The following table displays (a) scores from the Unit 1 test, (b) scores from the Unit 2 test, and (c) scores predicted for the Unit 3 test for each student this semester. Scatterplots display the same data.

Student	Test 1	Test 2	Test 3
1	20.5	41.0	52
2	27.5	48.5	
3	29.5	31.5	80
4	31.5	31.0	77
5	32.5	35.5	80
6	33.0	37.5	78
7	35.0	32.0	85
8	38.0	39.5	84
9	39.0	46.5	85
10	40.5	40.0	88
11	40.5	45.5	75
12	41.0	45.0	78
13	43.0	43.0	80
14	45.0	43.5	83
15	45.5	50.5	87
16	46.0	44.0	88
17	46.5	49.0	84
18	47.5	44.0	85
19	52.0	54.0	91
Means	38.63	42.18	81.11
Std Dv	8.04	6.52	8.46

1. Calculate Pearson's r, Spearman's rho, and Kendall's tau correlations between scores on Test 1 and Test 2.

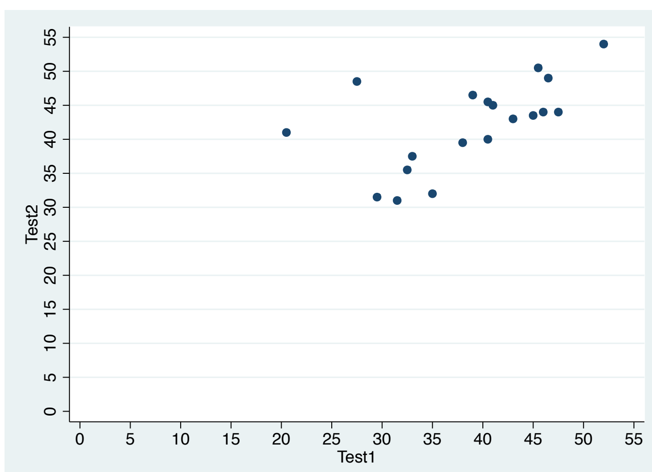
Pearson's r = __________

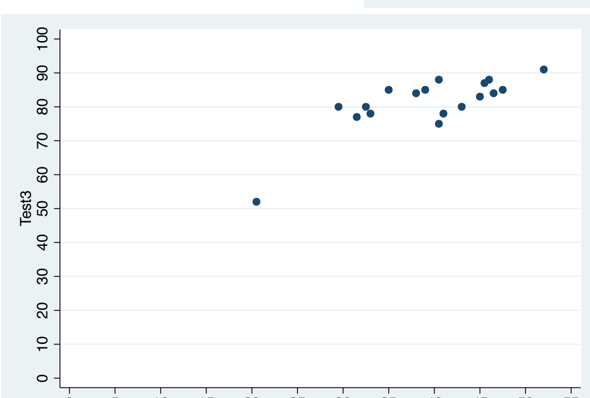
Spearman's rho = __________

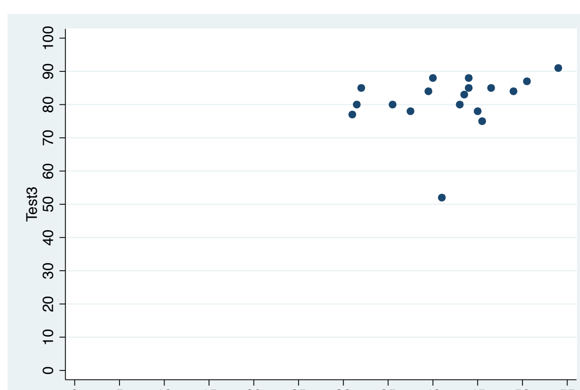
Kendall's tau = __________

2. It looks like scores from Test 1 and Test 2 might have a linear relationship. In the top scatterplot displayed below, roughly sketch the line that you think best fits the data. Guess the slope and y-intercept of that line and write its equation here:

Test 2 = __________ (Test 1) + _____________
(slope) (y-intercept)







3. If we want to find the equation of the line that best fits the Test 1 and Test 2 data, we use something called the least-squares criterion (which we'll learn in Activity #13).

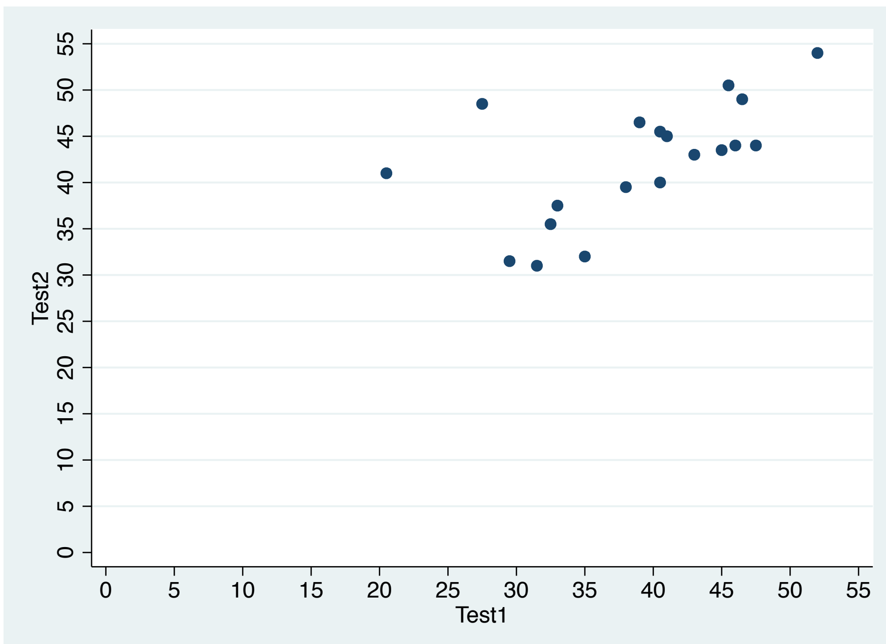
The line that best fits this data can be written as $Y = b_0 + b_1 X$**, where** b_0 = y-intercept **and** b_1 = slope **.**

To calculate the regression line by hand, we use the following: $b_1 = r\dfrac{S_y}{S_x}$ **and** $b_o = \overline{Y} - b_1\overline{X}$ **where**

r = Pearson's correlation coefficient**,** S_y = standard deviation of Y**,** S_X = standard deviation of X**,**

$\overline{Y}$ = mean of Y**, and** $\overline{X}$ = mean of X**. If we let X = Test #1 scores and Y = Test #2 scores, calculate the regression line for this data and sketch it on the scatterplot below.**

Regression Line: ____



4. Use your regression line to predict the Test 2 score for a student who earned a score of 45 on Test 1. What's your prediction for a student scoring 15 on Test 1? For which prediction do you have more confidence?

Predicted Test 2 score for student with Test1 = 45: ______________________

Predicted Test 2 score for student with Test1 = 15: ______________________

In which prediction do you have more confidence? Why? ___________________________________

____

6. When I had Stata compute correlations for our data, this is the output I received:

```
             |    Test1    Test2    Test3
-------------+---------------------------
       Test1 |   1.0000
       Test2 |   0.5886   1.0000
       Test3 |   0.7662   0.2626   1.0000
```

I then had Stata estimate the best-fitting line to predict Test 2 scores from Test 1 scores.

```
------------------------------------------------------------------------------
       Test2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Test1 |   .4772804   .1589839     3.00   0.008     .1418536    .8127072
       _cons |   23.74611   6.266479     3.79   0.001       10.525    36.96723
------------------------------------------------------------------------------
```

Use this output to verify your answers to questions 1 and 3. Interpret the slope and y-intercept values in this line. What do they represent?

Interpretation of slope in this example: ____

____

Interpretation of y-intercept in this example: ____

____

5. Soon, we'll also learn about the *coefficient of determination*, R^2. Calculate this coefficient for the Test 2 & Test 1 data by squaring your correlation coefficient. This coefficient can be interpreted in much the same way as we interpreted our eta-squared values in ANOVA. Go ahead and try to interpret your coefficient of determination.

$\mathbf{R^2}$ = _______________. Interpretation of $\mathbf{R^2}$ = ____

____