how many of them are expected to prefer that toothpaste brand. What is the variance?

b - If sample size is 125, how many of them are expected to prefer that toothpaste brand.

SOLUTION

a - $E(x) = n \times 0.32$ $V(x) = n \times 0.32 \times 0.68$

b - $E(x) = 125 \times 0.32 = 40$ Out of 125 questioned consumers 40 are expected to prefer that company's toothpaste.

32 - Probability that a sales representative of a credit card center of a bank that signs contract with merchants for collecting the credit card sales slips is 0.20 On a given day, what is the probability of signing '0' to 'n' contracts;

a - If she visits 5 merchants on a given day.

b - If she visits 7 merchants on a given day.

c - If the probability is 0.15 and if she visits 5 merchants on a given day.

d - What is the most probable number of merchants that sales representative signs contracts for each of the cases given above?

e - Compute the expected number and standard deviation of number of merchants whom contracts are signed with.

SOLUTION

Answers for part a, b and c are given in the following table;

Table 32.1 Binomial probabilities for different number of trials (**n**) and probability of success (**p** values)

	Part (**a**)	Part (**c**)
No. of merchants (x)	$n = 5$ $p = 0.20$ $P(x)$	$n = 5$ $p = 0.15$ $P(x)$
0	0.3277	0.4437
1	0.4096	0.3915
2	0.2048	0.1382
3	0.0512	0.0244
4	0.0064	0.0022
5	0.0003	0.0001

Part (**b**)	
No. of merchants (x)	$n = 7$ $p = 0.20$ $P(x)$
0	0.2097
1	0.3670
2	0.2753
3	0.1147
4	0.0287
5	0.0043
6	0.0004
7	0.0000

d - The most probable number of merchants that sales representative signs contracts are found by examining the above table 32.1.

In a, where $n = 5$ $p = 0.20$ the most probable number of merchant is 1. $P(x = 1) = 0.4096$

In b, where n = 7 p = 0.20 the most probable number of merchant is 1. P (x =1) = 0.3670
In c, where n = 5 p = 0.15 the most probable number of merchant is 0. P (x = 0) = 0.4437

e - n = 5 p = 0.20 E (x) = 5 x 0.20 = 1 Std. dev. = $\sqrt{5 \times 0.20 \times 0.80}$ = 0.8944

n = 7 p = 0.20 E (x) = 7 x 0.20 = 1.4 Std. dev. = $\sqrt{7 \times 0.20 \times 0.80}$ = 1.0583

n = 5 p = 0.15 E (x) = 5 x 0.15 = 0.75 Std. dev. = $\sqrt{5 \times 0.15 \times 0.85}$ = 0.7984

Attention: Parameters of binomial probability distribution are n (number of trials) and p (success probability). As seen in above example, for each pair of n and p different probability distributions are obtained.

33 - A company receives a shipment of hundred items. A random sample of four items is selected, and the shipment is rejected if any of these items proves to be unacceptable.

a - What is the probability of accepting a shipment containing four unacceptable items?
b - What is the probability of accepting a shipment containing two unacceptable items?
c - What is the probability of rejecting a shipment containing one unacceptable item?

SOLUTION

In this case 100 items are shipped but acceptance or rejection decision will be made on the basis of the 4 items which are selected from 100 items. Therefore n (number of trials) is 4 in this case. Each of the four items are controlled whether they are acceptable or not and if any of these items prove to be defective (in other words if number of defective items is equal to or more than 1) then all the shipment (100 items) is rejected.

Random variable x is defined as number of unacceptable items.

Decision rule: Accept the shipment if $x = 0$
Reject the shipment if $x \geq 1$

Proportion of the unacceptable items is uncertain for the buyer in advance and possible defective proportions are given as states of nature in parts a, b, and c.

a - $p = \frac{4}{100} = 0.04$ $n = 4$ $P(x=0) = \binom{4}{0} 0.04^0 \ 0.96^4 = 1 \,.\, 1 \,.\, 0.8493 = 0.8493$

If shipment contains 4 unacceptable items, probability of **accepting** the shipment is 0.8493

b - $p = \frac{2}{100} = 0.02$ $n = 4$ $P(x=0) = \binom{4}{0} 0.02^0 \ 0.98^4 = 1 \,.\, 1 \,.\, 0.9224 = 0.9224$

If shipment contains 2 unacceptable items probability of **accepting** the shipment is 0.9224

$$c - p = \frac{1}{100} = 0.01 \quad n = 4 \quad P(x \geq 1) = 1 - P(x = 0) = 1 - \binom{4}{0} 0.01^0 \; 0.99^4$$

$$= 1 - 0.9606 = 0.0394$$

If shipment contains 1 unacceptable item, probability of **rejecting** the shipment is 0.0394

As the shipment is accepted if $x = 0$, then probability of observing 0 defective item is $1 - P(x \geq 1)$

In other words, if shipment contains 1 unacceptable item;

probability of **rejecting** the shipment is; $1 - 0.9606 = 0.0394$
probability of **accepting** the shipment is; 0.9606

34 - The probability that a business administration graduate program student will complete the program in 2 years is 0.60. If 10 applicants will be accepted to the program this year.

a - What is the probability that half of the students will complete in 2 years?
b - What is the probability that at least half of the students will complete in 2 years?
c - What is the most probable number of students who will complete in 2 years?
d - What is the expected number of students who will complete in 2 years?
e - What is the standard deviation of students who will complete in 2 years?
f - What is the probability that at most 2 students will not complete in 2 years?

SOLUTION

a - Random variable x is defined as number of students who will complete the program within 2 years. Number of trials is the number of students accepted to the program and it is equal to10. The probability of success is the probability that a student will complete the program and it is equal to 0.60

$$P(x = 5) = \binom{10}{5} 0.60^5 \; 0.40^5 = 0.200658$$

b - $P(x \geq 5) = 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)]$

From cumulative binomial probability table $P(x \geq 5) = 1 - P(x < 4) = 1 - 0.6331 = 0.3669$

c - The most probable number of students who will graduate in 2 years can be simply obtained in the binomial probability table as 6. The highest probability is for 6 occurrences.

$P(x = 6) = 0.2505$

MARMARA ÜNİVERSİTESİ YAYINEVİ

d - $E(x) = n \cdot p = 10 \times 0.60 = 6$ students

e - $V(x) = n \cdot p \cdot q = 10 \times 0.60 \times 0.40 = 2.4$ Std. dev. $= \sqrt{2.4} = 1.5491$

f - Random variable x can be also defined as number of students who will not complete the program in 2 years but in that case probability of success should be consistent with the definition of random variable. That is; p should be 0.40 which is the probability that the student will not complete in two years.

In the following table random variable x is defined in two different ways. In fact; if number of students who will not complete in 2 years is 1 out of 10 applicants, it means that number of students who will complete in 2 years is 9. Similarly if the prior is 7 then the later is 3. Therefore; there is no difference between stating as "7 students who will not complete the program" or "3 students who will complete the program".

Table 33.1 Defining random variable in two different ways.

Number of students who will not complete	0	1	2	3	4	5	6	7	8	9	10
Number of students who will complete	10	9	8	7	6	5	4	3	2	1	0

In this part probability that at most 2 students will not complete the program is asked.

i) If x is defined as given in the first row of the table above, x will take the values of 0, 1 and 2.

ii) If x is defined as given in the second row of the table above, then x will take the values of 10, 9 and 8.

Solutions which lead the same result are given respectively below.

i) $P(x \leq 2) = [P(x = 0) + P(x = 1) + P(x = 2)]$

$$= \binom{10}{0} 0.40^0 \ 0.60^{10} + \binom{10}{1} 0.40^1 \ 0.60^9 + \binom{10}{2} 0.40^2 \ 0.60^8 =$$

ii) $P(x \geq 8) = [P(x = 10) + P(x = 9) + P(x = 8)]$

$$= \binom{10}{10} 0.60^{10} \ 0.40^0 + \binom{10}{9} 0.60^9 \ 0.40^1 + \binom{10}{8} 0.60^8 \ 0.40^2 =$$

As seen in the above solutions; first, second and the third terms are equal to each other. Combinations that are equal are given below. Obviously the rest of the terms are also equal.

$$\binom{10}{0} = \binom{10}{10}; \qquad \binom{10}{1} = \binom{10}{9}; \qquad \binom{10}{2} = \binom{10}{8}$$

35 - On the average 80 percent of the outputs produced by a machine is good. Outputs being good or defective is a Bernoulli Process.

a - What is the probability that the last item is good?

b - What is the probability that randomly selected two items are good?

c - What is the probability of observing two good items among randomly selected three items?

SOLUTION

a - Probability of the last item being good is 0.80. Probability of each product being good is 0.80 because in Bernouilli processes success probability is constant from trial to trial.

b - One of the characteristics of Bernouilli process is that outcomes of successive trials are independent. Therefore probability of selecting two good items is the product of individual probabilities of observing a good item. That is $0.80 \times 0.80 = 0.64$.

Part **b** can be also solved by using binomial formula;

$n = 2$, two items are selected.
$p = 0.80$ and
x is defined as number of good items.

$P(x = 2)$ is required.

$$P(x=2) = \binom{2}{2} 0.80^2 \; 0.20^0 = 1 \times 0.64 \times 1 = 0.64$$

c - $n = 3$, three items are selected.
$p = 0.80$ and
x is defined as number of good items.

Again probability of observing two good items is required but in this case among three products.

$$P(x=2) = \binom{3}{2} 0.80^2 \; 0.20^1 \; 3 \times 0.64 \times 0.20 = 0.384$$

36 - Para Bank has recently introduced a new type credit card called Silvercard targeted to a different group of customers. Past data show that 30 percent of applicants who have demanded previous credit card types, are rejected by bank as the applicants do not meet the requirements of the bank. Whether the applicant meets the requirement of bank or not is a bernouilli process. Suppose that on a given day a bank branch have received the application forms of 10 customers which they send to the credit evaluation center of the bank. Find the following probabilities.

a - Exactly 3 applicants will be rejected?
b - None of them will be rejected?
c - Less than 4 will be rejected?
d - More or equal to 6 will be accepted?

SOLUTION

$$\mathbf{a}\text{ - } P(x=3) = \binom{10}{3} 0.30^3 \; 0.70^7 = 0.2673$$

From binomial cumulative probability table

$$P(x=3) = P(x \le 3) - P(x \le 2) = 0.6496 - 0.3828 = 0.2673$$

$$\mathbf{b}\text{ - } P(x=0) = \binom{10}{0} 0.30^0 \; 0.70^{10} = 0.0282$$

From binomial cumulative probability table, the same result can be read directly from the table.

$$\mathbf{c}\text{ - } P(x<4) = [\,P(x=0) + P(x=1) + P(x=2) + P(x=3)\,]$$

From binomial cumulative probability table

$$P(x<4) = P(x \le 3) = 0.6496$$

d - Notice that in part d probability of more or equal to 6 applicants being accepted is asked. More or equal to 6 accepted applicants means less than 4 applicants being rejected. This probability is computed in part c which is equal to 0.6496.

37 - In an advertising program, it is claimed that 25 percent of consumers prefer brand SAF soaps. If this claim is true;

a - What is the probability that fewer than 3 consumers in a random sample of 10 prefer the advertiser's brand?
b - What is the probability that the number of consumers preferring is between 3 and 5 in a sample of 10 consumers?

c - Compute expected value and standard deviation of number of consumers who prefer that soap in a random sample of 60 consumers?

d - If among a sample of 60 consumers, only 10 preferred the brand, what inference would you make about the claim?

SOLUTION

a - $n=10\,;\,p=0.25\;\; P(x<3)=[P(x=0)+P(x=1)+P(x=2)]$

From binomial cumulative probability table;

$P(x<3)=P(x\leq 2)=0.5256$

b - $n=10\,;\,p=0.25\;\; P(3\leq x\leq 5)=[P(x=3)+P(x=4)+P(x=5)]$

From binomial cumulative probability table;

$P(3\leq x\leq 5)=\;P(\,x\leq 5)-P(x\leq 2)=0.4547$

c - $n=60\,;\,p=0.25$

$E(x)=n\,.\,p=60\times 0.25=15$ consumers out of 60 are expected to prefer that soap.

$V(x)=n\,.\,p.\,q=60\times 0.25\times 0.75=11.25$ Std. dev. $=\sqrt{11.25}=3.3541$

38 - A buyer checks large lots of electronic chips by inspecting a sample of 5 chips from each lot and if no defective item is found in sample, the lot enters to the production process without further testing. If at least 1 defective chip is observed, the chips in the lot is inspected one by one. What is the probability of inspecting all the chips in the lot if the proportion of defective chips in the lots is

a - 1 %

b - 10 %.

SOLUTION

All the chips in the lot are decided to be inspected in case of finding even one defective in a sample of 5 chips. Therefore; first step is to find out the probability of finding at least one defective item in a sample of 5 chips based on 1 % or 10 % defective item in the whole lot. Computed probability will be the probability of inspecting the whole lot.

a - $P(x\geq 1)\;=\;1-P(x=0)\;=1-\binom{5}{0}0.01^{0}\;0.99^{5}=0.95099$

b - $P(x\geq 1)\;=\;1-P(x=0)\;=1-\binom{5}{0}0.10^{0}\;0.90^{5}=0.59049$

39 - Marketing manager of a company has been revealed that the proportion of people favoring a new package design is 60 %. Suppose that manager has asked the opinion of 20 people selected randomly among their customers?

a - What is the probability that less than 9 will favor the new package design?

b - What is the probability that at most 6 will not favor the new package design?

c - What is the probability that exactly half of them will favor the new package design?

d - What is the expected value and the standard deviation of the persons favoring package design?

SOLUTION

$n = 20$

$p = 0.60$ ⇐ proportion of people who favor new package design

$q = 1 - p = 1 - 0.60 = 0.40$ ⇐ proportion of people who do not favor new package design

x ⇐ Number of people who favor new package design

a - Binomial probability tables are sometimes prepared up to $p = 0.50$ fortunately. Because of the symmetry of the distribution, while looking at the table, instead of 0.60, the probabilities computed for $p = 0.40$ can be used to find out the required probabilities. However the values of the random variable for which the probabilities are searched from the table should be consistent with p.

If p will be assumed as 0.40 percent which shows the proportion of people who do not favor the new design, then the variable should be defined as the number of people who do not favor the new design

Less than 9 will favor new design means at least 9 people (ie. equal or more than 9) will not favor the new package design.

From binomial cumulative table:

$P(x \geq 9) = 1 - P(x \leq 8) = 1 - 0.5956 = 0.4044$

b - At most 6 will not favor new design; $P(x \leq 6) = 0.25$

c - Half of them favors the new design means the other half will not favor the design.
From binomial cumulative table;

$P(x = 10) = P(x \leq 10) - P(x \leq 9) = 0.8725 - 0.7553 = 0.1172$

d - $E(x) = n \cdot p = 20 \times 0.60 = 12$

$V(x) = n \cdot p \cdot q = 20 \times 0.60 \times 0.40 = 4.8$ Std. dev. $= \sqrt{4.8} = 2.1909$

40 - 5 similar machines are operating independently in a production system. Probability of breakdown for each of these machines within 30 days is 0.30. Production system is not interrupted if at least two of the machines do not have breakdown.

a - What is the probability that exactly two of the machines will have breakdown within 30 days?

b - What is the probability that the production system will continue to operate for at least 30 days?

SOLUTION

$n = 5$ ⇐ total number of machines.
$p = 0.30$ ⇐ probability of a machine breakdown within 30 days.
x ⇐ number of machines that breakdown.

$$\mathbf{a} - P(x=2) = \binom{5}{2} 0.30^2\ 0.70^3 = 10\ .\ (0.09)\ (0.343) = 0.3087$$

b - Production system will continue to operate for at least 30 days if no machine breakdown occurs within 30 days.

$$P(x=0) = \binom{5}{0} 0.30^0\ 0.70^5 = 1\ .\ (0.09)\ (0.16807) = 0.015126$$

41 - In procurement of a part of machine, ABC Company's current acceptance sampling plan is based on a randomly selected sample of 10 parts. If the number of sample defectives is 0, shipments are accepted. Otherwise rejected.

a - What is the probability of accepting a shipment, if on the average 5 % is defective?

b - What is the probability of accepting a shipment, if on the average 1 % is defective?

c - What is the probability of rejecting a shipment, if on the average 10 % is defective?

SOLUTION

$$\mathbf{a} - P(x=0) = \binom{10}{0} 0.05^0\ 0.95^5 = 1\ .\ 1.\ (0.7738) = 0.77378$$

$$\mathbf{b} - P(x=0) = \binom{10}{0} 0.01^0\ 0.99^5 = 1\ .\ 1.\ (0.95099) = 0.95099$$

$$\mathbf{c} - P(x>0) = 1 - P(x=0) = 1 - \binom{10}{0} 0.10^0\ 0.90^5 = 1 - \{1\ .\ 1.\ (0.59049)\}$$

$$= 1 - 0.59049 = 0.40951$$

42 - An operator who enters sales slip input data in computer, incorrectly enters 5 slips out of 100 on the average. She enters 50 sales slip on a given day.

a - What is the probability that on a given day, she will enter all the slips correctly?

b - What is the probability that on a given day, she will correctly enter more than 90 % of the slips?

SOLUTION

$n = 25$

$p = 0.05$ ⇐ proportion of slips which are incorrectly entered

$q = 1 - p = 1 - 0.05 = 0.95$ ⇐ proportion of slips which are correctly entered

x ⇐ Number of slips which are incorrectly entered

a - When number of incorrectly entered sales slip is 0, that is; $x = 0$, then all the slips entered on a given day will be correct.

From binomial cumulative probability table; $n = 50$ $p = 0.05$ and $x = 0$

$P(x = 0) = 0.0769$

b - More than 90 % correctly entered slip correspond to more than 45 slips ($0.90 \times 50 = 45$). As x is defined as the number of incorrectly entered sales slip, more than 45 correctly entered sales slip correspond to less than 5 correctly entered sales slip.

From binomial cumulative probability table; $n = 50$ $p = 0.05$

$P(x < 5) = 0.8964$

43 - The product manager of a product claims that 50 % of their customers will not like the new package design proposed by their advertising agent. If this claim is true what is the probability that in a sample of 50 customers at least 33 will not like the package.

SOLUTION

$P(x \geq 33) = 1 - P(x \leq 32) = 1 - 0.9832 = 0.0164$

44 - A small restaurant in İstanbul Boshfours can serve up to 40 people. Reservation is required as they have only 40 seats. However among the people making dinner reservation 75 % actually attend dinner. For each night, restaurant makes reservation up to 45 people if that amount of reservation is received, because they don't expect all of them to attend. The probability distribution of number of seats reserved for each night is given in the following table.

Table 44.1 Probability distribution of number of seats reserved.

No.of res. seats (x)	35	36	37	38	39	40	41	42	43	44	45
Probability P(x)	0.16	0.2	0.15	0.11	0.09	0.07	0.07	0.06	0.04	0.03	0.02

a - What is the expected number and variance of reserved seats?

b - What is the expected number and variance of people who attend dinner if 45 people make reservation?

c - In what proportion of days does the number of people attending dinner exceed the number of available seats? (Assume independence between number of reservation made and the probability that a people will attend dinner)

SOLUTION

a - Table 44.2 Computation of expected number and variance of reserved seats.

No.of res. seats (x)	35	36	37	38	39	40	41	42	43	44	45	
Probability P(x)	0.16	0.20	0.15	0.11	0.09	0.07	0.07	0.06	0.04	0.03	0.02	E(x) =
x . P(x)	5.6	7.2	5.55	4.18	3.51	2.8	2.87	2.52	1.72	1.32	0.9	38.17
(x - E(x))	-3.17	-2.17	-1.17	-0.17	0.83	1.83	2.83	3.83	4.83	5.83	6.83	
$[x - E(x)]^2$	10.05	4.709	1.369	0.029	0.689	3.349	8.009	14.67	23.33	33.99	46.65	V(x) =
$P(x)[x - E(x)]^2$	1.608	0.942	0.205	0.003	0.062	0.234	0.561	0.88	0.933	1.02	0.933	7.3811

Expected number of reserved seats = 38.17 Variance of reserved seats = 7.3811

b - Expected number of people who attend dinner = $n.p = 45 \times 0.75 = 33.75$

Variance of people who attend dinner = $n.p.q = 45 \times 0.75 \times 0.25 = 36 = 8.4375$

c - In this part probability of joint occurrence of two events is asked. The number of people attending dinner can exceed the number of available seats if both the reservation is more than 40 and also if the number of people attending dinner is more than 40. Therefore the first step in solution should be defining these events.

Event A is defined as the number of reserved seats being greater than 40.
Event B is defined as the number of people actually attending dinner being greater than 40.

Event "A and B" is the joint occurrence of A and B combined by intersection. It means both the reserved seats will be greater than 40 and at the same time the number of people attending dinner will exceed the number of available seats (ie. 40).

If x shows the number of reserved seats;

$P(A) = P(x > 40) = 0.07 + 0.06 + 0.04 + 0.03 + 0.02 = 0.22$ (From the above table)

If x shows the number of people who are actually attending the dinner, the random variable x has a binomial probability distribution with $n = 45$ and $p = 0.75$.

$P(B) = P(x > 40) = 1 - P(x \leq 40) = 1 - 0.994055 = 0.005945$

As event A and B are assumed to be independent P (A and B) = P (A) . P (B)

MARMARA ÜNİVERSİTESİ YAYINEVİ

P (A and B) = 0.22 x 0.005945 = 0.001308 ⇒ Approximately 0.131 %.

Interpretation: The probability of receiving more than 40 reservation and the number of people that attend dinner exceeding 40 is only 0.131 %. That is, the probability is very low. In fact; in case of facing with such a situation, it will be difficult for the restaurant because of the shortage of seats and probably, the quality of the service provided decreases, but the probability that they face with such a situation is very low. On the other hand, the people who do not attend dinner although they make reservation is 25 % (1 - 75 % = 25 %). Therefore, they prefer overbooking for 5 more people rather than rejecting their reservation requirement to increase the number of people that attend.

45 - Number of orders for a spare part received by a firm is 14 per week on the average.

a - What is the probability of no order on a particular day?
b - What is the probability of one order on a particular day?
c - What is the probability of more than 2 orders on a particular day?
d - What is the probability of less than 3 orders on a particular day?

SOLUTION

Probability distribution which will be used to find out the required probabilities in this question is Poisson distribution. This distribution is a probability distribution for discrete random variables. Probability function of Poisson process is as follows:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

e =2.71828... is the base of the natural logarithms
Above probability function is applied to find out the probability of specific number of occurrences (x occurrences) in a given period of time, length, area. Average number of occurrences (λ) is known in a given period of time, length or area.

Characteristics of a Poisson process are as follows:

i) Average number of occurrences (λ) in a specific time interval, in a specific length, in a specific area is constant. In this eg. λ is 14 for a week. If average number of occurrences is considered for a day λ is 14/7 = 2.
ii)The number of occurrences in a time interval, etc. is independent of the occurrences in previous time intervals, etc. In this eg. whatever the number of occurrence in a given day, the probability of any number of occurrence is same for all days. (ie.Probability of 1 occurrence is same for each day of the week)
iii) In any short interval of time, probability of more than one occurrence is close to zero.

Solution to the problem based on the above explanation is as follows:

Average number of orders received is given as 14 and this average is **per week**. As specific number of orders is asked for **a day**, (λ) should be adjusted for a day.

Average number of orders per day (λ) should be taken as 2 (14 / 7 = 2).

Random variable x is defined as the number of orders

$$\mathbf{a}\text{ - } P(X=0) = \frac{e^{-2}2^0}{0!} = \frac{e^{-2}1}{1} = 0.135335$$

$$\mathbf{b}\text{ - } P(X=1) = \frac{e^{-2}2^1}{1!} = \frac{e^{-2}\ 2}{1} = 0.27067$$

$$\begin{aligned}\mathbf{c}\text{ - } P(x>2) &= 1 - P(x \le 2) = 1 - (P(x=0) + P(x=1) + P(x=2)) \\ &= 1 - (0.135335 + 0.27067 + 0.27067) \\ &= 1 - 0.676675 \\ &= 0.323325\end{aligned}$$

$$\begin{aligned}\mathbf{d}\text{ - } P(x<3) &= P(x=0) + P(x=1) + P(x=2) \\ &= 0.676675\end{aligned}$$

Note that $P(x<3) + P(x>2) = 1$ ($P\ x = 3$) is counted in the second term.
Because $x < 3$ includes $x = 0$, $x = 1$, and $x = 2$ and
$x > 2$ includes $x = 3$, $x = 4$, etc.

46 - A computer system in a company has one breakdown in a month (30 days) on the average. During heavy working days even a single failure in system leads loss of time and money.

a - What is the probability that no breakdown will occur in the next 30 days?

b - What is the probability that at least one breakdown will occur in the next 30 days?

SOLUTION

$\lambda = 1 \Rightarrow$ Average number of breakdown per month.
x = number of breakdowns

$$\mathbf{a}\text{ - } P(x=0) = \frac{e^{-1}1^0}{0!} = \frac{e^{-1}1}{1} = 0.367879$$

On the average; if 1 breakdown occurs in a month, probability that no breakdown (0 breakdown) occurs in a month is approximately 0.37.

$$\mathbf{b}\text{ - } P(x \ge 1) = 1 - P(x=0) = 1 - 0.367879 = 0.632121$$

MARMARA ÜNİVERSİTESİ YAYINEVİ

47 - The number of accidents in an intersection is a Poisson process. On the average 3 accidents occur in 4 weeks.

a - What is the probability of occurrence of a single accident in 4 weeks?
b - What is the expected number and variance of accidents in 4 weeks?
c - What is the probability of occurrence of a single accident in 2 weeks?
d - What is the expected number and variance of accidents in 2 weeks?

SOLUTION

a - Average number of accidents in 4 weeks is 3. Therefore $\lambda = 3$

$$P(x=1) = \frac{e^{-3}3^1}{1!} = 0.049787 \times 3 = 0.149361$$

b - In Poisson process both the expected number and variance of occurrences are equal to λ

$E(x) = \lambda = 3.$ $V(x) = \lambda = 3$ Standard deviation = $\sqrt{3} = 1.73205$

c - Average number of accidents in 2 weeks is $3 / 2 = 1.5$. Therefore $\lambda = 1.5$

$$P(x=1) = \frac{e^{-1.5}1.5^1}{1!} = 0.223130 \times 1.5 = 0.334695$$

d - $E(x) = \lambda = 1.5.$ $V(x) = \lambda = 1.5$ Standard deviation = $\sqrt{1.5} = 1.22474$

Attention: In Poisson process average number of occurrence is given for a specific time interval (or for a specific area, length, volume). While computing probability of observing any number of occurrence or expected number of occurrence during a specific time interval, the average number of occurrence λ should be adjusted for the required time interval.

48 - Suppose the distribution of number of calls to a police dispatcher between 11:00 and 12:00 pm. on Sundays follows a Poisson distribution with $\lambda = 5$.

a - What is the probability of no calls during this period?
b - What is the probability of more than 5 calls?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

Required probabilities can be obtained from the Poisson probability table in which probabilities of exactly x occurrences are given.

a - From the table, where $\lambda = 5$ $P(x=0) = 0.0067$

b - $P(x>4) = 1 - P(x \leq 4) = 1 - (0.0067 + 0.0337 + 0.0842 + 0.1404 + 0.1755)$

$= 1 - 0.4405 = \mathbf{0.5595}$

49 - PUTER computers are fully warranted during its first year. The number of service calls per computer during the warranty year is distributed according to Poisson with a mean of 0.8

a - What is the probability of no service calls during the warranty period?

b - What is the probability of more than 1 service call?

c - If cost of each service call for a distributor is 1 million TL on the average, what is the expected cost of warranty service calls per computer for this distributor?

SOLUTION

a - Average number of service calls during warranty period is 0.8. Therefore $\lambda = 0.8$.

$$P(x=0) = \frac{e^{-0.8}\,0.8^{0}}{0!} = \frac{e^{-2}\,1}{1} = 0.449329$$

b - Cumulative probability asked in part (b) can be obtained from the Poisson cumulative probabilities in which probabilities of less than or equal to x values are given.

$P(x>1) = 1 - P(x \leq 1) = 1 - 0.8088 = 0.1912$

c - Expected cost of warranty service calls per computer depends on the expected number of calls. That is equal to;

$E(x)$ * cost of each service call $= 0.8 * 1$ million $= 800.000$

MARMARA ÜNİVERSİTESİ YAYINEVİ

50 - NAC Company producing cans, reports that the number of failures per 8 hour shift on its assembly line follows a poisson process with a mean of 0.6. There are three shifts in the factory.

a - What is the expected value and standard deviation of the number of failures per shift?
b - What is the probability of exactly one failure during the midnight shift?
c - What is the probability of fewer than two failures during the afternoon shift?
d - What is the probability of no failure during three consecutive 8 hour shifts? (Assume that the machines operate independently across the shifts)

SOLUTION

a - $E(x) = \lambda = 0.6.$ $V(x) = \lambda = 0.6$ Standard deviation = $\sqrt{0.6} = 0.774596$

b - From Poisson distribution table in which probabilities of exactly x occurrences are given.

$P(x = 1) = 0.3293$

c -From Poisson cumulative probabilities in which probabilities of less than or equal to x are given.

$P(x < 2) = 0.9769$

d - It is the joint occurrence of three independent events. If events A, B, and C are defined as occurrence of no failure in each of the three shifts respectively. The probability of no failure during three consecutive 8 hour shifts is as follows;

$P(A) = P(x = 0) = 0.3293$ $P(B) = P(x = 0) = 0.3293$ $P(C) = P(x = 0) = 0.3293$

$P(A \text{ and } B \text{ and } C) = (0.3293)(0.3293)(0.3293) = 0.03570$

51 - An investment manager receives an average of 6 calls per half an hour from her clients. Assuming calls are Poisson process, what is the probability that she will not be interrupted by a call during a 5 minute coffee break?

SOLUTION

Average number of calls in half an hour is 6. Average number of calls in 5 minutes 6 / 6 = 1
Therefore $\lambda = 1$.

If there will be no call (x = 0) she will not be interrupted.

a - $P(x = 0) = \dfrac{e^{-1}1^0}{0!} = \dfrac{e^{-1}1}{1} = e^{-1} = 0.367879$

52 - A telephone order firm receives, 5 orders in five minutes on the average. However they can serve on the average 3 customers in 5 minutes due to insufficient number of lines and personnel to handle the orders. Assume that number of orders received and number of calls served are independent.

a - What is the probability that during any five minutes more than 3 orders will be receieved?

b - What is the probability that during any five minutes more than 3 customers will be served?

c - What is the probability that during any five minutes more than 3 orders will be received and they will be served?

SOLUTION

a - Random variable x is defined as the number of orders received in 5 minutes.

Average number of orders received in 5 minutes (λ) is 5

Solution by using Poisson probabilities table in which the probability of exactly 'x' occurrences are given;

$$P(x>3) = 1 - P(x \leq 2) = 1 - \{P(x=0) + P(x=1) + P(x=2)\}$$

$$= 1 - \left(\frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!}\right) = 1 - (0.0067 + 0.0337 + 0.0842)$$

$$= 1 - 0.1246 = 0.8754$$

b - Another random variable is considered in part **b**. X is defined as the number of customers served in 5 minutes.

Average number of customers served in 5 minutes (λ) is 3.

Solution by using Poisson probabilities table in which the probability of exactly 'x' occurrences are given;

$$P(x>3) = 1 - (0.0498 + 0.1494 + 0.2240) = 1 - 0.4232 = 0.5768$$

From Poisson cumulative probability table in which probabilities of less than or equal to 'x' are given, $P(x \leq 2)$ can be read directly from the table.

$$P(x>3) = 1 - P(x \leq 2) = 1 - 0.4232 = 0.5768$$

c - The probability that during any five minutes, more than 3 orders will be received and they will be served should be computed by using the multiplication rule of probability.

MARMARA ÜNİVERSİTESİ YAYINEVİ

Event A is receiving more than 3 orders. Computed as 0.8753 in part **a.**

Event B is serving more than 3 customers. Computed as 0.5768 in part **b.**

Probability that during any five minutes more than 3 orders will be received and they will be served is the probability of joint occurrence of event A and event B. Notice that they are assumed as independent events.

$P(A \text{ and } B) = (0.8753)(0.5768) = 0.504873$

53 - What is the probability of less than 2 arrivals in 10 minutes;

a - If on the average 2.4 arrivals occur in one hour?
b - If on the average 8.4 arrivals occur in one hour?

SOLUTION

a - If on the average 2.4 arrivals occur 1 hour;
in 10 minutes (2.4 / 6 = 0.4) arrival occurs.

Therefore λ is 0.4

From Poisson cumulative probabilities in which probabilities of less than or equal to 'x'

$P(x \leq 1) = 0.9384$

b - If on the average 8.4 arrivals occur 1 hour;
in 10 minutes 8.4 / 6 = 1.4 arrivals occur.
Therefore λ is 1.4

From Poisson cumulative probabilities in which probabilities of less than or equal to 'x'

$P(x \leq 1) = 0.5918$

Attention: Probability of less than 2 arrivals - when average occurrence is 0.4 - is higher than the probability of less than 2 arrivals - when average occurrence is 1.4. Note that when average is 1.4 probability of 0.9384 (which is close to the probability of less than or equal to 1 in case of $\lambda = 0.4$ correspond to 'less than 4 arrivals'. That is; when the average arrival rate (λ) increases, the probability of observing higher number of occurrences increases.

54 - Şaban Bekler is waiting for a hitchhiking to go his job in the morning at an intersection on the road of Bosphours Bridge. During the time he is waiting for a car, on the average 8 available cars pass by in 10 minutes. Suppose car arrivals are Poisson process. On a day if he can not get a car within 5 minutes, he will be late to his job.

a - What is the probability that Şaban will not be late to his job?

b - What is the probability that Şaban will wait from 2.5 minutes to 7.5 minutes?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

a - If he can get a car within 5 minutes he will not be late to his job. Arrival of at least one available car within 5 minutes is sufficient for him to go to his job on time. The random variable is the number of car arrivals and we are interested in the probability of arrival of at least one car $P(x \geq 1)$ within 5 minutes.

Average number of car arrivals should be also adjusted for 5 minutes. That is λ is $8 / 2 = 4$.

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \frac{e^{-4} 4^0}{0!} = 1 - e^{-4} = 1 - 0.0183 = 0.9817$$

b - Şaban will continue to wait if he can not get a car. In other words; in the question it is asked that he should wait for 2.5 minutes and he should not get a car in that time interval. Therefore the probability of 0 car arrival within 2.5 minutes should be considered. Besides this after 2.5 minutes in the next 5 minutes time interval (2.5 to 7.5 minutes) arrival of at least one car should be considered so that he will wait at most 7.5 minutes. As a result, there are two independent events here.

Event A is arrival of no. available car within 2.5 minutes

That is adjusted λ is $8 \times (2.5 / 10) = 2$.

$$P(A) = P(x = 0) = \frac{e^{-2} 2^0}{0!} = e^{-2} = 0.1353$$

Event B is arrival of at least one available car within 5 minutes (from 2.5 to 7.5 minutes).

That is, adjusted λ is $8 \times (5 / 10) = 4$.

$P(B) = P(x \geq 1) = 1 - P(x = 0)$ From part a, $P(x \geq 1)$ is found as 0.9817

Probability that Şaban will wait from 2.5 minutes to 7.5 minutes is P (A and B)

$P(A \text{ and } B) = P(A) . P(B)$ Events A and B are independent.

$P(A \text{ and } B) = (0.1353) \times (0.9817) = 0.132824$

Attention : In poisson process, the number of occurrences in a time interval does not depend on the occurrences in previous time intervals. That is; Poisson process has no memory. In the above problem, computation of probability of waiting for 5 minutes is not different for the beginning, the end or any middle 5 minutes of the 10 minutes period.

55 - Monthly rents of stores with size of approximately 100 meter square in a street is 300 million TL on the average with a standard deviation of 25 million TL. Assume that rents are normally distributed.

a - What is the z score corresponding to i) 375 ii) 300
b - What is the probability that rents are greater than 325 million?
c - What is the probability that rents are less than 275 million?
d - What is the probability that rents are between 275 million and 325 million?
e - What is the probability that rents are less than 250 million?
f - What is the probability that rents are greater than 350 million?
g - What is the probability that rents are between 250 million and 350 million?

SOLUTION

Normal distribution is a continuous probability distribution. It is a bell shaped and symmetrical around its mean. As it is a continuous probability distribution, the random variable may assume any value within an interval and the probability of the values within an interval is considered. These probabilities are represented by the areas under the normal curve. The parameters of the normal distribution is mean (μ) and standard deviation (σ). That is for each value of mean and/or standard deviation there is a different normal distribution. **N** (μ ; σ). In this example the random variable rent is assumed to have a normal distribution with mean equal 300 and standard deviation equal to 25. N (300 ; 25).

a - Z is a standard normal variable which has a standard normal distribution. The mean and standard deviation of this distribution is 0 and 1 respectively. The formula to convert the value of the random variable having a normal distribution to standard normal variable (z score) is as follows ;

$$z = (x - \mu)/\sigma$$

x is the value which the random variable assumes; μ and σ are the mean and standard deviation of the random variable. Z shows the distance between the value of the random variable and its mean ($x - \mu$) in terms of the standard deviation (σ) of the distribution

i) Z score corresponding to 375 is $z = \dfrac{375 - 300}{25} = 3$

A rent of 375 million TL is 3 standard deviation above the mean rent. In other words a rent of 375 million deviates from the mean rent by 3 standard deviation.

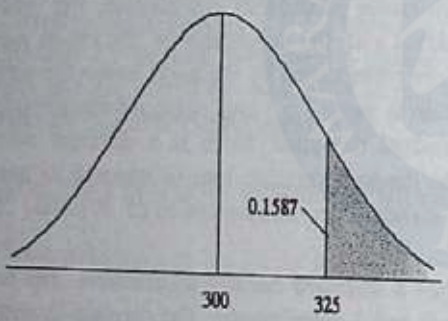
ii) Z score corresponding to 300 is $z = \dfrac{300 - 300}{25} = 0$

A rent of 300 million TL is equal to mean rent.

Attention: Suppose that the heights of the students in a class is normally distributed with mean equal to 1.68 m. and standard deviation equal to 4 cm. A student whose height is 1.80 meter is 12 cm above the average height (1.80 - 1.68 = 12). If this deviation from the mean is expressed in terms of standard deviation, he/she is 3 standard deviation (12/4 =3) above the mean. As a result a height of 1.80 is 3 standard deviation above its mean, a rent of 375 is 3 standard deviation above its mean etc. Note that by converting the values 1.80 and 375 to z scores, they are standardized.

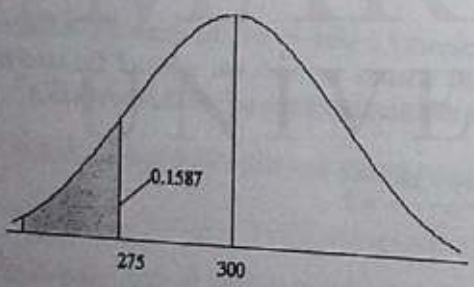
b - $P(x>325) \Rightarrow$ z score corresponding to 325 is $z = \dfrac{325-300}{25} = 1$

To find out the probabilities corresponding to the required intervals like x > 325, first step is to find out the corresponding z scores. Then, from the standard normal probability distribution table, probabilities corresponding to z scores can be directly read or adjusted according to the required interval. Here, half area table is used. That is; the probabilities given corresponding to z scores represent the area between z = 0 and the required z value.



$$P(x>325) = P(z>1)$$
$$= 0.5 - 0.3413$$
$$= 0.1587$$

c - $P(x<275) \Rightarrow$ z score corresponding to 275 is $z = \dfrac{275-300}{25} = -1$



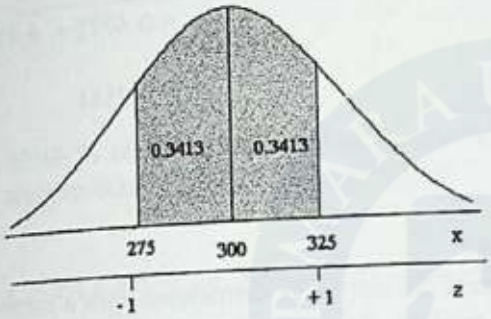
$$P(x<275) = P(z<-1)$$
$$= 0.5 - 0.3413$$
$$= 0.1587$$

A rent of 275 million TL is 1 standard deviation below the mean. As the normal distribution is symmetrical around its mean, the area (probability) corresponding to the interval between z = 0 and z = 1 is equal to the area corresponding to the interval between z = 0 and z = -1. Therefore in standard normal probability tables it is sufficient to have the areas corresponding to intervals between z = 0 and z = plus values.

MARMARA ÜNİVERSİTESİ YAYINEVİ

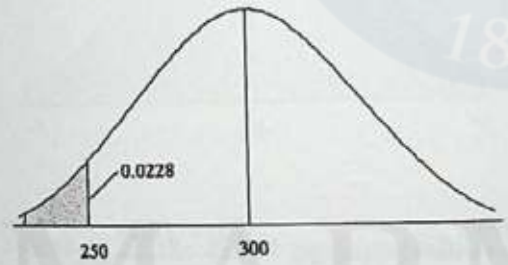
Attention: Standard normal probability tables can be also prepared as cumulative distribution function F (z). In such tables; F (z) shows the probability that a standard normal random variable is less than the value z. For example corresponding to z = 1, F (1) = 0.8413 is read directly from the table. This value is the probability of z being less than 1. P (z < 1) = 0.8413

d - P (275 < x < 325) ⇒ z score corresponding to 275 is -1
z score corresponding to 325 is 1



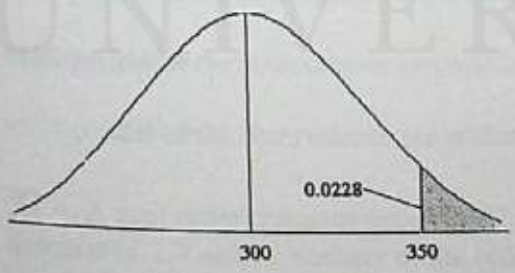
$$P(275 < x < 325) = P(-1 < z < 1)$$
$$= 0.3413 + 0.3413$$
$$= 0.6826$$

e - P (x < 250) ⇒ z score corresponding to 250 is $z = \frac{250 - 300}{25} = -2$



$$P(x < 250) = P(z < -2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

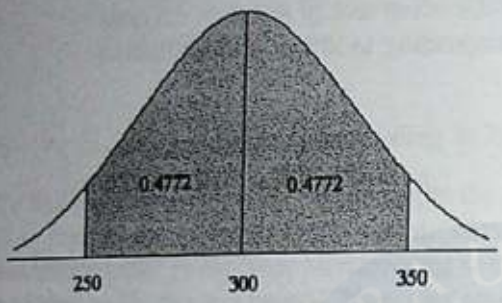
f - P (x > 350) ⇒ z score corresponding to 350 is $z = \frac{350 - 300}{25} = 2$



$$P(x > 350) = P(z > 2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

MARMARA ÜNİVERSİTESİ YAYINEVİ

g - $P(250 < x < 350) \Rightarrow$ z score corresponding to 250 is -2
z score corresponding to 325 is 2



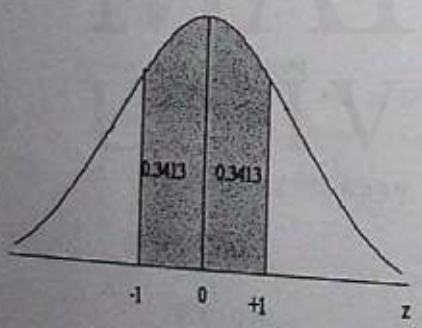
$$P(250 < x < 350) = P(-2 < z < 2)$$
$$= 0.4772 + 0.4772$$
$$= 0.9544$$

56 - The weights of packages of a brand of cereal are normally distributed with a mean of 2 kg. and a standard deviation of 50 gr. A package is selected randomly among this population. What is the probability that the package will weigh between;

a - 1.950 and 2.050 kg.
b - 1.900 and 2.100 kg.
c - 1.850 and 2.150 kg.

SOLUTION

a - $P(1.950 < x < 2.050) \Rightarrow Z_2 = \frac{1.950 - 2.000}{50} = -1 \quad Z_2 = \frac{2.050 - 2.000}{50} = 1$

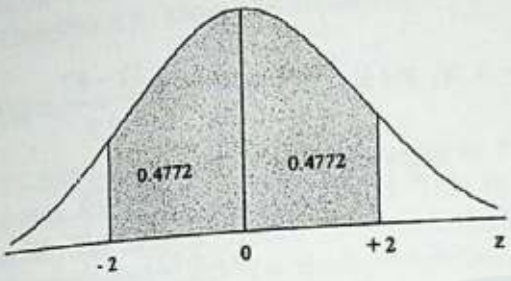


$$P(1.950 < x < 2.050) = P(-1 < z < 1)$$
$$= 0.3413 + 0.3413$$
$$= 0.6826$$

Note that probability that a package of cereal having a weight between 1.950 and 2.050 kg is 0.6826. In other words 68.26 % of the cereal packages which are assumed to have a normal distribution have weights between 1.950 and 2.050 kg.

MARMARA ÜNİVERSİTESİ YAYINEVİ

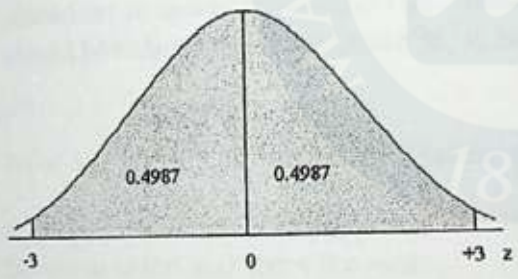
b - $P(1.900 < x < 2.100) \Rightarrow z_1 = \dfrac{1.900 - 2.000}{50} = -2 \qquad z_2 = \dfrac{2.100 - 2.000}{50} = 2$



$$P(1.900 < x < 2.100) = P(-2 < z < 2)$$
$$= 0.4772 + 0.4772$$
$$= 0.9544$$

95.44 % of the cereal packages which are assumed to have a normal distribution have weights between 1.900 and 2.100 kg.

c - $P(1.850 < x < 2.150) \Rightarrow z_1 = \dfrac{1.850 - 2.000}{50} = -3 \qquad z_2 = \dfrac{2.150 - 2.000}{50} = 3$



$$P(1.850 < x < 2.150) = P(-3 < z < 3)$$
$$= 0.4987 + 0.4987$$
$$= 0.9974$$

99.74 % of the cereal packages which are assumed to have a normal distribution have weights between 1.850 and 2.150 kg.

In normal distribution;

68.26 percent of the observations are within 1 standard deviation of the mean.

95.44 percent of the observations are within 2 standard deviation of the mean.

99.74 percent of the observations are within 3 standard deviation of the mean.

57 - A mail order company receives 47 orders on the average per week with a standard deviation of 2.7 orders. Number of the orders are normally distributed.

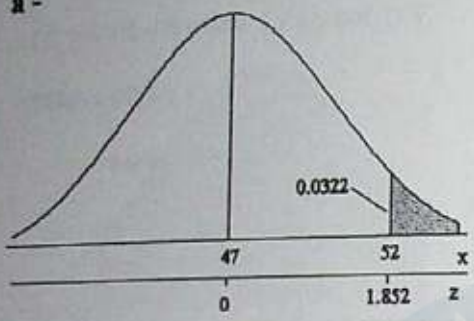
a - During a week, what is the probability that the company will receive i) more than 52, ii) less than 50, iii) between 39 and 53.

b - During 90 percent of the week, within what interval around mean will weekly orders be?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

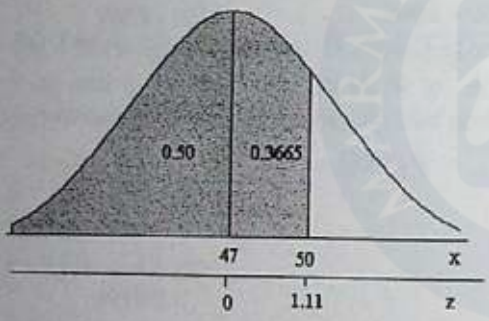
a -



i) $P(x > 52) \Rightarrow \quad z = \dfrac{52 - 47}{2.7} = 1.852$

$P(z > 1.85) = 0.5 - 0.4678$

$= 0.0322$

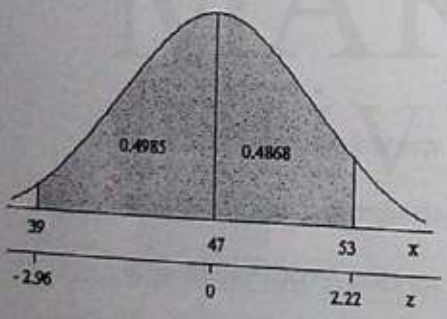


ii) $P(x < 50) \Rightarrow \quad z = \dfrac{50 - 47}{2.7} = 1.11$

$P(z < 1.11) = 0.5 + 0.3665$

$= 0.8665$

iii) $P(39 < x < 53) \Rightarrow \quad z = \dfrac{39 - 47}{2.7} = -2.963 \qquad z = \dfrac{53 - 47}{2.7} = 2.222$



$P(39 < x < 53) = P(-2.96 < z < 2.22)$

$= 0.4985 + 0.4868$

$= 0.9853$

b - The required interval around mean is the range of values of the random variable "sales" which has an occurrence probability of 0.90. In other words;

$P(x_1 < x < x_2) = 0.90$ is given, the values of x_1 and x_2 are asked.

MARMARA ÜNİVERSİTESİ YAYINEVİ

First step is to find out the z scores corresponding to these values from the standard normal probability distribution table as the corresponding probability is given. The table used here is the half area table which shows the probabilities from mean ($z = 0$) to any z score. Therefore 0.90 should be divided by 2 (0.90 / 2 = 0.4500). That is;

$P(z_1 < z < z_2) = 0.90 \quad P(z_1 < z < 0) = 0.45$ and $P(0 < z < z_2) = 0.45$

To find out the z score corresponding to a given probability, firstly the nearest probability within the table should be found. This is the inverse use of standard normal probability table. The nearest probabilities to 0.4500 in the table are 0.4495 and 0.4505 and the corresponding z scores to these probabilities are 1.64 and 1.65 respectively. That is;

$P(0 < z < 1.64) = 0.4495 \qquad P(0 < z < 1.65) = 0.4505$

We can take the average of 1.64 and 1.65 and assume that;

$P(0 < z < 1.645) = 0.4500 \quad P(-1.645 < z < 0) = 0.4500$

Attention: It should be noticed that minus sign in front of z value found from the table should be recorded as z_1, because z_1 is below the mean.

$P(-1.645 < z < 1.645) = 0.90$ This equality is read as follows;

In 90 percent of the time (here time unit is week) the sales level is between 1.645 standard deviation below and above the mean.

The next step is to find out the values which are 1.645 standard deviation below and above the mean. As the mean sales level is given as 47 and standard deviation is 2.7, by subtracting $z\,\sigma$ (1.645 x 2.7) from mean and adding $z\,\sigma$ (1.645 x 2.7) to mean x_1 and x_2 is computed.

$$-1.645 = \frac{x_1 - 47}{2.7} \qquad 1.645 = \frac{x_2 - 47}{2.7}$$

$x_1 = 47 - 1.645 (2.7) = 42.56 \qquad x_2 = 47 + 1.645 (2.7) = 51.44$

$P (42.56 < x < 51.44) = 0.90$

Note that in part **a**, the probabilities corresponding to more than, less than a z score or between two z scores are found. In part **b**, z scores corresponding to the probabilities are found.

MARMARA ÜNİVERSİTESİ YAYINEVİ

58 - Pınl corporation has developed a new detergent and will make pricing decision. Sales forecast will effect the determination of profit margin. Initial estimate of daily sales is 5000 kg on the average with a standard deviation of 850. However forecasts are revised later and daily sales level on the average is expected as 4500 kg. Sales are normally distributed.

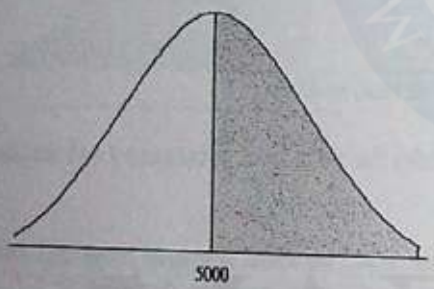
a -What is the probability that daily sales will be greater than 5000 according to the initial forecast?

b -What is the probability that daily sales will be greater than 5000 according to the revised forecast?

c -What is the probability that daily sales will be between 4000 and 5000 according to the revised forecast, given that; i) standard deviation is same, ii) standard deviation is 700 iii) standard deviation is 950?

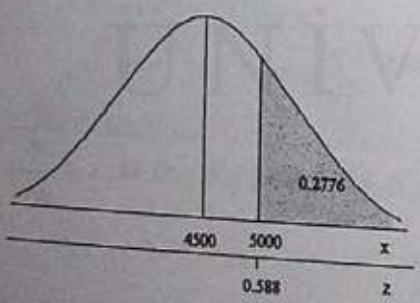
SOLUTION

a - N (5000 ; 850)



$$P(x > 5000) \Rightarrow P(z > 0) = 0.5$$

b - N (4500; 850)



$$P(x > 5000) \Rightarrow \quad z_1 = \frac{5000 - 4500}{850} = 0.588$$

$$P(z > 0.59) = 0.5 - 0.2224$$

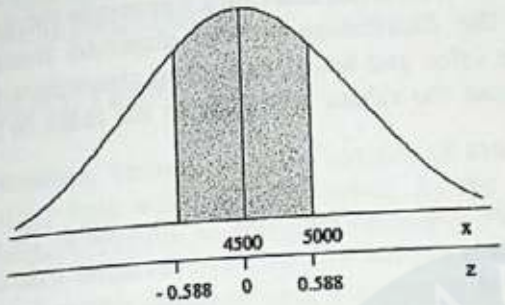
$$= 0.2776$$

Note that according to the initial estimate, probability that the sales level will be greater than 5000 is 50 percent. However, after the forecasts are revised, this probability is decreased to 27.76 percent

MARMARA ÜNİVERSİTESİ YAYINEVİ

c - i) N (4500 ; 850)

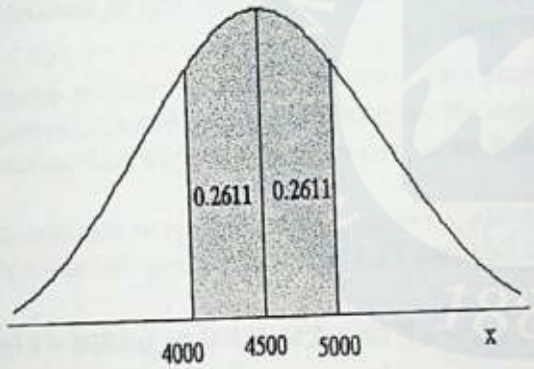
$P(4000<x<5000) \Rightarrow Z = \frac{4000-4500}{850} = -0.588 \quad Z = \frac{5000-4500}{850} = 0.588$



$P(-0.59 < x < 0.59) = 0.2224 + 0.2224$

$= 0.4448$

ii) N (4500; 700)



$P(4000 < x < 5000) \Rightarrow$

$Z = \frac{4000-4500}{700} = -0.714$

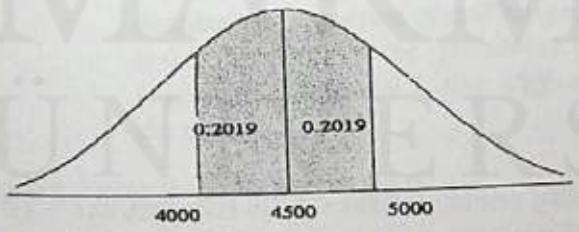
$Z = \frac{5000-4500}{700} = 0.714$

$P(-0.71 < x < 0.71) = 0.2611 + 0.2611$

$= 0.5222$

iii) N (4500; 950)

$P(4000 < x < 5000) \Rightarrow Z = \frac{4000-4500}{950} = -0.526 \quad Z = \frac{5000-4500}{950} = 0.526$



$P(-0.53 < x < 0.53) = 0.2019 + 0.2019$

$= 0.4038$

Note that; when standard deviation of the distribution is 850; the probability that the sales level being within 4000 and 5000 is 0.4448. When standard deviation decreases to 700 with the same mean, the probability of the same interval becomes 0.5222. That is the distribution gets narrower when standard deviation decreases and the probability of the same interval increases. In the third part standard deviation increases with the same mean and the probability of the same interval becomes 0.4038. That is the distribution becomes more dispersed when standard deviation increases and the probability of the same interval decreases.

Attention: Refer to the interpretation and the computation of the standard deviation. Remember that standard deviation is a measure of dispersion and in a way it measures an average deviation of the values from its mean. Therefore if the values are more dispersed, standard deviation will come out a larger value and less amount of observation will fall within a defined range fortunately (in above problem case iii). However if the distribution is less dispersed standard deviation will come out a smaller value and larger amount of observation will fall within a defined range. Because the values are closer to the mean in this case (in above problem case ii).

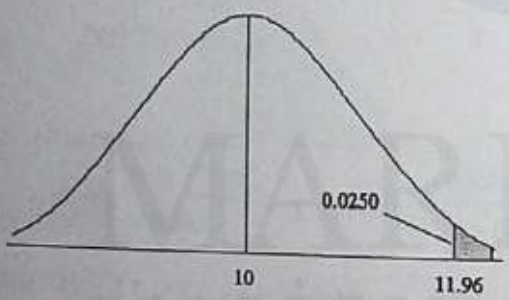
59 - Weekly average sales level of Koriflam cosmetics company is 10 million TL per salesman. Suppose that the distribution of sales is normal with standard deviation 1 million TL. Human resources management department proposed a new incentive system for the salesmen who sell more than 11.96 million TL.

a - What proportion of the salesmen will benefit from the new incentive system?

b - If the sales amount is less than 7.5 million TL. salesmen is regarded as unsuccessful, what proportion of the salesmen are unsuccessful?

SOLUTION

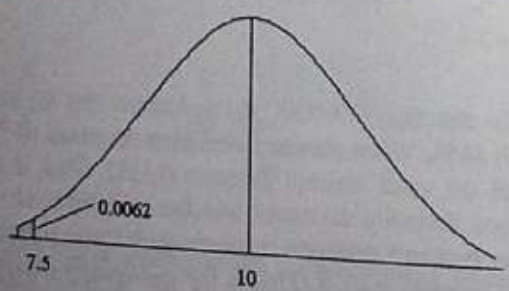
a - $P(x > 11.96) \Rightarrow z = \dfrac{11.96 - 10}{1} = 1.96$



$P(z > 1.96) = 0.5 - 0.4750 = 0.0250$

2.5 % of the salesman will benefit from the new incentive system.

b - $P(x < 7.5) \Rightarrow z = \dfrac{7.5 - 10}{1} = -2.5$



$P(z < -2.5) = 0.5 - 0.4938 = 0.0062$

0.6 % of the salesman will be regarded as unsuccessful.

MARMARA ÜNİVERSİTESİ YAYINEVİ

60 - The amount of tomato paste filled in one kg. cans is known to have a normal distribution with variance equal to 400 grams. The cans filled by amount of less than 950 grams of paste are not acceptable. What should be the average amount of tomato paste for which the machine should be adjusted so that no more than 1 % of the cans of tomato paste will be unacceptable?

SOLUTION

Random variable x is the amount of tomato paste filled by machine. If $x < 950$ grams, those cans will not be accepted. As the portion of the cans which are not accepted is required to be no more than 1 %; $P(x < 950) = 0.01$.

First step is to find out the corresponding z score to the x value given as 950 gram. 1 % area is at the left tail of the distribution below the mean.

$P(z < z_1) = 0.01$ In half area table, this corresponds to
$P(z_1 < z < 0) = 0.5 - 0.01 = 0.49 \Rightarrow z_1 = -2.23$ This z score correspond to the nearest probability which is 0.4901. This probability is found within the table. That is 950 gram is 2.23 standard deviation below the mean. Note that minus z score is considered here, because from the given probability it is understood that it is below the mean.

Second step is to find out the mean by adding 2.23 standard deviation to the 950 as in the first step 950 gram is found as 2.23 standard deviation below the mean.

$\sigma^2 = 400 \quad \sigma = \sqrt{400} = 20 \qquad \mu = x + z\,\sigma \qquad \mu = 950 + 2.23\ (20) = 994.6$

Note that $z = (x - \mu)/\sigma$ is used. $\Rightarrow \quad -2.23 = \dfrac{950 - \mu}{20}$

Attention: If the probability less than a value is smaller than 0.5 like the case given here, $P(x < 950) = 0.01$, that value is below the mean. Because in normal distribution proportion of observations less than a mean or greater than a mean is 0.5. Similarly; if the probability of a random variable greater than a value is given as less than 0.5 like $P(X > x) = 0.01$ then, that value should be above the mean.

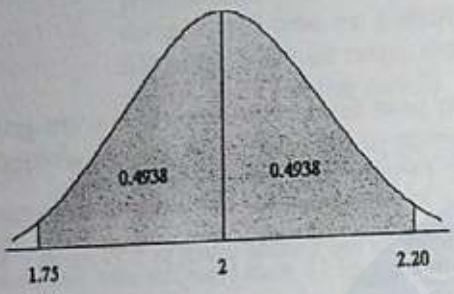
61 - The diameters of a circular machinery parts are normally distributed with mean 2 cm and standard deviation 0.1 cm. If the diameters are less than 1.75 and greater than 2.20 it is impossible to use them in the assembly.

a - What is the proportion of good parts and scrap parts?
If only 0.5 % of the parts are allowed to have diameters greater than 2.20:
b - What should be the mean diameter with the same standard deviation?
c - What should be the standard deviation with the same mean?

MARMARA ÜNİVERSİTESİ YAYINEVİ

SOLUTION

a - $P(1.75 < x < 2.20) \Rightarrow Z = \frac{1.75-2}{0.1} = -2.5 \quad Z = \frac{2.20-2}{0.1} = 2.5$



$P(1.75 < x < 2.20) = P(-2.5 < z < 2.5)$

$= 0.4938 + 0.4938$

$= 0.9876$

98.76 % is good parts, $1 - 0.9876 = 0.0124$ 1.24 % is scrap parts.

b - μ will be determined, σ will be kept as same and $P(x > 2.20) = 0.005$ is given.
From half area table, z score corresponding to 2.20 is computed as follows;

$$P(0 < z < z_1) = 0.5 - 0.005 = 0.495 \Rightarrow z_1 = 2.58$$

$$2.58 = \frac{2.20-\mu}{0.1} \qquad \mu = 2.20 - 2.58\,(0.1) = 1.942$$

The machine which drills holes to parts should be adjusted to drill with a mean of 1.942 given that the standard deviation is 0.1 cm. so that only 0.5 % of the parts will have diameters over 2.20

c - μ will be kept as same, σ will be changed $P(x > 2.20) = 0.005$ is given.
In part b z score corresponding to 2.20 is computed as 2.58

$$2.58 = \frac{2.20-2.00}{\sigma} \qquad \sigma = (2.20 - 2.00) / 2.58 = 0.0775$$

The machine which drills holes to parts should have a standard deviation of $\cong 0.08$ cm given that the mean is 2.00 cm. so that only 0.5 % of the parts will have a diameter over 2.20.

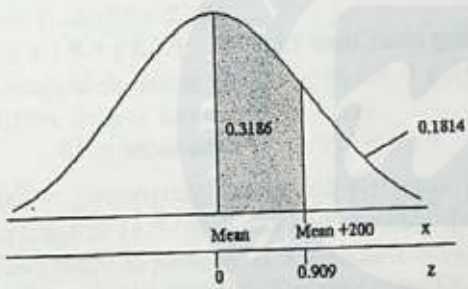
MARMARA ÜNİVERSİTESİ YAYINEVİ

62 - Lifetime of a specific type of lightbulb is normally distributed with a standard deviation of 220 hours. What is the probability that the lifetime of a randomly chosen battery exceeds the mean lifetime by more than 200 hours?

SOLUTION

Mean is unknown, standard deviation is given as 220. Exceeding mean by more than 200 means, the random variable lifetime x will take value more than $\mu + 200$.

That is probability that is asked is $P(\mu + 200)$. To find out this probability, we do not need to know the mean because in normal distributions it is sufficient to know the distance of the value of random variable to its mean that is given as more than 200 in the question.



$$P(x > \mu+2) \Rightarrow z = \frac{(\mu + 200) - \mu}{220} = 0.909$$

$$z = \frac{200}{220} = 0.909$$

$$P(z > 0.91) = 0.5 - 0.3186 = 0.1814$$

Note that z score shows the distance of the random variable value to its mean in terms of standard deviation. In this problem the distance is given as more than 200 hours to the mean and this is sufficient to find out z score.

63 - Two sales representatives of Bank White install POS (Point-of-sale) machine to their member merchants. First representative installs 12 machines on the average in a week with a standard deviation of 2. Second installs 15 with a standard deviation of 2. Assume that distributions are normal and independent.

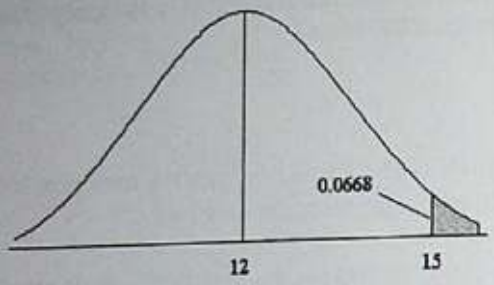
a - What is the probability that both sales representative will achieve to install more than 15 machines in a week?

b - What is the probability that both sales representative will achieve to install machines at least 1 standard deviation above their means in a week?

SOLUTION

a - Number of machines installed by first representative is normally distributed with mean 12 and standard deviation 2. $N(12; 2)$

Let event F be the first representative installing more than 15 POS. $P(F) = P(x > 15)$



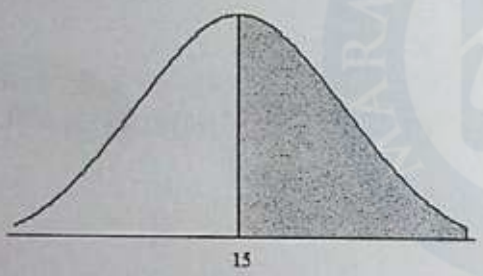
$$P(x > 15) \Rightarrow \quad z = \frac{15 - 12}{2} = 1.5$$

$$P(z > 1.5) = 0.5 - 0.4332 = 0.0668$$

$$P(F) = 0.0668$$

Number of machines installed by second representative is normally distributed with mean 15 and standard deviation 2. $N(15 ; 2)$

Let event S be the second representative installing more than 15 POS. $P(S) = P(x > 15)$



$P(x > 15) \Rightarrow$ As the mean is 15

$P(z > 0) = 0.5 \Rightarrow P(S) = 0.5$

Both sales representative will achieve to install more than 15 machines in a week is the joint occurrence of event F and event S. As they are independent events their joint occurrence is;

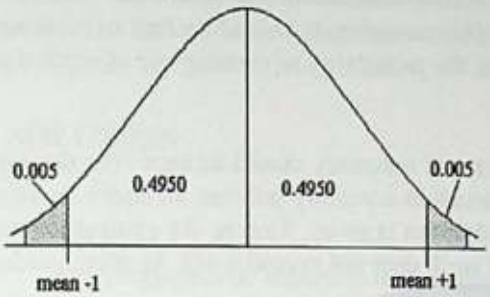
$P(F \text{ and } S) = P(F) . P(S) = (0.0668)(0.5) = 0.0334.$

64 - ÇAG Automation buys precision parts for a control device from MIS Machine. Each part has a specified mean dimension μ. A part having dimension which differs from μ by more than 1 mm is classified as defective. ÇAG requires that no more than 1 percent of parts should be defective. If MIS is to satisfy the requirement, what is the maximum variability (std. dev.) that can be permitted? Assume that the dimensions of parts are normally distributed

SOLUTION

Being different from μ by no more than 1 mm corresponds to the interval within 1 mm to the mean. It may be at most 1 mm less or 1 mm more than mean. As the parts which differ by more than 1 mm is required to be 1 % and both 1 mm less than and 1 mm more than mean values are considered as defective, two tails of the distribution should be considered and 1 % should be divided to 2. (0.01/ 2 =0.005);

MARMARA ÜNİVERSİTESİ YAYINEVİ



$P[(\mu - 1) < x < (\mu + 1)] = 0.99$

$P(0 < z < z_1) = 0.4950$

From the half area table, $z_1 = 2.575$ This z score is the average of 2.57 and 2.58 with the corresponding probabilities 0.4949 and 0.4951

$P(-2.575 < z < 2.575) = 0.99$

$z = (x - \mu)/\sigma \Rightarrow 2.575 = 1/\sigma$

$\sigma = 1/2.575 = 0.388.$

Standard deviation should be 0.388 if ÇAG requires that no more than 1 percent of parts differs from μ by more than 1 mm.

65 - Incorrectly assembled electronic parts in a plant is normally distributed with an average of 14.5 for a day and standard deviation is 1.25. If the number of parts which are incorrectly assembled is 2 standard deviation above the mean, it is the signal of some problem in the assembly line that should be examined to take some precautions.

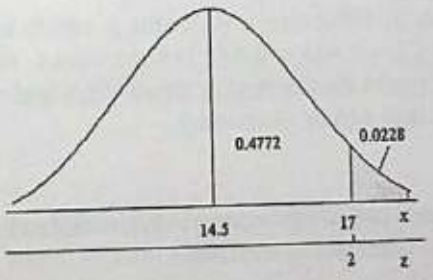
a - What is the number of incorrectly assembled parts that is determined as the tolerance limit?

b - What is the probability of exceeding tolerance limit?

SOLUTION

a - The number of incorrectly assembled parts which correspond to 2 standard deviation above the mean is the limit that is tolerated.

$x = \mu + z\sigma = 14.5 + 2\ (1.25) = 17$ If the number of incorrectly assembled parts in a given day is more than 17, according to the given decision rule the assembly line should be controlled.



b - $P(x > 17) = P(z > 2)$

$= 0.5 - 0.4772$

$= 0.0228$

MARMARA ÜNİVERSİTESİ YAYINEVİ

66 - Average weekly demand for a product which is normally distributed has a mean of 250 units and standard deviation is 80 units. How many units should be kept in inventory for a week if the company wants to ensure that the probability of running out of stock does not exceed 0.025?

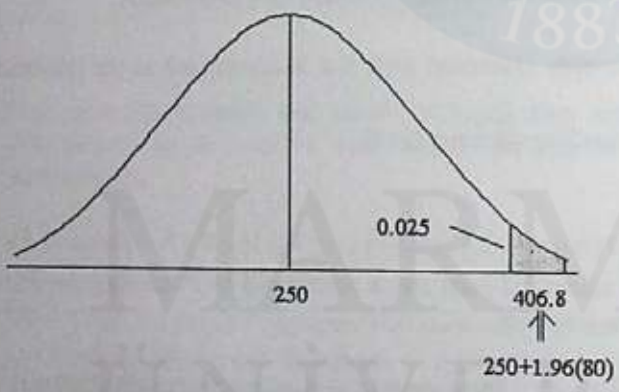
For not running out of a stock, sufficient amount of inventory should be kept. The decision on how much to keep is one of the basic decisions in inventory policies. In above question one of the criteria which is used to make this decision is given. That is; the company wants to ensure that the probability of running out of stock does not exceed 0.025. In other words;

If x is the demand level, the amount of x which will be kept in inventory is such a value that the probability of demand being greater than that value should be 0.025.

$P(X > x) = 0.025$ In half area table z score of x, which satisfies this condition is z_1;

$P(0 < z < z_1) = 0.5 - 0.025 = 0.4750 \Rightarrow z_1 = 1.96$

$x = \mu + z\sigma = 250 + 1.96\,(80) = 250 + 156.8 = 406.8$



If the amount of product in the inventory for a week is 406.8 they will be able to satisfy the demand up to this demand level and in 97.25 % (1 - 0.025 = 0.975) of the weeks the demand level is up to this amount. In 2.5 % of the weeks the demand is above 406.8 and as they do not keep more than that amount for a week they will be in shortage.

The proportion of customers whose demand are met is called customer service level. In this example customer service level is 97.25 %.

MARMARA ÜNİVERSİTESİ YAYINEVİ

67 - In a meeting of a committee of 50 members, 75 % are expected to attend. What is the probability that at most 18 members will not attend to the meeting?

SOLUTION

$n = 50$

$p \Rightarrow$ proportion of members who do not attend $= 0.25$

$q \Rightarrow$ proportion of members who attend $= 0.75$

$x \Rightarrow$ Number of members who do not attend.

From binomial cumulative probability table

$P(x \leq 18) = 0.9713$

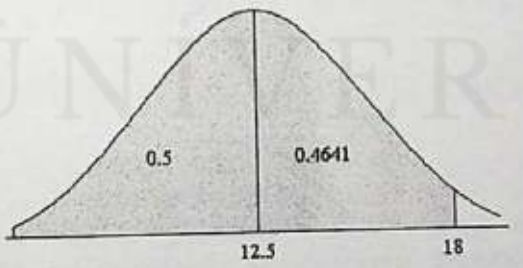
If both $np \geq 5$ and $nq \geq 5$, Binomial distribution can be approximated by Normal distribution. $E(x) = n.p = \mu$ $V(x) = n.p.q$

When p is 0.50, binomial distribution becomes a symmetrical distribution. Hence if p is closer to 0.50 and if n is large, binomial probability distribution is better approximated by normal probability distribution.

Expected number and standard deviation of the members who do not attend to the meeting is;

$E(x) = 50 \times 0.25 = 12.5$ $V(x) = 50 \times 0.25 \times 0.75 = 9.375$ $\sigma = \sqrt{9.375} = 3.06$

$P(x < 18) \Rightarrow$ z score corresponding to 18 is $z = \dfrac{18 - 12.5}{3.06} = 1.797 \cong 1.80$



$P(x < 18) = P(z < 1.80)$

$= 0.5 + 0.4641$

$= 0.9641$

The probability found as 0.9641 by normal approximation is very close to the probability which is 0.9713 obtained by binomial distribution.

68 - It has been observed in accounting department of a company that, out of 100 payroll check, 10 are prepared incorrectly. If 250 payroll checks will be prepared in a given month, what is the probability that less than 20 payroll checks will be incorrect?

SOLUTION

Proportion of incorrectly prepared payroll checks is $p = \frac{10}{100} = 0.10 \quad n = 250$

By using binomial probability function;

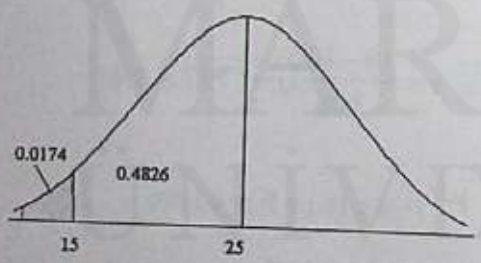
$P(x < 15) = 0.017508$

However as $np \geq 5$ and $nq \geq 5$; it is possible to approximate it by normal distribution

$n.p = 250 \times 0.10 = 25$ and $n.q = 250 \times 0.90 = 225$.

$E(x) = n.p = \mu = 25 \quad V(x) = n.p.q = 250 \times 0.10 \times 0.90 = 22.5 \quad \sigma = \sqrt{22.5} = 4.743$

$P(x < 15) \Rightarrow$ z score corresponding to 15 is $z = \frac{15 - 25}{4.74} = -2.11$



$P(x < 15) = P(z < -2.11)$

$= 0.5 - 0.4826$

$= 0.0174$

69 - While packaging one kg cube sugars, due to a failure in packaging machine 5 % of the one kg cube sugars are incorrectly packaged. Suppose that in a production run, 2000 one kg cube sugar packages are produced.

a - What is the expected number and standard deviation of incorrect packages?
b - What is the probability that no more than 80 are out of standards?

MARMARA ÜNİVERSİTESİ YAYINEVİ

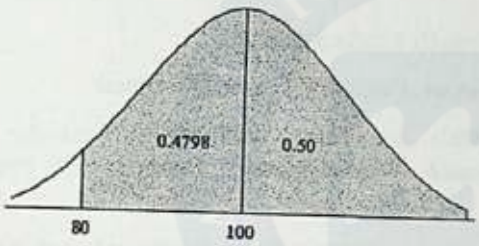
SOLUTION

a - $E(x) = 2000 \times 0.05 = 100$ Std. dev. $= \sqrt{2000 \times 0.05 \times 0.95} = 9.745$

b - Binomial probability distribution can be approximated by normal distribution.

$E(x) = n \times p = \mu = 100$ $V(x) = n \times p \times q = \sigma^2 = 95$ $\sigma = \sqrt{95} = 9.746$

$P(x > 80) \Rightarrow$ z score corresponding to 80 is $z = \dfrac{80 - 100}{9.746} = -2.05$



$P(x > 80) = P(z > -2.05)$

$= 0.5 + 0.4798$

$= 0.9798$

70 - Liptea company has distributed 1000 small tea packages attached lottery coupons for promotion purposes. The proportion of coupons received back is assumed as 75 %.

a - How many coupons are expected to be received back?

b - What is the probability of receiving at least 780 coupons?

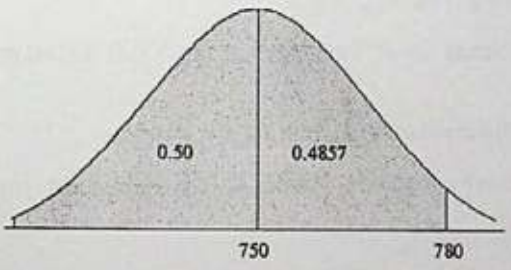
SOLUTION

a - $E(x) = 1000 \times 0.75 = 750$ coupons are expected to be received back.

b - Binomial probability distribution can be approximated by normal distribution.

$E(x) = \mu = 750$ $V(x) = \sigma^2 = 95$ $\sigma = \sqrt{187.5} = 13.69$

$P(x < 780) \Rightarrow$ z score corresponding to 780 is $z = \dfrac{780 - 750}{13.69} = 2.19$



$P(x < 780) = P(z < 2.19)$

$= 0.5 + 0.4857$

$= 0.9857$

MARMARA ÜNİVERSİTESİ YAYINEVİ

REFERENCES

Anderson, D.R., Sweeney, D.J., Williams, T.A. (1996). *Statistics for Business and Economics*. West Publishing Company. St. Paul. USA.

Anderson, T.W., Sclove, S.L. (1978). *An Introduction to the Statistical Analysis of Data*. Houghton Mifflin Company. Boston, USA.

Bowen, E. K., Starr M.K. (1994). *Basic Statistics for Business and Economics*. McGraw - Hill Inc., Litcratür Yayıncılık, İstanbul.

Gülçür, Fazıl K. (1979). *İstatistik Metodları*. Met-Er Matbaası İstanbul.

Köksal, Bilge A. (1985). *İstatistik Analiz Metodları*, Çağlayan Basımcvi, İstanbul.

Kvanli, A.H., Guynes, C.S., Pavur, R.J. (1996). *Introduction to Business Statistics; A Computer Integrated, Data Analysis Approach*. West Publishing Company. St. Paul. USA.

Lapin, L.L. (1990). *Statistics for Modern Decisions*. Harcourt Brace Jovanovich, Inc. San Fransisco, USA.

Lewin, R.I., (1981). *Statistics for Management*. Prentice Hall International, Inc. New Jersey. USA.

Mendenhall, W., Reinmuth, J.E., Beaver, R. (1989) *Statistics for Management and Economics*. PWS-KENT Publishing Company. USA.

Newbold, P. (1988). *Statistics for Business and Economics*. Prentice-Hall International Limited. New Jersey. USA.

Parsons, R. (1974). *Statistical Analysis: A Decision Making Approach*. Harper & Row Publishers, Inc. N.Y. USA.

Serper, Özer, Necmi Gürsakal (1989). *Araştırma Yöntemleri*. Filiz Kitabevi, İstanbul

Serper, Özer, Mustafa Aytaç (1988). *Örnekleme*. Filiz Kitabevi, İstanbul

Serper, Özer (1992). *Uygulamalı İstatistik 1*. Filiz Kitabevi, İstanbul.

Serper, Özer (1993). *Uygulamalı İstatistik 2*. Filiz Kitabevi, İstanbul.

SPSS for Windows, Advanced Statistics, Release 5 (1992). U.S.A.

Vuran, Ateş (1981). *İstatistik III*, İ.İ.T.İ.A. Nihad Sayar Yayın ve Yardım Vakfı Yayınları, İstanbul.

Vuran, Ateş (1982). *İstatistik II: İhtimaller, İstatistiksel Takdir*. İ.İ.T.İ.A. İşletme

Yamane, Taro. (1964). *Statistics, An Introductory Analysis*. Harper & Row Publishers, Inc. N.Y. USA.

MARMARA ÜNİVERSİTESİ YAYINEV

MARMARA ÜNİVERSİTESİ 1883

MARMARA
ÜNİVERSİTESİ

MARMARA ÜNİVERSİTESİ YAYINEVİ

Marmara Üniversitesi İletişim Fakültesi
Basın Yayın Teknikleri Anabilim Dalı
tarafından basılmıştır.

ISBN 975 - 400 - 146 - 4