SUBJECT and GRADE	Mathematics Grade 12	
TERM 3	Week 3	
TOPIC	Statistics	
AIMS OF LESSON	• Determine the trend of a scatterplot. • How to find the equation of the regression line using your calculator. • Drawing the regression line. • Determine the correlation coefficient between two variables.	
RESOURCES	***Paper based resources***	***Digital resources***
	Please go to your Statistics chapter in your textbook.	Where you see this icon in the lesson you can click on it to see a video on how to use your calculator. Where you see this icon in the lesson you can click on it to see a video on how to use your calculator.

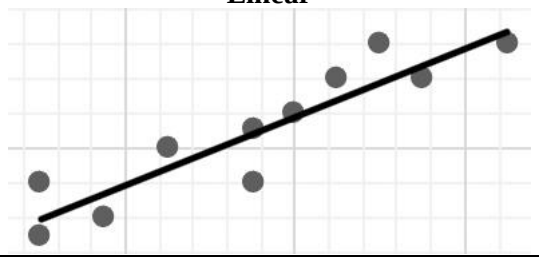
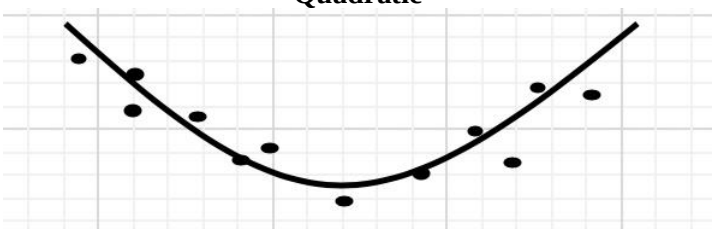
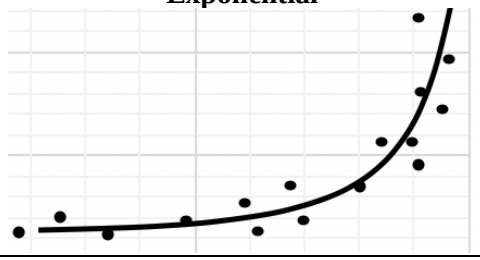
INTRODUCTION

Dear learner in previous grades you have learnt how to find the equation of a linear function. The standard form of a linear function is $y = mx + c$. In order to find the equation one would need to find the values of m and c. It will be useful to practice how to determine the equation of a linear function. In this week's lesson we will focus on finding the equation of the regression line, drawing a regression line and determining the correlation coefficient.

CONCEPTS AND SKILLS

Scatterplots

A scatterplot is a graph used to determine whether there is a relationship between paired data. A scatter plot diagram is a powerful tool for researchers to determine if there is any association between two variables. The data on the scatterplot could follow the following trends: linear, quadratic or exponential.

Linear	Quadratic	Exponential
		

Regression line

The regression line is basically an accurate line of best fit. It uses the least squares method to calculate the gradient and y intercept of the regression line.

Standard form of regression line: $y = A + Bx$

How to determine the equation of the regression line with a Casio(fx-82 ZA) calculator:

1. Press MODE and select STAT.
2. Select $A + Bx$.
3. Enter the data points.
Column(X) type = after each data point
Column(Y) type = after each data point
4. After entering all the data points press AC.
5. Press SHIFT then press 1.
6. Press 5: Reg
7. Press 1 then =: to find A
8. Press SHIFT then press 1.
9. Press 5: Reg
10. Press 2 then =: to find B.

https://youtu.be/4YdKyRVmISk

How to determine the regression line with a SHARP (EL-W535HT) calculator

1.< $MODE$ >[1:STAT][1:LINE]
2. Enter x-values and the y-values together.
3. Press CHANGE after each (x; y)
4. Find the value for a , the y-intercept [ALPHA][( ] [=]
5. Find the value for b , the gradient [ALPHA] [ )] [=]

Line of best fit

Refers to a line through a scatter plot of data points that best represents the trend of the data. The line of best fit is not that accurate, but it helps us to see if there is a trend in the data set.

How to draw a line of best fit:

- Try to have the same amount of data points above and below the line.

Exclude any outliers

How to draw the regression line:

If your scale stars with the origin of (0;0)

- Plot the A value of the regression line i.e the y-intercept.
- Determine the mean point($\overline{x}$; $\overline{y}$) and plot the mean point.
- Draw a line from point A through the mean point.

How to draw a regression line if the scale of the graph do not start with (0; 0)

In order to draw the regression line, substitute any two x-values that lie between the minimum and maximum x- values into the equation of the regression line, plot the two points and then join them up

How to determine the mean point using a calculator:

- Make sure your data is entered in your calculator.
- Press SHIFT then Press 1.
- Press 4: Var
- Press 2 then =: $\overline{x}$
- Press SHIFT then press 1.

Press 4: Var Press 5 then = : $\overline{y}$

Outliers

A value that "lies outside" (is much smaller or larger than) most of the other values in a set of data. Do not add an outlier to the dataset when drawing a line of best fit or calculating the regression line. The reason for this is that the outlier is not part of the trend and will influence the trend lines and hence any future predictions from these trend lines.

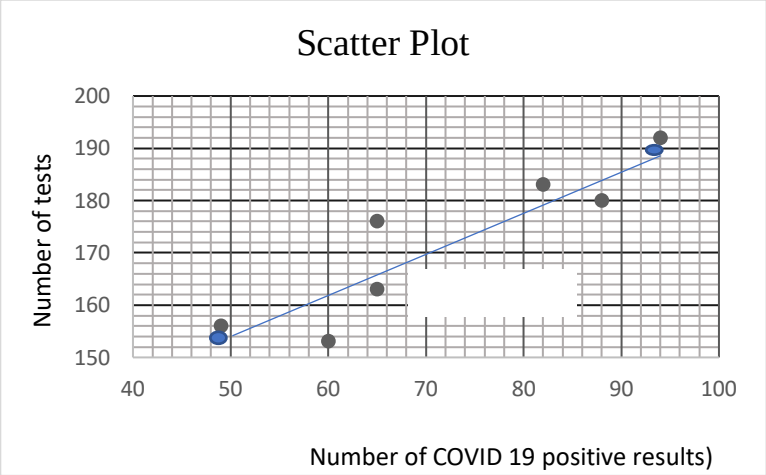
Example 1: At a clinic where Covid19 tests were conducted the following data was recorded for 7 days in June

https://video.tutonic.org/T12leastsquare

Number of COVID 19 positive results	49	65	82	60	65	94	88
Number of tests	156	176	183	153	163	192	180

1. Draw a scatterplot of the given data.
2. Determine the equation of the least squares regression line.
3. Draw the least squares regression line.
4. How many tests needs to be done to record approximately 100 COVID 19 positive results per day.

Solution:



2. $y = A + Bx$

$y = 113{,}47 + 0{,}81x$

3. $y = 113{,}47 + 0{,}81(\mathbf{49}) = \mathbf{153.16}$ $\therefore$ (49; 153.16)

$y = 113{,}47 + 0{,}81(\mathbf{94}) = \mathbf{189{,}61}.$ $\therefore$ (94; 189,61)

4. $y = 113{,}47 + 0{,}81(100)$ $y \approx 195$ tests

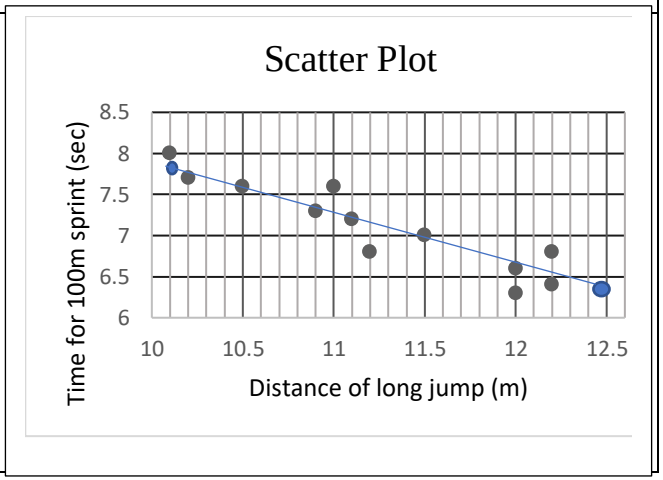
CAN YOU?

The table below shows the time taken by 12 athletes to run 100m sprint and their best distance for long jump

Time for 100m sprint (sec)	10,1	10,2	10,5	10,9	11	11,1	11,2	11,5	12	12	12,2	12,5
Distance of long jump (m)	8	7,7	7,6	7,3	7,6	7,2	6,8	7	6,6	6,3	6,8	6,4

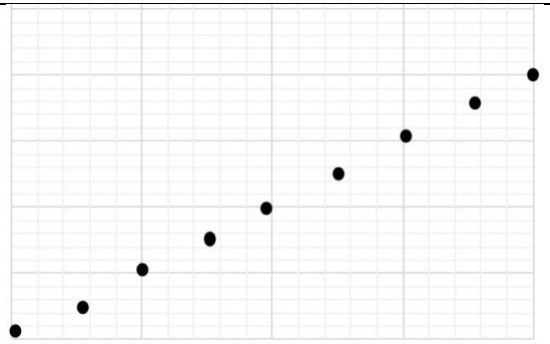
1. Represent the data as a scatter plot.
2. Use the data to calculate the equation of the least squares regression line.
3. Draw the least squares regression line on your scatter plot.
4. An athlete runs the 100m in 11,7 seconds, use the formula to predict the distance of this athlete's jump.

2. $y = 14{,}34 - 0{,}64x$
3. (10,1; 7,8)
   (12,5; 6,3)
4. y=6,852m

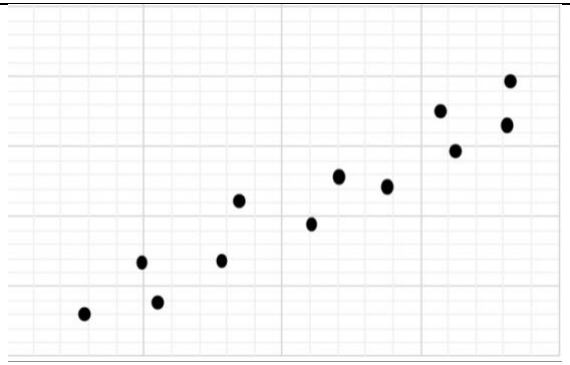


Correlation

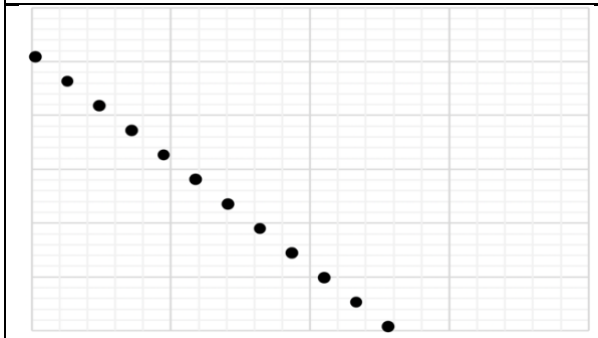
The correlation coefficient is a statistical measure of the strength of the relationship between two variables. Correlation coefficient is denoted by (r) and is between -1 and 1. The closer the data points are to the regression line the stronger the relationship. That means the closer r is to 1 or -1. The further the data points are away from the regression line, the weaker the relationship, and the closer r is to 0. If the gradient of the regression line is positive, then the data set has a positive correlation and if the gradient of the regression line is negative, then the data has a negative correlation. If the correlation coefficient is greater than 0,9 we say there is a very strong positive correlation. If the correlation coefficient is smaller than -0,9 we say there is a very strong negative correlation. The strength of the association is determined by the correlation coefficient **(**r).



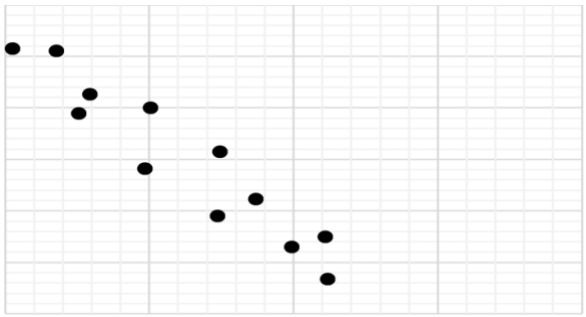
Perfect positive linear association.



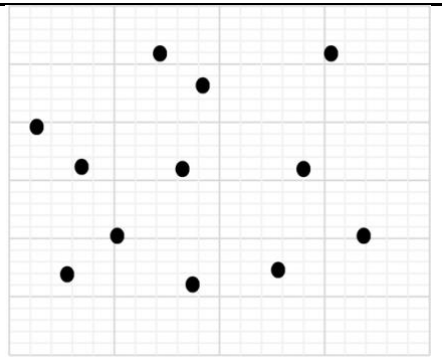
Very Strong positive linear association.



Perfect negative association.



Very Strong negative association.



No correlation.

Example 2:

The data in the table represents the results of the trial examination and the final examination mark of 11 Gr 12 learners

Trial exam	80	68	94	72	74	83	56	78	65	75	88
Final exam	78	71	96	77	77	80	58	83	71	80	92

- Use the given table to find the correlation coefficient r and describe the relationship of the trial exam and the final examination.
  Comment on the strength of the correlation of the two variables.
  Solution: $= 0{,}96$ $\therefore$ strong positive linear correlation between variables..

How to calculate the correlation coefficient using your calculator:

- Make sure your data is entered correctly in your calculator. See page 2.
- Press shift then press 1.
- Press the number next to "Reg"
- Press the number next to r then press =.

https://youtu.be/n7NAHB8_rSQ

CAN YOU?

A group of 12 learners have been asked to measure their resting heart rate (beats per minute) and the time (in minutes) that they exercise in a week. The data below was gathered.

Minutes of exercise per week	30	40	60	90	140	180	270	350	360	420	440	500
Resting heart rate (BPM)	82	77	75	70	68	67	60	58	52	50	48	45

- Calculate the correlation coefficient for the given data and describe the relationship between minutes of exercise per week and resting heart rate.

Answers : -0,984987 Strong negative linear correlation between variables

Consolidation

- Remember to exclude outlier.
- Regression lines is used to show the general trend which a set of data follows
- The correlation coefficient tells us about the strength of the relationship between the variables (always between -1 and 1)
- Practice by working out old question papers to get acquainted with the way question are asked in exams.

ACTIVITY	**Mind Action Series**	**Via Africa**	**Siyavula**	Classroom mathematics	Platinum
	Ex 1 & Ex 2 pg. 295 – 303	Mixed Exercise (nr. 3 – 5) pg. 106	End of chapter exercise pg. 395 - 399	Ex 12.5 pg 324 Ex 12.4 pg 318	Ex 2 pg 249 - 252