

AP Statistics

Unit 02 – Bivariate Data

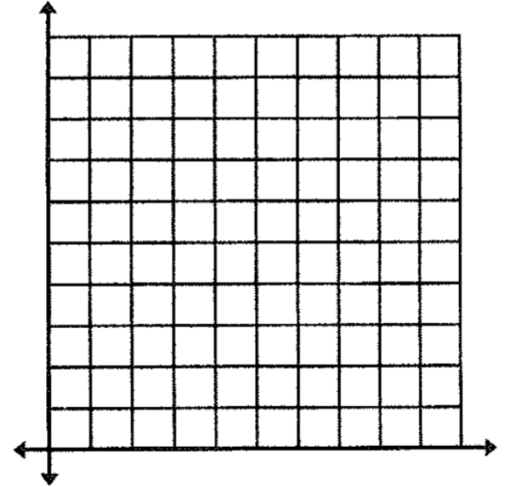
Day 2 Homework: Correlation & Regression Practice

Name _____

Period _____

- 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.

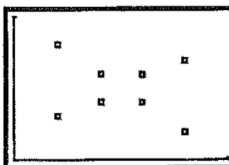
Country	Marijuana	Other Drugs
Czech Republic	22	4
Denmark	17	3
England	40	21
Finland	5	1
Ireland	37	16
Italy	19	8
Northern Ireland	23	14
Norway	6	3
Portugal	7	3
Scotland	53	31
USA	34	24



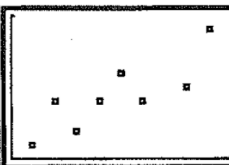
- Construct a scatterplot.
- Write a brief description of the association between the percent of teens who have used marijuana and the percent who have used other drugs. Include form, direction, and strength.
- What is the correlation between these two variables?
- Do these results confirm that marijuana is a “gateway drug,” that is, that marijuana use leads to the use of other drugs? Explain.

- 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.

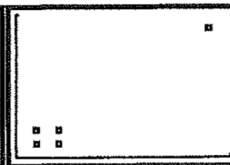
a.



b.



c.



d.



Correlations:

$r = -.48$

$r = .98$

$r = .82$

$r = -.17$

$r = 1$

$r = .17$

$r = -1$

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 from a test (Score1) to predict the score on another test (Score2).

$$\text{Score2} = 1.12 + 0.218 \text{ Score1}$$

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

$$s = 0.127419 \quad R\text{-Sq} = 95.7\% \quad R\text{-Sq}(\text{adj}) = 95.1\%$$

- Which test score is the explanatory variable and which is the response variable?
- What is the meaning of the Coef column of the Constant row?
- What is the meaning of the Coef column of the Score1 row?
- What percent of the variation in the Score2 variable is explained by the regression?
- Is the linear regression a good fit? Explain.

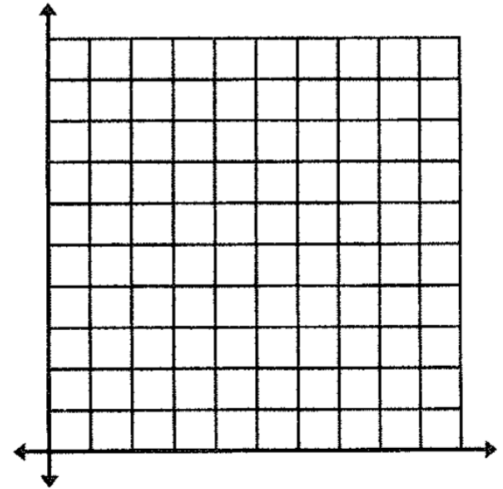
4. Is there a relationship between test anxiety and exam score performance?

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:



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?

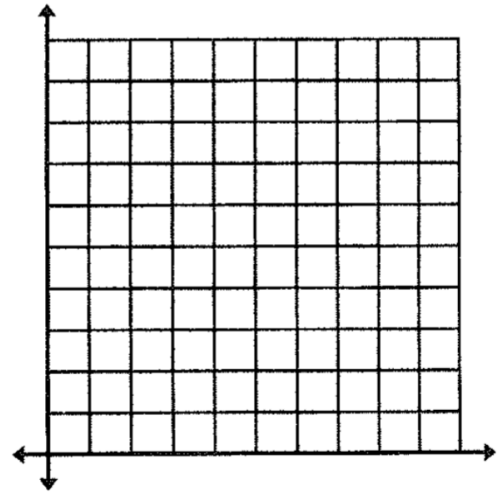
c. Use your calculator to determine the least-squares regression line (LSRL). Write the equation and plot this line on your graph. Be sure to show what information you're using to plot the line.

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

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

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



- Construct a scatterplot of the data on the graphs above. Which variable is the explanatory and which is the response?
- Use your calculator to find the least-squares regression line that describes the relationship between the number of transplants and the year.
- 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?

x	y	\hat{y}	residual = $y - \hat{y}$

