

Here is the data from assignment that used the Data Flyer to find the best fitting equation $y = mx + b$ for the 15 different (x, y) pairs. The variable x actually represents the **birthrate** (the number of births per 1000 population in 1992), whereas the variable y represents **femaleLifeExpectancy** in 1992. The countries represented range from France (birthrate = 13, femaleLifeExpectancy = 82 years) to Somalia (birthrate = 46, femaleLifeExpectancy = 55). Netherlands (13, 81) and Tanzania (50, 55) or Burkina Faso (50, 53) are nearly identical.

How can you speed up getting the EQUATION that best fits a line to the above data?

For $x = 13$, we have 81 and 82 as the two different values for y. Netherlands and France.

Let point p1 be $(x_1, y_1) = (13, 81.5)$ or $p_1 = (13, 81)$ or $p_1 = (13, 82)$ would work out fine too.

Let point p2 be $(x_2, y_2) = (50, 54)$ or point p2 could be either (50, 53) or (50, 55) would be fine too.

The rise is then $(y_2 - y_1) = (54 - 81.5) = -27.5$

and

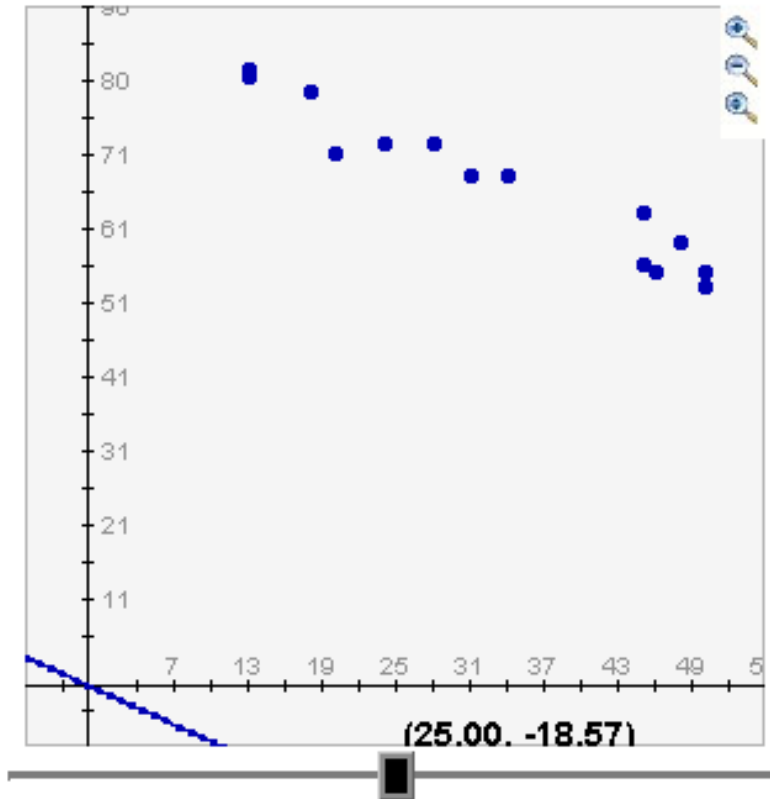
The slope $m = -27.5 / 37 = -0.743$

the run is going to be $(x_2 - x_1) = (50 - 13) = 37$

So, this is a good initial value to try for the SLOPE of the line, so we make $f(x) = -.743 * x$ and then do SET FUNCTION.

Here is a screen snapshot of the function now after doing the SET FUNCTION. It is easily seen that the INTERCEPT needs to be up around 85 or 90 instead of down at 0. It appears that the SLOPE is in the same tilt as the set of 15 points.

Sum of squares of deviations: 1.256e5



$$f(x) = -0.743 * x$$

 Change Function

☐ Show Squares

Data:

45	63
18	79
28	72
24	72
34	68
20	71
28	72
13	82
13	81
50	55
50	53

Plot Data

Clear Data

☒ Auto Scale

☐ Show Vertical Asymptotes

Show Tabular Data

☒ No Grid

☐ Light Grid Lines

So, now if we do the CHANGE FUNCTION operation, we can type in the + 90.0 and see what the function $f(x) = -.743 * x + 90.0$ looks like. The sum of the squares of the deviations should come down from 12,560 to a much, much smaller value.

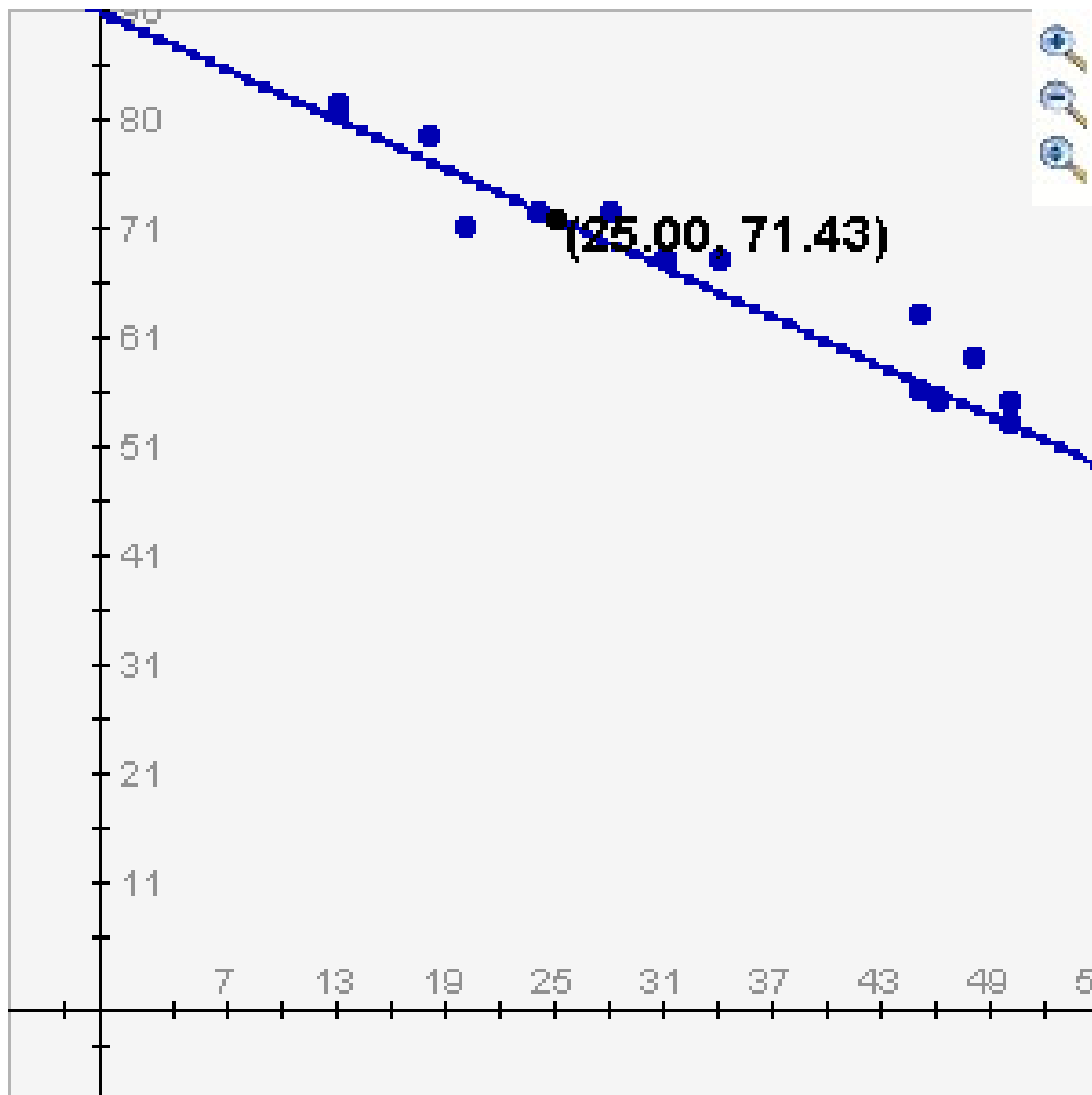
As you can see on the next page, the value dropped from 12,560 down to 122.225, so it was reduced to about 1/100th of its original amount by adjusting the INTERCEPT (changing it from 0 to +90).

It appears as if there are 9 points above the line and 3 below the line, with the other 3 points or countries being on the line. But there are two points out on the right edge of the graph that are easily the most distance from the line. So instead of adjusting the INTERCEPT to a bigger number, what if we adjust the SLOPE to not be quite so steep?

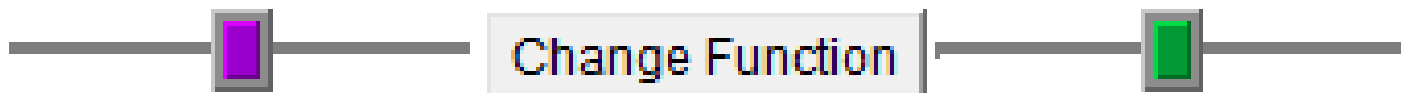
Try slope of -0.711 instead of -0.743 to move it closer to those two points. The deviations sum of squares drops to 86.992. Progress. Turn on SHOW SQUARES and notice the 3 largest squares could be reduced some by decreasing the SLOPE further. Let us try -7.00 instead of -0.711 and see what happens. 83.77 is a reduction of the sum of the squares of the deviations by more than 3 units. More progress. Now using the PURPLE and the GREEN sliders to fine tune the linear equation, we end up with:

Sum of squares of deviations = **83.672** with the linear equation $y = -0.698 * x + 90.01$ as the best fitting equation for the model of the linear relationship between birth rate and female life expectancy for our collection of 15 countries..

Sum of squares of deviations: 122.225



$$f(x) = -0.743 * x + 90.0$$



Sum of squares of deviations: 83.672

☒ Show Squares

Data:

45	63
18	79
28	72
24	72
34	68
20	71
28	72
13	82
13	81
50	55
50	53

Plot Data

Clear Data

☒ Auto Scale

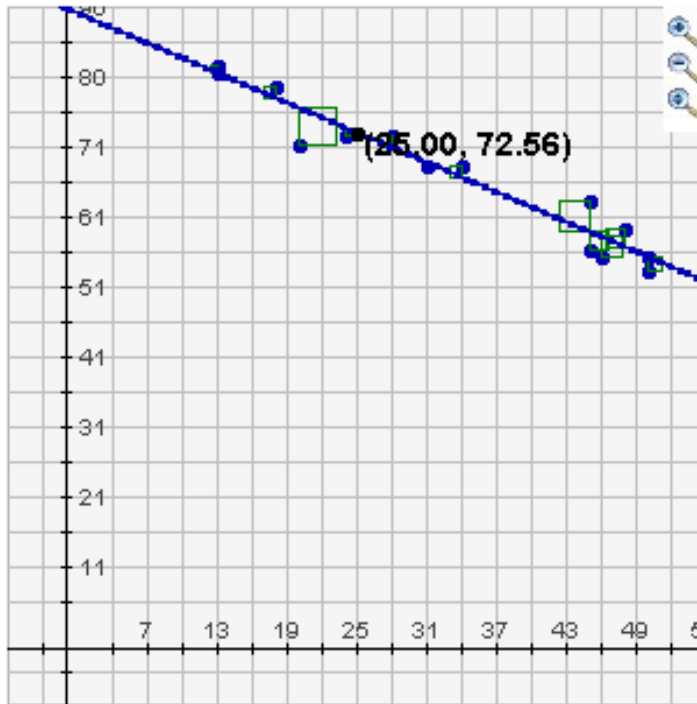
☐ Show Vertical Asymptotes

Show Tabular Data

☐ No Grid

☒ Light Grid Lines

☐ Dark Grid Lines



$$f(x) = -0.698 * x + 90.01$$

Change Function

Reset Sliders

Finally, here is the SPSS statistical software output for the same data:

Can you find the value for the SLOPE m in the SPSS output here?

Can you find the value that SPSS gave for the INTERCEPT b ?

Can you find the number that the SPSS statistical software calculated for the sum of the squares of the deviations? What does it call or label that value in the output?

What does SPSS call the m and the b from the linear equation values $y = mx + b$?

SPSS = Statistical Package for the Social Sciences.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.968 ^a	.938	.933	2.537

a. Predictors: (Constant), Births per 1000 population, 1992

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1259.263	1	1259.263	195.653	.000 ^a
	Residual	83.671	13	6.436		
	Total	1342.933	14			

a. Predictors: (Constant), Births per 1000 population, 1992

b. Dependent Variable: Female life expectancy 1992

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	89.985	1.765		50.995	.000
	Births per 1000 population, 1992	-.697	.050	-.968	-13.988	.000

a. Dependent Variable: Female life expectancy 1992