Chapter 15 - Descriptive Statistics

Content:

1. Data Types

New statistical terms are in **bold red**.

Statistical Research: Based on the research question to be answered we must determine:

- what is the **population** to which the question applies
- what **data** should be collected (**variables**)
- how to collect the data (census or one of the sampling methods)
- methods for the data analysis

Different variables have different characteristics (data types) and based on that allow different calculations. For that reason, variable data types are also called the Levels of Measurement

Variable	**Data Type (Level of Measurement)**	**Operations**	**Example**
Qualitative / Categorical	Nominal	Collected data values **can only be counted**. Only counts can be used in computations. In reports, **the data is presented in any order** (report first number for males and then females or opposite).	Gender, race…
	Ordinal	Collected data values **can only be counted** but the data is naturally ordered. In reports, **the data is presented in order** (freshmen, sophomores, juniors, seniors)	Seniority at school.
Quantitative / Numerical	Interval	Collected data values can be counted, added, or subtracted BUT cannot be multiplied or divided. Zero can be a value and negative values are possible.	Daily temperature in some city.
	Ratio	Collected data values can be counted, +, - , *, and / Zero indicates the absence of something. Negative values are not possible.	Scores on the test.

It is very important to accurately determine the data type of each variable because that tells us what calculations we can do with it and how to represent it in the reports. For example, daily temperatures can be counted (how many days were below 32F) BUT they can also be added and subtracted too because the daily temperature is an interval variable.

Quantitative (numerical) variables can be:

1. continuous (the time spent on homework)
2. discrete (the score on test)

Qualitative (categorical) variables (examples: gender, race, and major at the college)

The categorical data are **often coded as numbers** but these numbers are arbitrary and **cannot be used for computations**.

Example: Values for the variable "gender" of students in one class can be coded as M and F or 1 for males and 2 for females. In both cases, it makes sense only to count M-s and F-s (or count 1s and 2s).
It does not make sense to sum 1s and sum 2s – that would mean nothing in the given context.

Whenever the variable (gender in this case) allows us to code data in different ways it is a **qualitative variable.**
We typically count the instances of different values and use these counts in computations.

Example: In the class for variable "gender" we code males as 1s and females as 2s. We count 20 males (twenty 1s) and 30 females (30 2s). We need to determine the percentage of males and females in the class.

Based on counts we conclude that the total number of students is $20 + 30 = 50$

Now we can find the percentages:

$$\frac{20}{50} = 0.4 = 40\%\ are\ males \quad and \qquad \frac{30}{50} = 0.6 = 60\%\ are\ females$$

Observe that all computations are done using counts, we did not use actual values (1s and 2s).

Quantitative (numerical) variables (examples: daily temperature in some city or scores on the test)

When we deal with some quantitative variable we can count the instances of some values BUT we can also do calculations with collected values.

Example: If we collect a time to do the homework from 22 students in one class, we can:

- count how many of the collected times are less than an hour and how many are greater; (say, 15 students worked less than one hour and 7 worked more)
- add all times to get a" total time spent on the homework" in one class (for example, a total of 25.2 hours). If we know that there were 22 students we can find an average time spent on homework: $25.2\ /\ 22 \approx 1.16$)

In other words, for both qualitative and quantitative data, we count the values and do computations using counts BUT with **quantitative data,** we can also add/subtract (interval and ratio types) and we can multiply/divide (ratio types).

2. Graphical Representation of the Data

Statistical Research: Based on the research question to be answered we must determine:

- what is the **population** to which the question applies
- what **data** should be collected (**variables**)
- how to collect the data (census or one of the sampling methods)
- methods for the data analysis (descriptive and inferential)

After the data is collected it needs to be analyzed. **Methods to analyze the data may include:**

- summaries and graphs (this is **descriptive statistics,** *(statistics as the practice or science of collecting and analyzing the data, not statistic-s which are numbers based on the sample of the data)*

- the use of statistic-s (data from the sample) to **estimate population parameters** (this is **inferential statistics**). The inferential statistics produces good results ONLY if the sample is ***representative of the population*** (if the sampling method was appropriate).

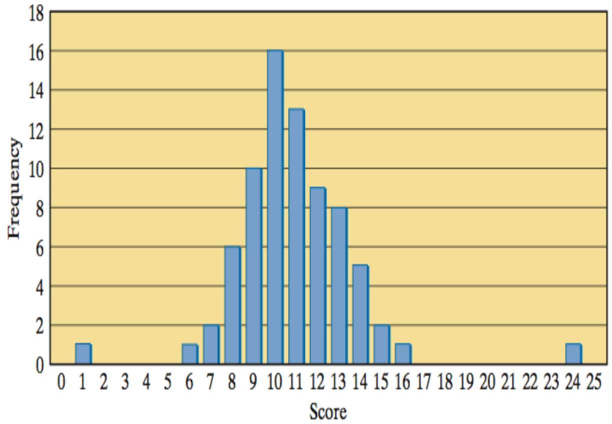
The summaries of the data can be organized in:

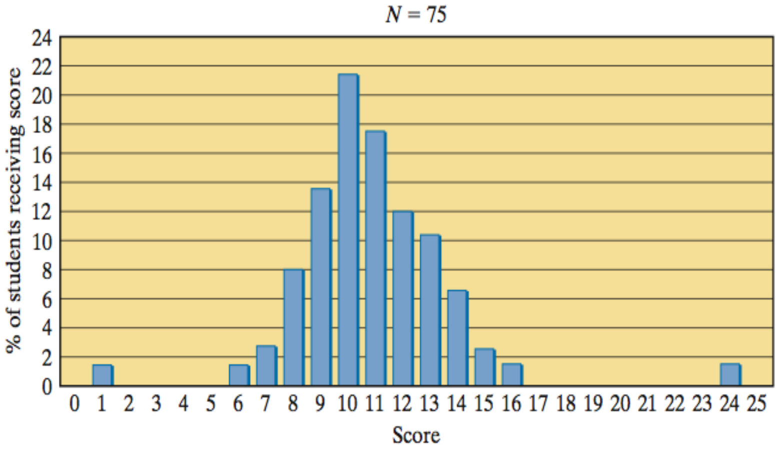
- frequency bar graphs (count instances)
- relative frequency bar graphs (frequencies are given as percentages)
- pictograms (bar graphs that use icons or pictures instead of bars)
- pie charts (**use only for nominal data** AND when the number of categories is small)
- tables
- histogram (for the intervals defined on continuous data **– the bars must touch or be very close**)

Example 1. The scores on the exam in one class are given in the **frequency bar graph** (on the left) and a **relative frequency bar graph** (on the right). Observe that:

- **relative frequency bar graph must indicate the total number of students (N=75).** Otherwise, the graph can be misleading, especially if the N is small. **Always put N** (or n for sample data) **on ALL graphs.**

- as both graphs show the same data we can draw the same conclusions from both but, in some cases, we may need additional calculations (for example to calculate percentages when we read frequency bar graph).





How many students scored 9 points? Read from the frequency bar graph: it is 10 students.

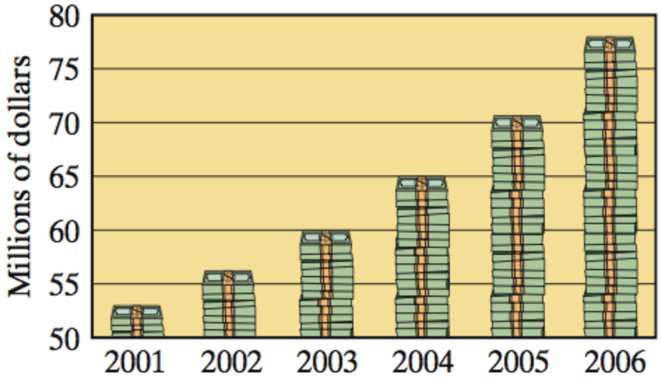
What is the percentage of students who scored 9 points? $\frac{10}{75} = 0.1754 \approx 17.5\%$

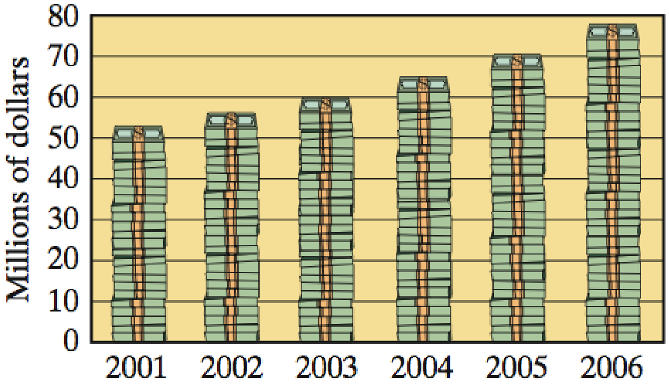
Watch for the possible misleading actions done in graphs:

1) stretching the scale of the vertical axis

2) "cheating" on the choice of starting value on the vertical axis (graph does not start with 0)

Example 2: Two **pictograms** below represent the same data. In the graph on the right, we see the real growth while the graph on the left leads us to believe that the growth was much higher (the scale is stretched and it does not start with 0).



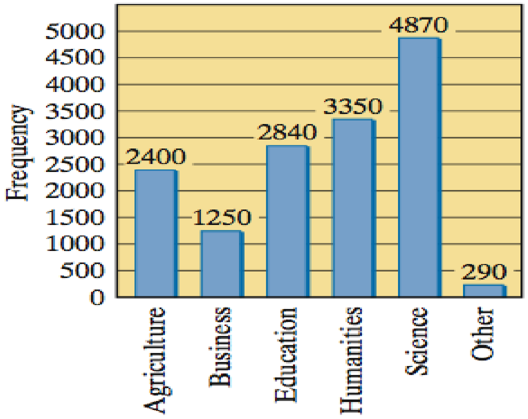


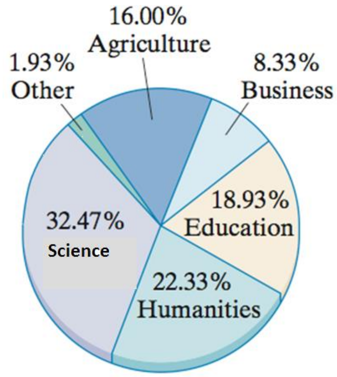
The pie chart should be used **only for nominal data**. The number of categories should be relatively small (up to 10).

Example 3. The number of students per major is represented both in the bar graph and in the pie-chart.

Observe that the **order of bars in the bar graph is arbitrary.** It is in alphabetical order in this case.

The bars could have been ordered according to some other criteria. For example, from the smallest majors to the largest ones or opposite. That is possible because the major is a nominal variable. Observe that bars do not touch.





Only for the nominal data, the order of bars in the bar graphs is arbitrary.

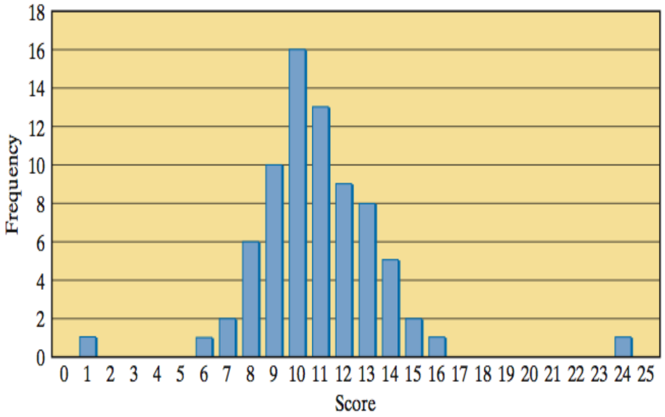
In all other cases (**ordinal, interval, and ratio data**) bars must be represented in the order that makes sense.

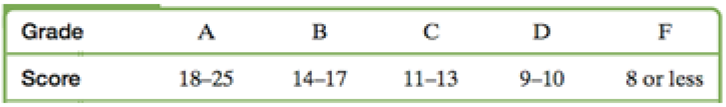
When the number of bars is bigger than 20, a bar graph becomes hard to read.

How many bars are recommended in a bar graph? Between 5 and 20 bars.

Example 4. Quiz scores for one class are presented in the frequency bar graph below.

Create a frequency bar graph with only 5 bars and compare the two graphs for readability.





Grade	A	B	C	D	F
Score	18–25	14–17	11–13	9–10	8 or less

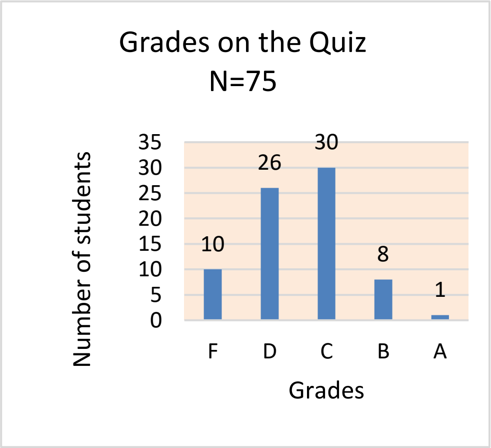
To get only 5 bars we must group scores into 5 categories (A-F). This is entirely our choice: one option is in the table above

Then we count the number of students in each category by reading the graph.

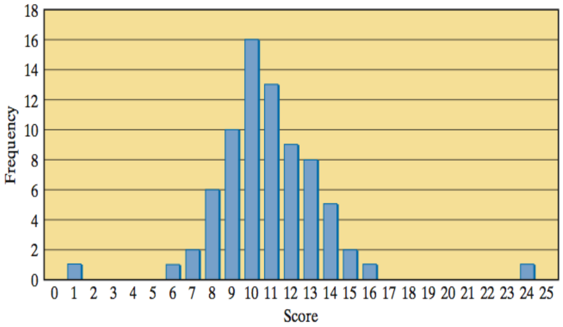
- A: Students who scored 18-25: 1
- B: Students who scored 14-17: 5+2+1=8
- C: Students who scored 11-13: 13+9+8=30
- D: Students who scored 9-10: 10+16=26
- F: Students who scored 8 or less: 6+2+1+1=10

Now we can create a table and make a bar graph based on it.

Quiz scores for one class					
Grade	A	B	C	D	E
Score	18-25	14-17	11-13	9-10	8 or less
Number of students	1	8	30	26	10



Compare the readability of the 5-bar graph to the original 25-bar graph:

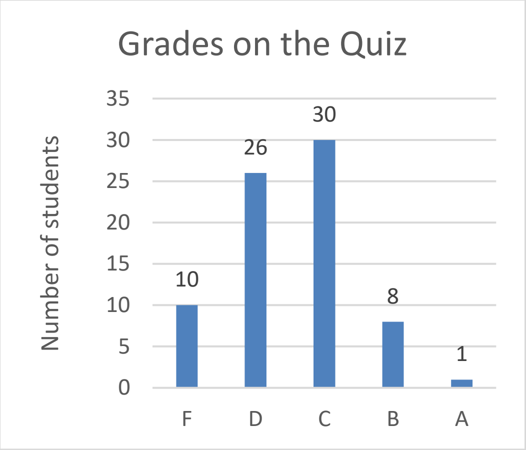


How to turn the frequency bar graph into a relative frequency bar graph?

Example: The data from the table below are represented in the bar graph.

We want to make a frequency bar graph out of this data (to represent the data in percentages).

Quiz scores for one class					
Grade	A	B	C	D	E
Number of students	1	8	30	26	10



First we must find total number of students: $1 + 8 + 30 + 26 + 10 = 75$

Then we must find percent of students for each grade:

A: $\frac{1}{75} = 0.0133 = 1.33\%$ B: $\frac{8}{75} = 0.106\overline{6} = 10.67\%$

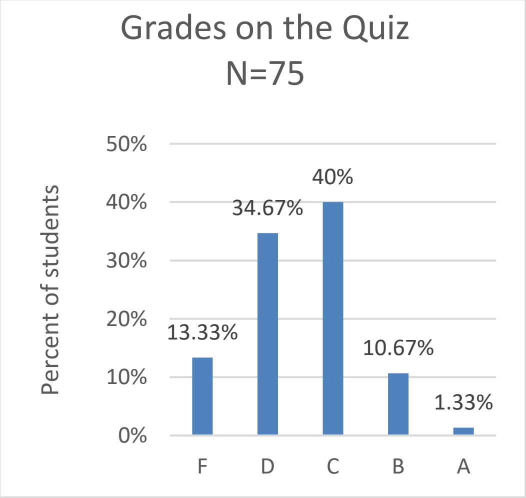
C: $\frac{30}{75} = 0.40 = 40\%$ D: $\frac{26}{75} = 0.3467 = 34.67\%$

F: $\frac{10}{75} = 0.1333 = 13.33\%$

Quiz scores for one class					
Grade	A	B	C	D	E
Number of students	1	8	30	26	10
% of students	1.33%	10.67%	40%	34.67%	13.33%

Now we can make a frequency bar graph using percentages instead of counts.

It is mandatory to show $\boldsymbol{N}$ (if the graph is about population) or $\boldsymbol{n}$ (if the graph is about the sample**).**



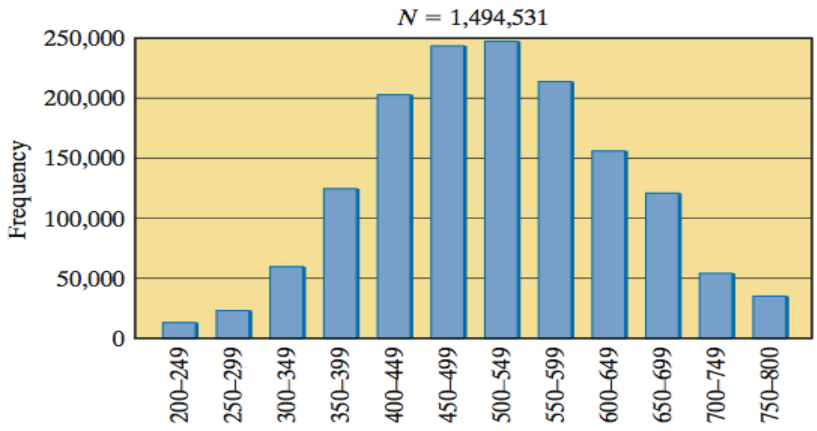
Histogram

When the number of possible variable values is great (SAT scores of 800) or the **variable is continuous**, it is not practical to represent each individual value by the bar graph.

In such cases, the data (variable values) are grouped in intervals as we did in the previous example.

When the variable is continuous, such a bar graph has a special name: it is called **histogram**.

The histogram should also have 5-20 intervals preferably of the equal length and the bars should be either very close or touch.



Student's SAT scores.

What is the total number of students?

N = 1,494,531 students

How many students scored between 500 and 549?

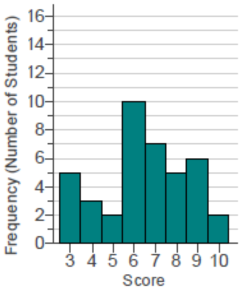
250,000 students

What is the percent of students who scored between 500 and 549?

$$\frac{250{,}000}{1{,}494{,}531} \approx 0.167 = 16.7\%$$

Solved examples from HWK:

The bar graph describes the scores of a group of students on a 10-point math quiz



1. How many students took a quiz?

Add numbers from bars: 5 + 3 + 2 + 10 + 7 + 5 + 6 + 2 = 40

What percentage of students scored 2? 0 students score 2

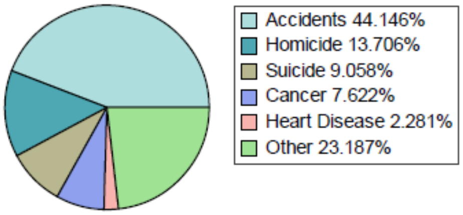
Percent is: 0 /40 = 0 so 0%

What percentage of students scored 6 or more?

Scores of 6 or more: 10 + 7 + 5 + 6 + 2 =30

Percent is: 30 /40 = 0.75 = 75%

Possible Cause of Death in a Population Among 18 to 22-Year-Olds, N = 19,548



2. Use the pie chart on to answer the questions:

a. Is the Couse of Death qualitative or quantitative variable?
b. How many deaths were due to an accident? (round to the integer)

a. Causes of death are categories so Cause of Death is categorical / qualitative variable. Only qualitative data can be represented in a pie chart.

b. Accidents = 44.146% = 0.44146 Total number of deaths is N = 19,548

Compute 19548 (0.44146) = 8629.66008 round to 8630

3. Numerical Summaries (Medians, Means, Percentiles, and Box-plots)

Some measures of location that describe quantitative data are:

- **Average** (also called **a mean**) is arithmetic average
- **Minimum** - the smallest value in the data set
- **Maximum** - the largest value in the data set
- **Median (50th percentile)**
- **Quartiles (1st and 3rd quartiles – those are 25th percentile and 75th percentile)**

An Average (A) of a set of *N* numbers is found by adding the numbers and dividing the total by *N*

$$A = \frac{d_1+d_2+d_3+\cdots+d_N}{N}$$

Observe that average is also called a mean.

Practice: Find the minimum, maximum, and mean for {3, 1, 5, -4, -5} Answer: $Min = -5 \quad Max = 5 \quad A = 0$

Practice: Find the minimum, maximum, and mean for {-3, 0, 1, 0} Answer: $Min = -3 \quad Max = 1 \quad A = -0.5$

Observe that the calculated average, A, is often NOT the element of the set.

Example: To find an average ***A*** of a data set **given by the frequency table** such as table below we do the following:

Score	5	10	15	20
Frequency	3	0	4	1

1. Read the frequency table: there are 3 instances of the score 5; 0 instances of the score 10…

   The frequency table is shorthand for writing out each element.

   In this example the scores without the use of frequency table would look like: {5, 5, 5, 15, 15, 15, 15, 20}

   We can find the sum of all scores by adding them up $S = 5+5+5+15$… = 95 or we can sum them up using the frequencies from the table $S = 5(3)+10(0)+15(4)+20(1) = 15+60+20 = 95$

2. The actual number of elements (scores in this case), is: $N = 3+0+4+1 = 8$

3. The average is computed by dividing the sum by a number of elements: $A = \frac{S}{N} = \frac{95}{8} = 11.875$

General observation: Different calculators may do rounding in slightly different ways so sometimes your results may be slightly different from what we computed in the class or what you see in the book.

Observe that to compute other measures of location (**minimum, maximum, median, Q1 and Q3**) **the data must be sorted first.**

Steps to compute median:

1. Sort the data set from smallest to largest. Assume that $d_1, d_2, d_3, \ldots \quad d_N$ represent the **sorted data**. N is the number of elements.
2. If **N is odd** then the median **is an element *d*** with the index $\frac{N+1}{2}$ *(this is an element from the data set)*
3. If **N is even** then the median is an ***average of elements*** d_i ***with the indexes*** $\frac{N}{2}$ ***and*** $\frac{N}{2}+1$

Consider the data set { -5.6, 2.5, -4.7, 4.9, 4.9, -0.7, -4.7, -0.4 }

a. Find the average.

b. Find the median.

c. Consider the data set { -5.6, 2.5, -4.7, 4.9, 4.9, -0.7, -4.7 } having one less data point than the original set. Find the average and the median of this data set.

$$a.\quad A=\frac{-5.6+2.5-4.7+4.9+4.9-0.7-4.7-0.4}{8}=\frac{-3.8}{8}$$

$$=-0.475$$

b. For the median, we **must order** the data set and number the elements

{ -5.6,	-4.7 ,	-4.7,	-0.7,	-0.4,	2.5,	4.9,	4.9 }
1	2	3	4	5	6	7	8

There are 8 elements (**N is even**) so compute $\frac{N}{2}=\frac{8}{2}=4$ and

average of 4th and 5th elements: $\frac{-0.7-0.4}{2}=\frac{-1.1}{2}=-0.55$

$$c.\quad A=\frac{-5.6+2.5-4.7+4.9+4.9-0.7-4.7}{7}$$

$$=\frac{-3.4}{7}=-0.4957\ldots$$

For the median, we **must order** the data set

{ -5.6,	-4.7 ,	-4.7,	-0.7,	-2.5,	4.9,	4.9 }
1	2	3	4	5	6	7

There are 7 elements (**N is odd**) so compute

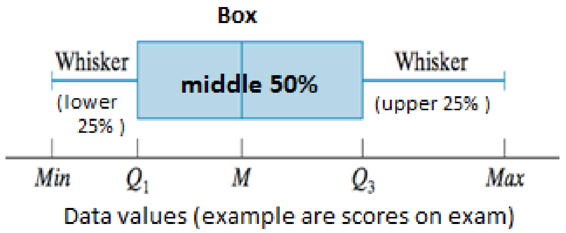
$\frac{N+1}{2}=\frac{7+1}{2}=4$ select the 4th element: -0.7

Box Plots - They are NOT always symmetrical around the mean!

The five-number summary

Min	-	Minimum	
Q_1	-	1st quartile	(25%)
Q_2 or M	-	Median	(50%)
Q_3	-	3rd quartile	(75%)
Max	-	Maximum	(100%)

Box Plots represent visually five-number summary and frequencies.

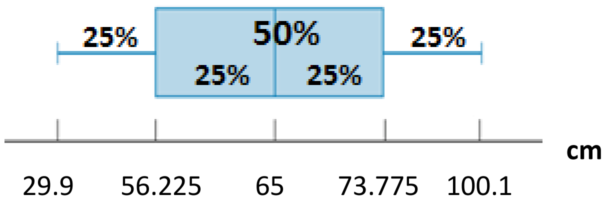


Frequencies (example: 50% of students scored between Q1 and Q3)

What percent of the data is between 56.2 and 73.7? 50%

What percent of the data is between 29.9 and 73.7? 75%

What percent of the data is between 56.2 and 100.1? 75%



What percent of the data is above 29.9 cm? _________ 100%

What percent of the data is below 56.2 cm? _________ 25%

What percent of the data is above 56.2 cm? _________ 75%

What percent of the data is above 65 cm? __________50%

What percent of the data is above 73.7 cm? _________ 25 %

What percent of the data is above 100.1 cm? _________ 0%

0% of the data is below ________ 29.9 cm

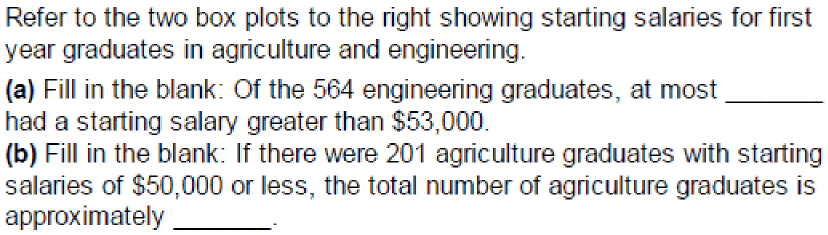
25% of the data is below _______ 56.2 cm

50% of the data is below________ 65 cm

75% of the data is below________ 73.7 cm

0% of the data is above ________ 100.1 cm

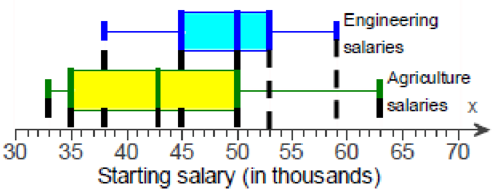
25% of the data is above________ 73.7 cm



Refer to the two box plots to the right showing starting salaries for first year graduates in agriculture and engineering.

(a) Fill in the blank: Of the 564 engineering graduates, at most ________ had a starting salary greater than $53,000.

(b) Fill in the blank: If there were 201 agriculture graduates with starting salaries of $50,000 or less, the total number of agriculture graduates is approximately ________.



a) $53,000 = Q3 **Above Q3** is 25% of salaries so we compute 564 (0.25) = 141 engineering graduates

b) $50,000 = Q3 **Below Q3** is 75% of salaries. We do not know a total number of agricultural graduates so we set up the percentile formula as:

$$\mathrm{N}(0.75) = 201 \quad \text{and compute} \quad \mathrm{N} = \frac{201}{0.75} = 268 \text{ agriculture graduates}$$

4. The measures of spread: Interquartile range (IQR) and standard deviation (SD or σ)

The **range, R** is computed as: **R = Max – Min**

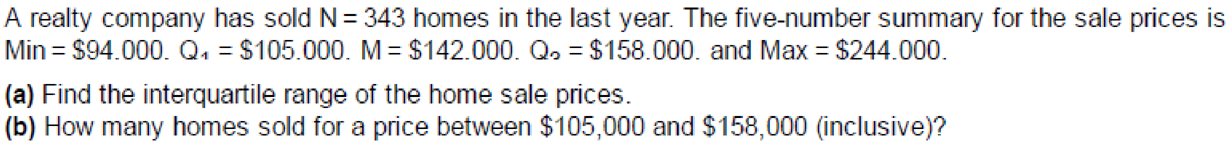
Example: For the set {-6, -2, 0, 7, 8} $\mathbf{Min} = -\mathbf{6}$ $\mathbf{Max} = \mathbf{8}$ $\mathbf{R} = \mathbf{8} - (-\mathbf{6}) = \mathbf{8} + \mathbf{6} = \mathbf{14}$

When extreme values of Min and Max (outliers) exist this gives a misleading impression about the spread of the data.

To eliminate the effect of outliers we can focus on **the spread of the middle 50%** (the blue box in the box plots).

The middle 50% are in the **Interquartile range (IQR)**:

$$\mathbf{IQR} = \mathbf{Q_3} - \mathbf{Q_1}$$



A realty company has sold N = 343 homes in the last year. The five-number summary for the sale prices is Min = $94.000. Q_1 = $105.000. M = $142.000. Q_3 = $158.000. and Max = $244.000.

(a) Find the interquartile range of the home sale prices.

(b) How many homes sold for a price between $105,000 and $158,000 (inclusive)?

First, make a box plot like one on the right.

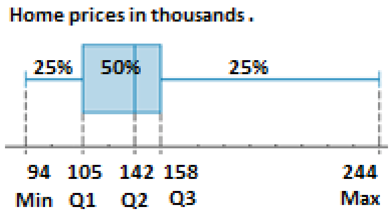
Observe that in this case, a box-plot is NOT symmetrical. The left and right tails are VERY different in length.

Also, the lengths Q2 – Q1 and Q3 – Q2 are different.

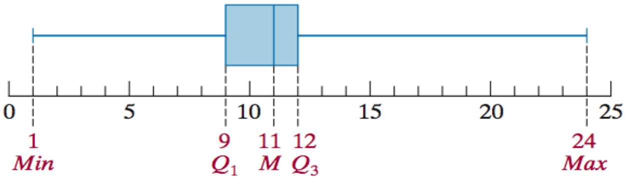
a. IQR = Q3 – Q1 = 158 - 105 = 53 IQR is $53,000

b. N = 343 and 50% sold between $105,000 and $158,000

Compute 343(0.50) = 171.5 round to 172



For the box-plot below $IQR = 12 - 9 = 3$ (observe that the box is **often NOT symmetrical around the mean**).



Middle 50% of the data is in the interval [9, 12]

There are large whiskers on both sides so the range is: $\boldsymbol{R = 24 - 1 = 23}$
The data is in the interval [1, 24]

We may **want to eliminate outliers and make whiskers smaller.**

One way to **eliminate the outliers is to compute**:

$\mathbf{Q_3 + 1.5\ (IQR)}$ called **upper fence**

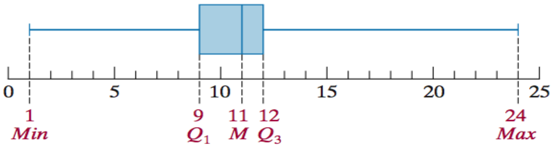
$\mathbf{Q_1 - 1.5\ (IQR)}$ called **lower fence**

Outliers are:

data values > $\mathbf{Q_3 + 1.5\ (IQR)}$

data values < $\mathbf{Q_1 - 1.5\ (IQR)}$

Outliers are often eliminated from further statistical processing.



$Q_3 = 12 \quad Q_1 = 9 \quad \text{IQR} = 12 - 9 = 3$

Compute the upper fence for the box-plot above: $Q_3 + 1.5\ (\text{IQR}) = 12 + 1.5(3) = 12 + 4.5 = 16.5$

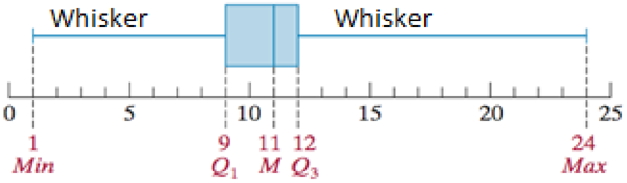
***Interpret**: 16.5 is upper fence and data values above 16.5 will be eliminated as outliers.*

Compute the lower fence for the box plot above: $Q_1 - 1.5\ (\text{IQR}) = 9 - 1.5(3) = 9 - 4.5 = 4.5$

***Interpret**: 4.5 is lower fence and data values below 4.5 will be eliminated as outliers.*

No outliers have been eliminated.

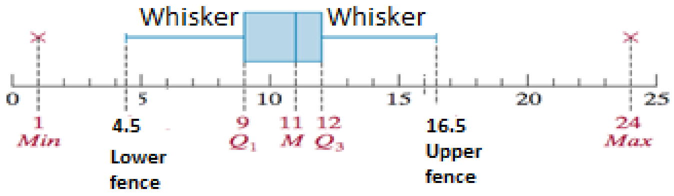
The data is in the interval: [1, 24]



- **25% of data is in the interval [1, 9]**
  i.e. 25% of the data is between 1 and 9
- **50% of the data is in the interval (9, 12]**
  i.e. 50% of the data is between 9 and 12
- **25% of the data is in the interval (12, 24]**
  i.e. 25% of the data is between 12 and 24

Outliers have been eliminated (whiskers are smaller).

The data is in the interval [4.5, 16.5]



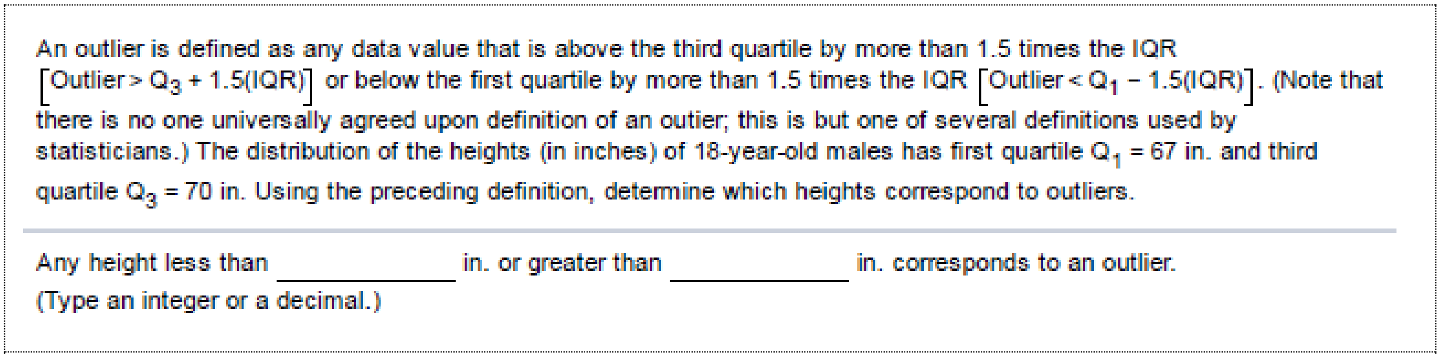
- **25% of data is in the interval [4.5, 9]**
  i.e. 25% of the data is between 4.5 and 9
- **50% of the data is in the interval (9, 12]**
  i.e. 50% of the data is between 9 and 12
- **25% of the data is in the interval (12, 16.5]**
  i.e. 25% of the data is between 12 and 16.5

So, the original data in the box plot is in the interval $[1, \ 24]$ with the range: $\mathbf{R = 24 - 1 = 23}$

After eliminating the outliers the data is in the interval $[4.5, \ 16.5]$ with the range: $\mathbf{R = 16.5 - 4.5 = 12}$

Any further statistical processing will include only data values from the interval $[4.5, \ 16.5]$.

There are other ways to eliminate outliers but we will use this one.



An outlier is defined as any data value that is above the third quartile by more than 1.5 times the IQR $\left[\text{Outlier} > Q_3 + 1.5(\text{IQR})\right]$ or below the first quartile by more than 1.5 times the IQR $\left[\text{Outlier} < Q_1 - 1.5(\text{IQR})\right]$. (Note that there is no one universally agreed upon definition of an outier; this is but one of several definitions used by statisticians.) The distribution of the heights (in inches) of 18-year-old males has first quartile $Q_1 = 67$ in. and third quartile $Q_3 = 70$ in. Using the preceding definition, determine which heights correspond to outliers.

Any height less than ____________ in. or greater than ____________ in. corresponds to an outlier.
(Type an integer or a decimal.)

In tis case we have Q1 = 67 in and Q3 = 70 in

IQR = Q3 – Q1 = 70 – 67 = 3 Q1 – 1.5(IQR) = 67 – 1.5(3) = 67 – 4.5 = 62.5 in

Q3 + 1.5(IQR) = 67 + 1.5(3) = 70 + 4.5 = 74.5 in

The Standard Deviation (denoted as σ or SD)

Another way to look at the spread of the data is by **focusing on the average and the spread around the average**.

If ***A*** is the average (mean) of the data set and ***X*** **is one value of the set,**

$$x's \ \ deviation \ from \ the \ mean = \ x - A$$

So, now instead of the elements of the set, we will use their deviations (differences $x - A$).

Calculating the average of these deviations will give us 0 because negative and positive deviations will cancel each other out. This makes the average useless in this case.

The cancellation of positive and negative deviations can be avoided by **squaring each deviation** making it positive.

When all squared deviations are added and divided by the number of elements N (the average of squared deviations is computed) the result is called **variance** and denoted by **V**. $Observe \ that \ V \geq 0$

Finally, we take the square root of the variance and get the **standard deviation (SD)**.
SD is also denoted by the Greek letter σ (read sigma). We can write that $\sigma = \sqrt{V}$

The steps to compute the standard deviation.

1. **find the average (mean) of all elements in the dataset. x denotes any element from that set,** A **is average**
2. **find the deviation for each element x from the average and square it:** $(x - A)^2$
3. **sum all squared deviations and compute their average; the result is a variance** (denoted as V)
4. **find the square root of variance; this is a standard deviation** (denoted as σ *o r SD*)

Standard deviation (SD) indicates how the data is spread around the mean.

SD = 0 means that all data values are the same. Larger SD indicates that the data is more varied (spread).

The SD of a data set is measured in the same units as the original data.

a. For example, if we deal with a set of scores on a test in points, then the SD is also given in points.

b. If the SD is given in dollars, then we can conclude that the original data must have been money such as prices or salaries. The data could not have been, for example, test scores.

For data sets that are based on the same underlying scale, a comparison of standard deviations can tell us which dataset has a larger spread of the data.

Standard deviations of data sets that are given in different units can be compared only after the data is standardized using z-scores. We will not cover z-scores in this course.

5. Compute SD

Here we show how SD is computed. Manual computation of SD will not be on quizzes or tests but you should have an idea of how it is computed.

Chapter 15, problem 55 – **Find the standard deviation (SD) for the problems a), b), and c)**			
Steps	a) Compute SD for **{5, 5, 5, 5}**	b) Compute SD for **{0, 5, 5, 10}**	c) Compute SD for **{0, 10,10, 20}**
1. Find an average $A = \frac{Sum}{N}$	$A = \frac{5+5+5+5}{4} = \frac{20}{4} = 5$	$A = \frac{0+5+5+10}{4} = \frac{20}{4} = 5$	$A = \frac{0+10+10+20}{4} = \frac{40}{4}$ $= 10$
2. Find the **deviation for each element from the A** and square it: $(x-A)^2$	$5-5=0 \quad 0^2=0$ $5-5=0 \quad 0^2=0$ $5-5=0 \quad 0^2=0$ $5-5=0 \quad 0^2=0$	$0-5=-5 \quad (-5)^2=25$ $5-5=0 \quad 0^2=0$ $5-5=0 \quad 0^2=0$ $10-5=5 \quad 5^2=25$	$0-10=-10 \quad (-10)^2=100$ $10-10=0 \quad 0^2=0$ $10-10=0 \quad 0^2=0$ $20-10=10 \quad 10^2=100$
3. Find the average of squared deviations (this is called **variance V**)	$V = \frac{0+0+0+0}{4} = 0$	$V = \frac{25+0+0+25}{4} = \frac{50}{4}$	$V = \frac{100+0+0+100}{4} = \frac{200}{4}$ $= 50$
4. **Standard deviation (SD or σ)** is the square root of Variance: $SD = \sqrt{V}$	$SD = \sqrt{V} = \sqrt{0} = 0$	$SD = \sqrt{V} = \sqrt{\frac{50}{4}} = \sqrt{\frac{25*2}{4}}$ $= \frac{5\sqrt{2}}{2}$	$SD = \sqrt{V} = \sqrt{50} = \sqrt{25*2}$ $= 5\sqrt{2}$