USING A CALCULATOR TO INVESTIGATE WHETHER A LINEAR, QUADRATIC OR EXPONENTIAL FUNCTION BEST FITS A SET OF BIVARIATE NUMERICAL DATA

Jackie Scheiber
RADMASTE, Wits University
Jackie.scheiber@wits.ac.za

Target audience: FET Band Mathematics teachers
Duration: 2-hour workshop

CURRICULUM REFERENCES

1) GRADE 9 (5.2 Represent data)

Draw a variety of graphs by hand/technology to display and interpret data including:

- bar graphs and double bar graphs
- histograms with given and own intervals
- pie charts
- broken-line graphs
- scatter plots

2) GRADE 12 (10. Statistics)

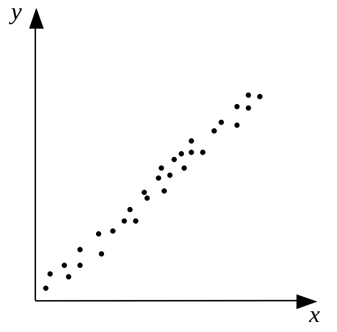
a) Represent bivariate numerical data as a scatter plot and suggest intuitively and by simple investigation whether a linear, quadratic or exponential function would best fit the data.
b) Use a calculator to calculate the linear regression line which best fits a given set of bivariate numerical data.
c) Use a calculator to calculate the correlation co-efficient of a set of bivariate numerical data and make relevant deductions.

Reference:

Pike, M. Classroom Mathematics Grade 12 (2013) Heinemann Publishers

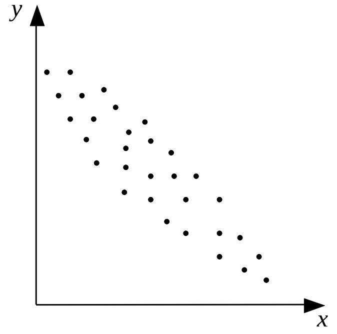
BIVARIATE DATA

- When each item in a population has **two** measurements associated with it (e.g. the heights and the masses of the members of a rugby team), then the data can be represented by points on a two-dimensional **scatter plot** or **scatter diagram** or **scatter graph** or **scatter chart**. The scatter plot can show whether there is an **association** (or correlation) between the two variables.



Positive linear correlation

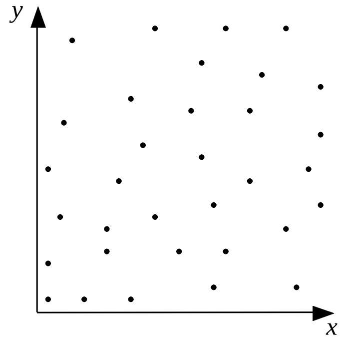
- If the points are roughly in a line, then we can describe the correlation as being **linear**. A linear correlation can be described as **positive** if, as one quantity increases, so does the other **OR negative** if, as one quantity increases, the other decreases.



Negative linear correlation

- When the points in a scatter plot have a significant or very strong negative or strong positive correlation, we can draw a line of best fit, also called a least squares regression line or just a regression line. The line of best fit is a line about which the points appear to be clustered.

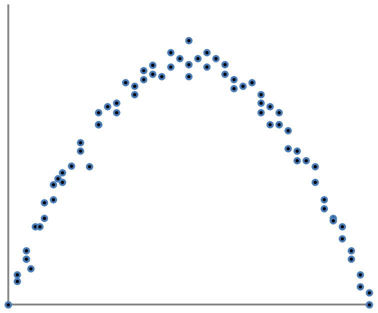
- Calculators and statistical software can be used to find the equation of the line of best fit or regression line. The bivariate data items are used as input. The calculator or computer then given the values of the unknowns in the equation.



No correlation

- As can be seen alongside, in some cases a curve will fit a set of bivariate data better than a line can. We can use a calculator to calculate the equation of an appropriate non-linear function that models the data. At school level the two non-linear functions that we use are the quadratic and exponential functions.

- One of the difficulties of using a curve to model the data is that the curve may fit the data within the given range, but may otherwise be inappropriate. It is always best to examine the scatter plot carefully and to then fit a curve that looks close to the data points and also seems to explain the relationship in the context of the data.



Non-linear correlation

TASK 1

The table lists the world records for the 100 m men's sprint from 1983 to 2009.

Athlete	Nationality	Year	Time (s)
1. Willie Williams	USA	1956	10,1
2. Armin Hary	West Germany	1960	10,0
3. Jim Hines	USA	1968	9,95
4. Calvin Smith	USA	1983	9,93
5. Carl Lewis	USA	1988	9,92
6. Leroy Burrell	USA	1991	9,90
7. Carl Lewis	USA	1991	9,86
8. Leroy Burrell	USA	1994	9,85
9. Donovan Bailey	Canada	1996	9,84
10. Maurice Greene	USA	1999	9,79
11. Tim Montgomery	USA	2002	9,78
12. Asafa Powell	Jamaica	2005	9,77
13. Asafa Powell	Jamaica	2007	9,74
14. Usain Bolt	Jamaica	2008	9,72
15. Usain Bolt	Jamaica	2008	9,69
16. Usain Bolt	Jamaica	2009	9,58

(Source: http://www.rss.org.uk/uploadedfiles/documentlibrary/803.pdf)

Asafa Powell powers to the 9,77 second World record in Athens in 2005 (AFP/Getty Images)

1) Use the data to draw a scatter plot on the given set of axes.

2) Describe the correlation between the two variables in words. Is there a direct cause and effect relationship between the variables or is it possible that the relationship between the two variables may be a coincidence?

3)
  a) Use your calculator to determine the *linear regression function (y = A + Bx)* that best models the data.

  b) Use the calculator and the equation of the linear regression line to complete the following table:

Year	1960	1968	1988	1999	2005	2008
Time *(s)*						

  c) Use the values in the table to draw the linear regression function on the scatter plot.

4)

a) Use your calculator to determine the *quadratic regression function* $(y = A + Bx + Cx^2)$ that best models the data.

b) Use the calculator and the equation of the quadratic regression function to complete the following table:

Year	1960	1968	1988	1999	2005	2008
Time *(s)*						

c) Use the values in the table to draw the quadratic regression function on the scatter plot.

5)

a) Use your calculator to determine the *exponential regression function* $(y = A.B^x)$ that best models the data.

b) Use the calculator and the equation of the exponential regression function to complete the following table:

Year	1960	1968	1988	1999	2005	2008
Time *(s)*						

c) Use the values in the table to draw the exponential regression function on the scatter plot

6) Complete the following table by filling in the sets of values that you have calculated on it.

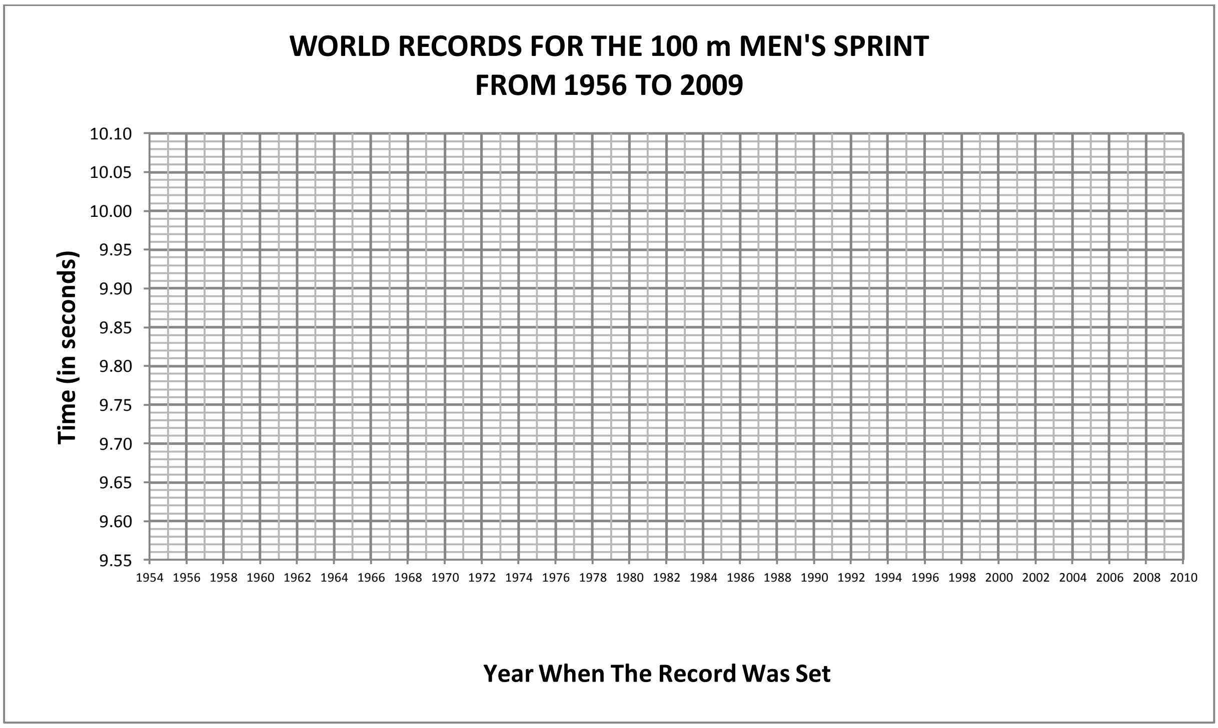
Year	**Actual time (s)**	**Value of linear regression function *(s)***	**Value of quadratic regression function *(s)***	**Value of exponential regression function *(s)***
1960	10,0			
1968	9,95			
1988	9,92			
1999	9,79			
2005	9,77			
2008	9,69			

7) Which regression function suits the context best? Give reasons for your answer.

SOLUTION TO TASK 1

1)

	Athlete	Nationality	Year	Time (s)
1.	Willie Williams	USA	1956	10,1
2.	Armin Hary	West Germany	1960	10,0
3.	Jim Hines	USA	1968	9,95
4.	Calvin Smith	USA	1983	9,93
5.	Carl Lewis	USA	1988	9,92
6.	Leroy Burrell	USA	1991	9,90
7.	Carl Lewis	USA	1991	9,86
8.	Leroy Burrell	USA	1994	9,85
9.	Donovan Bailey	Canada	1996	9,84
10.	Maurice Greene	USA	1999	9,79
11.	Tim Montgomery	USA	2002	9,78
12.	Asafa Powell	Jamaica	2005	9,77
13.	Asafa Powell	Jamaica	2007	9,74
14.	Usain Bolt	Jamaica	2008	9,72
15.	Usain Bolt	Jamaica	2008	9,69
16.	Usain Bolt	Jamaica	2009	9,58



2) Describe the correlation that exists between the two variables:

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

3)

a) Calculate the *linear regression function (y = A + Bx)* as follows:

PROCEDURE	KEY SEQUENCE
Go to the Stats Mode	[MODE] [2:STAT] [2: A + BX]
Enter all the X-values	1956 [=] 1960 [=] 1968 [=] 1983 [=] …
Go to the top of the Y-value column	[▼] [▶]
Enter all the Y-values	10.1 [=] 10.0 [=] 9.95 [=] 9.93 [=] …
Clear the screen	Press [AC]
Go to the Stats Calculation Screen and then find the values of A and B	[SHIFT] [1] (STAT) [5: Reg] [1: A] [=] [SHIFT] [1] (STAT) [5: Reg] [2: B] [=]

A = …………………………………….. (correct to 2 decimal places

B = …………………………………….. (correct to 3 decimal places)

Equation of the linear regression function: ……………………………………………

b) Use your calculator to determine the time (correct to 2 decimal places) for the given years as follows:

PROCEDURE	KEY SEQUENCE
Determine the time in 1960	1960 [SHIFT] [1] (STAT) [5: Reg] [5: $\hat{y}$]
Determine the time in 1968	1968 [SHIFT] [1] (STAT) [5: Reg] [5: $\hat{y}$],
Continue like this to determine the rest of the values	

Year	1960	1968	1988	1999	2005	2008
Time (to 2 decimal places)						

c) Plot these values onto the scatter plot and use them to draw the linear regression function on your scatter plot.

4)

a) Calculate the *quadratic regression function* $(y = A + Bx + Cx^2)$ as follows:

PROCEDURE	KEY SEQUENCE
NOTE: you don't have to enter the data again.	
Change to the quadratic regression function	[SHIFT] [1] (STAT) [1: Type] [3: _ + CX^2]
Clear the screen	Press [AC]
Find the values of A, B and C	[SHIFT] [1] (STAT) [5: Reg] [1: A] [=] [SHIFT] [1] (STAT) [5: Reg] [2: B] [=] [SHIFT] [1] (STAT) [5: Reg] [3: C] [=]

A = …………………………………….. (correct to 2 decimal places

B = ……………………………………... (correct to 2 decimal places)

C = ……………………………………... (correct to 4 decimal places)

Equation of the quadratic regression function:

……………………………………………………………………………………………

b) Use your calculator to determine the time (correct to 2 decimal places) for the given years as follows:

PROCEDURE	KEY SEQUENCE
Determine the time in 1960	1960 [SHIFT] [1] (STAT) [5: Reg] [6: $\hat{y}$]
Determine the time in 1968	1968 [SHIFT] [1] (STAT) [5: Reg] [6: $\hat{y}$],
Continue like this to determine the rest of the values	

Year	1960	1968	1988	1999	2005	2008
Time (to 2 decimal places)						

c) Plot these values onto the scatter plot and use them to draw the graph of the quadratic regression function.

5)

a) Calculate the *exponential regression function* ($y = A.B^x$) as follows:

PROCEDURE	KEY SEQUENCE
NOTE: you don't have to enter the data again.	
Change to the exponential regression function	[SHIFT] [1] (STAT) [1: Type] [6: A . B^X]
Clear the screen	Press [AC]
Find the values of A and B	[SHIFT] [1] (STAT) [5: Reg] [1: A] [=] [SHIFT] [1] (STAT) [5: Reg] [2: B] [=]

A = …………………………………….. (correct to 2 decimal places

B = ……………………………………... (correct to 3 decimal places)

Equation of the exponential regression function:

……………………………………………………………………………………

b) Use your calculator to determine the time (correct to 2 decimal places) for the given years as follows:

PROCEDURE	KEY SEQUENCE
Determine the time in 1960	1960 [SHIFT] [1] (STAT) [5: Reg] [5: $\hat{y}$]
Determine the time in 1968	1968 [SHIFT] [1] (STAT) [5: Reg] [5: $\hat{y}$],
Continue like this to determine the rest of the values	

Year	1960	1968	1988	1999	2005	2008
Time (to 2 decimal places)						

c) Plot these values onto the scatter plot and use them to draw the graph of the exponential regression function.

6) Fill in your calculated values on this table:

Year	Actual time (s)	Value of linear regression function	Value of quadratic regression function	Value of exponential regression function
1960	10,0			
1968	9,95			
1988	9,92			
1999	9,79			
2005	9,77			
2008	9,69			

7) Which regression function suits the context best? Give reasons for your answer.

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

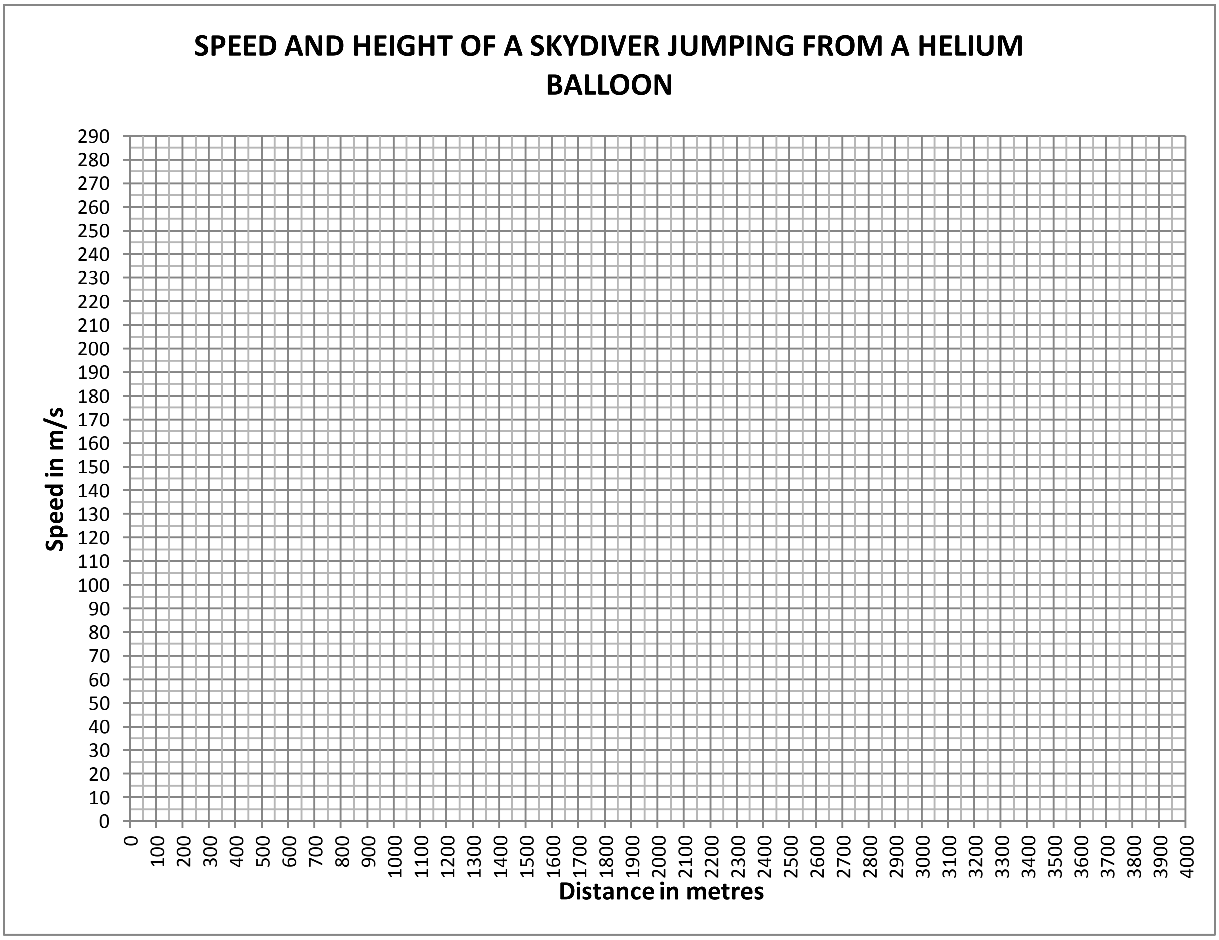
TASK 2

- Felix Baumgartner of Austria holds the record for high altitude skydiving. In October 2012 he skydived an estimated 39 km.
- His highest speed reached before the air resistance started slowing him down, was 1 342 km/h or 373 m/s which is faster than the speed of sound.
- The following table gives the speed, v in m/s, at which a skydiver is moving after having travelled a through a distance, s in m, from a helium balloon which was at a great height when the skydiver jumped.

http://www.theguardian.com/media/2012/oct/15/felix-baumgartner-skydive-youtube

Distance s (m)	130	530	1 050	1 540	2 710	3 750
Speed v (m/s)	52	103	145	175	233	274

1) Draw a scatter plot of the data
2) Use your calculator to fit the following regression functions to the data:
   a) Linear
   b) Exponential
3) Sketch the graphs of the functions obtained in 2) on to your scatter plot using different colours for each.
4) Use the two models and the calculator to determine the speed of the skydiver after falling through a distance of
   a) 1 000 m   b) 2 700 m   c) 3 700 m
5) Which of the two models fits the data better? Give reasons for your choice.


SPEED AND HEIGHT OF A SKYDIVER JUMPING FROM A HELIUM BALLOON
Speed in m/s
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
3200
3300
3400
3500
3600
3700
3800
3900
4000
Distance in metres