Descriptive Statistics

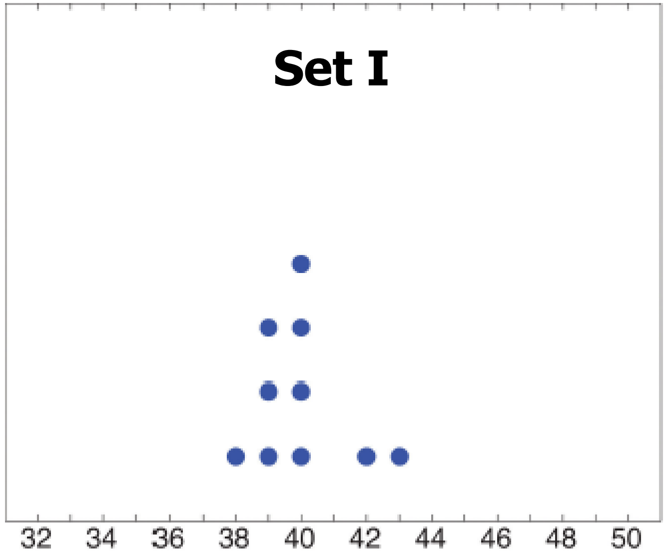
Continue within another topic of the Descriptive Statistics:

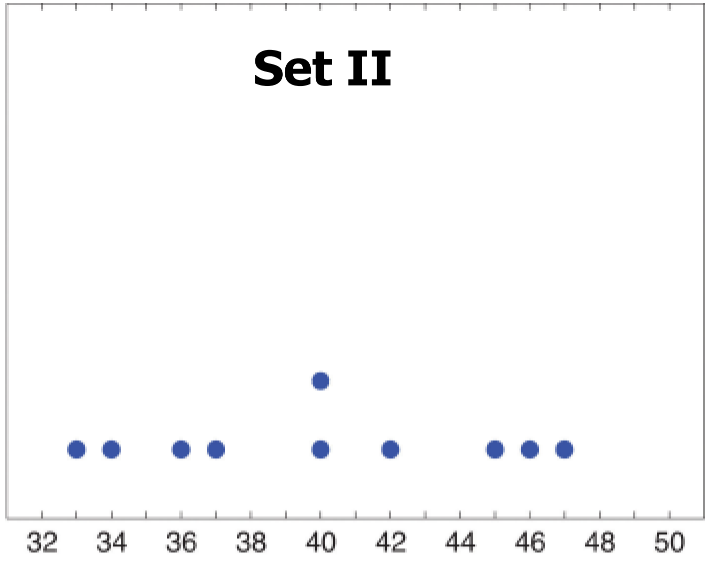
Measures of Variability

- There are three measures of the variability of a data set: **the range, the variance, and the standard deviation.**
- To learn the concept of the variability of a data set. Let us look the following case (1):

Case (1) Look at the two data sets in Table below "Two Data Sets" and the graphical representation of each, called a dot plot, in Figures below "Dot Plots of Data Sets".

Data Set I:	40	38	42	40	39	39	43	40	39	40
Data Set II:	46	37	40	33	42	36	40	47	34	45





Dot Plots of Data Sets

Solution:

The two sets of ten measurements each center at the same value: they both have mean, median, and mode 40.

Nevertheless a glance at the figures shows that they are markedly different.

In Data Set I the measurements vary only slightly from the center, while for Data Set II the measurements vary greatly. Just as we have attached numbers to a data set to locate its center, we now wish to associate to each data set numbers that measure quantitatively how the data either scatter away from the center or cluster close to it. These new quantities are called measures of variability, and we will discuss three of them.

(1) The Range is the simplest measure of variability .

The range is a measure of variability because it indicates the size of the interval over which the data points are distributed. A smaller range indicates less variability (less dispersion) among the data, whereas a larger range indicates the opposite.

Definition

The **range**[9] *of a data set is the number R defined by the formula*

$$R = x_{\max} - x_{\min}$$

where $x_{\max}$ *is the largest measurement in the data set and* $x_{\min}$ *is the smallest.*

Case (2) Find the range of each data set in previous Table "Two Data Sets".

Solution:

For Data Set I the maximum is 43 and the minimum is 38, the range is (R = 43 – 38 = 5) and the range for Data Set II equals 14.

In Data Set I the measurements vary only slightly from the center, while for Data Set II the measurements vary greatly. Just as we have attached numbers to a data set to locate its center, we now wish to associate to each data set numbers that measure quantitatively how the data either scatter away from the center or cluster close to it. These new quantities are called measures of variability, and we will discuss three of them.

(2) The Variance and the Standard Deviation are other two measures of variability that we will consider are more elaborate and also depend on whether the data set is just a sample drawn from a much larger population or is the whole population itself (that is, a census).

Although the first formula in each case looks less complicated than the second, the latter is easier to use in hand computations, and is called a shortcut formula.

Definition

The **sample variance** *of a set of n sample data is the number s^2 defined by the formula*

$$s^2 = \frac{\Sigma(x-\overline{x})^2}{n-1}$$

which by algebra is equivalent to the formula

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}$$

The **sample standard deviation**[10] *of a set of n sample data is the square root of the sample variance, hence is the number s given by the formulas*

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}}$$

Case (3) Find the sample variance and the sample standard deviation of Data Set II in the previous Table "Two Data Sets".

Solution:

To use the defining formula (the first formula) in the definition we first compute for each observation x its deviation $x - \bar{x}$ from the sample mean.

Since the mean of the data is $\bar{x} = 40$, we obtain the ten numbers displayed in the second line of the supplied table.

x	46	37	40	33	42	36	40	47	34	45
$x-\bar{x}$	6	−3	0	−7	2	−4	0	7	−6	5

Then

$$\Sigma(x-\bar{x})^2 = 6^2 + (-3)^2 + 0^2 + (-7)^2 + 2^2 + (-4)^2 + 0^2 + 7^2 + (-6)^2$$

so

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{224}{9} = 24.\overline{8}$$

and

$$s = \sqrt{24.\overline{8}} \approx 4.99$$

Can be computed ten deviations for Data Set I and verify that their squares add up to 20, so that the sample variance and standard deviation of Data Set I are the much smaller numbers:

$$s^2 = 20 \,/\, 9 = 2.\overline{2} \text{ and } \sqrt{20/9} \approx 1.49.$$

Case (4) Find the sample variance and the sample standard deviation for the following data set:

1.90 3.00 2.53 3.71 2.12 1.76 2.71 1.39 4.00 3.33

Solution:

The sample variance has different units from the data. For example, if the units in the data set were inches, the new units would be inches squared, or square inches. It is thus primarily of theoretical importance and will not be considered further in this text, except in passing.

If the data set comprises the whole population, then the *population* standard deviation, denoted σ (the lower case Greek letter sigma), and its square, the *population* variance σ^2, are defined as follows.

$$\Sigma x = 1.90 + 3.00 + 2.53 + 3.71 + 2.12 + 1.76 + 2.71 + 1.39 + 4.00 + 3.33$$

and

$$\begin{aligned}\Sigma x^2 &= 1.90^2 + 3.00^2 + 2.53^2 + 3.71^2 + 2.12^2 + 1.76^2 \\ &\quad + 2.71^2 + 1.39^2 + 4.00^2 + 3.33^2 \\ &= 76.7321\end{aligned}$$

the shortcut formula gives

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1} = \frac{76.7321 - \frac{(26.45)^2}{10}}{10-1} = \frac{6.77185}{9} = .75242\overline{7}$$

and

$$s = \sqrt{.75242\overline{7}} \approx .867$$

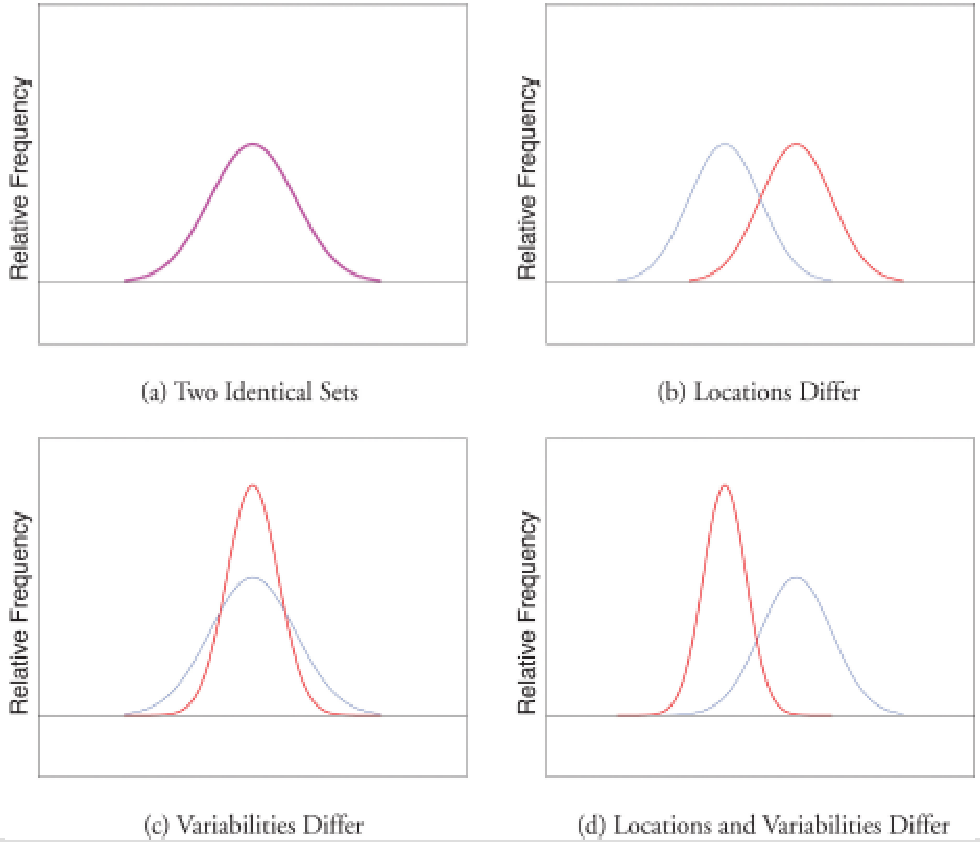
Definition

The **population variance** *and* **population standard deviation**[11] *of a set of N population data are the numbers* σ^2 *and* σ *defined by the formulas*

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \quad and \quad \sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

Note that the denominator in the fraction is the full number of observations, not that number reduced by one, as is the case with the sample standard deviation. Since most data sets are samples, we will always work with the sample standard deviation and variance.

Finally, in many real-life situations the most important statistical issues have to do with comparing the means and standard deviations of two data sets. The following figure “Difference between Two Data Sets" illustrates how a difference in one or both of the sample mean and the sample standard deviation are reflected in the appearance of the data set as shown by the curves derived from the relative frequency histograms built using the data.



(a) Two Identical Sets (b) Locations Differ (c) Variabilities Differ (d) Locations and Variabilities Differ

Key: The range, the standard deviation, and the variance each give a quantitative answer to the question “How variable are the data?”

EXERCISES

BASIC

1. Find the range, the variance, and the standard deviation for the following sample.

$$1 \quad 2 \quad 3 \quad 4$$

2. Find the range, the variance, and the standard deviation for the following sample.

$$2 \quad -3 \quad 6 \quad 0 \quad 3 \quad 1$$

3. Find the range, the variance, and the standard deviation for the following sample.

$$2 \quad 1 \quad 2 \quad 7$$

4. Find the range, the variance, and the standard deviation for the following sample.

$$-1 \quad 0 \quad 1 \quad 4 \quad 1 \quad 1$$

5. Find the range, the variance, and the standard deviation for the sample represented by the data frequency table.

x	1	2	7
f	1	2	1

6. Find the range, the variance, and the standard deviation for the sample represented by the data frequency table.

x	−1	0	1	4
f	1	1	3	1

APPLICATIONS

7. Find the range, the variance, and the standard deviation for the sample of ten IQ scores randomly selected from a school for academically gifted students.

$$132 \quad 162 \quad 133 \quad 145 \quad 148$$
$$139 \quad 147 \quad 160 \quad 150 \quad 153$$

8. Find the range, the variance and the standard deviation for the sample of ten IQ scores randomly selected from a school for academically gifted students.

$$142 \quad 152 \quad 138 \quad 145 \quad 148$$
$$139 \quad 147 \quad 155 \quad 150 \quad 153$$

ADDITIONAL EXERCISES

9. Consider the data set represented by the table

x	26	27	28	29	30	31	32
f	3	4	16	12	6	2	1

   a. Use the frequency table to find that $\Sigma x = 1256$ and $\Sigma x^2 = 35{,}926$.
   b. Use the information in part (a) to compute the sample mean and the sample standard deviation.

10. Find the sample standard deviation for the data

x	1	2	3	4	5
f	384	208	98	56	28

x	6	7	8	9	10
f	12	8	2	3	1

11. A random sample of 49 invoices for repairs at an automotive body shop is taken. The data are arrayed in the stem and leaf diagram shown. (Stems are thousands of dollars, leaves are hundreds, so that for example the largest observation is 3,800.)

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

For these data, $\Sigma x = 101{,}100$, $\Sigma x^2 = 244{,}830{,}000$.

   a. Compute the mean, median, and mode.
   b. Compute the range.

c. Compute the sample standard deviation.

12. What must be true of a data set if its standard deviation is 0?

13. A data set consisting of 25 measurements has standard deviation 0. One of the measurements has value 17. What are the other 24 measurements?

14. Create a sample data set of size $n = 3$ for which the range is 0 and the sample mean is 2.

15. Create a sample data set of size $n = 3$ for which the sample variance is 0 and the sample mean is 1.

16. The sample $\{-1,0,1\}$ has mean $\bar{x} = 0$ and standard deviation $s = 1$. Create a sample data set of size $n = 3$ for which $\bar{x} = 0$ and s is greater than 1.

17. The sample $\{-1,0,1\}$ has mean $\bar{x} = 0$ and standard deviation $s = 1$. Create a sample data set of size $n = 3$ for which $\bar{x} = 0$ and the standard deviation s is less than 1.

18. Begin with the following set of data, call it Data Set I.

$$5 \quad -2 \quad 6 \quad 14 \quad -3 \quad 0 \quad 1 \quad 4 \quad 3 \quad 2 \quad 5$$

a. Compute the sample standard deviation of Data Set I.
b. Form a new data set, Data Set II, by adding 3 to each number in Data Set I. Calculate the sample standard deviation of Data Set II.
c. Form a new data set, Data Set III, by subtracting 6 from each number in Data Set I. Calculate the sample standard deviation of Data Set III.
d. Comparing the answers to parts (a), (b), and (c), can you guess the pattern? State the general principle that you expect to be true.

ANSWERS

1. $R = 3$, $s^2 = 1.7$, $s = 1.3$.

3. $R = 6$, $s^2 = 7.\overline{3}$, $s = 2.7$.

5. $R = 6$, $s^2 = 7.3$, $s = 2.7$.

7. $R = 30$, $s^2 = 103.2$, $s = 10.2$.

9. $\bar{x} = 28.55$, $s = 1.3$.

11.
a. $\bar{x} = 2063$, $\tilde{x} = 2000$, mode $= 2000$.
b. $R = 3400$.
c. $s = 869$.

13. All are 17.

15. $\{1,1,1\}$

17. One example is $\{-.5,0,.5\}$.

Relative Position of Data

When you take an exam, what is often as important as your actual score on the exam is the way your score compares to other students' performance. If you made a 70 but the average score (whether the mean, median, or mode) was 85, you did relatively poorly. If you made a 70 but the average score was only 55 then you did relatively well. In general, the significance of one observed value in a data set strongly depends on how that value compares to the other observed values in a data set.

In this topic can be learned:

- The meaning of the relative position of an element of a data set.
- The meaning of each of two measures, the percentile rank and the z-score, of the relative position of a measurement and how to compute each one.
- The meaning of the three quartiles associated to a data set and how to compute them.-
- The meaning of the five-number summary of a data set, how to construct the box plot, and how to interpret the box.

(1) Percentiles and Quartiles

Any exam you passed can be evaluated by "only score" and/or "percentile ranking of that score". In example may be that score is 63 and its 85th of percentile. An actual score is 63 while 85% of the exam scores are less than or/and equal to 63.

Definition: Given an observed value X in a data set, X is the Pth percentile of the data if the percentage of the data that are less than or equal to X is P. The number P is the percentile rank of X.

Case (5) What is the percentile for the values 1.39 and 3.33 in the data set below.

1.90 3.00 2.53 3.71 2.12 1.76 2.71 1.39 4.00 3.33

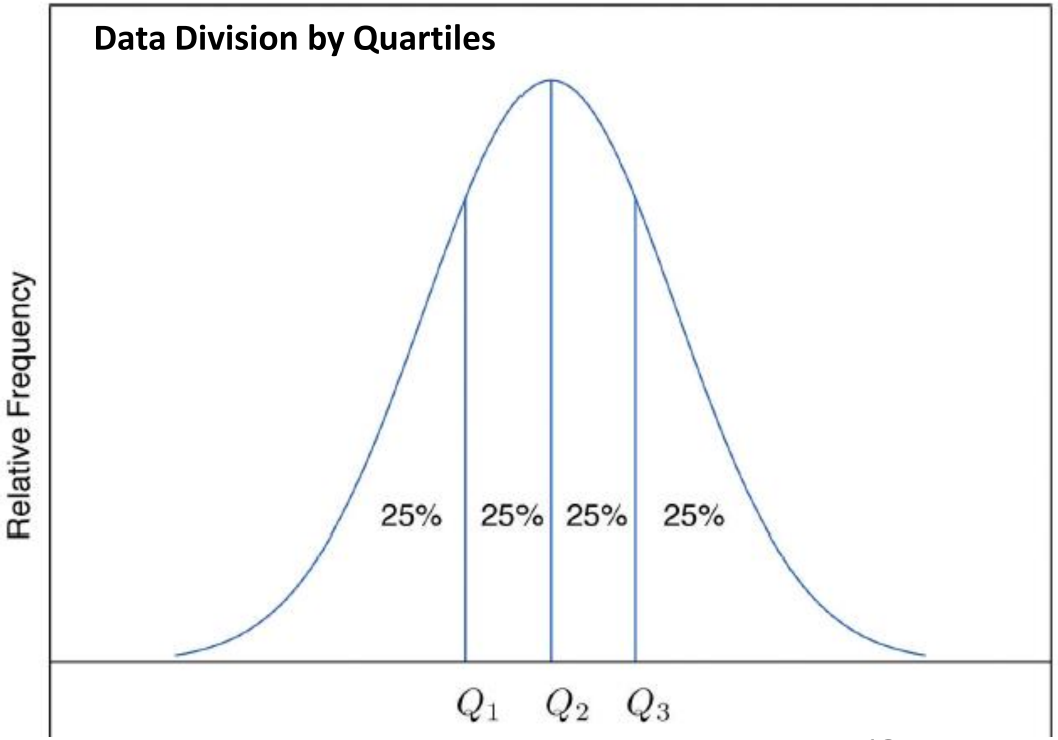
Solution:

First the data should be written in increasing order: 1.39 1.76 1.90 2.12 2.53 2.71 3.00 3.33 3.71 4.00

The only data value that is less than or equal to 1.39 is 1.39 itself. Since 1 is 1⁄10 = 0.10 or 10% of 10 total values, the value 1.39 is the 10th percentile. Eight data values are less than or equal to 3.33. Since 8 is 8⁄10 = 0.80 or 80% of 10 total values, the value 3.33 is the 80th percentile.

Quartiles:

The *P*th percentile cuts the data set in two so that approximately *P*% of the data lie below it and (100 − *P*)% of the data lie above it. In particular, the three percentiles that cut the data into fourths, as shown in the Figure below are called the **quartiles**.



Definition:

1. *The* **second quartile** Q_2 *of the data set is its median.*
2. *Define two subsets:*
   1. *the lower set: all observations that are strictly less than* Q_2;
   2. *the upper set: all observations that are strictly greater than* Q_2.
3. *The* **first quartile** Q_1 *of the data set is the median of the lower set.*
4. *The* **third quartile** Q_3 *of the data set is the median of the upper set.*

Case (6) Find the quartiles for the following data set: 1.90 3.00 2.53 3.71 2.12 1.76 2.71 1.39 4.00 3.33

Solution:

- First the data should be written in increasing order: 1.39 1.76 1.90 2.12 2.53 2.71 3.00 3.33 3.71 4.00

- This data set has n = 10 observations. Since 10 is an even number, the median is the mean of the two middle observations:

$\bar{x}$ = (2.53+2.71)/2 = 2.62 (Thus the second quartile is*Q*2 = 2.62).

- Lower: L = {1.39,1.76,1.90,2.12,2.53} and

- Upper: U = {2.71,3.00,3.33,3.71,4.00}

Each has an odd number of elements, so the median of each is its middle observation. Thus the first quartile is *Q*1 = 1.90, the median of *L*, and the third quartile is *Q*3 = 3.33, the median of *U*.

Case (7) Adjoin the observation 3.88 to the data set of the previous case and find the quartiles of the new set of data.

Solution:

- The data should be written in increasing order:

1.39 1.76 1.90 2.12 2.53 2.71 3.00 3.33 3.71 3.88 4.00

- This data set has 11 observations. The second quartile is its median, the middle value 2.71. Thus $Q2 = 2.71$.

- The lower and upper subsets are now

Lower: $L = \{1.39, 1.76, 1.90, 2.12, 2.53\}$

Upper: $U = \{3.00, 3.33, 3.71, 3.88, 4.00\}$

The lower set L has median the middle value 1.90, so $Q1 = 1.90$. The upper set has median the middle value 3.71, so $Q3 = 3.71$.

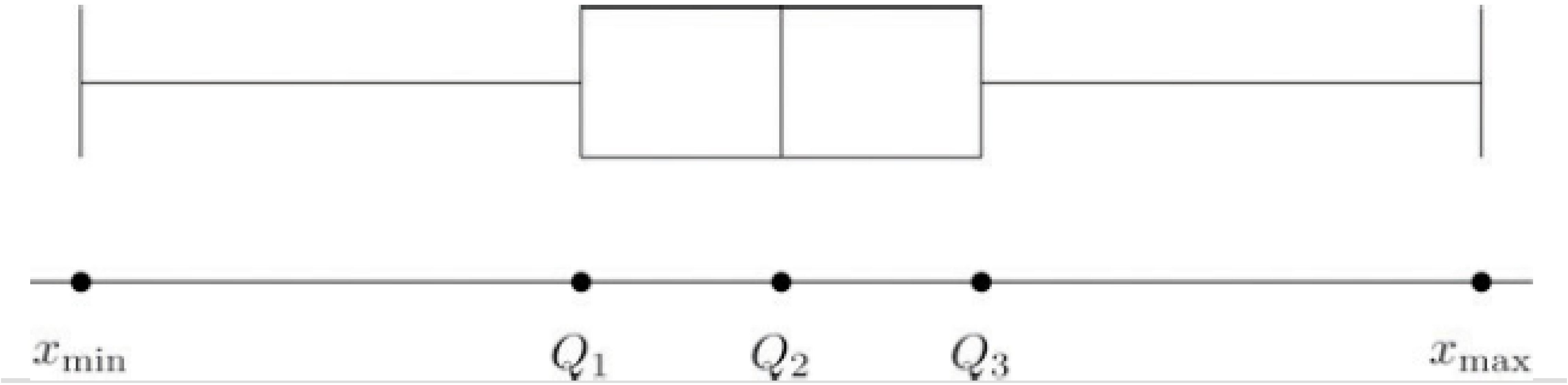
In addition to the three quartiles, the two extreme values, the minimum x_{min} and the maximum x_{max} are also useful in describing the entire data set. Together these five numbers are called the **five-number summary** of the data set:

$$\{x_{min} , Q1 , Q2 , Q3 , x_{max}\}$$

The five-number summary is used to construct a **box plot** as in the following figure. Each of the five numbers is represented by a vertical line segment, a box is formed using the line segments at *Q*1 and *Q*3 as its two vertical sides, and two horizontal line segments are extended from the vertical segments marking *Q*1 and *Q*3 to the adjacent extreme values.

The two horizontal line segments are referred to as “whiskers,” and the diagram is sometimes called a “box and whisker plot.”.

There are other types of box plots that differ somewhat from what presented here, although all are based on the three quartiles.



Note that the distance from Q1 to Q3 is the length of the interval over which the middle half of the data range. Thus it has the following special name.

2) The interquartile range (IQR) is the quantity measure.

Definition: $IQR = Q3 - Q1$

Case (8) Construct a box plot and find the IQR for the data presented in Case (6).

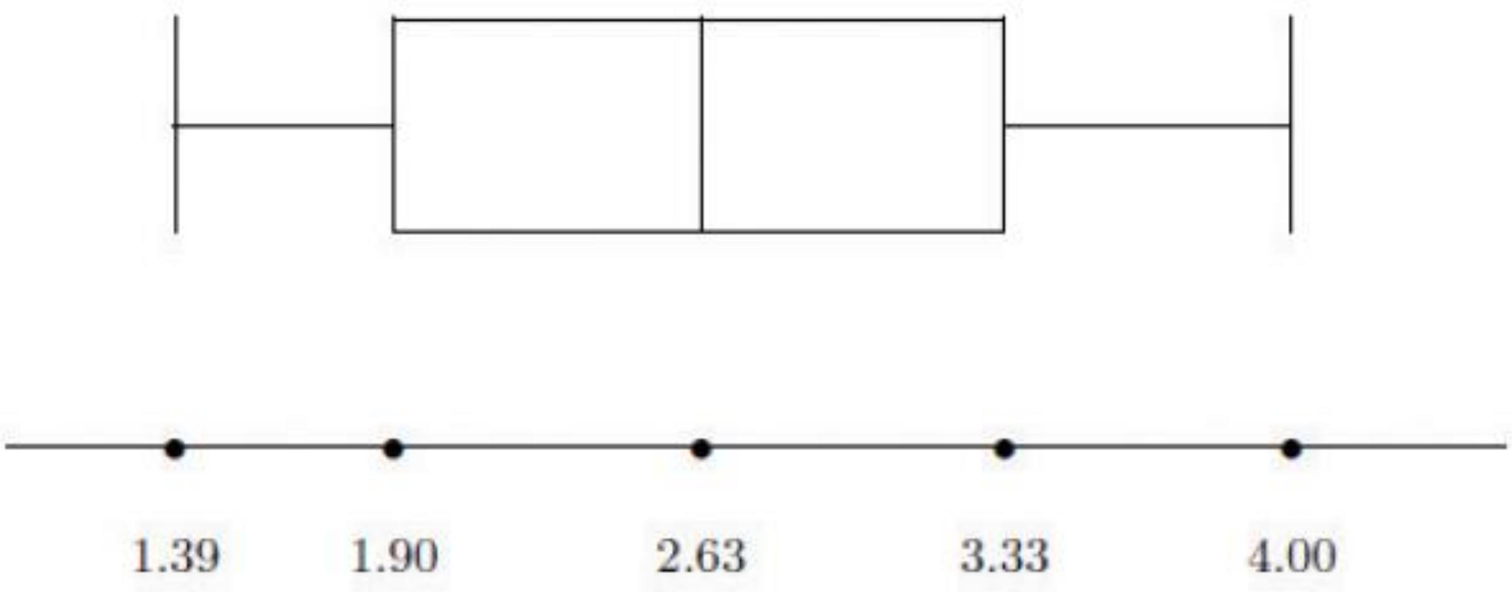
1.39 1.76 1.90 2.12 2.53 2.71 3.00 3.33 3.71 4.00

Solution:

From the solution of previous Case (6) we know that the five-number summary is

$x_{\min} = 1.39$ $Q1 = 1.90$ $Q2 = 2.62$ $Q3 = 3.33$ $x_{\max} = 4.00$

The box plot is



The interquartile range is IQR = 3.33 − 1.90 = 1.43.

3) *z*-scores is another way to locate a particular observation *x* in a data set is to compute its distance from the mean in units of standard deviation.

Definition: *The **z-score**[18] of an observation x is the number z given by the computational formula*

$$z = \frac{x - \bar{x}}{s} \quad or \quad z = \frac{x - \mu}{\sigma}$$

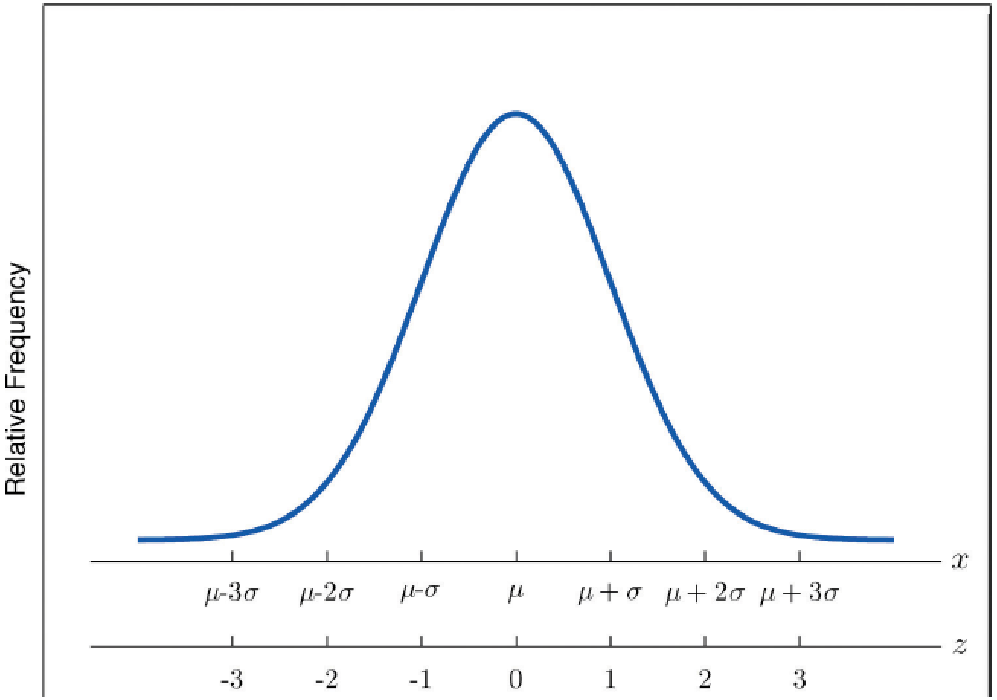
according to whether the data set is a sample or is the entire population.

The formulas in the definition allow us to compute the z-score when x is known. If the z-score is known then x can be recovered using the corresponding inverse formulas:

$x = \bar{x} + sz$ OR $x = \mu + \sigma z$

The z-score indicates how many standard deviations an individual observation x is from the center of the data set, its mean. If z is negative then x is below average.

If z is 0 then x is equal to the average. If z is positive then x is above average (See next figure).



Case (9) Find the *z*-scores for all ten observations for the following data set:

1.90 3.00 2.53 3.71 2.12 1.76 2.71 1.39 4.00 3.33

Solution: Calculated $\bar{x}$ and s for these data: $\bar{x}$ = 2.645 and s = 0.8674.

The first observation x = 1.9 in the data set has z-score: $\mathbf{z} = \frac{x - \bar{x}}{s} = \frac{3.00 - 2.645}{0.8674} = 0.4093$

which means that *x* = 3.00 is 0.4093 standard deviations *above* the sample mean. Repeating the process for the remaining observations gives the full set of *z*-scores: −0.86 0.41 −0.13 1.23 −0.61 −1.02 0.07 −1.45 1.56 0.79

Case (10) Suppose the mean and standard deviation of data set μ = 2.70 and σ = 0.50. The *z*-scores of two values of data set are z = −0.62 and z = 1.28. What are their two original values (x) from data set?

Solution: Using the second formula:

First Value of $x = \mu + z\,\sigma = 2.70 + (-0.62)(0.50) = 2.39$

Second Value of $x = \mu + z\,\sigma = 2.70 + (1.28)(0.50) = 3.34$

Keys:

- Percentile rank and *z*-score of a measurement indicate its relative position with regard to the other measurements in a data set.
- The three quartiles divide a data set into fourths.

• The five-number summary and its associated box plot summarize the location and distribution of the data.

EXERCISES

BASIC

1. Consider the data set

69	92	68	77	80
93	75	76	82	100
70	85	88	85	96
53	70	70	82	85

   a. Find the percentile rank of 82.
   b. Find the percentile rank of 68.

2. Consider the data set

8.5	8.2	7.0	7.0	4.9
9.6	8.5	8.8	8.5	8.7
6.5	8.2	7.6	1.5	9.3
8.0	7.7	2.9	9.2	6.9

   a. Find the percentile rank of 6.5.
   b. Find the percentile rank of 7.7.

3. Consider the data set represented by the ordered stem and leaf diagram

10	0 0
9	1 1 1 1 2 3
8	0 1 1 2 2 3 4 5 7 8 8 9
7	0 0 0 1 1 2 4 4 5 6 6 6 7 7 7 8 8 9
6	0 1 2 2 2 3 4 4 5 7 7 7 7 8 8
5	0 2 3 3 4 4 6 7 7 8 9
4	2 5 6 8 8
3	9 9

   a. Find the percentile rank of the grade 75.
   b. Find the percentile rank of the grade 57.

4. Is the 90th percentile of a data set always equal to 90%? Why or why not?

5. The 29th percentile in a large data set is 5.
   a. Approximately what percentage of the observations are less than 5?
   b. Approximately what percentage of the observations are greater than 5?

6. The 54th percentile in a large data set is 98.6.
   a. Approximately what percentage of the observations are less than 98.6?
   b. Approximately what percentage of the observations are greater than 98.6?

7. In a large data set the 29th percentile is 5 and the 79th percentile is 10. Approximately what percentage of observations lie between 5 and 10?

8. In a large data set the 40th percentile is 125 and the 82nd percentile is 158. Approximately what percentage of observations lie between 125 and 158?

9. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the stem and leaf diagram in Figure 2.2 "Ordered Stem and Leaf Diagram".

10. Find the five-number summary and the IQR and sketch the box plot for the sample explicitly displayed in Note 2.20 "Example 7" in Section 2.2 "Measures of Central Location".

11. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the data frequency table

x	1	2	5	8	9
f	5	2	3	6	4

12. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the data frequency table

x	−5	−3	−2	−1	0	1	3	4	5
f	2	1	3	2	4	1	1	2	1

13. Find the z-score of each measurement in the following sample data set.

$$-5 \quad 6 \quad 2 \quad -1 \quad 0$$

14. Find the z-score of each measurement in the following sample data set.

$$1.6 \quad 5.2 \quad 2.8 \quad 3.7 \quad 4.0$$

15. The sample with data frequency table

x	1	2	7
f	1	2	1

has mean $\overline{x} = 3$ and standard deviation $s = 2.71$. Find the z-score for every value in the sample.

16. The sample with data frequency table

x	−1	0	1	4
f	1	1	3	1

has mean $\overline{x} = 1$ and standard deviation $s = 1.67$. Find the z-score for every value in the sample.

17. For the population

$$0 \quad 0 \quad 2 \quad 2$$

compute each of the following.

a. The population mean μ.
b. The population variance σ^2.
c. The population standard deviation σ.
d. The z-score for every value in the population data set.

18. For the population

$$0.5 \quad 2.1 \quad 4.4 \quad 1.0$$

compute each of the following.

a. The population mean μ.
b. The population variance σ^2.
c. The population standard deviation σ.
d. The z-score for every value in the population data set.

19. A measurement x in a sample with mean $\overline{x} = 10$ and standard deviation $s = 3$ has z-score $z = 2$. Find x.

20. A measurement x in a sample with mean $\overline{x} = 10$ and standard deviation $s = 3$ has z-score $z = -1$. Find x.

21. A measurement x in a population with mean $\mu = 2.3$ and standard deviation $\sigma = 1.3$ has z-score $z = 2$. Find x.

22. A measurement x in a sample with mean $\mu = 2.3$ and standard deviation $\sigma = 1.3$ has z-score $z = -1.2$. Find x.

APPLICATIONS

23. The weekly sales for the last 20 weeks in a kitchen appliance store for an electric automatic rice cooker are

20	15	14	14	18
15	19	12	13	9
15	17	16	16	18
19	15	15	16	15

a. Find the percentile rank of 15.
b. If the sample accurately reflects the population, then what percentage of weeks would an inventory of 15 rice cookers be adequate?

24. The table shows the number of vehicles owned in a survey of 52 households.

x	0	1	2	3	4	5	6	7
f	2	12	15	11	6	3	1	2

a. Find the percentile rank of 2.
b. If the sample accurately reflects the population, then what percentage of households have at most two vehicles?

25. For two months Cordelia records her daily commute time to work each day to the nearest minute and obtains the following data:

x	26	27	28	29	30	31	32
f	3	4	16	12	6	2	1

Cordelia is supposed to be at work at 8:00 a.m. but refuses to leave her house before 7:30 a.m.

a. Find the percentile rank of 30, the time she has to get to work.
b. Assuming that the sample accurately reflects the population of *all* of Cordelia's commute times, use your answer to part (a) to predict the proportion of the work days she is late for work.

26. The mean score on a standardized grammar exam is 49.6; the standard deviation is 1.35. Dromio is told that the z-score of his exam score is −1.19.

a. Is Dromio's score above average or below average?
b. What was Dromio's actual score on the exam?

27. A random sample of 49 invoices for repairs at an automotive body shop is taken. The data are arrayed in the stem and leaf diagram shown. (Stems are

thousands of dollars, leaves are hundreds, so that for example the largest observation is 3,800.)

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

For these data, $\Sigma x = 101{,}100$, $\Sigma x^2 = 244{,}830{,}000$.

a. Find the z-score of the repair that cost $1,100.
b. Find the z-score of the repairs that cost $2,700.

28. The stem and leaf diagram shows the time in seconds that callers to a telephone-order center were on hold before their call was taken.

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

a. Find the quartiles.
b. Give the five-number summary of the data.
c. Find the range and the IQR.

ADDITIONAL EXERCISES

29. Consider the data set represented by the ordered stem and leaf diagram

10	0 0
9	1 1 1 1 2 3
8	0 1 1 2 2 3 4 5 7 8 8 9
7	0 0 0 1 1 2 4 4 5 6 6 6 7 7 7 8 8 9
6	0 1 2 2 2 3 4 4 5 7 7 7 7 8 8
5	0 2 3 3 4 4 6 7 7 8 9
4	2 5 6 8 8
3	9 9

a. Find the three quartiles.
b. Give the five-number summary of the data.
c. Find the range and the IQR.

30. For the following stem and leaf diagram the units on the stems are thousands and the units on the leaves are hundreds, so that for example the largest observation is 3,800.

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

a. Find the percentile rank of 800.
b. Find the percentile rank of 3,200.

31. Find the five-number summary for the following sample data.

x	26	27	28	29	30	31	32
f	3	4	16	12	6	2	1

32. Find the five-number summary for the following sample data.

x	1	2	3	4	5	6	7	8	9	10
f	384	208	98	56	28	12	8	2	3	1

33. For the following stem and leaf diagram the units on the stems are thousands and the units on the leaves are hundreds, so that for example the largest observation is 3,800.

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

a. Find the three quartiles.
b. Find the IQR.
c. Give the five-number summary of the data.

34. Determine whether the following statement is true. "In any data set, if an observation x_1 is greater than another observation x_2, then the *z*-score of x_1 is greater than the *z*-score of x_2."

35. Emilia and Ferdinand took the same freshman chemistry course, Emilia in the fall, Ferdinand in the spring. Emilia made an 83 on the common final exam that she took, on which the mean was 76 and the standard deviation 8. Ferdinand made a 79 on the common final exam that he took, which was more difficult, since the mean was 65 and the standard deviation 12. The one who has a higher *z*-score did relatively better. Was it Emilia or Ferdinand?

36. Refer to the previous exercise. On the final exam in the same course the following semester, the mean is 68 and the standard deviation is 9. What grade on the exam matches Emilia's performance? Ferdinand's?

37. Rosencrantz and Guildenstern are on a weight-reducing diet. Rosencrantz, who weighs 178 lb, belongs to an age and body-type group for which the mean weight is 145 lb and the standard deviation is 15 lb. Guildenstern, who weighs 204 lb, belongs to an age and body-type group for which the mean weight is 165 lb and the standard deviation is 20 lb. Assuming *z*-scores are good measures for comparison in this context, who is more overweight for his age and body type?

ANSWERS

1. a. 60.
   b. 10.

3. a. 59.
   b. 23.

5. a. 29.
   b. 71.

7. 50%.

9. $x_{\min} = 25, Q_1 = 70, Q_2 = 77.5, Q_3 = 90, x_{\max} = 100$, $\text{IQR} = 20$

11. $x_{\min} = 1, Q_1 = 1.5, Q_2 = 6.5, Q_3 = 8, x_{\max} = 9, \text{IQR} = 6.5$

13. −1.3, 1.39, 0.4, −0.35, −0.11.

15. $z = -0.74$ for $x = 1$, $z = -0.37$ for $x = 2$, $z = 1.48$ for $x = 7$.

17. a. 1.
    b. 1.
    c. 1.
    d. $z = -1$ for $x = 0$, $z = 1$ for $x = 2$.

19. 16.

21. 4.9.

23. a. 55.
    b. 55.

25. a. 93.
    b. 0.07.

27. a. −1.11.
    b. 0.73.

29. a. $Q_1 = 59, Q_2 = 70, Q_3 = 81$.
    b. $x_{\min} = 39, Q_1 = 59, Q_2 = 70, Q_3 = 81, x_{\max} = 100$.
    c. $R = 61$, $\text{IQR} = 22$.

31. $x_{\min} = 26, Q_1 = 28, Q_2 = 28, Q_3 = 29, x_{\max} = 32$.

33. a. $Q_1 = 1450, Q_2 = 2000, Q_3 = 2800$.
    b. $\text{IQR} = 1350$.
    c. $x_{\min} = 400, Q_1 = 1450, Q_2 = 2000, Q_3 = 2800$, $x_{\max} = 3800$.

35. Emilia: $z = .875$, Ferdinand: $z = 1.1\overline{6}$.

37. Rosencrantz: $z = 2.2$, Guildenstern: $z = 1.95$. Rosencrantz is more overweight for his age and body type.