زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

Comparative Study Of
Measure Of Central tendency

Research Project

Submitted to the department of (Mathematic) in partial fulfillment of the requirements for the degree of BSc. in (Mathematics)

By:

Savin Faraidun Pasha

Supervised By:

Dr. Ivan Subhy Latif

April-2024

Certification of the Supervisor

I certify that this work was prepared under my supervision at the Department of Mathematics / College of Education / Salahaddin University- Erbil in partial fulfillment of the requirements for the degree of Bachelor of philosophy of Science in Mathematics.

Signature:

Supervisor **: Dr. Ivan Subhy Latif**

Scientific grade**: Assistant Lecturer**

Date**: 7/ 4/2024**

In view of available recommendations, I forward this word for debate by the examining committee.

Signature:

Name**: Dr. Rashad Rashid Haji**

Scientific grade: **Assistant Professor**

Chairman of the Mathematics Department

Date**: 7 / 4 /2024**

ACKNOWLEDGMENT

First of all I would like to thanks God for helping me to complete this project with success. Secondly, I would like to express my special thanks to my supervisor (Dr.Ivan Subhy Latif) it has been great honor to her student.

It is great pleasure for me to undertake this project I have taken efforts however it would not have been possible without the support and help of many individuals.

Also I would like to express my gratitude towards my parents and my friends

My thanks appreciations go to Mathematical Department and all my valuable teachers.

ABSTRACT

In this project we find and study the Comparative Study Of Measure Of Central tendency and we distinguish . The three measures of central tendency are discussed in this article: the mode, the median, and the mean. These measures of central tendency describe data in different and important ways, in relation to the level of measurement (nominal, ordinal, interval, or ratio) used to obtain the data

Table of Contents

List Of Figures

INTRODUCTION

a measure of Central tendency is typical value around which other figures aggregate according to Croxton and Cowden an average is a single value with in the range of the data that is used to represent all the values in the series. since an average is somewhere within the range of data. it is sometimes called a measure of Central value according to proof bowely measures of Central tendency averages are statistics constants which enable us to comprehend in a single effort the significance of the whole.

Chapter one

Basic Definitions:-

1.1 Central Tendency (s.manikandan, 2023) :

Central tendency is defined as “the statistical measure that identifies a single value as representative of an entire distribution. It aims to provide an accurate description of the entire data. It is the single value that is most typical/representative of the collected data. The term “number crunching” is used to illustrate this aspect of data description. The mean, median and mode are the three commonly used measures of central tendency.

1.2 Mean (s.manikandan, 2023):

The mean represents the average value of the dataset. It can be calculated as the sum of all the values in the dataset divided by the number of values. Mean is the most commonly used measure of central tendency. There are different types of mean, viz. arithmetic mean, weighted mean, geometric mean (GM) and harmonic mean (HM). If mentioned without an adjective (as mean), it generally refers to the arithmetic mean.

1.3 Arithmetic Mean (Weisstein, Eric W, 2020):

The most common measure of central tendency is the arithmetic mean. In layman’s terms, the mean of data indicates an average of the given collection of data. It is equal to the sum of all the values in the group of data divided by the total number of values.

1.4 Harmonic Mean ((Weisstein, Eric , 2023):

The Harmonic Mean is defined as the reciprocal of the average of the reciprocals of the data values.. It is based on all the observations, and it is rigidly defined. Harmonic mean gives less weightage to the large values and large weightage to the small values to balance the values correctly. In general, the harmonic mean is used when there is a necessity to give greater weight to the smaller items. It is applied in the case of times and average rates.

1.5 Geometric Mean (Anon., 2019):

the geometric mean is defined as the nth root of the product of n numbers. It is noted that the geometric mean is different from the arithmetic mean. Because, in arithmetic mean, we add the data values and then divide it by the total number of values. But in geometric mean, we multiply the given data values and then take the root with the radical index for the total number of data values. For example, if we have two data, take the square root, or if we have three data, then take the cube root, or else if we have four data values, then take the 4th root, and so on.

1.6 Weighted Mean (madansky, 2017) :

A weighted average is a calculation that takes into account the varying degrees of importance of the numbers in a data set. In calculating a weighted average, each number in the data set is multiplied by a predetermined weight before the final calculation is made. A weighted average can be more accurate than a simple average in which all numbers in a data set are assigned an identical weight.

1.7 Root Mean Square (Quadratic Mean) (Alan R . jones, 2018):

Statistically, the root mean square (RMS**)** is the square root of the mean square, which is the arithmetic mean of the squares of a group of values. RMS is also called a quadratic mean and is a special case of the generalized mean whose exponent is 2. Root mean square is also defined as a varying function based on an integral of the squares of the values which are instantaneous in a cycle

1.8 Median (hogg , R.V and craig, 1995):

the median is defined as the middle value of a sorted list of numbers. The middle number is found by ordering the numbers. The numbers are ordered in ascending order . Once the numbers are ordered, the middle number is called the median of the given data set.

1.9 Mode (A.M. Mood, F.A. Graybil, 1963):

In statistics, the mode is the value that is repeatedly occurring in a given set. We can also say that the value or number in a data set, which has a high frequency or appears more frequently, is called mode or modal value.

1.10 Range (E. J. Gumbel., 1947):

The range in statistics for a given data set is the difference between the highest and lowest values. For example, if the given data set is {2,5,8,10,3}, then the range will be 10 – 2 = 8.

Chapter two

Measures Of Central Tendency

The measures of central tendency help you understand the typical or central value with in a data set . which can be useful for making comparisons and drawing conclusions about the data .

It's important to choose the appropriate measure based on the nature of your data and the specific questions you're trying to answer .

For example , if your data contains extreme outliers, the median may be a more robust measure of central tendency than the mean .

The goal of central tendency is to identify the single value that is the best representative for the entire set of data

Generally , the central tendency of a data set can be described using the following measures :

- Mean (average)
- Median
- Mode

Mean :

The mean represents the average value of the data set . it can be calculated as the sum of all the values in the data set divided by the number of values in general it is considered as the arithmetic mean

The concept of the mean can be traced back to ancient civilizations such as Egypt , Greece , and China . These early civilizations used various methods to calculate averages of numerical data , often for practical purpose like astronomy , Taxation , and Trade .

Some other measures of mean used to find the central tendency are as follows :

Arithmetic Mean :

For ungrouped data ;

$$\bar{X} = (x_1 + x_2 + x_3 + \ldots + x_n)/n$$

For grouped data :

$$\bar{X} = \frac{\Sigma(fX)}{n}$$

Geometric Mean :

For ungrouped data :

$$\bar{X} = \sqrt[n]{(x1 * x2 * x3 * \ldots * xn)}$$

$$\bar{X} = Antilog \frac{(\sum \log x_i)}{n}$$

For grouped data :

$$\bar{X} = \sqrt[\Sigma f_i]{{x_1}^{f_1} * {x_2}^{f_2} * {x_3}^{f_3} * \ldots * {x_n}^{f_n}}$$

$$\bar{X} = Antilog \frac{(\sum f_i \log x_i)}{\Sigma f_i}$$

Harmonic Mean :

For ungrouped data :

$$\bar{X} = \frac{n}{\Sigma(\frac{1}{x})}$$

For grouped data :

$$\bar{\bar{X}}=\frac{\sum f}{\sum\left(\frac{f}{x}\right)}$$

Quadratic Mean :

For ungrouped data :

$$\bar{X} = \sqrt{\frac{\sum x_i^2}{n}}$$

For grouped data :

$$\bar{X} = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i}}$$

Weighted Mean :

For ungrouped and grouped data :

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i . x_i}{\sum_{i=1}^{n} f_i}$$

Root Mean Square :

For ungrouped data :

$$\bar{X} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}}$$

For grouped data:

$$\bar{X} = \sqrt{\frac{\sum_{i=1}^{n} f_i . x_i^2}{N}}$$

It’s observed that if all the values in the dataset are the same then all geometric , harmonic , and arithmetic values are the same . if there is variability in the data , then the mean vale differs .

Median :

Is the middle value of the data set in which the data set is arranged in the ascending order or in descending order . when the data set contains an even number of the data set can be found by taking the mean of the middle two values .
The concept of finding the middle value in data set can be traced back to ancient civilizations like the Egyptians and babylonians they used simple methods to find the middle value in a set of numbers .
During the middle ages , scholars in europa contained to develop mathematical and statistical concepts . the idea of the median was used in various applications , including , astronomy and land surveys
Today , the median is a fundamental concept in statistics and is widely is widely used in various fields, including economics , social sciences , and data analysist .
Median is valuable measure of central tendency particularly when dealing with data sets that may contain outliers or extreme values , as its less affected by such values than the mean . the historical development of the median reflects the ongoing evolution of statistical techniques and their application to real world problems.

- For un grouped data :

1. If the total number of observations (n) is odd, then the median is (n+1)/2 th observation.

2. If the total number of observations (n) is even, then the median will be average of n/2th and the (n/2)+1 th observation.

- **For grouped data :**

$$\textbf{Median}=\boldsymbol{\iota}+\left(\frac{\frac{\sum f_i}{2}-cf}{f}\right)*\boldsymbol{h}$$

l is the lower limit of the median class

n is the number of observations

f is the frequency of median class

h is the class size

cf is the cumulative frequency of class preceding the median class.

Mode :

The mode represents the frequently occurring value in the data set sometimes the data set may contain multiple modes and in some cases it does not contain any mode at all

the concept of the mode can be treated back to ancient civilization like the Egyptian and Greeks who used it in various aspects of their daily lives including and surveys and taxation

during the European Renaissance in the 16th and 17th centuries there was a growing interest in collecting and analyzing data scholars like in girolamo cardano and Galileo Galilei discussed the concept of the mode in their work

today the most remains a fundamental statistical concept and is widely used it in various Fields including mathematics, economics ,social sciences and natural sciences, to describe the central or typical value in a data set of data it is especially useful when dealing with categorical or discrete data

- For grouped data :

The formula to find the mode of the grouped data is:
Mode $= \boldsymbol{l_{m_0}} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) * \boldsymbol{h}$

- For ungrouped data :

To find the mode for ungrouped data, find the observation that occurs the maximum number of times.

COMPARISON

1. mean provides a measure of the center by considering all values but it can be skewed by extreme values. median on other hand is not affected by extreme values, making it more appropriate for skewed data set or when outliers are present.

2. the mode suitable for identifying the most common values in a data set but it may not be applicable or useful for continuous or continuous like data.

3. the choice of measure of Central tendency depends on the nature of the data and the specific question you want to answer for normally distributed data ,the mean is often used for skewed data or when dealing with outliers, the median is a better choice the mode is most commonly used for categorical data or to identify the most frequent category in a data set

4. it's also possible to use these measures to gain a more comprehensive understanding of the data's central tendency ,especially when the data distribution is complex or not behaved

Chapter three

Examples:-

1. What is the mean of 2, 4, 6, 8 and 10?

Solution:

First, add all the numbers.

2+ 4 + 6 + 8 + 10 = 30

Now divide by 5 (total number of observations).

Mean = 30/5 = 6

2. Find the mean for the following distribution.

x_i	11	14	17	20
f_i	3	6	8	7

Solution:

For the given data, we can find the mean using the direct method.

x_i	f_i	f_ix_i
11	3	33
14	6	84
17	8	136
20	7	140
	$\sum f_i = 24$	$\sum f_i\, x_i = 393$

Mean = $\sum f_ix_i / \sum f_i$ = 393/24 = 16.4

3. **The following data represents the survey regarding the heights (in cm) of 51 girls of Class x. Find the median height.**

Height (in cm)	**Number of Girls**
Less than 140	4
Less than 145	11
Less than 150	29
Less than 155	40
Less than 160	46
Less than 165	51

Solution:

To find the median height, first, we need to find the class intervals and their corresponding frequencies.

4 girls are below 140. Therefore, the frequency of class intervals below 140 is 4.
11 girls are there with heights less than 145, and 4 girls with height less than 140
Hence, the frequency distribution for the class interval 140-145 = 11-4 = 7
Likewise,
the frequency of 145 -150= 29 – 11 = 18
Frequency of 150-155 = 40-29 = 11
Frequency of 155 – 160 = 46-40 = 6
Frequency of 160-165 = 51-46 = 5

Therefore, the frequency distribution table along with the cumulative frequencies are given below:

Class Intervals	Frequency	Cumulative Frequency
Below 140	4	4
140 – 145	7	11
145 – 150	18	29
150 – 155	11	40
155 – 160	6	46
160 – 165	5	51

Here, n= 51.
Therefore, n/2 = 51/2 = 25.5
Thus, the observations lie between the class interval 145-150, which is called the median class.
Therefore,
Lower class limit = 145
Class size, h = 5
Frequency of the median class, f = 18
Cumulative frequency of the class preceding the median class, cf = 11.
We know that the formula to find the median of the grouped data is:

Median=$\iota + \left(\frac{\frac{n}{2} - cf}{f}\right) * h$

Now, substituting the values in the formula, we get
Median=145+(25.5−1118)×5
Median = 145 + (72.5/18)
Median = 145 + 4.03
Median = 149.03.
Therefore, the median height for the given data is 149. 03 cm.

4. **The following table gives the frequency distribution of the number of orders received each day during the past 50 days at the office of a mail-order company. Calculate the mean.**

Numbers of order	f
10-12	4
13-15	12
16-18	20
19-21	14
	n=50

Solution

Numbers of order	x	f	fx
10-12	11	4	44
13-15	14	12	168
16-18	17	20	240
19-21	20	14	280
	N=50		=832

X is the midpoint of the class. It is adding the class limits and divide by 2.

$$\bar{x} = \frac{\sum fx}{n} = \frac{832}{50} = 16.64$$

5. **Based on the grouped data below, find the median:**

Time to travel to work	frequency
1-10	8
11-20	14
21-30	12
31-40	9
41-50	7

Solution:

1st Step: Construct the cumulative frequency distribution

Time to travel to work	frequency	Cumulative Frequency
1-10	8	8
11-20	14	22
21-30	12	34
31-40	9	43
41-50	7	50

$$\frac{n}{2} = \frac{50}{2} = 25 \rightarrow class\ median\ is\ the\ 3rd\ class$$

So, $cf = 22$, $f_m = 12$, $l_m = 20.5$, $and\ h = 10$

Therefore,

Median= $\iota + \left(\frac{\frac{n}{2} - cf}{f}\right) * h$

=20.5+$\left(\frac{25-22}{12}\right) * 10$

=24

Thus, 25 persons take less than 24 minutes to travel to work and another 25 persons take more than 24 minutes to travel to work.

Using the same method of calculation as in the Median, we can get Q1 and Q3 equation as follow

$$Q_1 = l_{q_1} + \left(\frac{\frac{n}{4} - cf}{f_{q_1}}\right) * h \qquad Q_3 = l_{q_3} + \left(\frac{\frac{3n}{4} - cf}{f_{q_3}}\right) * h$$

- Class $Q_1 = \frac{n}{4} = \frac{50}{4} = 12.5$

  Class Q1 is the 2nd class Therefore,

$$Q_1 = l_{q_1} + \left(\frac{\frac{n}{4} - cf}{f_{q_1}}\right) * h$$

$$=10.5+(\frac{12.5-8}{14}) * 10$$

$$=13.7143$$

- Class $Q_3 = \frac{3n}{4} = \frac{3*50}{4} = 37.5$
  Class Q3 is the 4th class Therefore,

$$Q_3 = l_{q_3} + \left(\frac{\frac{3n}{4}-cf}{f_{q_3}}\right) * h$$

$$=30.5+(\frac{37.5-34}{9}) * 10$$

$$=34.3889$$

Interquartile Range

$IQR = Q_3 - Q_1$
=34.3889-13.7143
=20.6746

6. Based on the grouped data below, find the mode

Time to travel to work	frequency
1-10	8
11-20	14
21-30	12
31-40	9
41-50	7

Solution:

Mode =$\boldsymbol{l_{m_0} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) * h}$

$\boldsymbol{\Delta_1 = 14 - 8 = 6}$

$\boldsymbol{\Delta_2 = 14 - 12 = 2}$

$\boldsymbol{l_{m_0} = 10.5}$

h=10

mode =10.5+$\boldsymbol{(\frac{6}{6-2}) * 10 = 17.5}$

7. **The following table is a collecting information about the weight of ((40)) students of mathematic department in salahaddin University , which are recorded to nearest kilogram**

   40 , 50 , 55 , 67 , 50 , 65 , 82 , 52 , 65 , 44 , 45 , 52 , 50 , 81
   57 , 44 , 65 , 76 , 79 , 52 , 49 , 50 , 78 , 61 , 69 , 68 , 56 , 77
   74 , 67 , 72 , 72 , 73 , 63 , 80 , 47 ,48 , 53 , 42 , 60

Solution:

No,of class = 2.5*$\sqrt[4]{40} \approx 7$

Range = 81-40 = 41

Size=$\frac{range}{No,of\ class} = \frac{41}{7} \approx 6$

No. of class	**class**	**Class mark**	**Frequency f_i**	**Class boundary**	**Less than**		**More than**		rf_i	pf_i
1	39-45	42	5	38.5-45.5	39	0	39	40	0.125	12.5
2	46-52	49	10	45.5-52.5	46	5	46	35	0.25	25
3	53-59	56	4	52.5-59.5	53	15	53	25	0.1	10
4	60-66	63	6	59.5-66.5	60	19	60	21	0.15	15
5	67-73	70	7	66.5-73.5	67	25	67	15	0.175	17.5
6	74-80	77	6	73.5-80.5	74	32	74	8	0.15	15
7	81-87	84	2	80.5-87.5	81	38	81	2	0.05	5
			$\sum f_i = 40$		88	40	88	0		

Arithmetic mean :

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{5(42)+10(49)+4(56)+6(63)+7(70)+6(77)+2(84)}{40} = \frac{2422}{40} = 60.55$$

Geometric mean :

$$\bar{X} = \sqrt[\Sigma f_i]{{x_1}^{f_1} * {x_2}^{f_2} * {x_3}^{f_3} * \ldots * {x_n}^{f_n}}$$

$$= \sqrt[40]{(42)^5 * (49)^{10} * (56)^4 * (63)^6 * (70)^7 * (77)^6 + (84)^2} = 5.06015$$

Harmonic mean :

$$\bar{X} = \frac{\Sigma f}{\Sigma(\frac{f}{x})} = \frac{40}{\frac{5}{42}+\frac{10}{49}+\frac{4}{56}+\frac{6}{63}+\frac{7}{70}+\frac{6}{77}+\frac{2}{84}} = 57.84296$$

Quadratic mean (RMS) :

$$\sqrt{\frac{\sum f_i x_i^2}{\sum f_i}} = \sqrt{\frac{(5*40^2)+(10*49^2)+(4*56^2)+(6*63^2)+(7*70^2)+(6*77^2)+(2*84^2)}{40}} = 9724.57$$

Median :

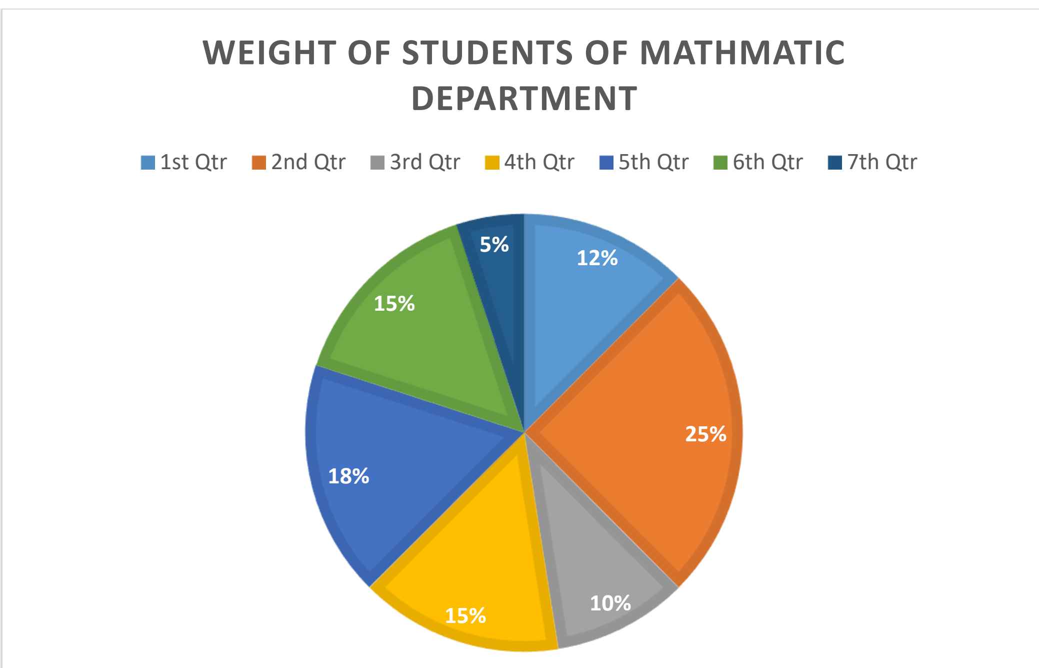
$$\textbf{Median}=\iota + \left(\frac{\frac{\sum f_i}{2}-cf}{f}\right) * \boldsymbol{h}= 59.5+(\frac{20-19}{6}) * 7=60.6$$

Mode :

$$\text{Mode} = \boldsymbol{l_{m_0}} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) * \boldsymbol{h} = 45.5+(\frac{10}{6+10}) * 7=49.875$$

Pie chart :

Figure 1



Graphical representation :

Figure 2

✓ Histogram :

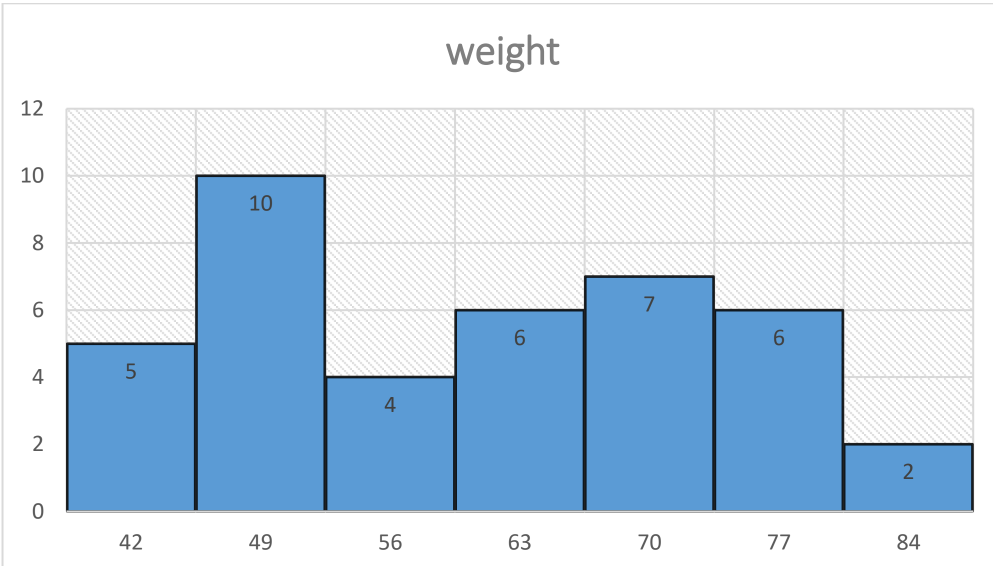
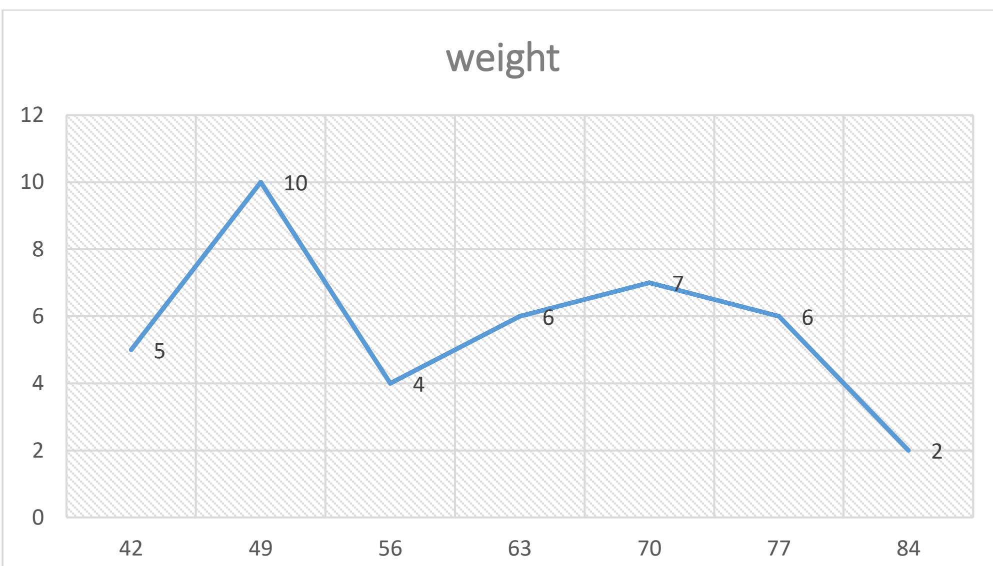


Figure 3

✓ Frequency polygon :



References

1. (Weisstein, Eric , 2023. *harmonic mean,* s.l.: mathworld.wolfram.com.
2. A.M. Mood, F.A. Graybil, 1963. *Introduction to the theory of statistics,* s.l.: McGraw-Hill.
3. Alan R . jones, 2018. *Probability, Statistics and Other Frightening Stuff.* london and newyork: routledge.
4. Anon., 2019. *"2.5: Geometric Mean",* s.l.: Statistics LibreTexts.
5. dr. awaz kakamam, 2022. frequency distribution. In: *probability and statistics.* s.l.:s.n.
6. E. J. Gumbel., 1947. The Distribution of the Range. In: *the annals of mathematical statistics.* new york: s.n., pp. 384-412.
7. hogg , R.V and craig, 1995. *introduction to mathematical statistics.* 5th ed. newyork: macmillan.
8. Hogg, Mckean , Craig, 2019. *introduction to mathematical statistics.* 8th ed. boston: pearson.
9. madansky, D., 2017. *http://www.analyticalgroup.com/download/WEIGHTED_MEAN.pdf,* s.l.: s.n.
10. s.manikandan, 2023. frequency distribution. *journal of pharmacology and pharmacotherapeutics,* 14(2).
11. Weisstein, Eric W, 2020. *"Arithmetic Mean",* s.l.: mathworld.wolfram.com.

پوختە

لەم پرۆژەیەدا ئێمە لێکۆڵینەوەی بەراوردکاریی پێوانەکردنی مەیلی ناوەندی دەدۆزینەوە و لێی دەکۆڵینەوە و جیاوازی دەکەین . سێ پێوەرەکەی مەیلی ناوەندی لەم بابەتەدا باس دەکرێن: شێواز، ناوەندی و مامناوەند. ئەم پێوەرانەی مەیلی ناوەندی داتاکان بە شێوازی جیاواز و گرنگ وەسف دەکەن، لە پێوەندی لەگەڵ ئاستی پێوانەکردن (ناوی، ڕیزبەندی، نێوان، یان ڕێژە) کە بۆ بەدەستهێنانی داتاکان بەکاردەهێنرێن