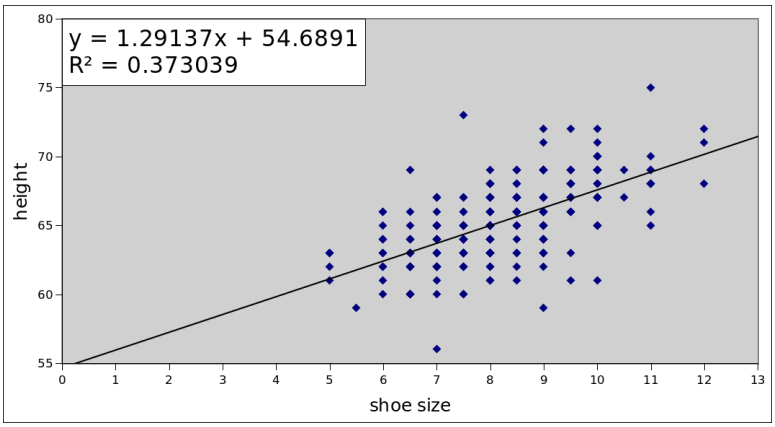
Tue, Mar 24 - Slide 227

This is a **scatterplot** of height versus shoe size for a sample of female C-N students.

- the horizontal axis is x: the shoe size (American adult sizes in increments of $\frac{1}{2}$)
- the vertical axis is y: the girl's height (in inches)



Things to know:

- There is a **linear** relationship because the points roughly hover around a straight line.
- The **linear regression equation** is $y = 54.7 + 1.29x$.
  This is a **model** of the relationship (or association) between x and y.
- The **slope** is 1.29, the coefficient of x.
- The **y-intercept** is 54.7.
- The **correlation** is $r = .611$. Excel usually reports with a capital R and squares it. $(.611)^2 = .373$.
- Since the slope is positive, the correlation is positive. If a graph slopes down, then correlation r is negative.
- If you found a size 10.5 shoe in the hall, a good guess for the height of the girl is

  $$\hat{y} = 54.7 + 1.29(10.5) = 68.25$$

  The "hat" indicates that this is the model's prediction.

Tue, Mar 24 - Slide 230, 231, 232

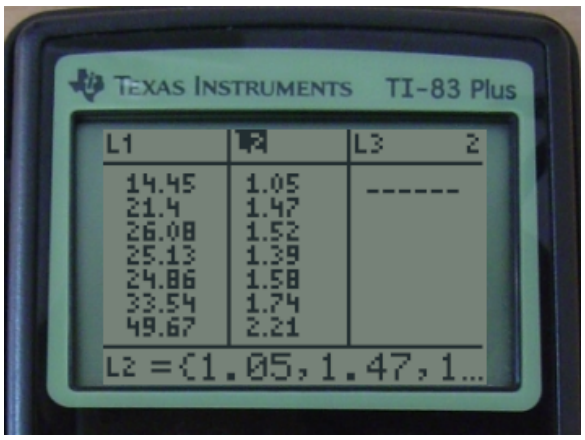
Linear correlation r is a number between -1 and 1.

- If measures how tight the scatterplot is around a straight line.
- If the line has negative slope, then correlation is also negative.
- Play a few rounds of this game until you get the idea.
  `http://www.istics.net/Correlations/`
- You should check the scatterplot to make sure a linear model is appropriate.

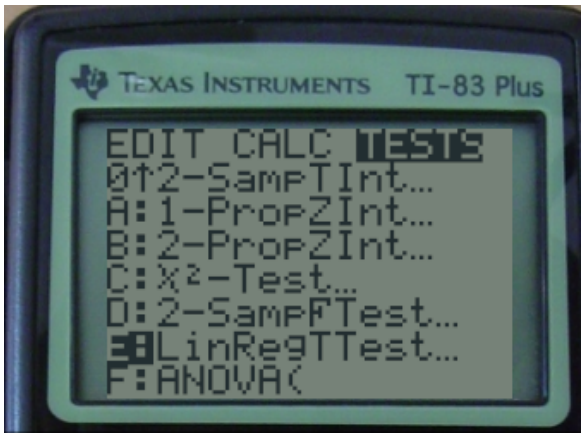
Tue, Mar 24 - Slide 234

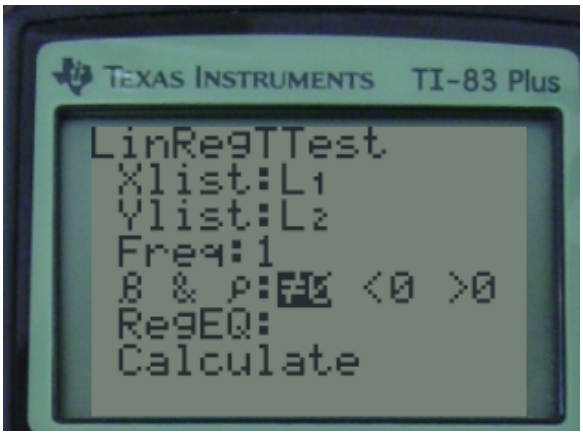
Usually we let some computer software like Excel compute the linear regression and correlation. Here's how to do it on your TI calculator:

- Enter the x values in L1, and the y values in L2.

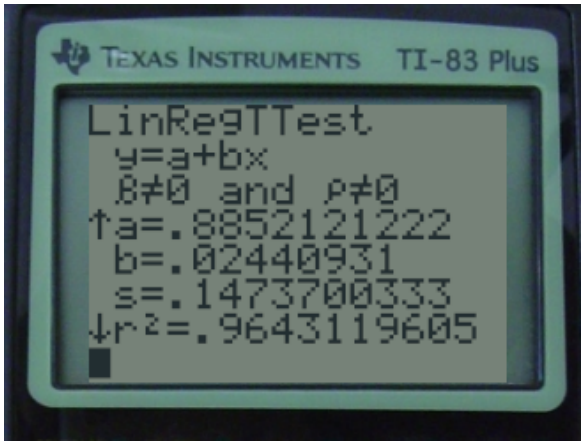


- Hit STAT-TESTS, and scroll down to select LinRegTtest





- Scroll down as needed. Read off the linear model $y = a + bx$, and the correlation r.



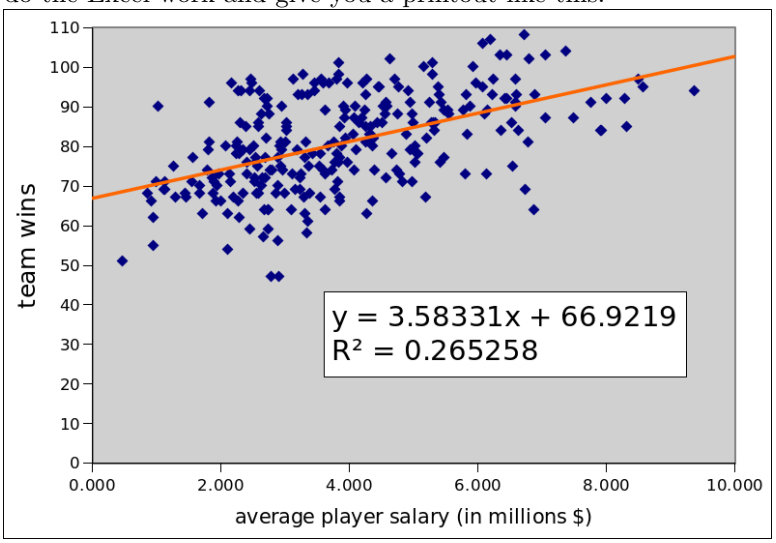
1. With x the \$ price of oil, and y the \$ price of gas, the regression line is $y = .885 + .0244x$.
   - The y-intercept says that gas would cost about 89¢ even if oil were free.
   - The slope says that each \$ 1 increase in the price of oil is associated with about 2.44¢ increase in the price of gas.

2. The correlation is $r = .982$. This is very strong positive correlation (a very tight linear fit).
3. $x = 60$ is well within known x values, so this prediction is called **interpolation**
$$\hat{y} = .885 + .0244(60) = 2.35$$
4. $x = 200$ is outside known x values, so this prediction is called **extrapolation**
$$\hat{y} = .885 + .0244(200) = 5.77$$

It is interesting to note that the corona virus situation, combined with recent actions by Russia and Saudi Arabia have caused this relationship to break a little bit. The current price of oil is about $ 30, so the model predicts $\hat{y} = 1.62$, but last I checked, gas was at $ 2.04.

Tue, Mar 24 - Slide 238

I'll do the Excel work and give you a printout like this:



1. The linear regression equation is $y = 66.9 + 3.58x$, and the correlation is positive $r = +\sqrt{.265258} = .515$.
2. $\hat{y} = 66.9 + 3.58(1.50) = 72.3$ wins
3. Solve $90 = 66.9 + 3.58x$ to get $x = 6.45$ million $ per player.
   Solve $91 = 66.9 + 3.58x$ to get $x = 6.73$ million $ per player.
   So to expect one more win, you must pay $6.73 - 6.45 = .28$ million more per player. Times 25 players gives the cost of a win at about $7 million.

Tue, Mar 24 - Slide 243

It is not necessarily true that if y is correlated with x, that x **causes** y. Often there are **confounding factors** that explain the apparant linear relationship.

Maybe I'll find a good video for you to watch about that. Here's an example to think about: according to `https://www.citylab.com/life/2012/02/what-dollar-store-locations-reveal-about-america/1115/` The prevalence of dollar stores is correlated:

- negatively with education
- negatively with income
- positively with religion
- positively with obesity
- positively with violence

But obviously we couldn't reduce violence by banning dollar stores.

Tue, Mar 24 - Slide 244

In the graph, the correlation between global temperature and number of seafaring pirates is $r^2 = 0.775$. But you can see that the slope is negative, so $r = -\sqrt{.775} = -.88$.

This correlation could be said to be **coincident** with the passage of time. Over the last few centuries, the earth has gotten warmer, and also piracy has declined.