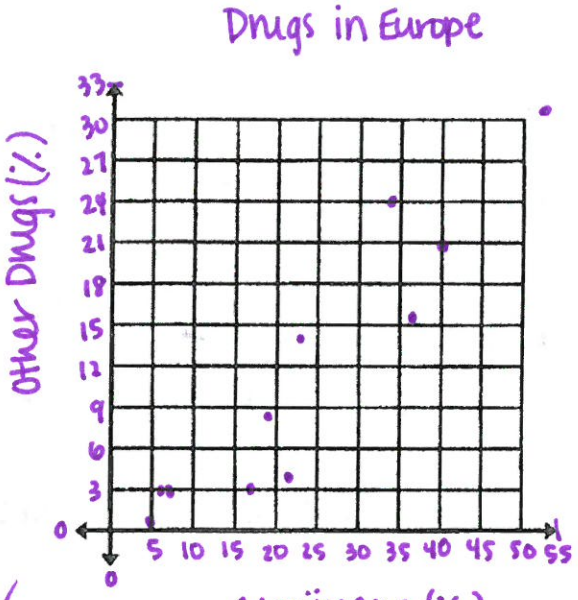
AP Statistics
Unit 02 – Bivariate Data
Day 2 Homework: Correlation & Regression Practice

Name Key
Period 0

1. A survey was conducted in the United States and 10 countries in Western Europe to determine the percentage of teenagers who had used marijuana and other drugs. The results are summarized in the table.

Country	Marijuana	Other Drugs
Czech Republic	22	4
Denmark	17	3
England	40	21
Finland	5	1
Ireland	37	16
Italy	19	8
Northern Ireland	23	14
Norway	6	3
Portugal	7	3
Scotland	53	31
USA	34	24



a. Construct a scatterplot. title, axes labels, scales, points ✓

b. Write a brief description of the association between the percent of teens who have used marijuana and the percent who have used other drugs. Include form, direction, and strength. There is a fairly weak, positive, linear association between marijuana use and other drug use.

c. What is the correlation between these two variables?

$r = 0.9341$

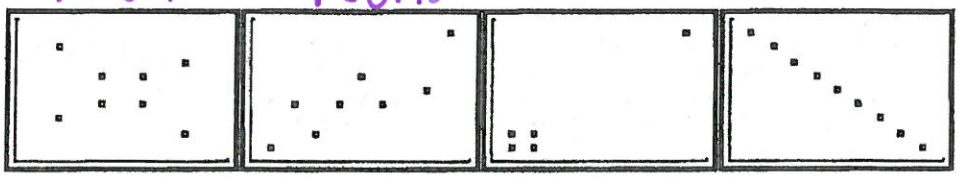
d. Do these results confirm that marijuana is a "gateway drug," that is, that marijuana use leads to the use of other drugs? Explain.

$\hat{y} = -3.07 + 0.6150x$ x = marijuana use, y = other drug use

the slope is positive and the correlation is close to 1 which implies that marijuana may lead to other drug use. they are closely correlated.

2. Match the following scatterplots with the appropriate correlation from the list. Note that not all of the correlations are used. The viewing window is the same in all four plots.

a. r = -0.17, r = -0.48 b. r = 0.82, r = 0.98 c. r = 0.17 ? d. r = -1



Correlations:
r = −.48 ✓ r = .98 ✓ r = .82 ✓ r = −.17 ✓ r = 1 X r = .17 ✓ r = −1 ✓

3. There are a number of software programs that will perform linear regression on a given set of data. Reading the output from these programs requires ignoring a lot of irrelevant information and focusing on the most important values.

   Below is the output from a program called Minitab that is widely used on large servers and other networked systems.

   In this example, a student is using the score form a test (Score1) to predict the score on another test (Score2).

```
Score2 = 1.12 + 0.218 Score1

Predictor      Coef     SE Coef       T       P
Constant     1.1177      0.1093   10.23   0.000
Score1      0.21767     0.01740   12.51   0.000

S = 0.127419   R-Sq = 95.7%   R-Sq(adj) = 95.1%
```

   a. Which test score is the explanatory variable and which is the response variable?

      Score1 is the explanatory variable

      Score2 is the response variable

   b. What is the meaning of the Coef column of the Constant row?

      this number represents the y-intercept, or the value of y when x=0. it is the value of a in the LSRL equation.

   c. What is the meaning of the Coef column of the Score1 row?

      this number represents the slope of the LSRL equation. For each increase in Score1 by 1, Score2 increases by 0.218.

   d. What percent of the variation in the Score2 variable is explained by the regression?

      R^2 = 95.7%. or 0.957

   e. Is the linear regression a good fit? Explain.

      We don't know the shape of the data, so we can't say for sure. However, if we take R^2 and find the square root to get r, r= 0.978 which is very close to 1. This means the LSRL is likely a good fit. Definitely if the data is linear in form.

4. Is there a relationship between test anxiety and exam score performance?

x = score on a measure of test anxiety
y = exam score

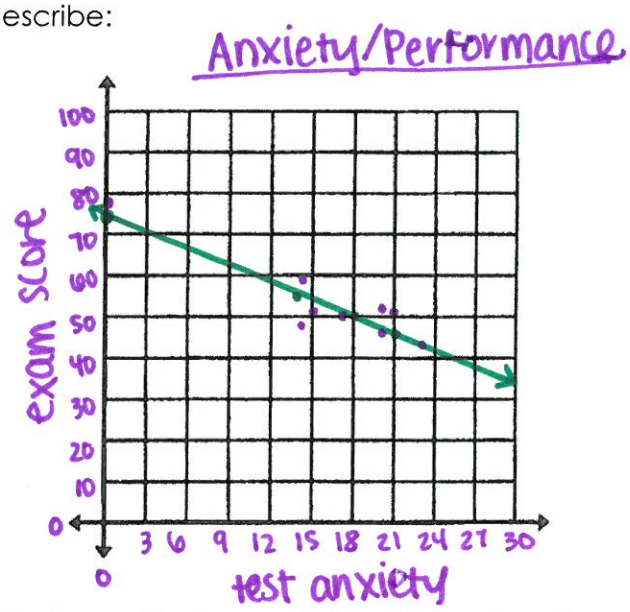
x	23	14	14	0	17	20	20	15	21
y	43	59	48	77	50	52	46	51	51

a. Construct a scatterplot of these points and describe:

F: fairly linear cluster between 14 and 23 with a lone point at (0,77)

D: there appears to be a negative association

S: pretty weak association because there are very few points and they form more of a cluster than a line.



b. Does there appear to be a linear relationship between the two variables? Based on the scatterplot, would you characterize the relationship as positive or negative? Strong or weak?

Yes, there appears to be a weak, negative, linear association.

c. Use your calculator to determine the least-squares regression line (LSRL). Write the equation and plot this line on your graph. Be sure to show what information you're using to plot the line.

$\widehat{\text{exam score}} = 74.46 - 1.34(\text{test anxiety})$

d. What is the correlation between x and y? Interpret this result.

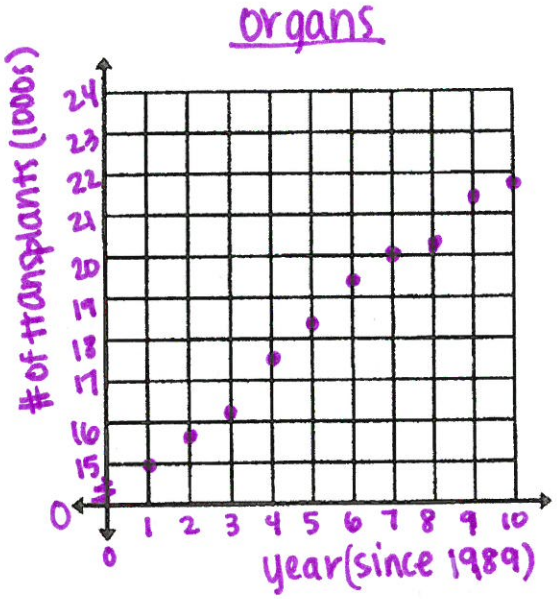
$r = -0.9124$ which would imply a fairly strong, negative association between test anxiety and exam score.

e. Is it reasonable to conclude that test anxiety caused poor exam performance? Explain.

Yes. The equation and graph both support the idea that the higher the test anxiety score, the lower the exam score. Other variables may have additional influence, but our conclusion is reasonable.

5. The following table gives the number of organ transplants performed in the United States each year from 1990 to 1999 (The Organ Procurement and Transplantation Network, 2003):

Year	# of Transplants (1000s)
1 (1990)	15.0
2	15.7
3	16.1
4	17.6
5	18.4
6	19.4
7	20.0
8	20.3
9	21.4
10 (1999)	21.8



a. Construct a scatterplot of the data on the graphs above. Which variable is the explanatory and which is the response?

the year is the explanatory variable
the # of transplants is the response variable

b. Use your calculator to find the least-squares regression line that describes the relationship between the number of transplants and the year.

(# of transplants in 1000s) = 14.23 + 0.7897(years since 1989)

c. Compute the 10 residuals and construct a residual plot. Describe the residual plot. Does it indicate that the linear model is a good fit for the relationship between the number of transplants and the year? the points appear to be random (could argue slight curve) and there are the same # of points above and below the LSRL

x	y	$\hat{y}$	residual = y - $\hat{y}$
1	15.0	15.02	15.0 - 15.02 = -0.02
2	15.7	15.79	15.7 - 15.79 = -0.09
3	16.1	16.58	16.1 - 16.58 = -0.43
4	17.6	17.37	17.6 - 17.37 = 0.23
5	18.4	18.16	18.4 - 18.16 = 0.14
6	19.4	18.95	19.4 - 18.95 = 0.45
7	20.0	19.74	20.0 - 19.74 = 0.26
8	20.3	20.53	20.3 - 20.53 = -0.23
9	21.4	21.32	21.4 - 21.32 = 0.08
10	21.8	22.11	21.8 - 22.11 = -0.31