

Chapter 3 Review!

1. Choose **explanatory** and **response** variable: For the following pairs of variables, identify the response variable and the explanatory variable.

- Number of square feet in a house and assessed value of a house
- Political party preference (Democrat, Independent, Republican) and gender
- Annual income and number of years of education
- Number of pounds lost on a diet and type of diet (low-fat, low carbohydrate)

2. **Life after death for males and females:** In a recent General Social Survey, respondents answered the question, "Do you believe in life after death?" The table shows the responses cross-tabulated with gender.

Opinion About Life After Death by Gender			
	Opinion about life after death		
Gender	Yes	No	TOTAL
Male	417 .77	125 .23	542
Female	541 .86	88 .14	629

- Construct a table of conditional proportions on gender.
- Summarize the results. Is there much difference between responses of males and females?

*There is not too much of a difference - females are slightly higher for yes

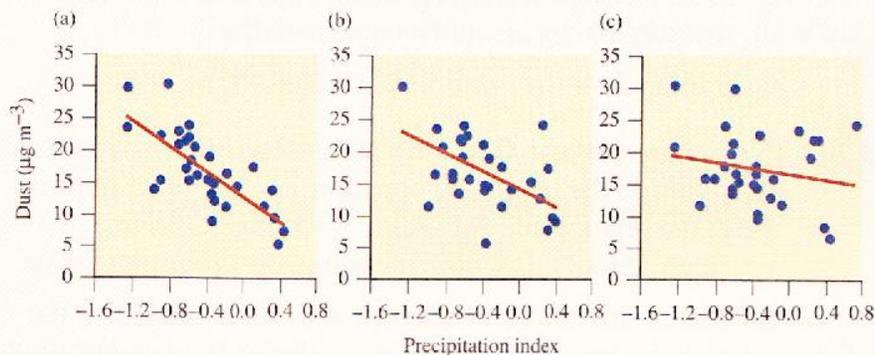
3. **Degrees and income:** The mean annual salaries earned in 2005 by year-round workers with various educational degrees are given in the table:

Degree	Mean Salary
No Diploma	\$19,964
High School Diploma	\$29,448
Bachelor's Degree	\$54,689
Master's Degree	\$67,898
Doctoral Degree	\$92,863
Professional Degree	\$119,009

Identify the **response** variable. Is it quantitative or categorical? **Q**

Identify the **explanatory** variable. Is it quantitative or categorical? **C**

4. **African droughts and dust:** Is there a relationship between the amount of dust carried over large areas of the Atlantic and Caribbean and the amount of rainfall in African regions? In an article (by J. Prospero and P. Lamb, *Science*, vol. 302, p 1024, Nov. 7 2003) the following scatterplots were given along with their corresponding equations. The precipitation index is a measure of rainfall.



Source: J. M. Prospero and P. J. Lamb, *Science*, 302: 1024–1027 (2003).

(iii)

(i)

(ii)

Match the following regression equations and correlations with the appropriate graph.

(i) $\hat{y} = 25.05 - 7.18x$; $r = -0.75$

(ii) $\hat{y} = 21.00 - 2.36x$; $r = -0.44$

(iii) $\hat{y} = 27.80 - 9.77x$; $r = -0.87$

Would you say there is a relationship? **Yes, negative relationship**

5. **Crime rate and urbanization:** For the sample data on crime in Florida, the regression line between $y =$ crime rate (number of crimes per 1000 people) and $x =$ percentage living in an urban environment is $\hat{y} = 24.5 + 0.56x$

a. Using slope, find the difference in predicted crime rates between counties that are 100% urban and counties that are 0% urban. Interpret.

$$y = 24.5 + .56(0) = 24.5 \quad y = 24.5 + .56(100) = 80.5 \quad 80.5 - 24.5 = 56$$

b. Interpret the correlation of 0.67 between these variables.

Moderately strong and positive

c. Show the connection between the correlations and the slope, using the standard deviations of 28.3 for crime rate and 34.0 for percentage urban.

$$b = .67(28.3/34) = .56$$

6. **Predict crime using poverty:** A recent analysis of data for the 50 U.S. states on $y =$ violent crime rate (measured as number of violent crimes per 100,000 people in the state) and $x =$ poverty rate (percent of people in the state living at or below the poverty level) yielded the regression equation, $\hat{y} = 209.9 + 25.5x$.

a. Interpret the slope.

For every 1% increase in poverty rate, there will be 25.5 more violent crimes

b. Would the correlation between these variables be positive or negative? Why?

Positive, the slope is positive

7. **Predicting Internet use from cell-phone use:** We use data from the "human development" data file and cell-phone and Internet use for 39 countries.

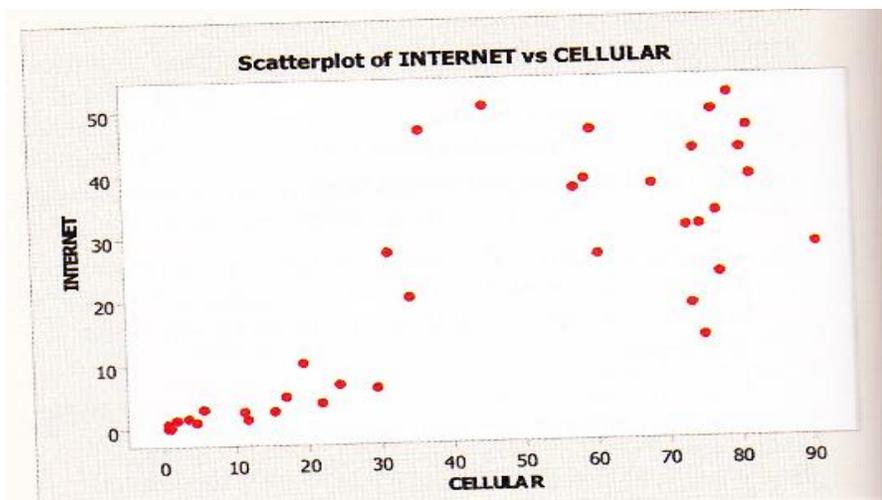
* Identify the approximate x - and y - coordinates for a nation that has less Internet use than you would expect given its level of cell-phone use.

***answers vary* (75, 15)**

*For the United States, $x = 45.1$ and $y = 50.15$. Find its predicted Internet use and residual. Interpret the large positive residual.

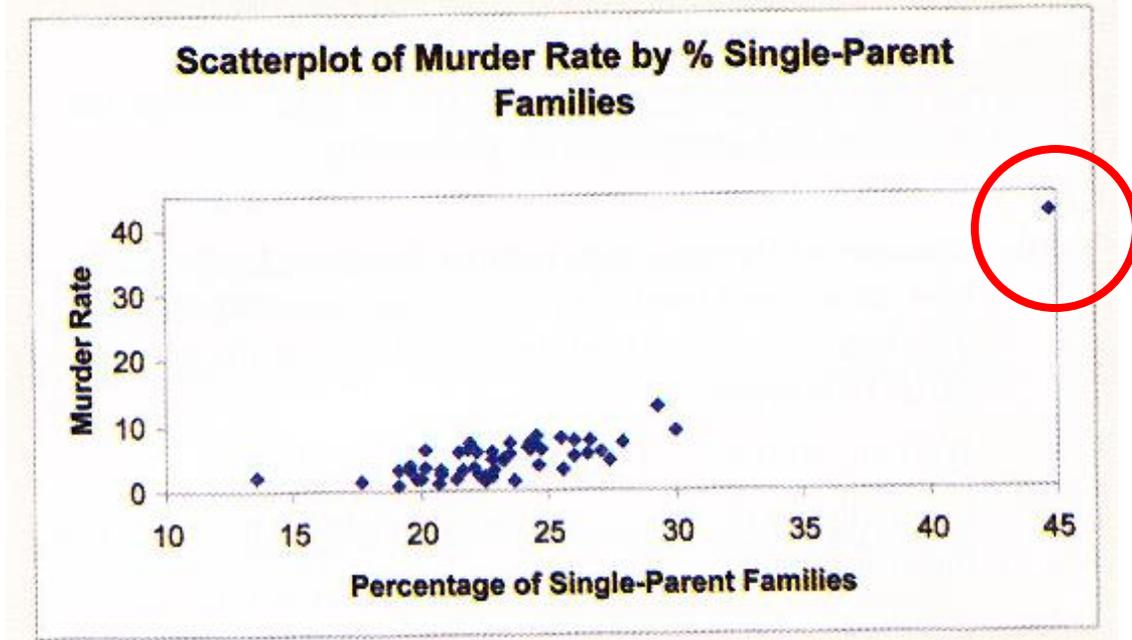
$$y \text{ hat} = 1.27 + .457(45.1) = 22.69$$

$$y - y \text{ hat} = 50.15 - 22.69 = 27.46 - \text{The US is higher than we would expect}$$



8. **Murder and single-parent families:** For Table 3.6 on the 50 states and D.C., the MINITAB figure below shows the relationship between the murder rate and the percentage of single-parent families. Identify D.C. on the scatterplot, and explain the effect you would expect it to have on a regression analysis.

This outlier will pull the regression line towards it causing a positive association



9. **Women's Olympic high jumps:** Example 11 in the textbook discussed how the winning height in the Olympic high jump changes over time. Using the "High Jump" data, it was found that: $\hat{y} = -11.48 + 0.0068x$ for predicting the women's height (in meters) using the year number.

a. Predict the winning Olympic high jump distance for women in (i) 2012, (ii) 3000.

$$y = -11.48 + .0068(2015) = 2.22 \quad y = -11.48 + .0068(3000) = 8.92$$

b. Do you feel comfortable making either prediction in (a)? Explain.

Probably for 2015, but not for 3000 - it's too far out

10. **Income and height.** A survey of *adults* revealed a positive correlation between the height of the subject and their income in the previous year. Explain how gender could be a potential lurking variable that could be responsible for the association.

Gender or athletes

11. **More sleep causes death?:** A study at the University of California at San Diego that reported, based on a nationwide survey, that those who averaged at least 8 hours of sleep a night were 12 percent more likely to die within six years than those who averaged 6.5 to 7.5 hours of sleep a night. Explain a lurking variable responsible for the observed association between sleeping and death rate. **Age - older people sleep longer & closer to death**