Correlation and Regression

What can you tell me about correlation and regresssion?

Use to answer:

Is there a relationship between two variables?

In what direction?

How strong?

A correlation study includes, in this order:

1. Scatterplot - *all graphing conventions apply*
2. Regression analysis - *state a, b, and r*
3. Hypothesis test - *do all 5 steps*
4. Line of best fit - *state it, plot points, graph it*
5. Summary - *all points*

The relationship is graphically depicted using a scatterplot.

The relationship is numerically quantified using regression analysis.

Fuel Efficiency

Use these data to make a scatterplot on your calculator.

Car Wt	mpg	Car Wt	mpg
3360	23	2900	28
3814	23	4035	19
3840	19	3545	24
3935	20	3050	31
2140	43	2540	34
4306	21	2877	25
2565	34	3192	28
3450	22	3810	22

What is the relationship between wt and mpg?
(Use your words.)

Perform the linear regression function on your calculator.

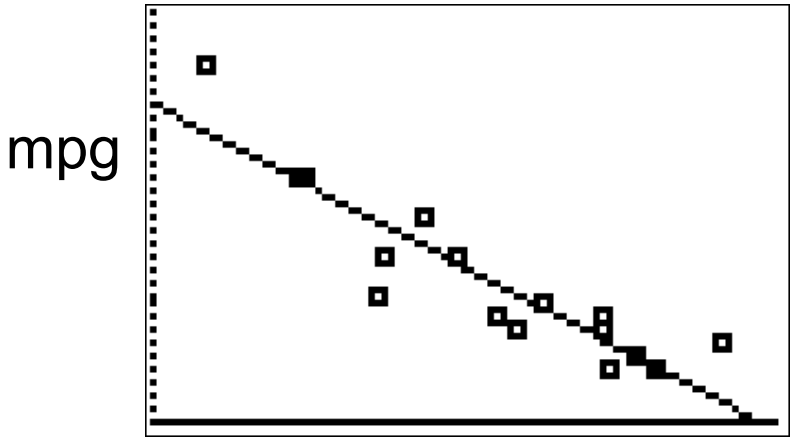
STAT - CALC - 4:LinReg(ax+b)

Write down the *a*, *b*, and *r* numbers.

Using the *a* and *b* results from your calculator accurately plot the regression line on your graph.

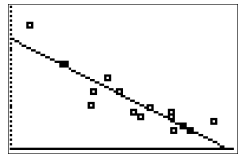
For the calc-saavy a and b can be found in VARS, 5, EQ
Enter aX+b for y1= and it will graph the regression line on your scatter plot

Screen shot of Wt vs mpg



$$mpg \approx -0.01(wt) + 59$$

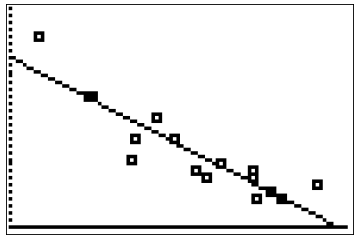
What is the direction of the relationship between weight and mpg?



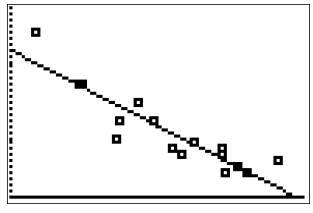
How strong is the relationship between weight and mpg?

What is the expected fuel usage for a car weighing 3700 lbs?

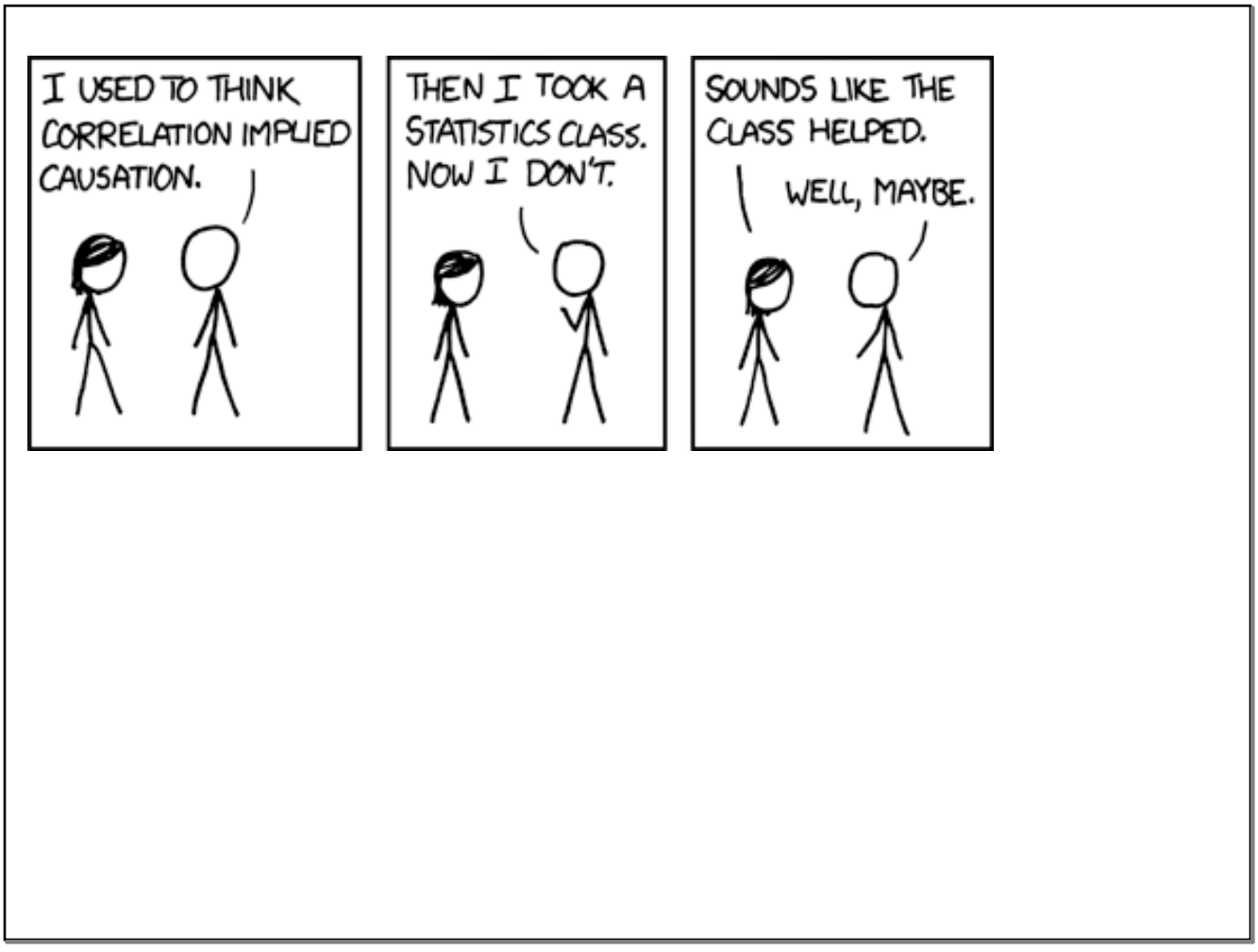
What is the expected weight of a car that gets 40 mpg?

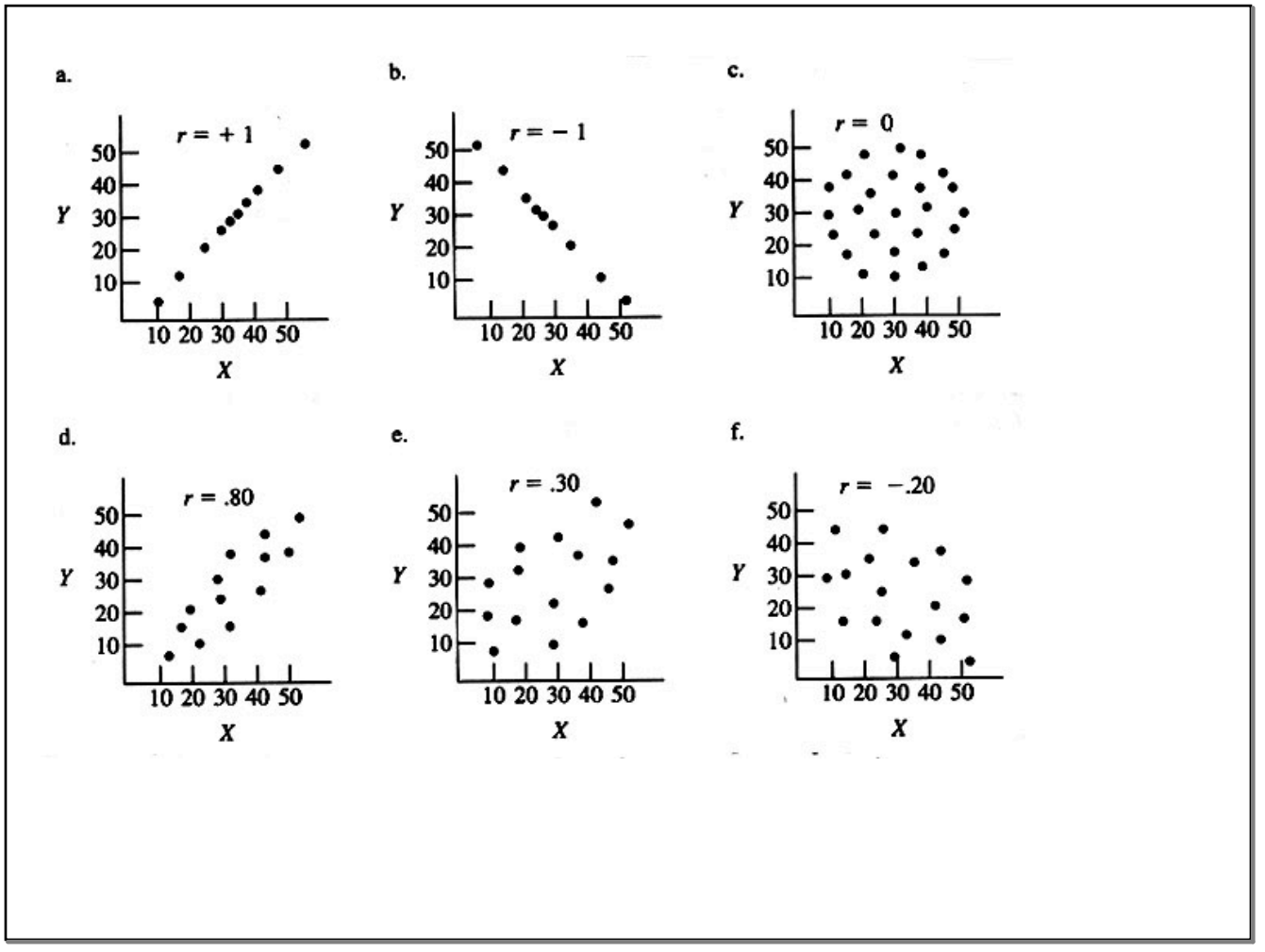


What is the expected weight of a car that gets 60 mpg?



October 28, 2014




a.
r = + 1
b.
r = − 1
c.
r = 0
d.
r = .80
e.
r = .30
f.
r = −.20
Y
X
10 20 30 40 50
10
20
30
40
50

A study was conducted with vegetarians to see whether the numbers of grams of protein each ate per day was related to diastolic blood pressure. The results are in the table. Is there is a significant relationship, predict the the diastolic pressure of a vegetarian who consumes 8 grams of protein per day.

Protein (g)	**4**	**6.5**	**5**	**5.5**	**8**	**10**	**9**	**8.2**	**10.5**
Pressure (mm/Hg)	**73**	**79**	**83**	**82**	**84**	**92**	**88**	**86**	**95**

Describe the relationship.

The Summary

You need to include the following:

- positive/negative
- strong.....weak
- correlation coefficient
- Outliers
- Relate to the real world, using statistical terms.

- Does coffee give you cancer?
- Does cough syrup work?
- Does chicken soup work better?
- Is the SAT a true measure of your college potential?
- Do cell phones give you cancer?
- Does childhood vaccines cause autism?

Does height affect income?

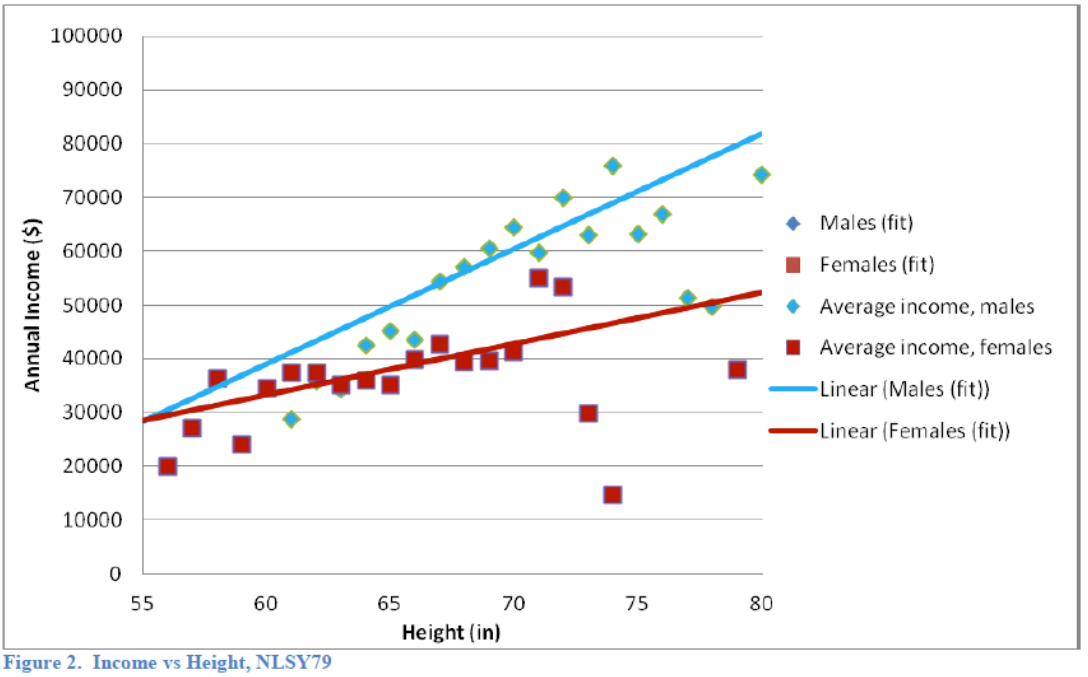
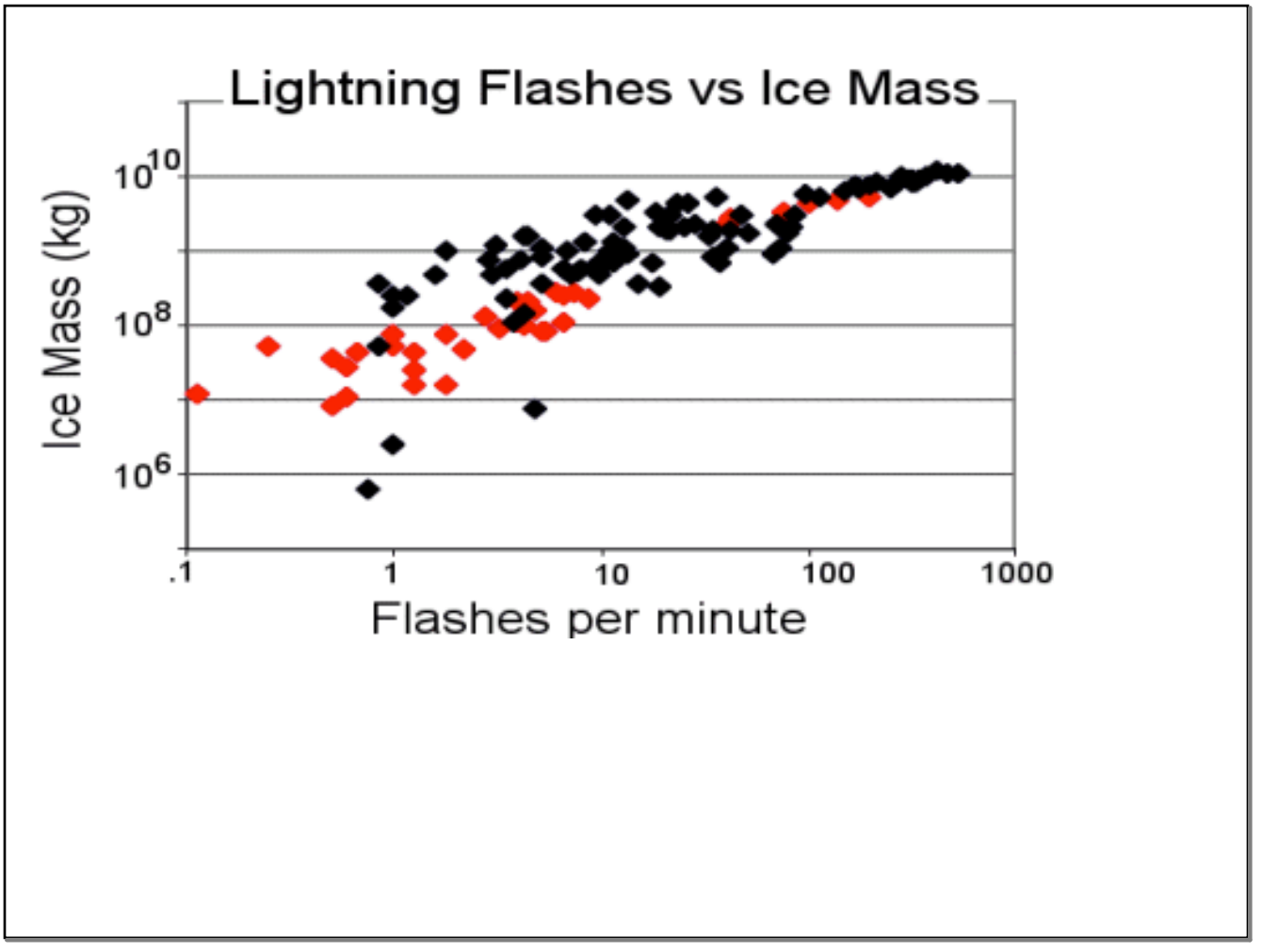


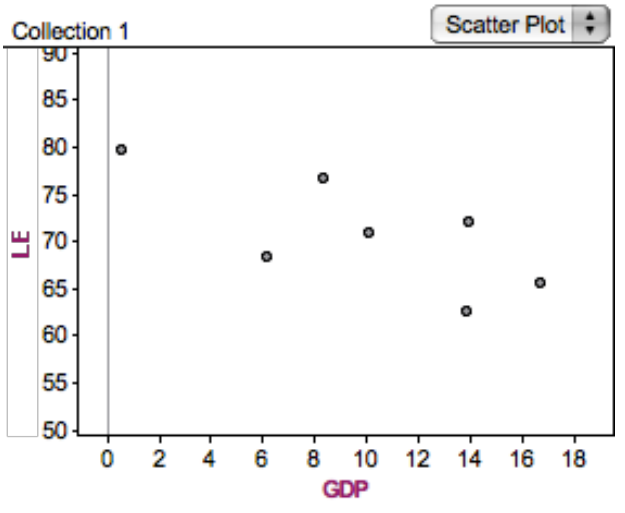
Figure 2. Income vs Height, NLSY79

About $2000/inch for males and $1000/inch for females.


Lightning Flashes vs Ice Mass
Ice Mass (kg)
10^{10}
10^{8}
10^{6}
.1
1
10
100
1000
Flashes per minute

Do you think Life Expectancy is related to a country's GDP?

GDP in (1000s)	LE (yrs)
600	79.8
16700	65.7
13800	62.6
10100	71.0
8300	76.7
13900	71.1
6200	68.5



Data taken from Index Mundi

Hypothesis Significance Test

Tests whether the observed correlation is real or due to chance.

5 step process – first in brief, then in detail.

Step 1. State the hypotheses.
Step 2. Find the critical values (cv).
Step 3. Calculate the test value (tv).
Step 4. Make the decision.
Step 5. Summarize the results.

Step 1. State the Hypotheses.

H_0: $\rho = 0$ *there is no relationship*

H_1: $\rho \neq 0$ *there is evidence for*

H_0 is the null hypothesis

H_1 is the alternate hypothesis

ρ (rho) is the probability the correlation coefficient is real.

Step 2. Find the Critical Values.

Use the t-distribution table with:

$t_{.95}$ column and

n - 2 degrees of freedom

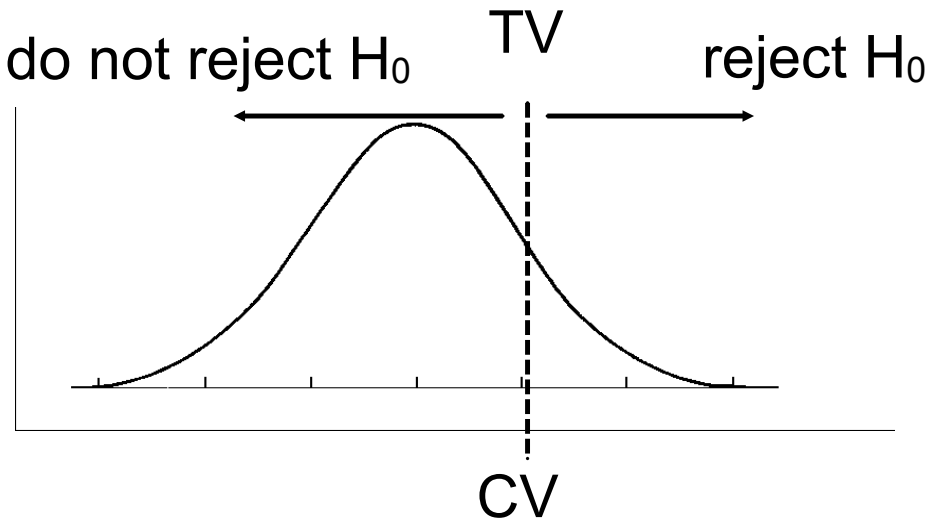
Step 3. Calculate the test value

$$tv = r\sqrt{\frac{n-2}{1-r^2}}$$

Step 4. Make the decision.

If t lies inside the critical region we will reject the null hypothesis. Therefore the observed correlation is real.

If t lies outside the critical region we will not reject the null hypothesis. Therefore the observed correlation is not real.



Step 5. Summarize.

If you reject H_0

"There is enough evidence to reject the null hypothesis, therefore there is enough evidence to suggest a relationship...."

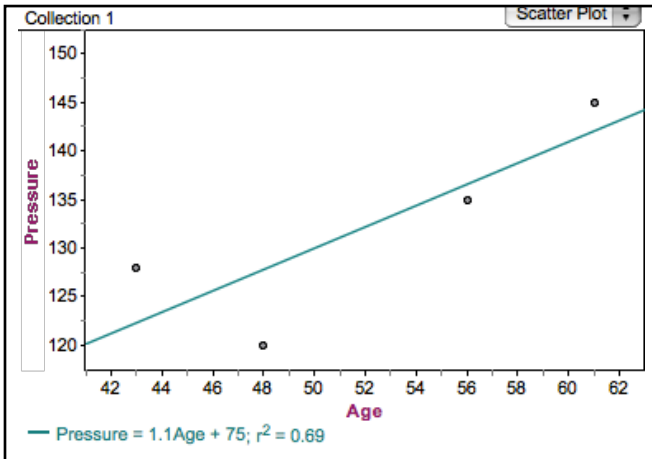
If you do not reject H_0

"There is not enough evidence to reject the null hypothesis, therefore there is not enough evidence to suggest a relationship...."

Try this.

Do the hypothesis test to determine if there is a relationship between age and systolic blood pressure.

Age (years)	Systolic Pressure (mm/Hg)
43	128
48	120
56	135
61	145



Age (years)	Systolic Pressure (mm/Hg)
43	128
48	120
56	135
61	145

$r = 0.8318$, $r^2 = 0.69$, $a = 1.1$, $b = 69$

$df = 2$, $CV = 2.92$, $TV = 2.37$

TV < CV so reject the null hypothesis. There is enough evidence to reject the null hypothesis therefore there is enough evidence to suggest a relationship between Age and Systolic Blood Pressure.

Putting it all together:

Expectations for a Correlation Study

A correlation study includes, in this order:

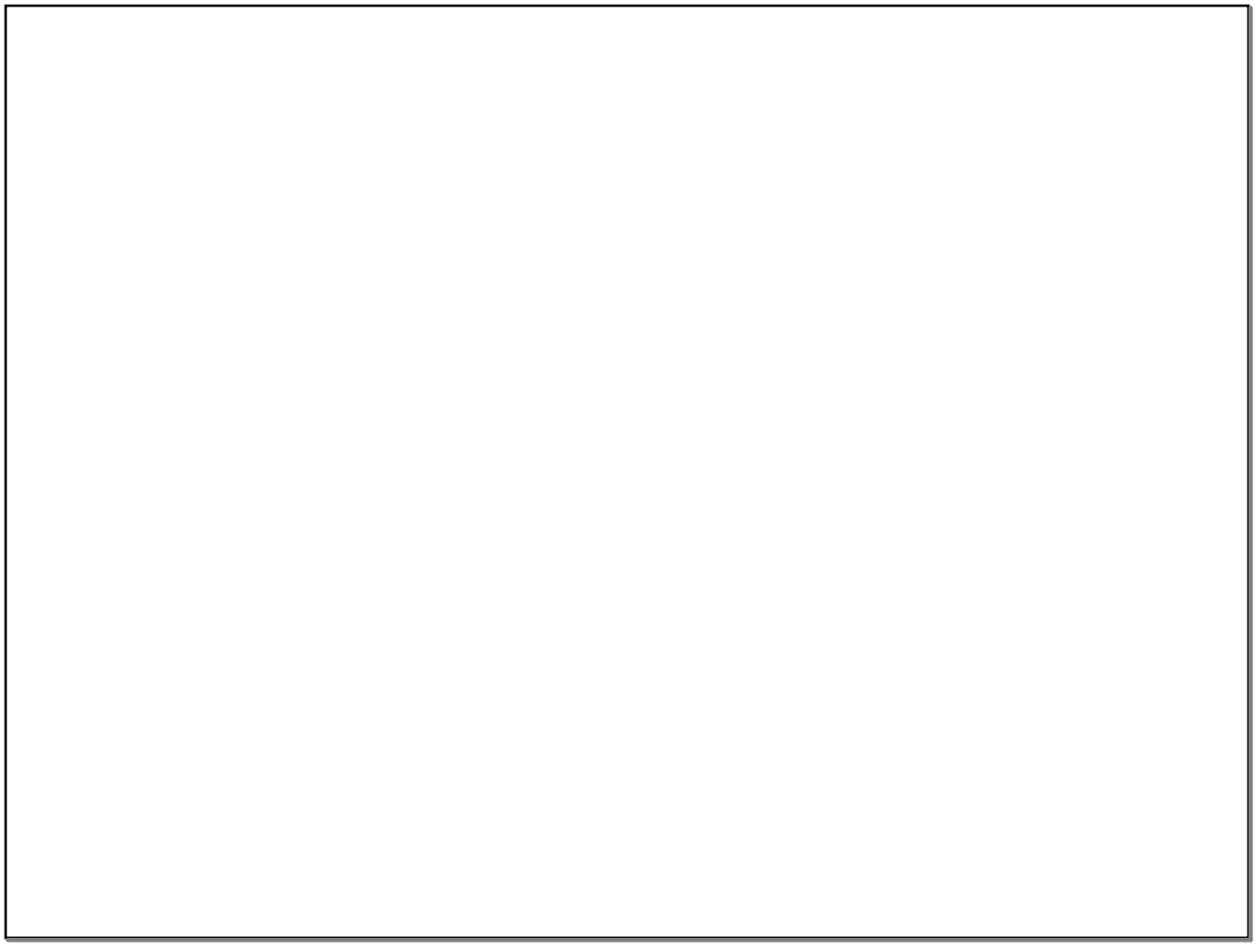
1. Scatterplot - *all graphing conventions apply*
2. Regression analysis - *state a, b, and r*
3. Hypothesis test - *do all 5 steps*
4. Line of best fit - *state it, plot points, graph it*
5. Summary - *all points*

The Summary

You need to include the following:

- positive/negative
- strong.....weak
- correlation coefficient
- Outliers
- Relate to the real world, using statistical terms.

Noisy Crickets

Complete and turn in.

October 28, 2014