Statistics 2

Statistics is the science of dealing with ***data***.

Frequency tables and ***bar graphs*** are a way of organizing and displaying data.

We now turn to numbers that ***summarize*** data.

- ***Average*** (also called, the ***mean***)
- ***Median*** — the "middle"
- ***Quartiles***—dividing the data into quarters

The average and the median are ***measures of central tendency***—the tendency of data to have a ***center*** in some sense.

Average

The ***average*** of a collection of data values is found by computing

$$\textbf{Average} = \frac{\text{Sum of data values}}{\text{Number of data values}}$$

Two other useful numbers:

- ***Minimum*** — the smallest data value.
- ***Maximum*** — the largest data value.

Example

The star forward of UAB s basketball team scored 29, 24, 22, 27, 39, 21, 25, 33, 3, and 18 points in 10 successive games. Find the minimum, maximum and average score.

Points to ponder

- Is the average itself a sensible data value?
- Is the average midway between the minimum and the maximum?
- Is the average a typical game score for this player?

Example — the contender

A high school basketball player that UAB is thinking of recruiting has the following record in 12 games: 20, 3, 22, 7, 30, 28, 4, 9, 26, 8, 28, 25. Find the minimum, maximum, and average scores. Is the average a typical score for this player?

Median

The **median** of a collection of data values is the "middle" value in the list of data values sorted into numerical order.

We say that the median M is at the 50^{th} percentile of the data set. At least 50% is at or below M and at least 50% of the data is at or above M.

Example

The star forward of UAB's basketball team scored 29, 24, 22, 27, 39, 21, 25, 33, and 3 points in 9 successive games. Find his median score.

Example

The star forward of UAB s basketball team scored 29, 24, 22, 27, 39, 21, 25, 33, 3, and 18 points in 10 successive games. Find his median score.

What is different about this example compared to the one before?

Median Odd versus Even

The ***median*** of a collection of data values is the “middle” value in the collection. This works fine if the collection has an **odd** number of data values.

But if the collection has an **even** number of data values, then the median lays halfway between the two middle data values.

Example

The star forward of UAB s basketball team on Prof. Blackbeard’s Stat 101 quizzes scored 7, 8, 5, 9, 10, 8, 7, 10, 10, and 6. What is his median quiz grade?

Suppose an 11th quiz is added to the list with a score of 1 (bad day). What is his median now?

Median Versus Average

Reconsider the quiz grades of the basketball star before the disastrous quiz.

5, 6, 7, 7, 8, 8, 9, 10, 10, 10

Compute the average A: A =

Observe how the disastrous quiz with grade 1 affects the average:

A =

Our player's quiz average has dropped from a ***B*** to a ***C***!

Moral of the story—the median is more *robust* than the average; that is, the median is more resistant to the effect of changes in individual data values.

Average from a Frequency Table

Consider the following generic frequency table:

Data Value	s_1	s_2	. . . .	s_k
Frequency	f_1	f_2	. . . .	f_k

The *average* of the data values is computed as follows:

- **Step 1.** Calculate the total T of the data values:

$$T = (s_1 \times f_1) + (s_2 \times f_2) + ... + (s_k \times f_k)$$

- **Step 2.** Calculate the number N of data values:

$$N = f_1 + f_2 + ... + f_k$$

- **Step 3.** Calculate the average:

$$\text{Average} = \frac{T}{N}$$

Prof. Blackbeard's Stat 101 MidTerm Exam Scores								
Score	1	6	7	8	9	10	11	**Total**
Frequency	1	1	2	6	10	16	13	
Score	12	13	14	15	16	24		
Frequency	9	8	5	2	1	1		**75**

Average:

Algorithm for Finding the Median

Assume we have a collection of N numbers.

- **Step 1.** Sort the N numbers by size.

- **Step 2.** Find the number located in the "middle" position. There are two cases:

  1. N is **odd**. The median is the data value (number) in the position $\frac{N+1}{2}$. ($\frac{N+1}{2}$ is $\frac{N}{2}$ rounded up.)
  2. N is **even**. The median is the number halfway between the numbers in positions $\frac{N}{2}$ and $\frac{N}{2}+1$.

We can see this in the previous basketball examples.

Example: Median from a Frequency Table

Bonus — If the numbers are in a frequency table, then they have already been sorted for us! We have only to determine whether to apply the algorithm for odd or even *N* to find the middle position.

10

Prof. Blackbeard's Stat 101 MidTerm Exam Scores								
Score	1	6	7	8	9	10	11	**Total**
Frequency	1	1	2	6	10	16	13	
Score	12	13	14	15	16	24		
Frequency	9	8	5	2	1	1		**75**

Middle position(s):

Median =

Quartiles

After the median, the next most commonly used set of percentiles are the 25th and 75th. The *first* and *third* quartiles, denoted Q_1 and Q_3, correspond to these percentiles. Thus, 25% of the data in a data set is less than or equal to Q_1, 25% of the data is greater than or equal to Q_3.

Algorithm for finding the Quartiles

Assume we have a *sorted* collection of N numbers.

- **Case 1.** If N/4 is not a whole number, we round up to find the position of Q_1 from the left and Q_3 from the right.

- **Case 2.** If N/4 is a whole number, we average the data values in positions $N/4$ and $N/4 + 1$ from the left to obtain Q_1 and average the data values in these positions coming from the right to find Q_3.

Example Revisited.
The star forward of UAB's basketball team in his first 10 games has the following *sorted* scores:

3 18 21 22 24 25 27 29 33 39

$Q_1 =$ $M = 24.5$ $Q_3 =$

Example. The star forward of UAB's basketball team in his first 12 games makes the following *sorted* scores:

3 18 20 21 22 24 25 26 27 29 33 39

$Q_1 =$ $M =$ $Q_3 =$

Example: Quartiles from a Frequency Table

Prof. Blackbeard's Stat 101 MidTerm Exam Scores								
Score	1	6	7	8	9	10	11	**Total**
Frequency	1	1	2	6	10	16	13	
Score	12	13	14	15	16	24		
Frequency	9	8	5	2	1	1		**75**

Previously found: Min = 1, Max = 24, Median = 11

Quartiles:

N = 75 N/4 = 18.75, not a whole number

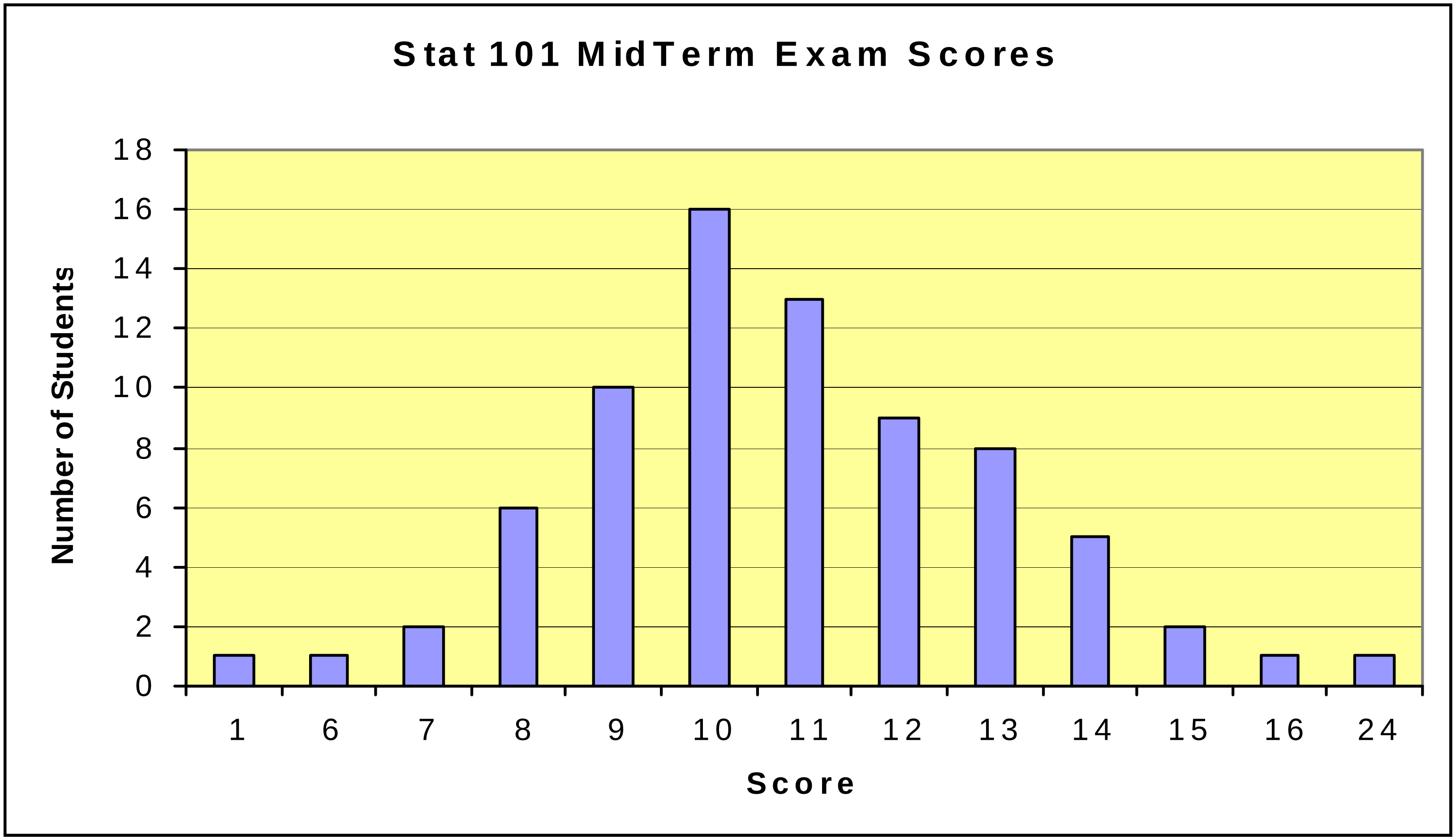
Then Q_1 is found in the 19th position from the left and Q_3 is found in the 19th position counting from the right.

Q_1=

Q_3=

Question. What if we started with a bar graph like the one below? How could we find the average, min, max, median, and quartiles?

Answer. Imagine it as a frequency table (see previous page) and calculate these from the table as previously done.



Computation of Stat 101 “Stats”

Minimum =

1st Quartile =

Median =

3rd Quartile =

Maximum =

Average =

Measures of Spread

Range — the difference between the maximum and minimum data values.

Interquartile Range (IQR) — the difference between the third quartile and first quartile.

In the last example,

Range = Max – Min =

IQR = Q_3 — Q_1 =