a) Total number of customers = (31 + 78 + 49 + 81 + 117 + 13) = 369

So, a total of 369 customers were served on this day at this gas station.

b) Table for class width and class midpoints,

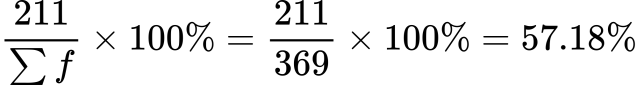
|  |  |  |
| --- | --- | --- |
| Class boundary | Class width | Class midpoint |
| 0 to <4  4 to <8  8 to <12  12 to <16  16 to <20  20 to <24 | 4  4  4  4  4  4 | 2  6  10  14  18  22 |

c) Table for relative frequency and percentage,

|  |  |  |  |
| --- | --- | --- | --- |
| Class Boundaries | f | Relative frequency | Percentage |
| 0 to <4  4 to <8  8 to <12  12 to <16  16 to <20  20 to <24 | 31  78  49  81  117  13 | 0.08  0.21  0.13  0.22  0.32  0.04 | 8.40  21.14  13.28  21.95  31.71  3.52 |
|  | Σf  = 369 | Σ  = 1 | Σ  = 100 |

d) Total number of customers purchased 12 gallons or more = (81 + 117 + 13) = 211

So, the percentage =



e) Since, 10 is not a boundary value, therefore, we cannot determine exactly how many customers purchased 10 gallons or less.

2.

1. Table for class limits,

|  |
| --- |
| Class limit |
| 1 to 200  201 to 400  401 to 600  601 to 800  801 to 1000  1001 to 1200 |

1. Table for class boundary and class midpoints,

|  |  |  |
| --- | --- | --- |
| Class | Class boundary | Class midpoint |
| 1 to 200  201 to 400  401 to 600  601 to 800  801 to 1000  1001 to 1200 | 0.5 to <200.5  200.5 to <400.5  400.5 to <600.5  600.5 to <800.5  800.5 to <1000.5  1000.5 to <1200.5 | 100.5  300.5  500.5  700.5  900.5  1100.5 |

3.

1. Minimum value of data = 2.2

Maximum value of data = 22.8

Let’s consider, class width = 5

So, number of classes =

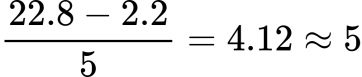


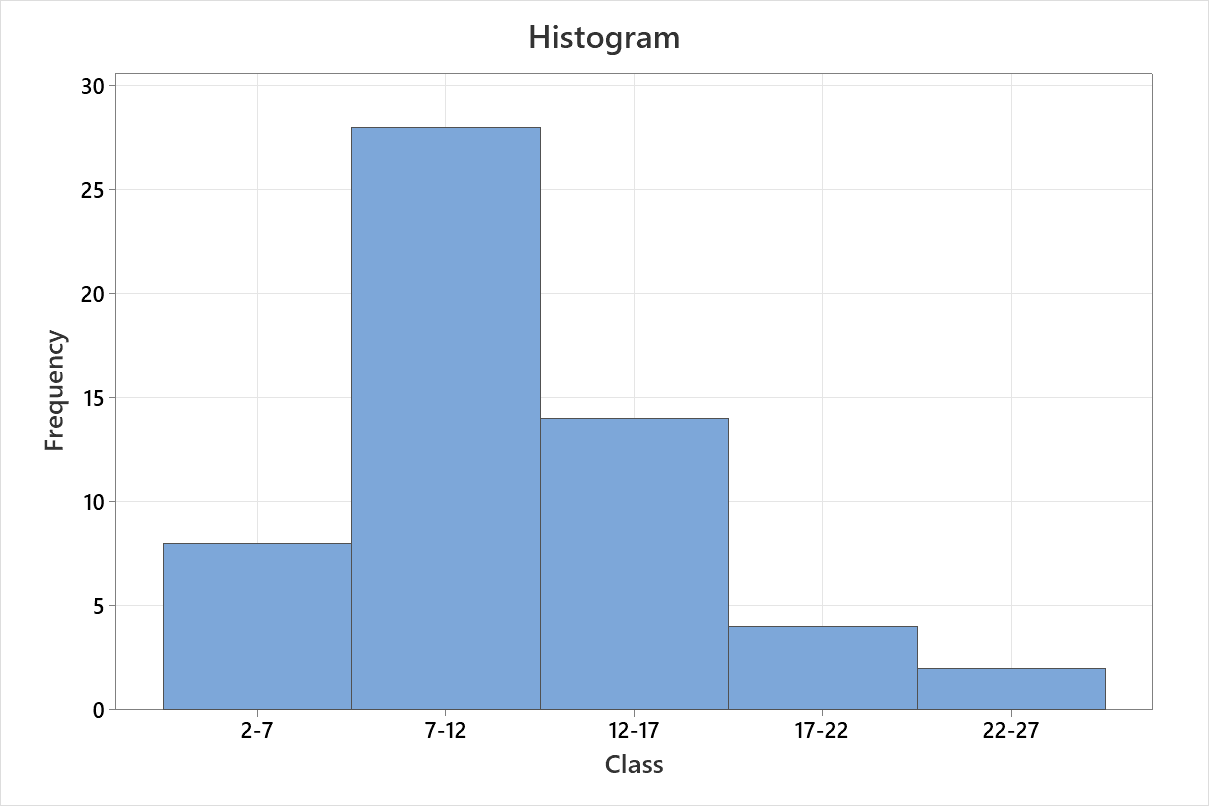
Table for frequency distribution,

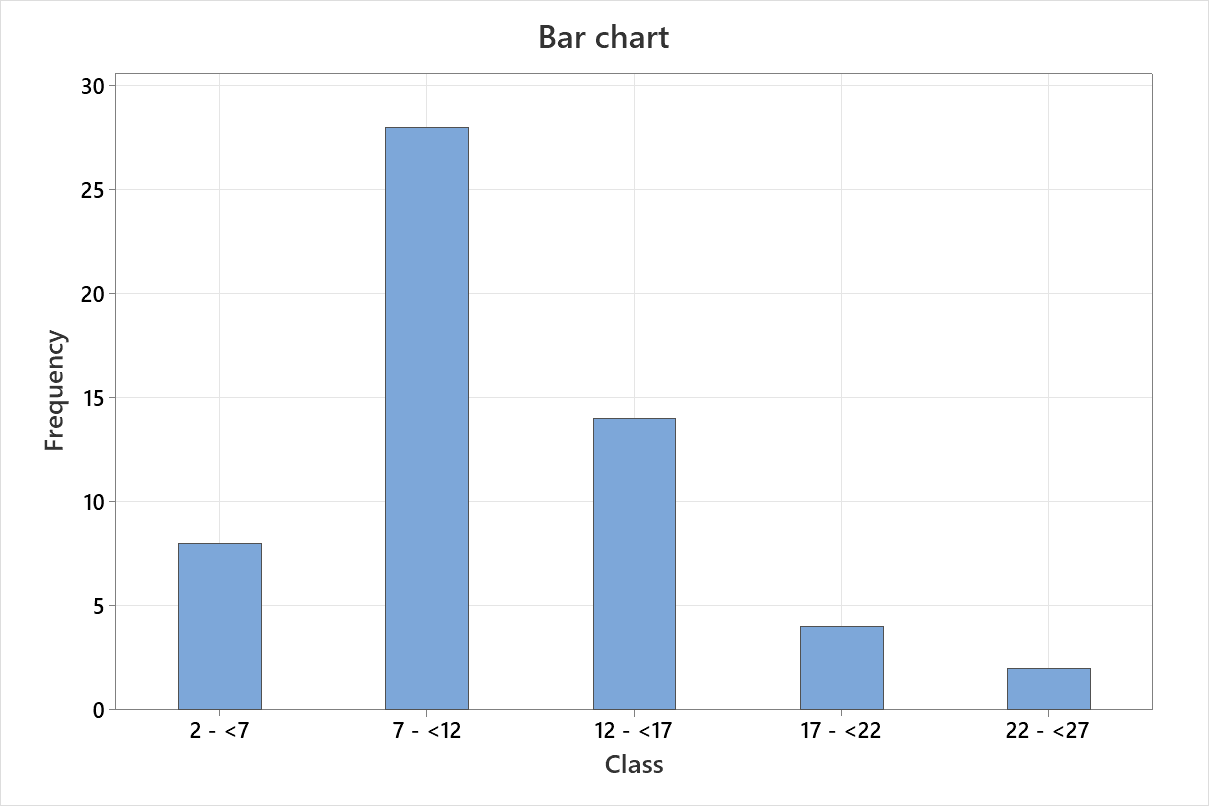
|  |  |  |
| --- | --- | --- |
| Class boundary | Tally | f |
| 2 to <7  7 to <12  12 to <17  17 to <22  22 to <27 | ~~||||~~ |||  ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ |||  ~~||||~~ ~~||||~~ ||||  ||||  || | 8  28  14  4  2 |
|  |  | Σf  = 56 |

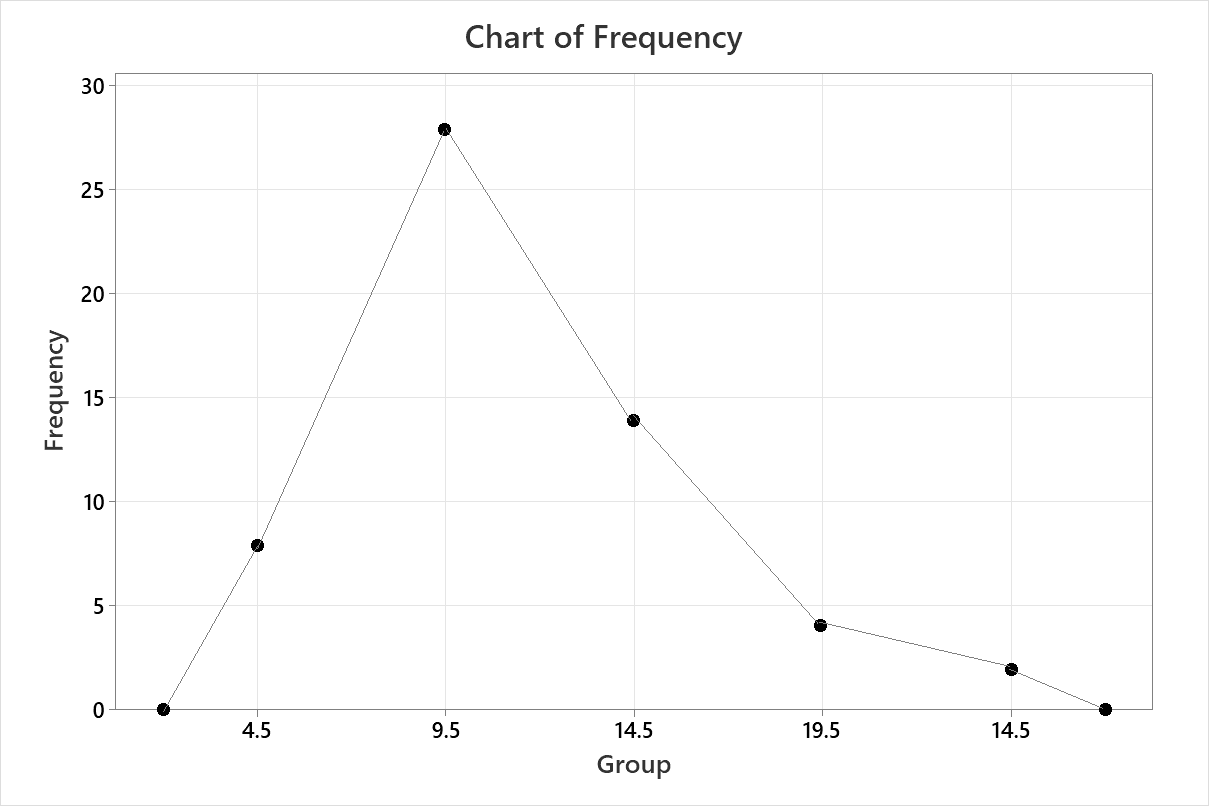
1. Table for relative frequency and percentage,

|  |  |  |  |
| --- | --- | --- | --- |
| Class boundary | f | Relative frequency | Percentage |
| 2 to <7  7 to <12  12 to <17  17 to <22  22 to <27 | 8  28  14  4  2 | 0.14  0.50  0.25  0.07  0.04 | 14.29  50.00  25.00  7.14  3.57 |
|  | Σf  = 56 | Σ  = 1 | Σ  = 100 |

1. Histogram, bar diagram and a polygon for the birth rate percentage distribution,







Polygon

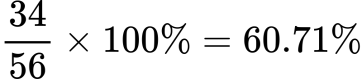
1. Data less than 11 =

2.2 3.1 4.6 5.4 5.6 5.6 6.5 6.6 7.4 7.7 7.8 8.1 8.2 8.5 8.6 8.8 8.9 9.3 9.4 9.6 9.7 9.7 9.8 9.8 9.9 10.1 10.2 10.2 10.3 10.5 10.5 10.8 10.9 10.9

Number of data less than 11 = 34

Total frequency, Σf = 56

percentage of the countries had a birth rate of less than 11 births per 1000 people =



4.

1. Smallest data = 1039

Largest data = 5490

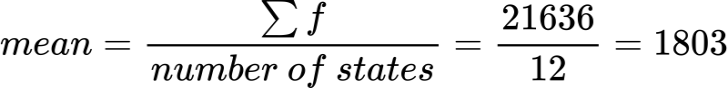
Let’s consider, class width = 1000

In one possible table,

Then 5 classes will hold all the data values if started from 1001.

Table for frequency (f) and cumulative frequency (F),

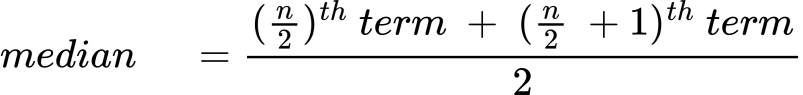
|  |  |  |
| --- | --- | --- |
| Class | f | F |
| 1001-2000  2001-3000  3001-4000  4001-5000  5001-6000 | 8  3  0  0  1 | 8  11  11  11  12 |
|  | Σf  = 12 |  |

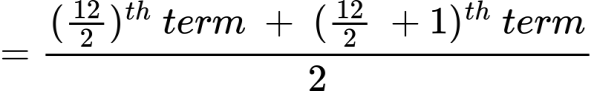


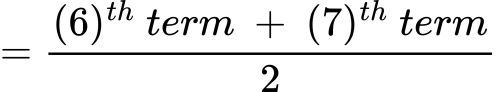
Data in ascending order =

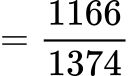
1039 1086 1113 1137 1139 1166 1374 1673 2009 2110 2300 5490

Number of states, n = 12 , an even number









wps

wps

1. From (b),

mean = 1803

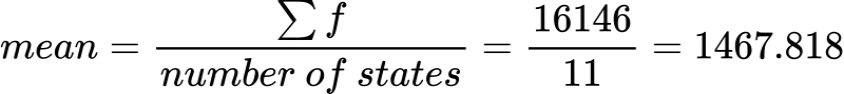
median = 1270

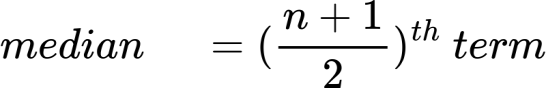
The outlier in the data set = 5490

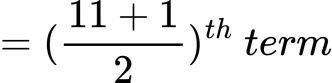
Data in ascending order after dropping the outlier =

1039 1086 1113 1137 1139 1166 1374 1673 2009 2110 2300

mean and median without the outlier;







wps

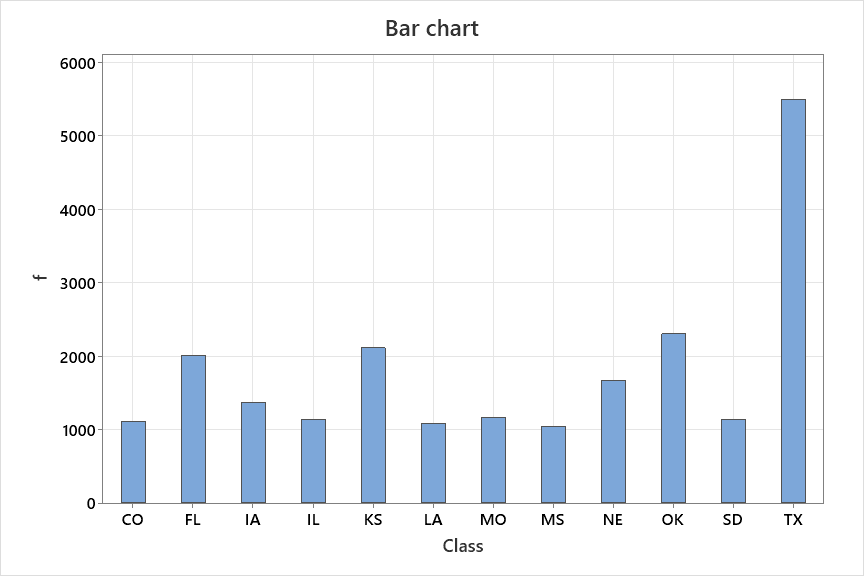
wps

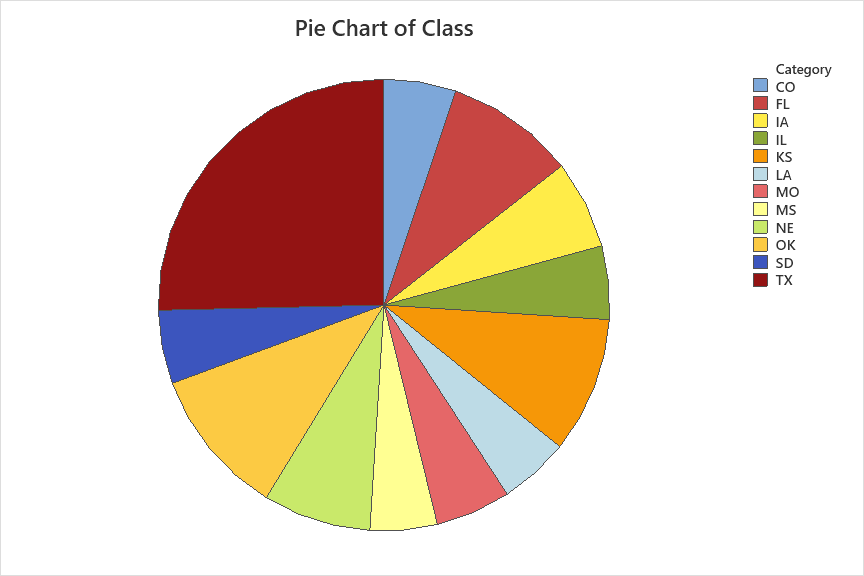
Difference in mean = 1803 - 1467.818 = 335.182

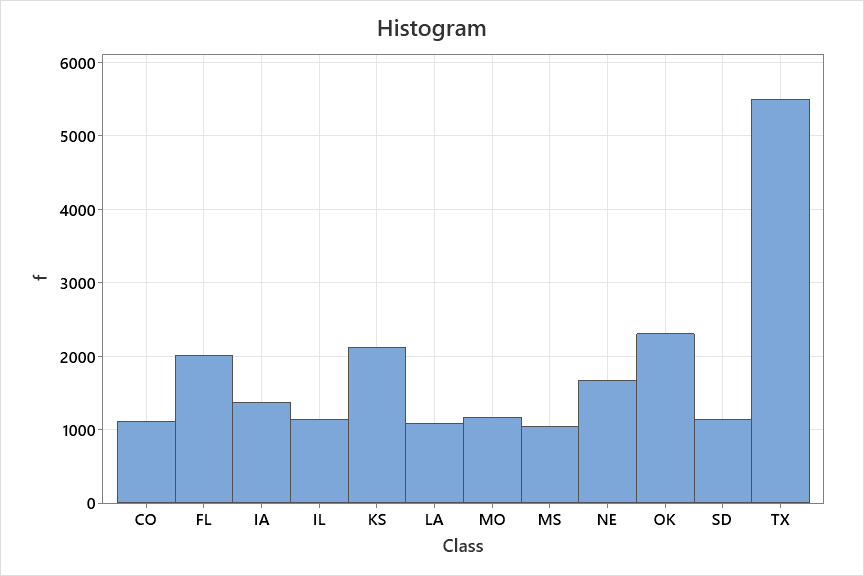
Difference in median = 1270 - 1166 = 104

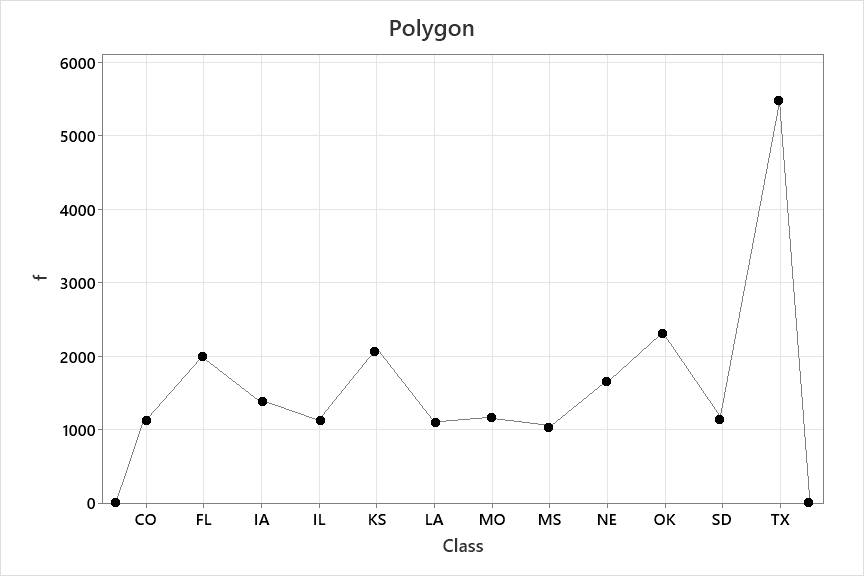
Therefore, mean is the summary measures changes by a larger amount when dropped the outlier.

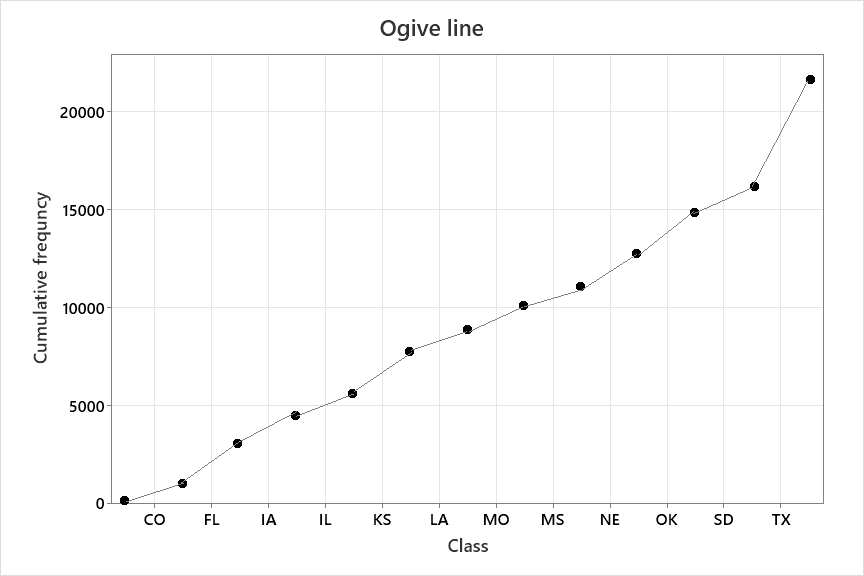
1. Since there is a significant outlier in the data set, therefore median is the better summary measure for these data than the mean and the mode.
2. bar diagram, pie chart, histogram, and polygon, ogive line











5.

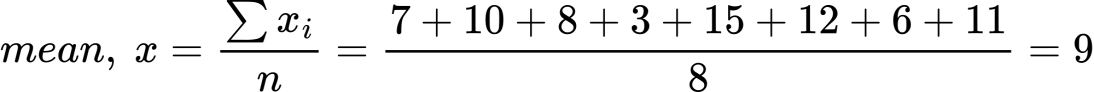


Table for deviations of the data values from the mean (xi - wps),

|  |  |
| --- | --- |
| x | x- wps |
| 7  10  8  3  15  12  6  11 | -2  1  -1  -6  6  3  -3  2 |
|  | Σ  = 0 |

Yes, the sum of these deviations is zero.

1. Smallest data value = 3

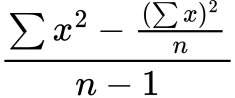
Largest data value = 15

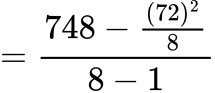
Range = Largest - Smallest = 15 - 3 = 12

Table for calculating variance,

|  |  |
| --- | --- |
| x | x2 |
| 7  10  8  3  15  12  6  11 | 49  100  64  9  225  144  36  121 |
| Σx  = 72 | Σx2  = 748 |

Number of sample, n = 8

variance, s2 =



wps

standard deviation, s =

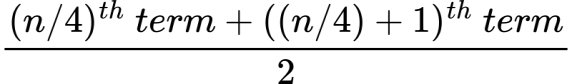
=

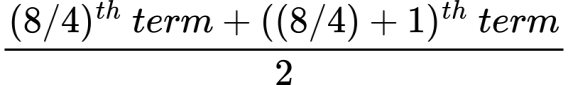
= 3.78

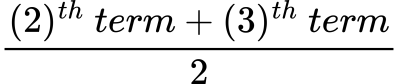
Data in ascending order,

3 6 7 8 10 11 12 15

Since, n = 8, which is divisible by 4

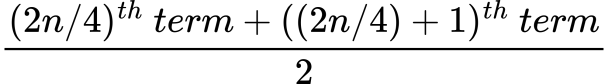
First quartile, Q1 = 

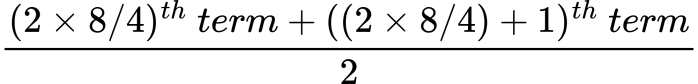
= 

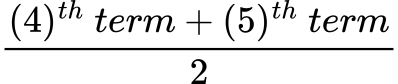
=

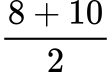
= 

= 6.5

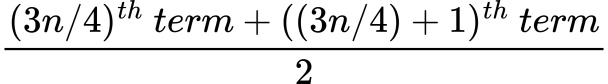
Second quartile, Q2 = 

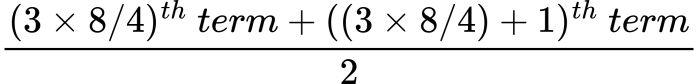
= 

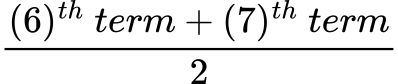
=

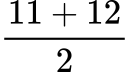
= 

= 9

Third quartile, Q2 = 

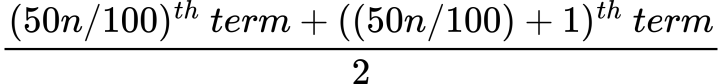
= 

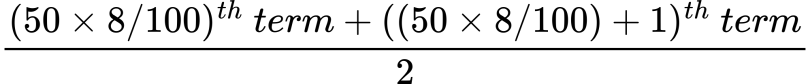
=

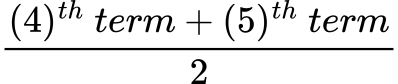
= 

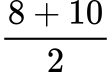
= 11.5

Since 50 n = 50 8 = 400 ; is divisible by 100

50th percentile, P50 = 

= 

= 

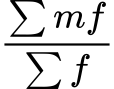
= 

= 9

6.

1. Table for determining mean,

|  |  |  |  |
| --- | --- | --- | --- |
| Class | m | f | mf |
| 10-20  20-30  30-40  40-50  50-60 | 15  25  35  45  55 | 5  8  12  7  4 | 75  200  420  315  220 |
|  |  | Σf  = 36 | Σmf  = 1230 |

mean = 

= 

= 34.1667

1. Table for determining median,

|  |  |  |
| --- | --- | --- |
| Class | f | F |
| 10-20  20-30  30-40  40-50  50-60 | 5  8  12  7  4 | 5  13  25  32  36 |
|  | Σf  = 36 |  |

Here,

l = 30

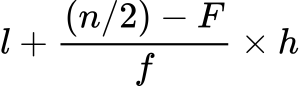
n = Σf = 36

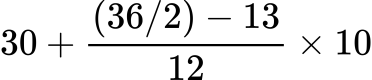
n/2 = 36/2 = 18

F = 13

f = 12

h = 20-10 = 10

median = 

= 

= 34.1667

1. Table for determining mode,

|  |  |
| --- | --- |
| Class | f |
| 10-20  20-30  30-40  40-50  50-60 | 5  8  12  7  4 |
|  | Σf  = 36 |

Here,

l = 30

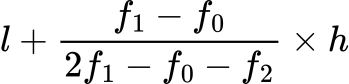
n = Σf = 36

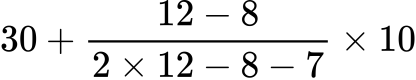
n/2 = 36/2 = 18

F = 13

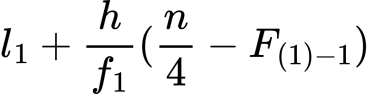
f = 12

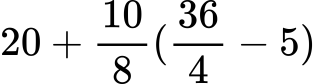
h = 20-10 = 10

mode = 

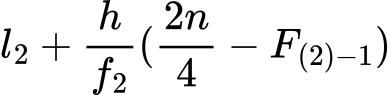
= 

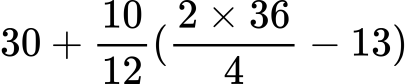
= 34.4444

1. First quartile, Q1 = 

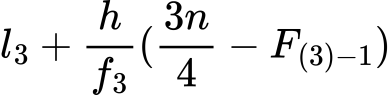
=

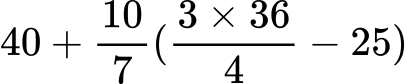
= 25

Second quartile, Q2 = 

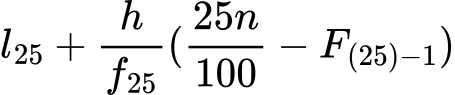
=

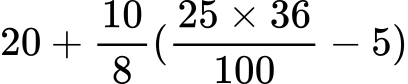
= 34.1667

Third quartile, Q3 = 

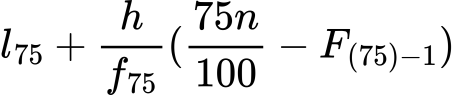
=

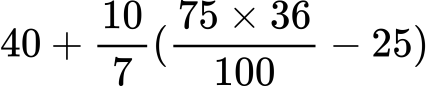
= 42.8571

25th percentile, P25  = 

=

= 25

75th percentile, P75  = 

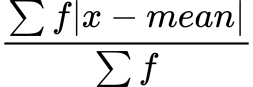
=

= 42.8571

1. Table for mean deviation and coefficient of mean deviation, range coefficient, quartiles deviation, variance, standard deviation,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | x | |x-mean| | f | f|x-mean| | f(x-mean)2 |
| 10-20  20-30  30-40  40-50  50-60 | 15  25  35  45  55 | 19.17  9.17  0.83  10.83  20.83 | 5  8  12  7  4 | 95.85  73.36  9.96  75.81  83.32 | 1837.44  672.71  8.27  821.02  1735.56 |
|  |  |  | Σf  = 36 | Σ  = 338.3 | Σ  = 5075 |

Here, mean = 34.17

mean deviation = 

=

= 9.3972

Coefficient of mean deviation = (mean deviation) / mean

= 9.3972 / 34.17

=0.28

Range coefficient =

=

= 0.57

Quartile deviation =

=

= 0.26

variance = (∑ f(x – mean)2)/n

= (5075)/36

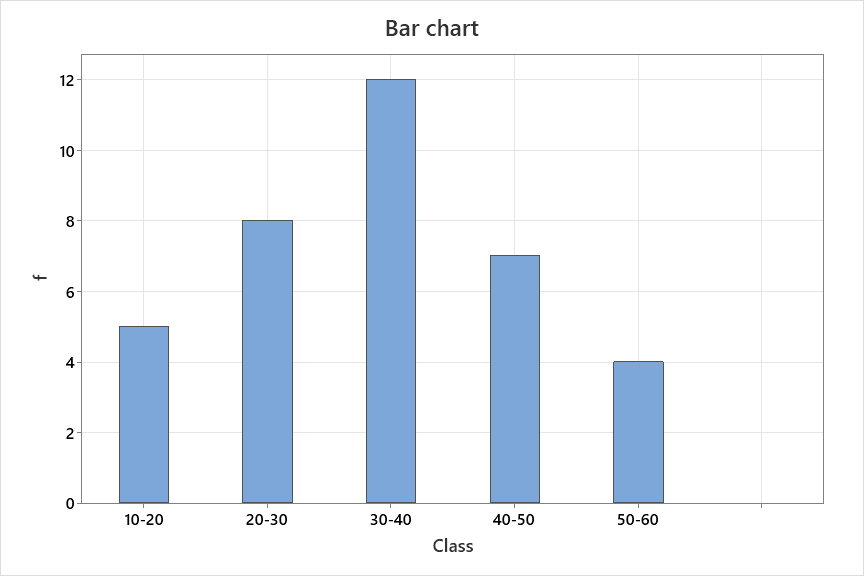
= 140.97

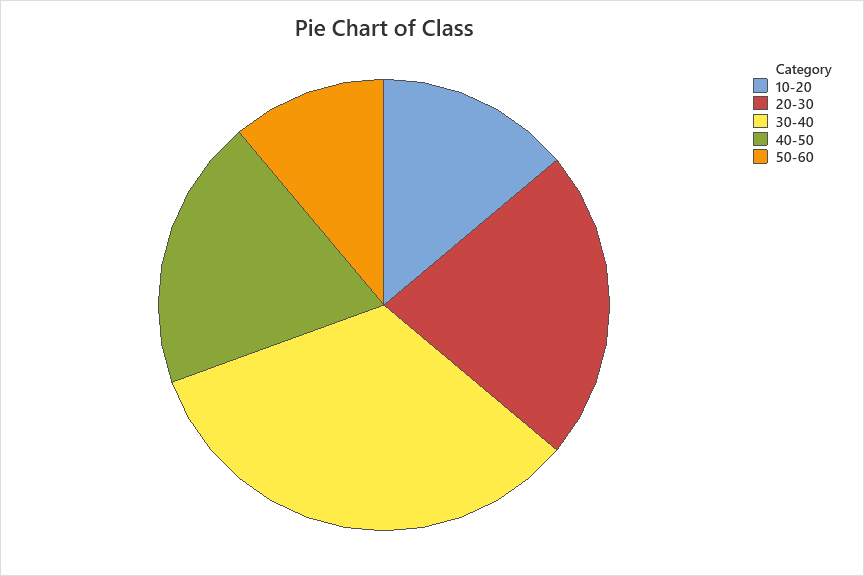
standard deviation =

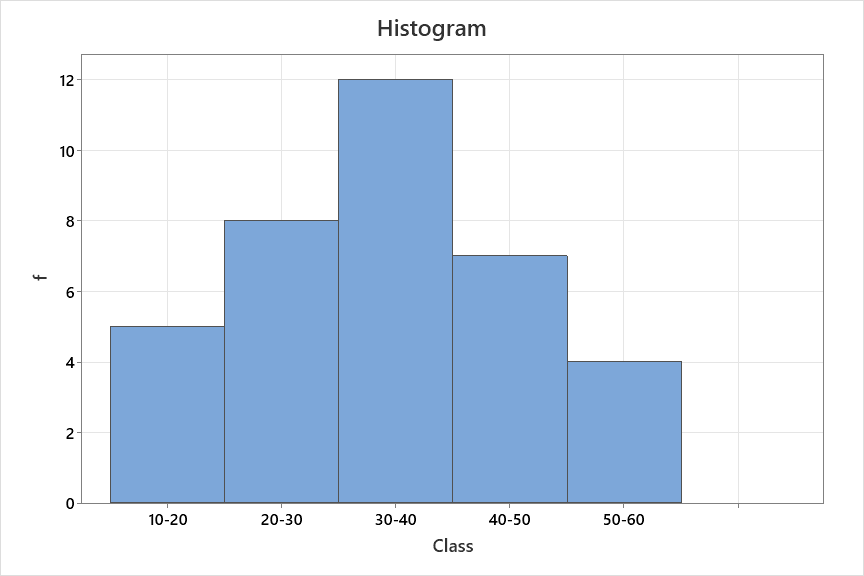
=

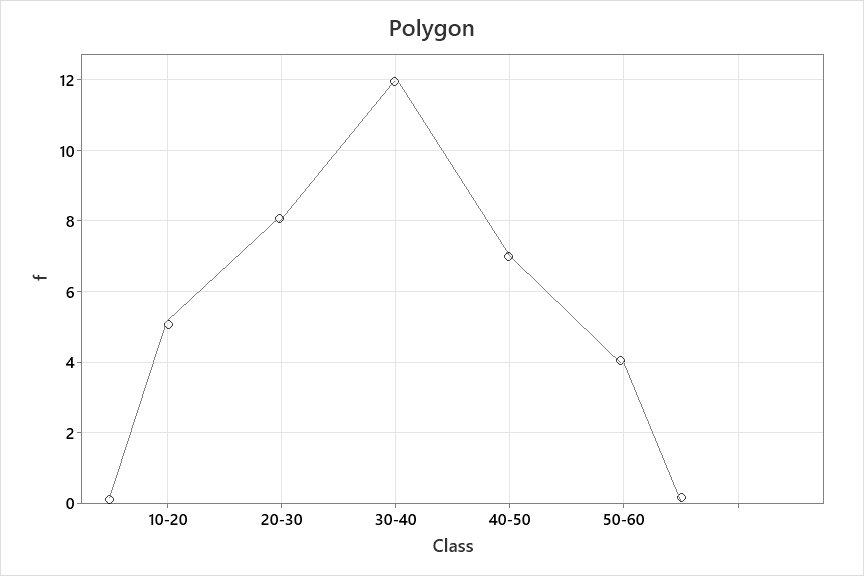
=11.87

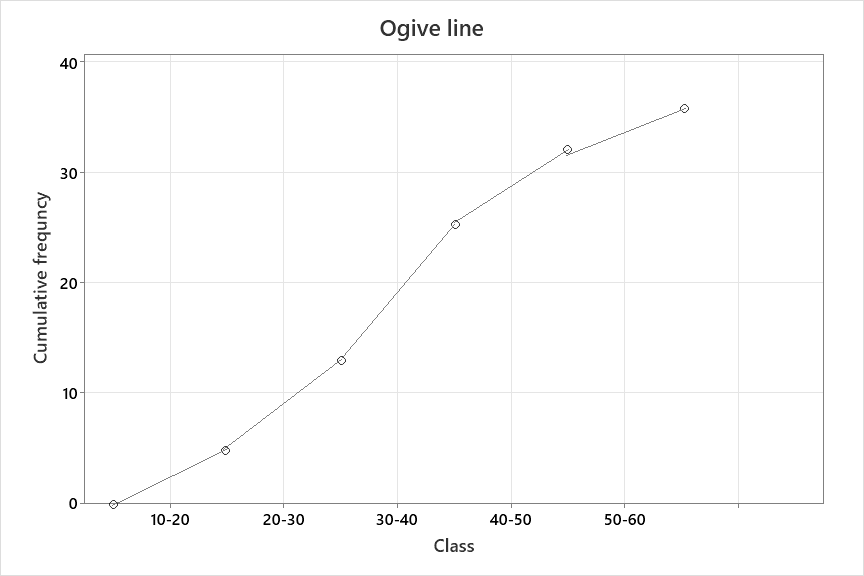
1. bar diagram, pie chart, histogram, and polygon, ogive line











7.

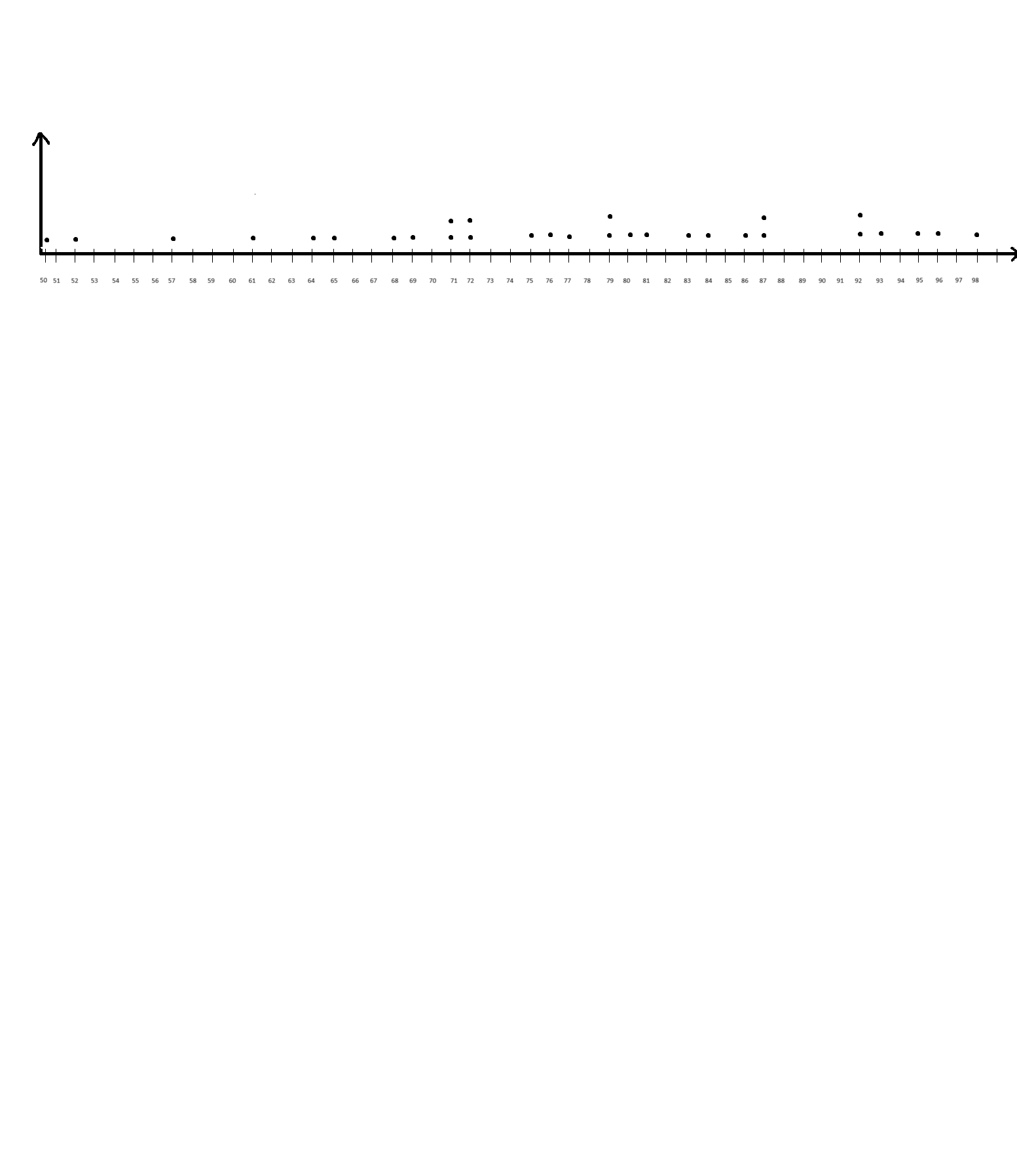
a)

Stem-and leaf display:

|  |  |
| --- | --- |
| 5  6  7  8  9 | 2 0 7  5 9 1 8 4  5 9 1 2 6 9 7 1 2  0 7 1 6 3 4 7  6 3 5 2 2 8 |

b)

Dot plot:



c)

The dot plot is more efficient for drawing conclusion about scattering, as we can easily visualize outliers, modes, minimum and maximum etc.

8.

Kurtosis is the degree of peakedness of a distribution, ususally taken in relation to a normal distribution. It quantifies how much the tails of a distribution differ from those of a normal distribution.

Here’s how kurtosis relates to understanding a frequency ditribution:

1. Hight kurtosis indicates that the distribution has heavy tails and a sharp peak, suggesting that the data has more outliers compared to a normal distribution.
2. Low kurtosis indicates that the distribution has lighter tails and is more flat-topped, suggesting that the data has fewer outliers compared to a normal distribution.

There are different measures of kurtosis, but the most commonly used is pearson’s moment coefficient of kurtosis, denoted by β2 .

a)

Given the first four moments of distribution,

μ’1 = -1.5

μ’2 = 17

μ’3 = -30

μ’4 = 108

So,

central moments,

μ1 = 0

μ2 = μ’2 - μ’12 = 17-(-1.5)2

= 14.75

μ3 = μ’3  - 3μ’1μ’2 + 2μ’13

= -30 - 3(-1.5)(17) + 2(-1.5)3

= 39.75

μ4 = μ’4  - 4μ’1μ’3  + 6μ’12 μ’2 + 3μ’14

= 108 - 4(-1.5)(-30) + 6(-1.5)2(17) + 3(-1.5)4

= 172.6875

b)

β1 = μ32 / μ23

= (39.75)2 / (14.75)3

= 0.4924 > 0 , so, positively skewed

β2 = μ4/ μ22

= (172.6875)2 / (14.75)2

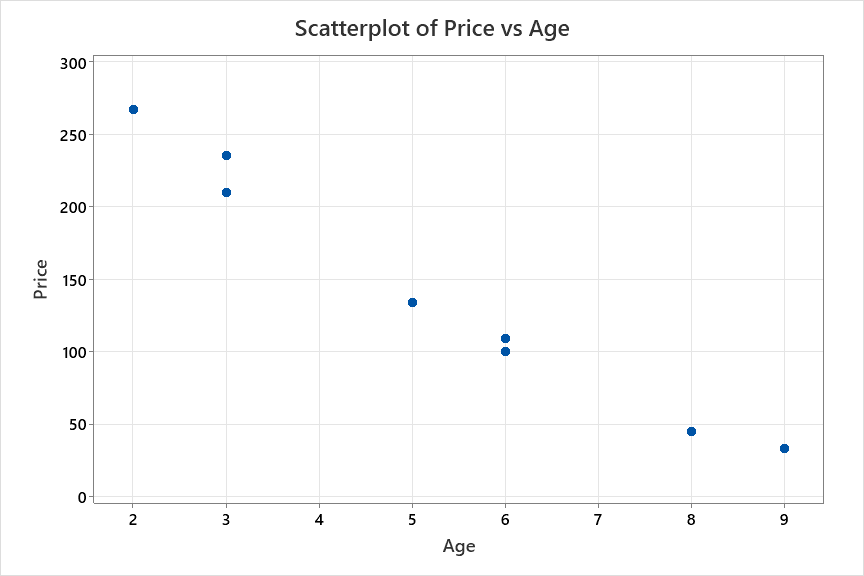
= 0.7937 < 3 , so, platykurtic

c)

While skewness tells us about the asymmetry of the distribution kurtosis provides additional information about the shape of the distribution’s tails. In this case, both measures are valuable for understandng the distribution’s charecteristics with kurtosis indicating a distribution with heavier tails and positive skewness indicating that the mode > median > mean.

9.

a)



by looking at the scatter diagram , we can observe that there exists a strong linear relationship between car age and price. If a straight line is drawn through the points, the points will be scattered closely around the line.

b)

|  |  |  |  |
| --- | --- | --- | --- |
| Age  x | Price  y | xy | x2 |
| 8  3  6  9  2  5  6  3 | 45  210  100  33  267  134  109  235 | 360  630  600  297  534  670  654  705 | 64  9  36  81  4  25  36  9 |
| Σx  = 42 | Σy  = 1133 | Σxy  = 4450 | Σx2  = 264 |

So,

x = Σx / n = 42 / 8 = 5.25

y = Σy / n = 1133 / 8 = 141.625

Now,

ssxy = Σxy - (Σx)(Σy) / n

= 4450 - (42)(1133) / 8

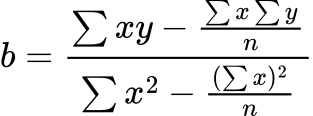
= -1498.25

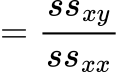
ssxx  = Σx2 - (Σx)2 / n

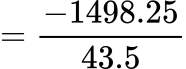
= 264 - (42)2  / n

= 43.5

So,







wps

C:/Users/BAB AL SAFA/AppData/Local/Temp/wps.PUiXGzwps

= 141.625 - (-34.44)(5.25)

= 322.435

Thus the regression line estimated by employing least-square method is,

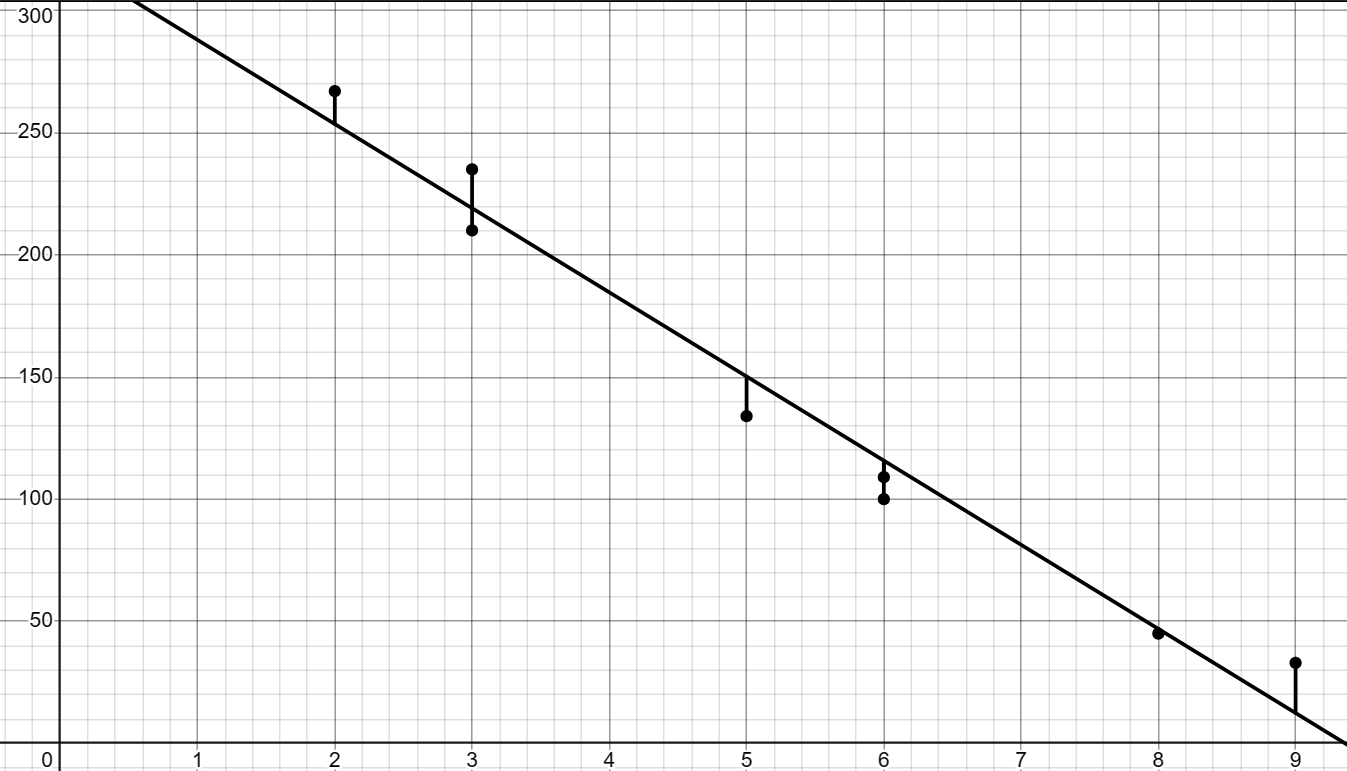
wps

c)

‘b’ represents an estimate of the average change in the value of the dependant variable y for each unit change in the independent variable x. In this particular case, the slope ‘b’ is negative which implies that as the age of cars (x) increases, the price decreases. So, the value b=-34.44 means that for an average increase of one year age of a car, the price would decrease on average by 34.44 hundred dollars.

The estimate of ‘a’ locates the regression line at point when x=0. Thus, if the age of a car is 0 years aka brand new, the average increase in price will be almost 322.435 hundred dollar.

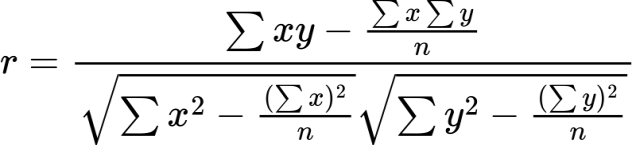
d)

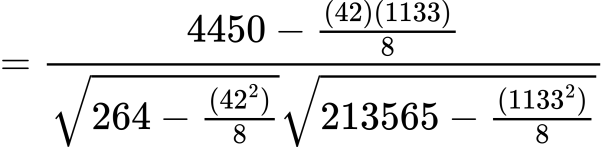


Graph of scatterplot and regression line and error shown with vertical lines.

Table for correlation coefficient,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age  x | Price  y | xy | x2 | y2 |
| 8  3  6  9  2  5  6  3 | 45  210  100  33  267  134  109  235 | 360  630  600  297  534  670  654  705 | 64  9  36  81  4  25  36  9 | 2025  44100  10000  1089  71289  17956  11881  55225 |
| Σx  = 42 | Σy  = 1133 | Σxy  = 4450 | Σx2  = 264 | Σy2  = 213565 |





= - 0.986 , value close to -1 indicates strong linear relationship with negative slope.

Since the regression model, wps

Regression coefficient = -34.44

Which indicates negative slope and also means that for an average increase of one unit in the independent variable, the value of dependant variable would decrease on average by 34.44 unit.

Therefore, data is not perfectly scattered.

e)

The best prediction of the price of a 7-year-old car of this model is,

wps

= 81.355

So, 81.355 hundreds of dollars or $8135.5

f)

The best prediction of the price of a 18-year-old car of this model is,

C:/Users/BAB AL SAFA/AppData/Local/Temp/wps.XIlrZrwps

= -297.485

But, price can’t be negative.

10.

a)

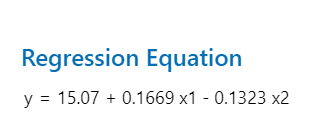


Figure 1: obtained from MINITAB

b)

The regression coefficients are 15.07, 0.1669 and − 0.1323

The value of a = 15.07 gives the value for ˆ y when x1 = 0 and x2 = 0. However, since x1 = 0 and x2 = 0 do not occur together in the sample data, the estimate is invalid.

The value b1 = 0.1669 gives the change in ˆ y for a one-unit change in x1 when x2 is held constant.

The value b2 = −0.1323gives the change in ˆ y for a one-unit change in x2 when x1 is held constant.

c)

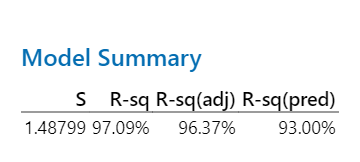


Figure 2: obtained from MINITAB

standard deviation of errors, se = 1.488

the coefficient of multiple determination, R2 = 0.9709

the adjusted coefficient of multiple determination, R2 = 0.9637

d)

y^ = 15.07 + (0.1669)x1 - (0.1323)x2 = 15.07 + .1669(87) − .1323(54) = 22.4461

e)

y^ = 15.07 + (0.1669)x1 - (0.1323)x2 = 15.07 + .1669(95) − .1323(49) = 24.4428

f)

df = n − k − 1 = 11 − 2 − 1 = 8

The 99% confidence interval for Β1 is

b1 ± t sb1 = .167 ± (3.355)(.034) = .167 ± .114 = .053 to .281

g)

H0: B2 = 0, H1: B2 < 0

Since σ∈ is unknown, use the t distribution.

For α = .01 with df = 8, the critical value of t is −2.896.

t = (b2 – B2)/ sb2 = −1.919

Do not reject H0 since −1.919 > −2.896.

So, B2 is not negative.