Economic Rockstar: How to Calculate the Measures of Central Tendency and Dispersion for a Grouped Frequency Distribution

Statistics for Econ 101

Economic Rockstar

Overview

In this series of lessons, you will be introduced to the **Measures of Central Tendency** and the **Measures of Dispersion**. These calculations, known collectively as *descriptive statistics*, help us to understand the data that we are studying. We can use these summary statistics to describe the data and to make comparisons with other similar data.

Objectives

The following lessons, complete with examples and video, will help you:

- ★ Understand the meaning of the mean, median, mode, standard deviation, interquartile range, symmetry, skewness and normal distribution.
- ★ Calculate the Measures of Central Tendency and the Measures of Dispersion.
- ★ Construct an Ogive and a Histogram.
- ★ Estimate the median, mode and interquartile range using the Ogive and Histogram and draw these diagrams.

Activities

- ★ Complete the lessons below and check out the YouTube videos related to each.
- ★ Subscribe to the Economic Rockstar channel and be notified of any new video releases.

Materials

You'll need a pen, paper and calculator or the use of excel.

Or you could take my word for it!

Other Resources

Economic Rockstar on YouTube.

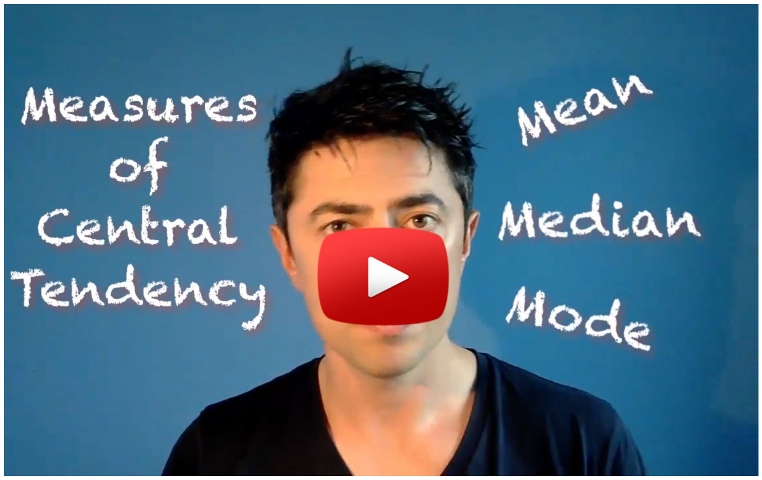
If you like economics, especially the real stuff that's not taught much at Uni, then check out the Economic Rockstar podcast on iTunes or visit the Economic Rockstar website.

About me:

I’m **Frank Conway** and I lecture economics, finance and statistics at 3rd level.

I’m also the host on the Economic Rockstar podcast – a Number 1 ‘New and Noteworthy’ podcast in both the Education and Business categories on iTunes.

I produce **video** content on mathematics, statistics, economics and finance-related topics, which can be found on YouTube.



I provide invaluable economics content that I believe is relevant, topical and real by interviewing economists who are actively engaged in academic research or who say it as it is.

This content can be viewed on www.economicrockstar.com and, for your convenience (if you’re like me), I’ve provided you with a **podcast** so that you can listen to this content on your smart phone, tablet or PC while walking, exercising or even washing the dishes!

Check out the Economic Rockstar podcast on iTunes (for iOS) or on Stitcher Radio (for Android).

Thanks for taking time out to read, watch and listen to the content that I’m providing you.

If you have any suggestions for further content that you would like me to provide you or if you want show your appreciation, then you can comment on the videos on YouTube or email me at frankconway@economicrockstar.com

Thanks,

Frank

Connecting Brilliant Minds In Economics and Finance

Video 010 **How to Calculate the Mean, Median and Mode for a Data Set Using Excel**

1) The Mean:

The mean is also known as the ***average***.

The mean of a set of data is the total of the data measurements (values) divided by the number of data points contained in the data set.

Mean: $\bar{x} = \frac{\Sigma x}{n}$ ***where:*** $\sum$ is the sum of the values in the data set

x represents the values in the data

n is number of values in the data set

Example 1a: **Calculate the mean given the data below:**

Data Set: 5, 7, 9, 10

Here, x are the values that we sum up (totaling 32) and then divide by n = 4.

Answer: **Mean =** $\frac{6 + 7 + 9 + 10}{4} = \frac{32}{4} =$ **8**

Example 1b: **Calculate the mean given the data below:**

Data Set: 10, 6, 5, 9, 8, 6, 4, 5, 7, 6

Answer: **Mean =** $\frac{10 + 6 + 5 + 9 + 8 + 6 + 4 + 5 + 7 + 6}{10} = \frac{66}{10} =$ **6.6**

2) The Median:

The median of any data set is the ***middle*** value when the measurements are arranged in ascending (or descending) order. It is this figure that divides the data into two equal parts.

Median: $\frac{n+1}{2}$ ***where:*** n is number of values in the data set.

1 is used to help find the median for an *even* or an *odd* data set.

2 halves the top answer to find the middle number.

Steps: 1) Rearrange the above numbers in ascending (or descending) order.

2) What is the middle number – the Median?

3) If the number of values (n) in the data set is an odd number, say 11, then using the formula above 11+1 =12/2 = 6. **The median is the 6^{th} number.**

4) If the number of values (n) in the data set is an even number, say 12, then using the formula above 12+1 =13/2 = 6½. **The median is the $6½^{th}$ number.** But since there is no such number, we take the average of the values of the 6^{th} and 7^{th} numbers. We add them together and divide by 2, just like the mean formula above. *(NB: Note that in this case the 2 comes from the number of values in the calculation! n = 2).*

Example 2a: **Calculate the median given the data below:**

Data Set: 23, 14, 7, 50, 8, 33, 19

Answer: When counting the number of values in the data set above we find that there is an ***odd*** number, i.e. 7 values in the data set.

Median = 7, 8, 14, 19, 23, 33, 50 ***Re-arrange***

$\frac{n+1}{2} \quad \frac{7+1}{2} = \frac{8}{2} = \mathbf{4^{th}}$ ***Middle number***

Median = 7, 8, 14, **19**, 23, 33, 50

Median = **19**

Example 2b: **Calculate the median given the data below:**

Data: 10, 6, 5, 9, 8, 6, 4, 5, 7, 6

Answer: When counting the number of values in the data set above we find that there is an ***even*** number, i.e. 10 values in the data set.

Median = 4, 5, 5, 6, 6, 6, 7, 8, 9, 10 — ***Re-arrange***

$\frac{n+1}{2} \quad \frac{10+1}{2} = \frac{11}{2} = 5\frac{1}{2}^{th}$ — ***Middle number***

Median = 4, 5, 5, 6, **6**, **6**, 7, 8, 9, 10 — ***Since a 5½th number doesn't exist, we choose the 5th and 6th number.***

Mean = $\frac{\Sigma x}{n} = \frac{6+6}{2} = 6$ — ***Even numbered data set***

Median = 6 — ***The average of the 5th and 6th number.***

Example 2c: **Find the median**

11, 3, 9, 7, 1, 5

Answer: ***Try it!***

3) The Mode:

The mode identifies the most popular item, i.e. the one that appears most frequently.

Example 3a: **Calculate the mode given the data below:**

Data: 3, 1, 4, 1, 2, 3, 5, 6, 1, 1

Answer: 3, **1**, 4, **1**, 2, 3, 5, 6, **1**, **1** **= 1**

Example 3b: **Find the mode given the data below:**

Data Set: 10, 6, 5, 9, 8, 6, 4, 5, 7, 6

Answer: ***Try it! (Check the YouTube Video 010 if you'd like to know the answer)***

Let's take a look at the following example, which I go through on a video in YouTube:

	Car Speeds (kph)
1	35
2	46
3	50
4	55
5	45
6	36
7	55
8	44
9	46
10	52
11	37
12	28
13	48
14	46
15	54
16	38
17	32
18	51
19	47
20	40
Total:	**885**
n	**20**

Arrange in Ascending Order
28
32
35
36
37
38
40
44
45
46
46
46
47
48
50
51
52
54
55
55

Mean: $\bar{x} = \frac{\Sigma x}{n} = \frac{885}{20} = 44.25$

Mode: By observation, we can see that the number 46 occurs most frequently (3 times). Therefore, the mode = 46

Median: $\frac{n+1}{2} = \frac{20+1}{2} = \frac{21}{2} = 10.5$

Choose the 10^{th} and 11^{th} numbers from the list above and find the average: $\bar{x} = \frac{\Sigma x}{n} = \frac{46}{2} = 46$

Video 011 Explaining the Standard Deviation Using the NFL Football Field

- ★ The **standard deviation** is a measure of the average deviation from the mean value.
- ★ It is the most common measure of ***dispersion***. (Note: dispersion = spread/variability).
- ★ It is used as a measure for comparing two similar types of data.

Let's take a look at the following football field:

There are 11 players on each team. Take a look at the position of both teams on the field. The blue team (Offense) is lined out differently to the orange team (Defense).

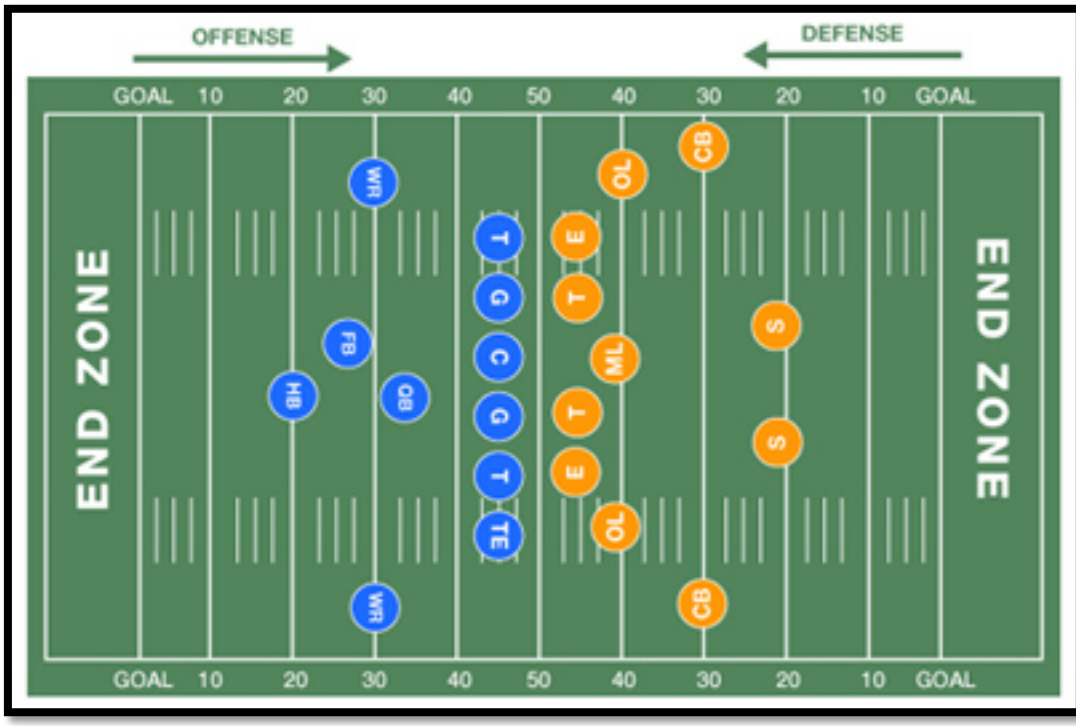


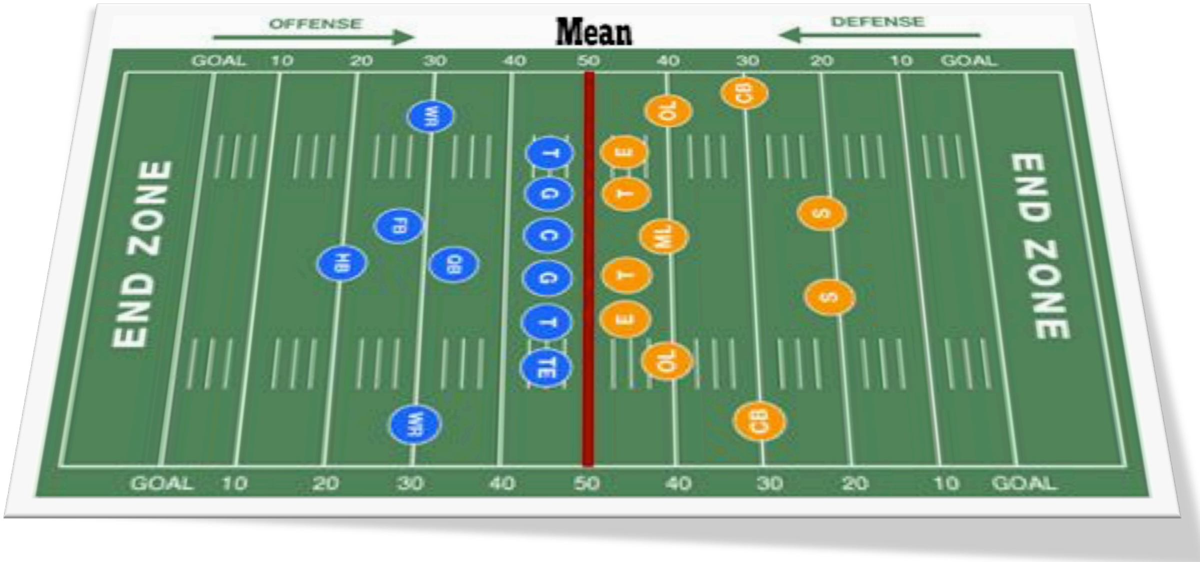
Image Source: http://shesasportsfan.com/tag/nose-guard/

- ★ The lines on the field are marked in 10-yard increments, representing the distance to travel to get from the center to the goal line.

- ★ From the center, it will take 50 yards to get to each goal line. Both teams therefore have the same distance to travel to reach the End Zone, making this a fair and equal game.

- ★ We call this type of layout a ***symmetrical distribution*** in maths! That is, if you're given a shape and find it's center, you can fold the shape along the center and both sides should be identical (if they're not, then the shape is non-symmetrical).

- ★ Okay! So **what is the standard deviation?** Firstly, let's find the mean. We'll assume that it's the center line in the field, i.e. 50. This is colored in red below.



- The distance between each long white line represents an area on the field, i.e. to go from the 40-yard line to the 50-yard line, the distance travelled is 10 yards.

- Since the field is ***symmetrical***, the 10-yard distance is identical for both the Offense and the Defense.

- At the 30-yard line, there is a distance of 20 yards to the mean. We could also say that there are ***'two10-yards'*** to get to the mean.

- We can continue this until we get to the End Zone. Therefore, there are ***'five 10-yards'*** to get to the center line.

- Since the distance between each line (marked by the numbers 10, 20, 30, 40 and 50) are all 10, then we can say that the **standard distance** between each line is **10 yards**.

Almost there!

- Let's use some data so that we can put the standard deviation into some mathematical perspective.

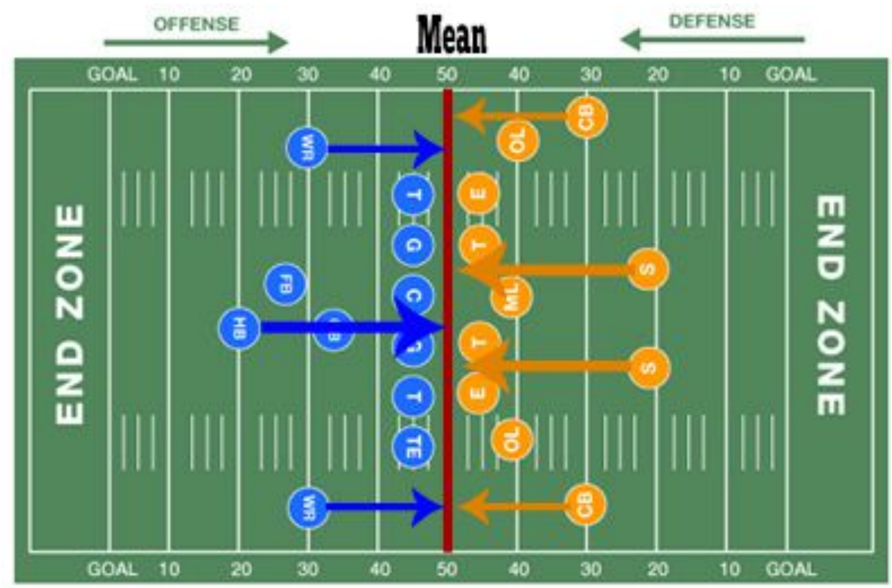
- But since we're talking about the NFL field, let's use football players as the data. The diagram has points representing the position of the players, which of course looks very similar to data points on any graph.

- To find the value of the **standard deviation**, we must find the deviation or distance of each point from the mean value.

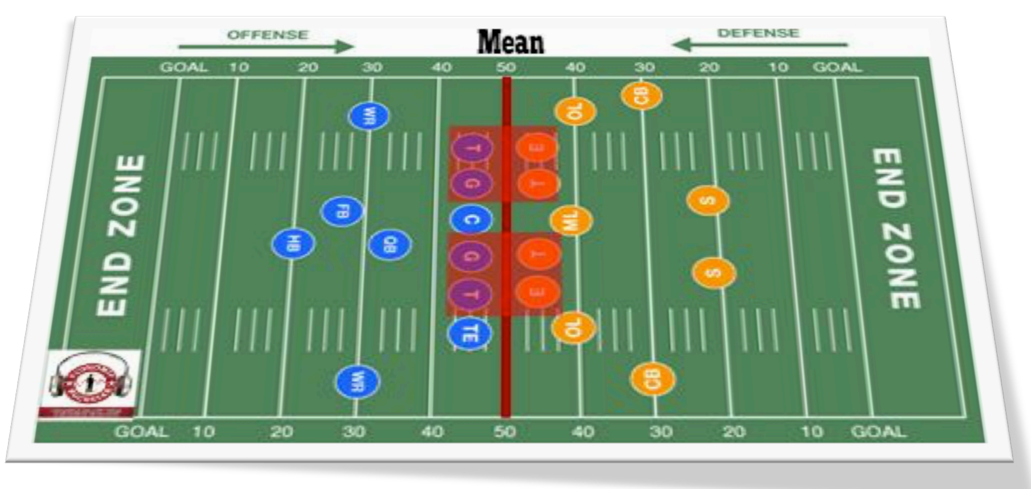
★ We can do this by observation. Firstly, there are no data points within the 10 and 20-yard areas. But, the Blue team has a player (data point) on the 20-yard line. This player is 30 yards from the center line or, using the mathematical expression, has a ***deviation*** of 30 yards from the mean of 50 (long blue arrow in image below).

★ The Orange team has two players almost on its 20-yard line, each with a deviation of say 28 yards from the center or mean (long orange arrows).

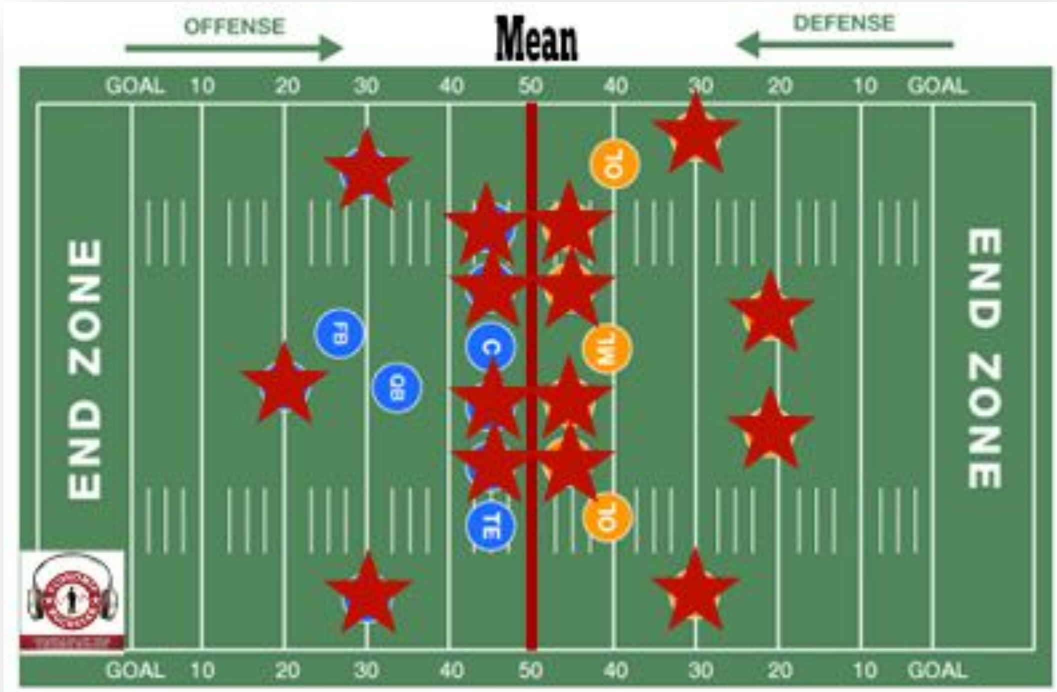
★ The Blue team has two players on the 30-yard line (a deviation of 20 yards from the mean), while the Orange team also has two players on its 30-yard line.



★ Other positions that are identical on both teams are highlighted below, giving both teams identical deviations from the mean (a deviation of 5 yards for each player from the mean):



Let's recap for the moment and identify what data points we've covered so far and the deviations of each point to the center:



- The Blue team have a total deviation of 90 yards [(30x1) + (20x2) + (5x4)] while the Orange team of a total deviation of 116 yards [(28x2) + (20x2) + (5x4)].

	Deviation from the Mean			
	Approx 30 yards (from mean)	20 yards (from mean)	5 yards (from mean)	Total Deviation
Blue team	1	2	4	90 yards
Orange team	2	2	4	116 yards

- At the moment there is a 26-yard difference between the two total deviations (116 – 90). But, we still have more data points to work on!
- The Blue team still has 4 players (points) on the field and the Orange team have 3 players.
- Can you tell what the deviations are for each point from the mean? Try it.
- The Orange team have three players almost on the 40 yard line, which is a deviation of almost 10 yards each to the mean. This gives a total deviation of, say, 26 yards (10 + 8 + 8).
- The Blue team are in different positions on the field. Two players are 5 yards from the line (a total of 10 yards), another player is 18 yards from the center (mean) and the final player is 24 yards from the center. This gives a total deviation of 52 yards from the center for these 4 players.

- ★ So, overall **the Blue team has a total deviation of 142 yards** (90 yards + 52 yards) and **the Orange team have a total deviation of 142 yards** (116 yards + 26 yards)!

Coincidence? Not when the data is distributed in a symmetrical fashion (i.e. a symmetrical distribution – another mathematical expression!)

So let's put all of the data points (distance of each player/point to the mean) in a table:

Blue Team		Orange Team	
Player	**Player**	**Player**	**Deviation**
1	30	1	28
2	20	2	28
3	20	3	20
4	5	4	20
5	5	5	5
6	5	6	5
7	5	7	5
8	5	8	5
9	5	9	10
10	18	10	8
11	24	11	8
Total	**142**	**Total**	**142**

Okay, I know there's a lot of numbers but I hope you're still with me.

- ★ Since, the Blue team are on the Offense, let's give their deviations a poistive sign (=) and since the Orange team are on the Defense, we'll give their deviations a negative sign (-).

Blue Team		Orange Team	
Player	**Player**	**Player**	**Deviation**
1	30	1	-28
2	20	2	-28
3	20	3	-20
4	5	4	-20
5	5	5	-5
6	5	6	-5
7	5	7	-5
8	5	8	-5
9	5	9	-10
10	18	10	-8
11	24	11	-8
Total	**142**	**Total**	**-142**

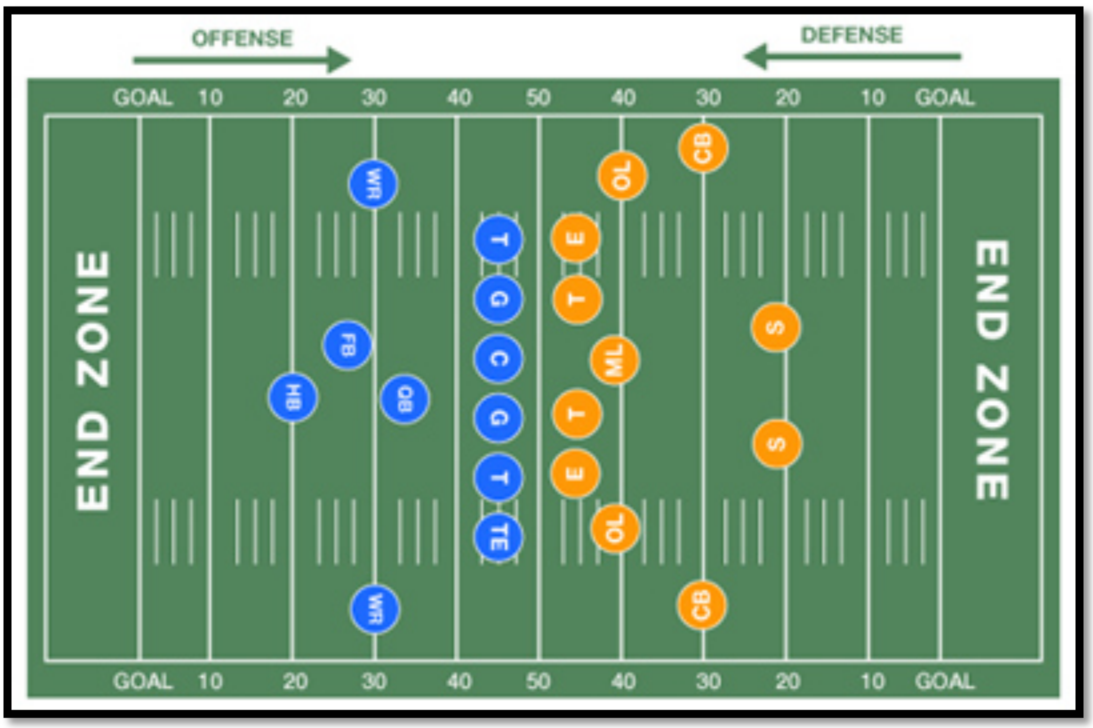
- ★ Adding the total deviations for both sides gives us a value of zero. Thus, the sum of the deviations from the mean when the data is distributed symmetrically (or normally distributed) is zero.

★ Now, we can find the standard devitaion of this data.

★ **But firstly, let's keep it simple!** We could make the assumption that, since the football field is split up into 5 equal sections of 10 yards in each half (not including the End Zone), then we could say that the standard deviation is 10 yards. Therefore, 1 standard deviation is 10 (yards), 2 standard deviations is 20 (yards), etc.

★ A standard deviation gives us a fixed (standard) figure that allows us to categorize the location of data points from the mean. So, there are 13 points (players) within 1 standard deviation of the mean. **(Note: 1 standard deviation refers to both sides of the mean!)**

★ There are 18 points (players) within 2 standard deviations of the mean and 22 points (players) within 3 standard deviations of the mean.



Okay, got that?

Back to the table so and we'll find the actual standard deviation for the data.

Steps: **1)** Square the deviations (to remove any negative numbers).

2) Sum up the squared deviations and divide by the number of data points.

3) Find the square root of your answer in step 2).

Blue Team		
Player	Deviation	Deviation^2
1	30	900
2	20	400
3	20	400
4	5	25
5	5	25
6	5	25
7	5	25
8	5	25
9	5	25
10	18	324
11	24	576
Orange Team		
Player	Deviation	Deviation^2
1	-28	784
2	-28	784
3	-20	400
4	-20	400
5	-5	25
6	-5	25
7	-5	25
8	-5	25
9	-10	100
10	-8	64
11	-8	64
22	Total	5446

- ★ Standard Deviation is the square root of the sum of the deviations divided by the total number of data points:

5446/22 = 247.54 and the square root of 247.54 is **15.73**.

- ★ Therefore, the true **Standard Deviation is 15.73 yards** as opposed to our assumed 10 yards.

Video 012:

How to Calculate the Mean and Standard Deviation for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Calculate the *mean* expenditure on food. **(8 marks)**

(b) Calculate the *standard deviation* expenditure on food. **(12 marks)**

(c) Calculate the *coefficient of variation*. **(4 marks)**

(d) *Interpret* the statistics found in parts *(a)*, *(b)* and *(c)* above. **(2 marks)**

(Total: 26 marks)

Formula:

★ **Arithmetic Mean:** $\bar{x} = \frac{\Sigma fx}{\Sigma f}$

★ **Standard Deviation:** $\sigma = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$

★ **Coefficient of Variation:** $cv = \frac{\sigma}{\bar{x}}$

Video 013

How to Calculate the Median for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Calculate the *median* expenditure on food using the appropriate formula. **(10 marks)**

(b) *Interpret* the statistics found in part *(a)* above. **(2 marks)**

(Total: 12 marks)

Formula:

★ **Median** : $L_m + \left[\frac{(\frac{N}{2} - F_{m-1})}{f_m}\right] c_m$

Video 014

How to Calculate the Mode for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Calculate the *mode* expenditure on food using the appropriate formula. **(10 marks)**

(b) *Interpret* the statistics found in part *(a)* above. **(2 marks)**

(Total: 12 marks)

Formula:

★ **Mode**: $L + \left[\frac{D_1}{D_1 + D_2} \right] xC$

Video 015

How to Calculate the Interquartile Range for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Calculate the *lower quartile* expenditure on food using the appropriate formula.

(8 marks)

(b) Calculate the *upper quartile* expenditure on food using the appropriate formula.

(8 marks)

(c) Calculate the *interquartile range* expenditure on food using the appropriate formula.

(4 marks)

(d) *Interpret* the statistics found above. **(5 marks)**

(Total: 25 marks)

Formula:

★ **Quartile:** $Q_i = L_Q + \left[\frac{(P_Q - F_{Q-1})}{f_Q}\right] c_Q$

★ **Interquartile Range:** $Q_3 - Q_1$

Video 016

How to Calculate Pearson's Measure of Skew and Represent the Measures of Central Tendency on a Normal Distribution Curve

(a) *Write out* the Mean, Median and Mode in the table below:

★ **Mean**	
★ **Median**	
★ **Mode**	

(b) Calculate *Pearson's Measure of Skew* and *interpret* your answer. **(4 marks)**

(c) Draw a *Normal Distribution Curve* to represent the Measures of Central Tendency.

(8 marks)

(b) *Interpret* the statistic and diagram found in part *(b)* and *(c)* above. **(6 marks)**

(Total:18 marks)

Formula

★ **Pearson's Skew:** $\frac{\text{Mean} - \text{Mode}}{\sigma}$

or

$$\frac{3(\text{Mean} - \text{Mode})}{\sigma}$$

Video 017

How to Estimate the Median and Interquartile Range with an Ogive for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Graphically estimate the *median* expenditure on food using an ogive. **(8 marks)**

(b) Graphically estimate the *interquartile range* expenditure on food using an ogive. **(8 marks)**

Video 018

How to Estimate the Mode with a Histogram for a Grouped Frequency Distribution Using Excel

The following frequency distribution table shows the amount spent on food in one particular shop in by customers:

Expenditure on Food (€)	No. of Customers
less than 10	4
10 and less than 20	8
20 and less than 30	20
30 and less than 40	60
40 and less than 50	30
50 and less than 60	6

(a) Graphically estimate the *modal expenditure on food* using a histogram.

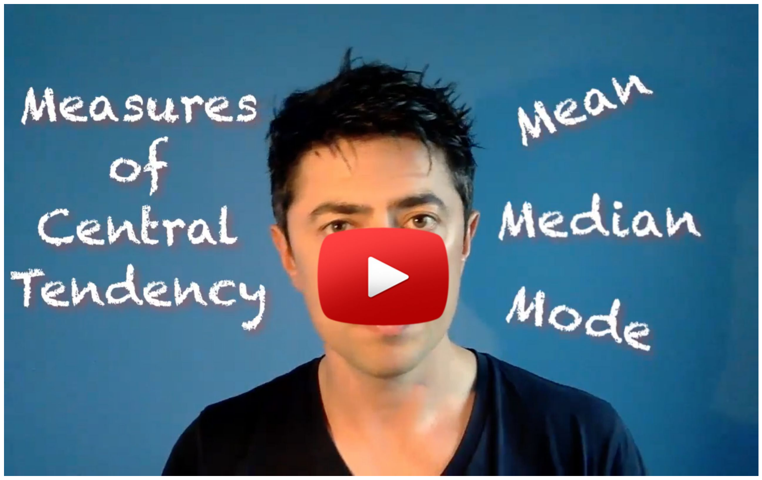
(12 marks)

About me:

I'm **Frank Conway** and I lecture economics, finance and statistics at 3rd level.

I'm also the host on the Economic Rockstar podcast – a Number 1 'New and Noteworthy' podcast in both the Education and Business categories on iTunes.

I produce **video** content on mathematics, statistics, economics and finance-related topics, which can be found on YouTube.



I provide invaluable economics content that I believe is relevant, topical and real by interviewing economists who are actively engaged in academic research or who say it as it is.

This content can be viewed on www.economicrockstar.com and, for your convenience (if you're like me), I've provided you with a **podcast** so that you can listen to this content on your smart phone, tablet or PC while walking, exercising or even washing the dishes!

Check out the Economic Rockstar podcast on iTunes (for iOS) or on Stitcher Radio (for Android).

Thanks for taking time out to read, watch and listen to the content that I'm providing you.

If you have any suggestions for further content that you would like me to provide you or if you want show your appreciation, then you can comment on the videos on YouTube or email me at frankconway@economicrockstar.com

Thanks,

Frank

Connecting Brilliant Minds In Economics and Finance

CONNECTING BRILLIANT MINDS IN ECONOMICS AND FINANCE