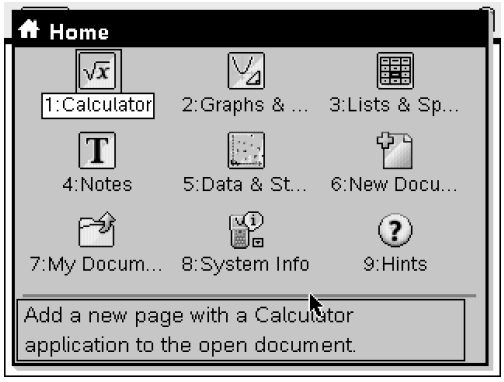
Everything you need to do with Linear Regression using the TI-Nspire: Olympic Swimming (part 1)

Here are the Olympic gold medalists and winning times from the men's 100 *m* Freestyle event.

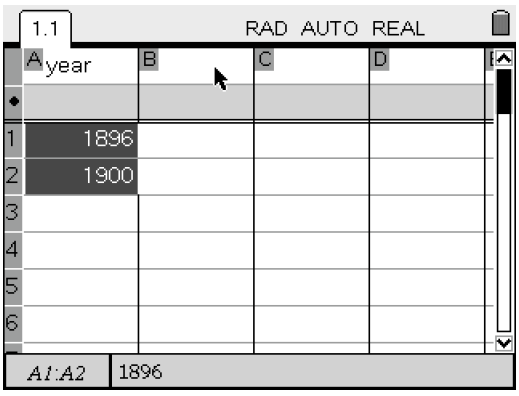
Year	Gold Medal Winner	time
1896	Alfred Hajos, Hungary	1:22.2
1904	Zoltan de Halmay, Hungary	1:02.8
1908	Charles Daniels, United States	1:05.6
1912	Duke Kahanamoku, United States	1:03.4
1920	Duke Kahanamoku, United States	1:01.4
1924	John Weissmuller, United States	59.0
1928	John Weissmuller, United States	58.6
1932	Yasuji Miyazaki, Japan	58.2
1936	Ferenc Csik, Hungary	57.6
1948	Wally Ris, United States	57.3
1952	Clark Scholes, United States	57.4
1956	Jon Henricks, Australia	55.4
1960	John Devitt, Australia	55.2
1964	John Schollander, United States	53.4
1968	Mike Wenden, Australia	52.2
1972	Mark Spitz, United States	51.22
1976	Jim Montgomery, United States	49.99
1980	Jorg Woithe, East Germany	50.40
1984	Rowdy Gaines, United States	49.80
1988	Matt Biondi, United States	48.63
1992	Aleksandr Popov, Unified Team	49.02
1996	Aleksandr Popov, Russia	48.74
2000	Pieter van den Hoogenband, Netherlands	48.30
2004	Pieter van den Hoogenband, Netherlands	48.17
2008	Alain Bernard, France	47.21

Determine the explanatory variable _____ and response variable _____

Enter the data and make a scatterplot: Get a new list page: Home→3: Lists and Spreadsheets. Use the first list (A) and name it "year" (no quotes needed).



Enter the years 1896 and 1900 to start the pattern in cells 1 and 2 (even though 1900 is not in the list). Drag the cursor over the two years to highlight them.

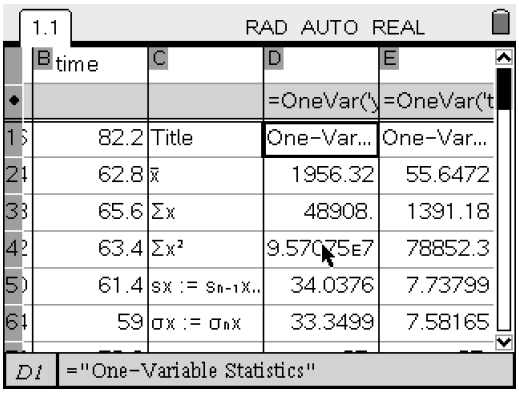


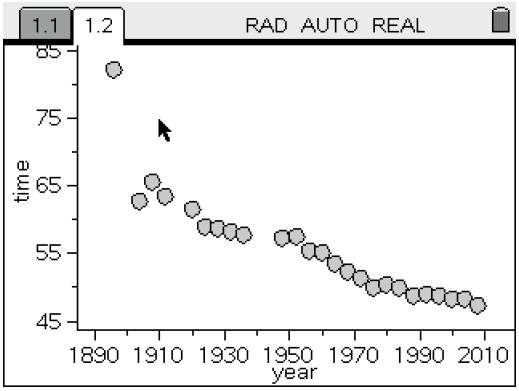
Go to menu→3: Data→3: Fill Down→hit enter. The cells will become outlined. Drag the darkened area down to cell 30. The pattern will "fill" the list "years" following your pattern.
Delete the years when there was no 100 *m* swim event or Olympic games—1900, 1916, 1940, 1944). Use ctrl→menu→5: Delete Cell (you can delete 1940 and 1944 by highlighting them both).

Label the second column (B) "time" and enter the times in <u>seconds</u> (sorry, there's no short cut).

With one cell of the "time" list selected, go to menu→4: Statistics→ 1: Stat Calculations…→1: One-Variable Statistics→ enter. Use 2 for Number of Lists→OK. Use "time" for X1 List→tab→Frequency List: 1→3 tabs→1st Result Column: c→tab→OK→enter.

In the second row you should see $\Sigma x = 1391.18$ (a good way to check data entry—unlikely to have two errors that cancel out).

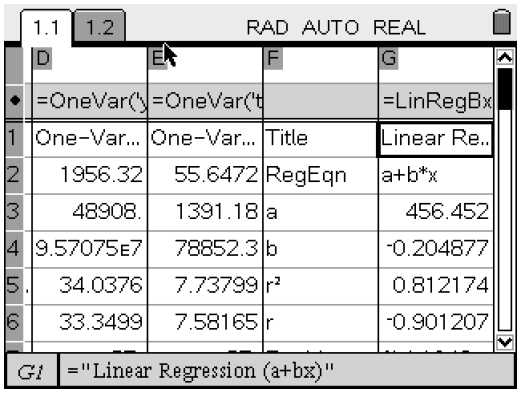




To graph the scatterplot: Insert a new page by going to home→5: Data & Statistics.

"Click to add variable"→select "year" for the explanatory variable →get a dot plot. Click on the center of the left edge of the screen to add the variable "time"→scatterplot. Is the data <u>approximately</u> linear? Positive or negative?

Go back to the Page 1 (ctrl→left click) and move to list E. Find the regression model: menu→4: Statistics→1: Stat Calculations→4: Linear Regression (a+bx). Choose "year" for X List and "time" for Y List→tab to OK.



	D	E	F	G
◆	=OneVar(\	=OneVar('t		=LinRegBx
1	One-Var...	One-Var...	Title	Linear Re..
2	1956.32	55.6472	RegEqn	a+b*x
3	48908.	1391.18	a	456.452
4	9.57075E7	78852.3	b	-0.204877
5	34.0376	7.73799	r²	0.812174
6	33.3499	7.58165	r	-0.901207

G1 ="Linear Regression (a+bx)"

Write the equation in context. Interpret *r*, *r*-squared, the *y*-intercept and slope.

predicted time = 456.45 – .2049 *(Olympic year)*

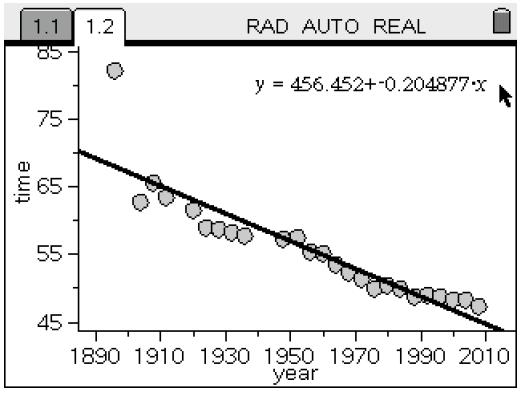
r = –0.9012 – strong, negative association between (Olympic) year and 100 *m* winning men's swim times.

r^2 = 0.8122 – approximately 81% of the variation in 100 *m* winning times can be attributed to the linear relationship with Olympic year

y-intercept = 456.45 – in year 0 AD, our model predicts a winning 100 *m* time of approximately 456.45 seconds (a nonsense value indicating our model does not hold at the *y*-intercept)

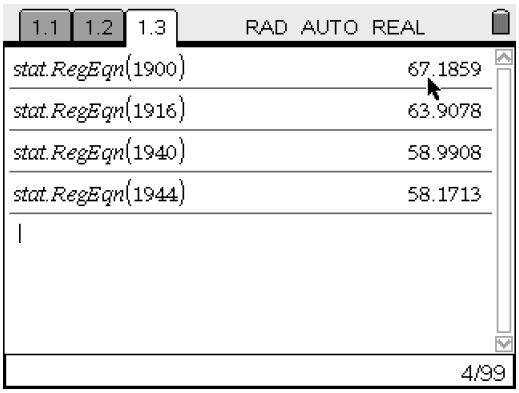
slope = –0.2049 – for every one year increase in Olympic date, the 100 *m* time <u>decreases</u> by 0.2049 sec. (or since Olympics are held every four years, for each successive Olympic game, our model predicts the 100 *m* time will decrease about 0.82 seconds)

Use ctrl→right to go back to the scatterplot page. Use menu→4: Analyze→
6: Regression→2: Show Linear (a+bx) to see the equation and plot the line on the graph.



No Olympic games were held in 1918, 1940, and 1944. Use the regression equation to predict the winning times in those years if the Olympics had been held (and 1900)

1900: 1916: 1940: 1944:



Use a calculator page for this: home→1: Calculator. ctrl→menu→stat.regeqn.

Insert the year inside the ( )→enter. To get additional years you can repeat the above or go up to the equation and use copy and paste: ctrl→menu→3: copy / paste

In 1980 an East German, Jorg Woithe, won with a time of 50.40 seconds (the US did not attend the Moscow Olympics). Calculate the value of his residual

(actual time – expected time) = $(y - \hat{y})$.

ctrl→menu→stat.regeqn. Insert 1980 inside the ( )→enter. Subtract the answer from 50.4 to find the residual: (50.4 – ctrl ans).

Find the dot on the graph for 1980 (appears as red in the software and as dark grey on the handheld). Is the dot above the line or below the line? Does that mean he ran faster or slower than was expected?