

Problem 1

The following are sample data on lengths (in minutes) of 13 movies:

90 102 120 180 90 92 90 137 120 151 97 93 120

Find the mean, median, first/lower quartile, third/upper quartile, and mode(s) of the data.

Answer:

The mean is

$$\bar{x} = \frac{90+102+120+180+90+92+90+137+120+151+97+93+120}{13} \text{ minutes} = \frac{1482}{13} \text{ minutes} = 114 \text{ minutes}$$

[Ref.: <https://edcommstatistics.blogspot.com/2019/07/introduction-to-two-measures-in.html>]

To find the median and quartiles, sort the data from the smallest to the highest, resulting in:

90 90 90 92 93 97 102 120 120 120 137 151 180

Name $x_1 = 90, x_2 = 90, x_3 = 90, x_4 = 92, \dots, x_{12} = 151, x_{13} = 180$

The number of data is $n = 13$. n is odd, so the median is $Me = X_{\frac{n+1}{2}}$.

$$\frac{n+1}{2} = \frac{13+1}{2} = 7$$

According to the naming stated above, we have $x_7 = 102$, so the median of these data is 102 minutes.

[Ref.: <https://edcommstatistics.blogspot.com/2019/07/the-median.html>]

To calculate the first quartile, begin with finding the value of L .

$$L = \frac{1}{4}(n+1) = \frac{1}{4}(13+1) = 3\frac{1}{2}$$

Note that $3 < 3\frac{1}{2} < 4$ (In this case, $k = 3$ and $k+1 = 4$)

To find the first quartile, use the formula $Q_1 = \frac{j}{4}x_k + \frac{4-j}{4}x_{k+1}$

Referring to <https://edcommstatistics.blogspot.com/2019/07/the-quartiles.html>, $j = 2$ because $3\frac{1}{2}$

is exactly halfway between 3 and 4. To continue, note that $x_3 = 90$ and $x_4 = 92$. Consequently,

$$Q_1 = \frac{2}{4} \cdot 90 + \frac{4-2}{4} \cdot 92 = \frac{1}{2} \cdot 90 + \frac{1}{2} \cdot 92 = 45 + 46 = 91$$

So, the first quartile is 91 minutes.

To calculate the third quartile, begin with finding the value of U .

$$U = \frac{3}{4}(n+1) = \frac{3}{4}(13+1) = 10\frac{1}{2}$$

Note that $10 < 10\frac{1}{2} < 11$ (In this case, $k = 10$ and $k+1 = 11$)

To find the third quartile, use the formula $Q_3 = \frac{j}{4}x_k + \frac{4-j}{4}x_{k+1}$

Referring to <https://edcommstatistics.blogspot.com/2019/07/the-quartiles.html>, $j = 2$ because $10\frac{1}{2}$ is exactly halfway between 10 and 11. To continue, note that $x_{10} = 120$ and $x_{11} = 137$. Consequently,

$$Q_3 = \frac{2}{4} \cdot 120 + \frac{4-2}{4} \cdot 137 = \frac{1}{2} \cdot 120 + \frac{1}{2} \cdot 137 = 60 + 68.5 = 128.5$$

So, the third quartile is 128.5 minutes.

To find the mode(s) of the data, search for the data occurring most frequently. In this case, 90 and 120 appear most frequently, each of them appears 3 times while others occur less than 3 times. So the data have two modes: $(Mo)_1 = 90$ minutes and $(Mo)_2 = 120$ minutes.

[Ref.: <https://edcommstatistics.blogspot.com/2019/07/the-mode.html>]

Problem 2

The values below represent the lengths (in seconds) of the songs on a CD.

227 248 228 233 241 266 231 279 261 305 254 305 206 302 263 282

Find the mode(s), median and both quartiles of the data.

Answer:

The mode is 305 seconds as 305 appears the most frequently. It appears twice while others appear less than twice.

To find the median and quartiles, sort the data from the smallest to the highest, resulting in:

206 227 228 231 233 241 248 254 261 263 266 279 282 302 305 305

Name $X_1 = 206$, $X_2 = 227$, $X_3 = 228$, ..., $X_{16} = 305$

The number of data is $n = 16$. n is even, so the median is $Me = \frac{X_{n/2} + X_{1+n/2}}{2}$

In this case, $n/2 = 16/2 = 8$ and $1 + n/2 = 1 + 8 = 9$. Consequently, $Me = \frac{X_8 + X_9}{2}$

According to the naming previously stated, $X_8 = 254$ and $X_9 = 261$

$$\text{Therefore, } Me = \frac{254+261}{2} = \frac{515}{2} = 257.5$$

So, the median of the data is 257.5 seconds.

To calculate the first quartile, begin with finding the value of L .

$$L = \frac{1}{4}(n + 1) = \frac{1}{4}(16 + 1) = 4\frac{1}{4}$$

Note that $4 < 4\frac{1}{4} < 5$ (In this case, $k = 4$ and $k+1 = 5$)

To find the first quartile, use the formula $Q_1 = \frac{j}{4}x_k + \frac{4-j}{4}x_{k+1}$

Referring to <https://edcommstatistics.blogspot.com/2019/07/the-quartiles.html>, $j = 3$ because $4\frac{1}{4}$ is closer to 4 than to 5. To continue, note that $x_4 = 231$ and $x_5 = 233$. Consequently,

$$Q_1 = \frac{3}{4} \cdot 231 + \frac{4-3}{4} \cdot 233 = \frac{3}{4} \cdot 231 + \frac{1}{4} \cdot 233 = 173.25 + 58.25 = 231.5$$

So, the first quartile is 231.5 seconds.

To calculate the third quartile, begin with finding the value of U.

$$U = \frac{3}{4}(n + 1) = \frac{3}{4}(16 + 1) = 12\frac{3}{4}$$

Note that $12 < 12\frac{3}{4} < 13$ (In this case, $k = 12$ and $k+1 = 13$)

To find the third quartile, use the formula $Q_3 = \frac{j}{4}x_k + \frac{4-j}{4}x_{k+1}$

Referring to <https://edcommstatistics.blogspot.com/2019/07/the-quartiles.html>, $j = 1$ because $12\frac{3}{4}$ is closer to 13 than to 12. To continue, note that $x_{12} = 279$ and $x_{13} = 282$. Consequently,

$$Q_3 = \frac{1}{4} \cdot 279 + \frac{4-1}{4} \cdot 282 = \frac{1}{4} \cdot 279 + \frac{3}{4} \cdot 282 = 69.75 + 211.5 = 281.25$$

So, the third quartile is 281.25 seconds.

Problem 3

Below is the sample data showing the distances traveled by 191 buses before their engines failed, in thousands of miles.

No.	Distance	Number of Buses
1	1 - 20	6
2	21 - 40	11
3	41 - 60	16
4	61 - 80	25
5	81 - 100	34
6	101 - 120	46
7	121 - 140	33
8	140 - 160	16
9	161 - 200	4
	TOTAL	191

Find the mean, median, quartiles, and mode of the distances traveled by the buses before their engines failed.

Answer:

Below is the table helping to calculate the mean.

i	Distance	f_i	M_i	$f_i M_i$
1	1 - 20	6	10.5 *)	63.0 †)
2	21 - 40	11	30.5 **)	335.5 ††)
3	41 - 60	16	50.5	808.0
4	61 - 80	25	70.5	1762.5
5	81 - 100	34	90.5	3077.0
6	101 - 120	46	110.5	5083.0
7	121 - 140	33	130.5	4306.5
8	140 - 160	16	150.5	2408.0
9	161 - 200	4	180.5	722.0
	$\sum_{i=1}^9 f_i$	191	$\sum_{i=1}^9 f_i M_i$	18565.5

$$*) M_1 = \frac{1+20}{2} = \frac{21}{2} = 10.5 \quad **) M_2 = \frac{21+40}{2} = \frac{61}{2} = 30.5$$

$$†) f_1 M_1 = 6 \times 10.5 = 63.0 \quad ††) f_2 M_2 = 11 \times 30.5 = 335.5$$

[Refer to <https://edcommstatistics.blogspot.com/2019/07/calculating-mean-of-grouped-data.html>

to get more complete details on calculating the mean of grouped data.]

$$\text{Then, } \bar{X} = \frac{\sum_{i=1}^9 f_i M_i}{\sum_{i=1}^9 f_i} = \frac{18565.5}{191} \approx 97.2$$

So, the mean of the distances traveled by the buses before their engines failed is 97.2 thousand miles.

To find the first quartile, add an extra column into which the Data Numbers will be entered.

[Ref.: <https://edcommstatistics.blogspot.com/2019/09/the-quartiles-and-median-of-grouped-data.html>]

Having entered the Data Numbers, we get the following table.

No.	Distance	f_i	Data Numbers
1	1 - 20	6	1 - 6
2	21 - 40	11	7 - 17
3	41 - 60	16	18 - 33
4	61 - 80	25	34 - 58
5	81 - 100	34	59 - 92
6	101 - 120	46	93 - 138
7	121 - 140	33	139 - 171
8	140 - 160	16	172 - 187
9	161 - 200	4	187 - 191
TOTAL		191	

In this case, $n = 191$ so $\frac{1}{4} \times n = \frac{1}{4} \times 191 = 47\frac{3}{4} = 47.75$

Referring to the table, data number $47\frac{3}{4}$ is in the fourth class, namely class 61 - 80. This is the first quartile class. The lower class boundary of the class is $L_1 = 61 - 0.5 = 60.5$. The frequency of the class is $f_1 = 25$. The class width is $c = 80.5 - 60.5 = 20$. There are 33 data in the preceding classes ($33 = 6 + 11 + 16$), so $(\Sigma f)_1 = 33$. Substituting the values into the formula $Q_1 = L_1 + \frac{n/4 - (\Sigma f)_1}{f_1} \cdot c$, we have:

$$Q_1 = 60.5 + \frac{47.75 - 33}{25} \times 20 = 60.5 + \frac{14.75}{25} \times 20 = 72.3$$

So the first quartile of the data is 72.3 thousand miles.

To calculate the third quartile, begin with finding $3n/4$. In this case, $3n/4 = 3 \times 191/4 = 143\frac{3}{4} = 143.25$.

Referring to the table above, data number $143\frac{3}{4}$ is in the seventh class. This class is the third quartile class.

No.	Distance	f_i	Data Numbers
1	1 - 20	6	1 - 6
2	21 - 40	11	7 - 17
3	41 - 60	16	18 - 33
4	61 - 80	25	34 - 58
5	81 - 100	34	59 - 92
6	101 - 120	46	93 - 138
7	121 - 140	33	139 - 171
8	140 - 160	16	172 - 187
9	161 - 200	4	187 - 191
TOTAL		191	

The lower class boundary of the class is $L_3 = 121 - 0.5 = 120.5$. The frequency of the class is $f_3 = 33$. The class width is $c = 140.5 - 120.5 = 20$. There are 138 data in the preceding classes ($138 = 6 + 11 + 16 + 25 + 34 + 46$), so $(\Sigma f)_3 = 138$. Substituting the values into the formula $Q_3 = L_3 + \frac{3n/4 - (\Sigma f)_3}{f_3} \cdot c$, we have:

$$Q_3 = 120.5 + \frac{143.25 - 138}{33} \times 20 = 120.5 + \frac{5.25}{33} \times 20 \approx 123.68$$

So the first quartile of the data is 123.68 thousand miles.

To calculate the median, begin with finding $n/2$. In this case, $n/2 = 191/2 = 95\frac{1}{2} = 95.5$. Referring to the table above, data number $95\frac{1}{2}$ is in the sixth class. This class is the median class.

No.	Distance	f_i	Data Numbers
1	1 - 20	6	1 - 6
2	21 - 40	11	7 - 17
3	41 - 60	16	18 - 33
4	61 - 80	25	34 - 58
5	81 - 100	34	59 - 92
6	101 - 120	46	93 - 138
7	121 - 140	33	139 - 171
8	140 - 160	16	172 - 187
9	161 - 200	4	187 - 191
TOTAL		191	

The lower class boundary of the class is $L_2 = 101 - 0.5 = 100.5$. The frequency of the class is $f_2 = 46$. The class width is $c = 120.5 - 100.5 = 20$. There are 92 data in the preceding classes ($92 = 6 + 11 + 16 + 25 + 34$), so $(\Sigma f)_2 = 92$. Substituting the values into the formula $Q_2 = L_2 + \frac{n/2 - (\Sigma f)_2}{f_2} \cdot c$, we have:

$$Q_2 = 100.5 + \frac{95.5 - 92}{46} \times 20 = 100.5 + \frac{3.5}{46} \times 20 \approx 102.02$$

So the first quartile of the data is 102.02 thousand miles.

Problem 4

The owner of a bookstore surveys 7 customers at random and asks how many books they bought last year. He gets the following data: 11, 12, 13, 8, 10, 11, 5. The survey is carried out to estimate the mean number of books bought by all customers last year. To do this, he has to find the standard deviation of the data. Calculate the variance and standard deviation of the data.

Answer:

$$\text{The mean is } \bar{X} = \frac{11+12+13+8+10+11+5}{7} = \frac{70}{7} = 10$$

Using the formula $s^2 = \frac{\sum_{i=1}^n X_i^2}{n-1} - \frac{n}{n-1} \bar{X}^2$ to find the sample variance, we have:

$$s^2 = \frac{11^2 + 12^2 + 13^2 + 8^2 + 10^2 + 11^2 + 5^2}{7-1} - \frac{7}{7-1} \cdot 10^2 = \frac{744}{6} - \frac{7}{6} \cdot 100 \approx 7.33$$

The standard deviation is the square root of the variance, i.e. $s = \sqrt{s^2}$, so $s = \sqrt{7.33} \approx 2.71$

So the variance is 7.33 and the standard deviation is 2.71.

Problem 5

A school has 3 math teachers and 4 English teachers. The years of experience of them are as follows.

Math teachers: 5, 8, 5

English teachers: 3, 3, 5, 5

Find the variance and standard deviation of each group's data. Which group varies more in terms of years of experience?

Answer:

To calculate the variance of the groups' data, use the formula $\sigma^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$

Math teachers:

$$\bar{X} = \frac{5+8+5}{3} \text{ years} = 6 \text{ years}$$

$$\text{The variance: } \sigma^2 = \left[\frac{5^2+8^2+5^2}{3} - 6^2 \right] \text{ years}^2 = \left(\frac{114}{3} - 36 \right) \text{ years}^2 = 2 \text{ years}^2$$

$$\text{The standard deviation: } \sigma = \sqrt{2} \text{ years} \approx 1.41 \text{ year}$$

English teachers:

$$\bar{X} = \frac{3+3+5+5}{4} \text{ years} = 4 \text{ years}$$

$$\text{The variance: } \sigma^2 = \left[\frac{3^2+3^2+5^2+5^2}{4} - 4^2 \right] \text{ years}^2 = \left(\frac{68}{4} - 16 \right) \text{ year}^2 = 1 \text{ year}^2$$

$$\text{The standard deviation: } \sigma = \sqrt{1} \text{ year} = 1 \text{ year}$$

Based on the variances, math teachers vary more in terms of years of experience. (The same conclusion is reached if the standard deviations are compared.)

COMPARE PROBLEM 4 WITH PROBLEM 5. Why were different formulae applied?

In Problem 4, we considered all the customers who bought books from the store last year as the population. The number of the customers might be hundreds or more, but only 7 data were taken from the population. So, in this case the sampling method was applied. Therefore, the sample variance and standard deviation formulae were used.

In Problem 5, there were two populations. The first population was all the math teachers of the school, which consisted of 3 teachers. The second was all the English teachers of the school, which consisted of 4 teachers. In each group, all the population data were taken. In other words, in each group the census method was applied. Therefore, in this case the population variance and standard deviation formulae were used.

A hint for deciding which formula to use

If the number of the data in hand equals to the number of population members, use the population variance or standard deviation. (See Problem 5.) If it is less than the number of population members, use the sample variance or standard deviation. (See Problem 4.)

To get more examples, visit <https://edcommstatistics.blogspot.com/2019/07/introduction-to-two-measures-in.html>

PROBLEM 6

A survey took 25 respondents as samples. They were asked how much they spent on watching movies last year. The following frequency distribution table summarizes the survey results.

No.	Spending (dollars)	The Number of Respondents
1.	5 - 11	2
2.	12 - 16	2
3.	17 - 21	6
4.	22 - 26	11
5.	27 - 37	4

Find the variance and standard deviation.

Answer:

Applying the same procedure as in Problem 3 to find the mean of the data, we get:

No.	Spending (\$)	f_i	M_i	$f_i.M_i$
1	5-11	2	8	16
2	12-16	2	14	28
3	17 - 21	6	19	114
4	22 - 26	11	24	264
5	27 - 37	4	32	128
	$\sum_{i=1}^5 f_i$	25	$\sum_{i=1}^5 f_i M_i$	550

From the relevant values in the table above, we get the mean: $\bar{X} = \$ \frac{550}{25} = \$ 22$

Subtract the mean from each midpoint (M_i), then we get the d_i of each class, i.e. $d_i = M_i - \bar{X}$. The resulting d_i 's are shown in the d_i column below.

No.	Spending (\$)	f_i	M_i	d_i
1	5-11	2	8	-14 *)
2	12-16	2	14	-8 **)
3	17 - 21	6	19	-3
4	22 - 26	11	24	2
5	27 - 37	4	32	10
	$\sum_{i=1}^5 f_i$	25	$\sum_{i=1}^5 f_i M_i$	

Note: *) -14 = 8 - 22 **) -8 = 14 - 22

For each class, calculate d_i^2 and $f_i d_i^2$. The results are summarized as follows.

No.	Spending (\$)	f_i	d_i	d_i^2	$f_i d_i^2$
1	5-11	2	-14	196 [*])	392 ^{†)}
2	12-16	2	-8	64	128
3	17 - 21	6	-3	9 ^{**)}	54 ^{††)}
4	22 - 26	11	2	4	44
5	27 - 37	4	10	100	400
	$n = \sum_{i=1}^5 f_i$	25		$\sum_{i=1}^5 f_i d_i^2$	1018

Note: ^{*}) $196 = (-14)^2$ ^{**)} $9 = (-3)^2$ ^{†)} $392 = 2 \times 196$ ^{††)} $54 = 6 \times 9$

Use $s^2 = \frac{\sum_{i=1}^k f_i d_i^2}{n-1}$ to find the sample variance, we get $s^2 = \frac{1018}{25-1} = \frac{1018}{24} \approx 42.42$

The sample standard deviation is the square root of its variance, resulting in $s = \sqrt{42.42} \approx 6.51$.

So the variance is \$² 42.42 and the standard deviation is \$ 6.51.

PROBLEM 7

A survey took 25 respondents as samples. They were asked how much they spent on watching movies last year. The following frequency distribution table summarizes the survey results.

No.	Spending (dollars)	The Number of Respondents
1.	7 - 11	2
2.	12 - 16	2
3.	17 - 21	6
4.	22 - 26	11
5.	27 - 31	4

Find the variance and standard deviation.

Answer:

As you may verify, the classes have an equal class width. The class width is $c = 5$.

[Refer to <https://edcommstatistics.blogspot.com/2019/08/anatomy-of-frequency-distribution-tables.html> to study more about the class width.]

As the classes are of the same width, this problem can be solved faster using the coding method.

[Ref.: <https://edcommstatistics.blogspot.com/2019/08/calculating-variance-without.html>]

Applying the method, we get the following table.

No.	Spending (dollars)	f_i	u_i	u_i^2	$f_i \cdot u_i$	$f_i u_i^2$
1.	7 - 11	2	-2	4	-4	8
2.	12 - 16	2	-1	1	-2	2
3.	17 - 21	6	0	0	0	0
4.	22 - 26	11	1	1	11	11
5.	27 - 31	4	2	4	8	16
	SUM	25		SUM	13	37

The appropriate formula to use is

$$s^2 = \left[\frac{\sum_{i=1}^k f_i u_i^2}{n-1} - \frac{(\sum_{i=1}^k f_i u_i)^2}{n(n-1)} \right] c^2$$

Then, the variance is $s^2 = \$^2 \left[\frac{37}{25-1} - \frac{13^2}{25(25-1)} \right] \cdot 5^2 = \$^2 \left[\frac{37}{24} - \frac{169}{25 \cdot 24} \right] \cdot 25 = \$^2 31.5$ and the standard deviation is $s = \$ \sqrt{31.5} \approx \$ 5.61$.

PROBLEM 8

A municipal government office has a complete list of the heights of hotel buildings in the city. The frequency distribution table below shows the data.

No.	Height (meters)	The Number of Buildings
1.	81 - 90	3
2.	91 - 100	5
3.	101 - 110	5
4.	111 - 120	9
5.	121 - 130	6
6.	131 - 140	2

Determine the variance and standard deviation.

Answer:

As the table shows the complete list of the heights of hotel buildings in the city, what we have are population data, so the suitable formula to be used is $\sigma^2 = \frac{\sum_{i=1}^f f_i d_i^2}{n}$ where $d_i = M_i - \bar{X}$. But because all classes have equal width, i.e. $c = 10$, we may apply the coding method, using $\sigma^2 = \left[\frac{\sum_{i=1}^k f_i u_i^2}{n} - \left(\frac{\sum_{i=1}^k f_i u_i}{n} \right)^2 \right] \cdot c^2$. For the sake of quick and easy solution, the coding method is being applied.

No.	Height (meters)	f_i	u_i	u_i^2	$f_i u_i$	$f_i u_i^2$
1.	81 - 90	3	-2	4	-6	12
2.	91 - 100	5	-1	1	-5	5
3.	101 - 110	5	0	0	0	0
4.	111 - 120	9	1	1	9	9
5.	121 - 130	6	2	4	12	24
6.	131 - 140	2	3	9	6	18
	SUM	30			16	68

$$\sigma^2 = \left[\frac{68}{30} - \left(\frac{16}{30} \right)^2 \right] \cdot 10^2 m^2 \approx 198.22 m^2$$

$$\sigma = \sqrt{198.22} m \approx 14.08 m$$

So the variance is $198.22 m^2$ and the standard deviation is $14.08 m$.

PROBLEM 9

Similar to Problem 8 but with the following data.

No.	Height (meters)	The Number of Buildings
1.	41 - 90	3
2.	91 - 100	5
3.	101 - 110	5
4.	111 - 120	9
5.	121 - 130	6
6.	131 - 150	2

Answer:

In this problem the classes are not of equal widths. So we may not apply the coding method. Instead, we apply the procedure as discussed in Problem 6. The mean of the data is calculated as demonstrated at the table below.

No.	Height (m)	f_i	M_i	$f_i M_i$	$d_i = M_i - \bar{X}$	d_i^2	$f_i d_i^2$
1.	41 - 90	3	65.5	196.5	-43.7	1909.69	5729.07
2.	91 - 100	5	95.5	477.5	-13.7	187.69	938.45
3.	101 - 110	5	105.5	527.5	-3.7	13.69	68.45
4.	111 - 120	9	115.5	1039.5	6.3	39.69	357.21
5.	121 - 130	6	125.5	753.0	16.3	265.69	1594.14
6.	131 - 150	2	140.5	281.0	31.3	979.69	1959.38
	SUM	30		3275.0			10646.70

$$\bar{X} = \frac{\sum_{i=1}^6 f_i M_i}{\sum_{i=1}^6 f_i} = \frac{3275}{30} m \approx 109.2 m$$

Using the population variance formula $\sigma^2 = \frac{\sum_{i=1}^k f_i (M_i - \bar{X})^2}{n}$, we get:

$$\sigma^2 = \frac{10646.70}{30} m^2 \approx 354.89 m^2$$

The population standard deviation is $\sigma = \sqrt{\sigma^2}$. Using the formula, we get $\sigma = \sqrt{354.89} \approx 18.84 m$.

So the variance is $354.89 m^2$ and the standard deviation is $18.84 m$.