

INSTRUCTIONS: Read the questions carefully and completely. Answer each question and show work in the space provided. Partial credit will not be given if work is not shown. When asked to explain, describe, or comment, do so within the context of the problem. Be sure to include units when dealing with quantitative variables.

1. [20 pts] Short answer.
- a) [3] Statistics is about ... _____. (Fill in the blank with one word.)
 - b) [3] Indicate whether each variable is categorical/qualitative or numerical/quantitative. For numerical/quantitative variables, give the appropriate units.

Variable	Categorical/ Qualitative	Numerical/ Quantitative (units)
College Major		
Commute Time		

- c) [4] For a set of 100 observations, the distance from the minimum to the lower quartile is 1.8, from the lower quartile to the median is 0.9, from the median to the upper quartile is 0.4 and from the upper quartile to the maximum is 0.1. Describe the shape of the distribution and explain why you choose that shape.
- d) [3] The median of 11 numbers is 20. If one of those 11 numbers is changed from 23 to 43, what is the new median?
- e) [4] For the following question identify the population, variable, type of variable and parameter. What is the average size of the high school graduating class for students attending Iowa State University?

Population:

Variable:

Type of variable:

Parameter:

- f) [3] The mean of 10 numbers is 20, if one of those numbers is changed from 23 to 43, what is the new mean?
2. [15 pts] The following is excerpted from an article written by Paul Recer, Associated Press, that appeared in the February 6, 2002 Des Moines Register.

Study: Tanning raises cancer risk.

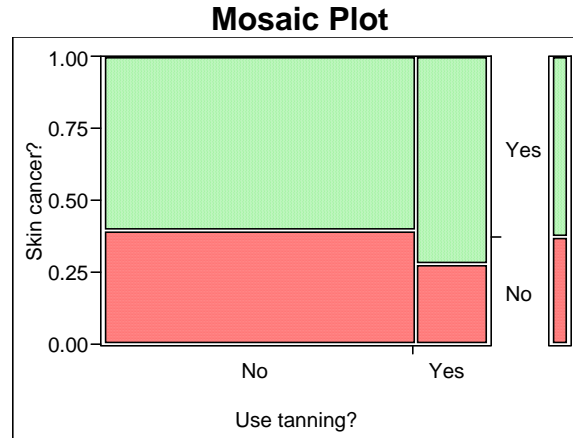
In a study in today's *Journal of the National Cancer Institute*, researchers found that people who used tanning devices were 1.5 to 2.5 times more likely to have common kinds of skin cancer than were people who did not use the devices.

Researchers looked at the medical records and interviewed 1436 people. Of these, 896 had basil cell or squamous cell skin cancer. The other 540 people did not have either cancer.

Among the skin cancer patients, 190 reported using tanning devices at some time. Of those without skin cancer, only 75 had used tanning devices.

- a) [4] Identify the variables in this study and indicate their types.
- b) [4] Why is this an observational study and not an experiment? Explain briefly.
- c) [3] Is this a retrospective or prospective study? Explain briefly.

- d) [4] Below is a mosaic plot of the summarized data. Does this plot agree with the reported finding “that people who used tanning devices were 1.5 to 2.5 times more likely to have common kinds of skin cancer than were people who did not use the devices.”? Explain briefly.



3. [14 pts] The following is excerpted from an article that appeared in the January 21, 2007 Des Moines Register.

Broccoli, tomatoes work better together.

Broccoli and tomatoes – two vegetables known to help fight cancer – are more effective against prostate cancer if they’re eaten together as part of a daily diet than if they’re eaten alone, a new study with rats suggests.

Researchers fed a diet containing 10% broccoli powder and 10% tomato powder to a group of rats that had been implanted with prostate cancer cells. Another group of rats received only the 10% broccoli powder while a third group of rats received only the 10% tomato powder.

After 22 weeks, the researchers found that the combined broccoli/tomato diet was the most effective in reducing the prostate cancer (measured as % reduction in prostate tumor size).

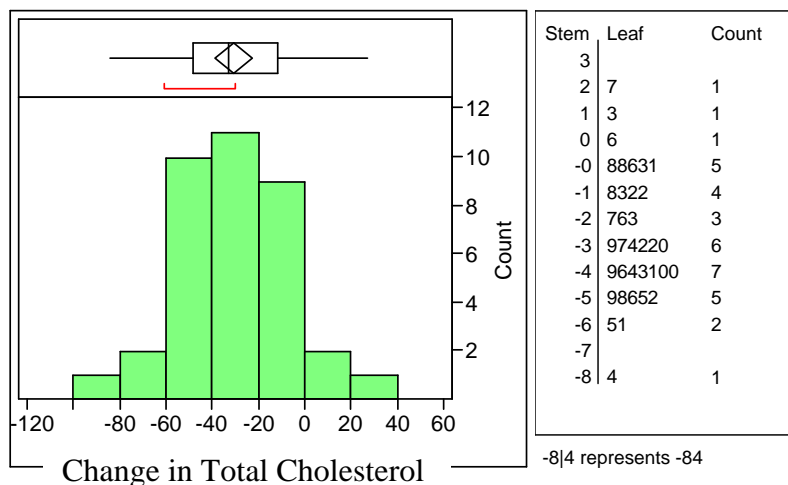
- a) [5] Why was this an experiment and not an observational study? Explain briefly.

b) [3] Identify the response variable. What type of variable is this? Qualitative/Categorical or Quantitative/Numerical?

c) [3] What were the treatments? Be specific.

d) [3] Was there a control group in this experiment? Explain briefly.

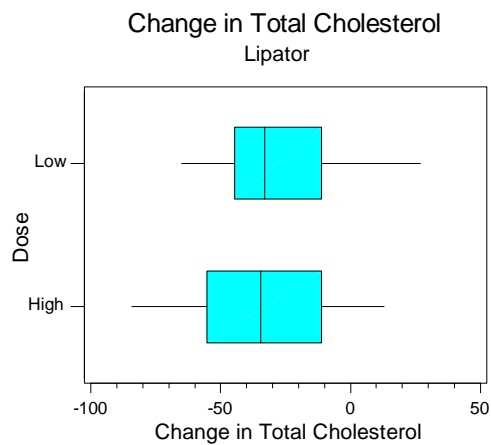
4. [26 pts] In class we discussed an experiment looking at the efficacy of a drug prescribed to lower cholesterol. 36 individuals are involved in the experiment and the response is the change in their total cholesterol level. A negative value indicates that their cholesterol went down by that much. A positive value indicates that their cholesterol actually went up while taking the drug. Below is JMP output for the change in total cholesterol while taking the drug Lipator™.



Mean	-30.8
Std Dev	24.27
N	36

- a) [4] Describe the shape of the distribution of change in total cholesterol.
- b) [4] Are there any apparent outliers? Support your answer by referring to the appropriate plot.
- c) [4] What is the value of the sample mean? the sample standard deviation?
- d) [4] Compute the value of the sample median.

There were two treatment groups in the experiment. One group of 18 randomly assigned participants were given a low dose of Lipitor™, while a second group of 18 randomly assigned participants were given a high dose of Lipitor™.



- e) [2] For the high dose of Lipator™ what is the approximate value of the median change in total cholesterol?
- f) [3] For the high dose of Lipator™ what is the approximate value of the InterQuartile Range (IQR)?
- g) [5] Miguel, who took a high dose of Lipator, lowered his total cholesterol by 37 points (change in cholesterol = -37). Alphonse, who took a low dose of Lipator, lowered his total cholesterol by 34 points (change in cholesterol = -34). Which of the two had a bigger **standardized** change?

Change in Total Cholesterol			
	High dose		Low dose
Mean	-32.0	Mean	-29.6
Std Dev	26.21	Std Dev	22.85
N	18	N	18

5. [25 pts] There is some indication that consumption of moderate amounts of wine can help prevent heart attacks. Below is a plot and summary of yearly wine consumption (liters of alcohol from drinking wine, per person) and yearly death rates from heart disease (deaths per 100,000 people) for a sample of 8 European countries.

Country	1	2	3	4	5	6	7	8
Alcohol from wine, x	3.9	2.9	2.9	9.1	7.9	1.8	5.8	1.3
Heart disease death rate, y	167	131	220	71	107	167	115	286

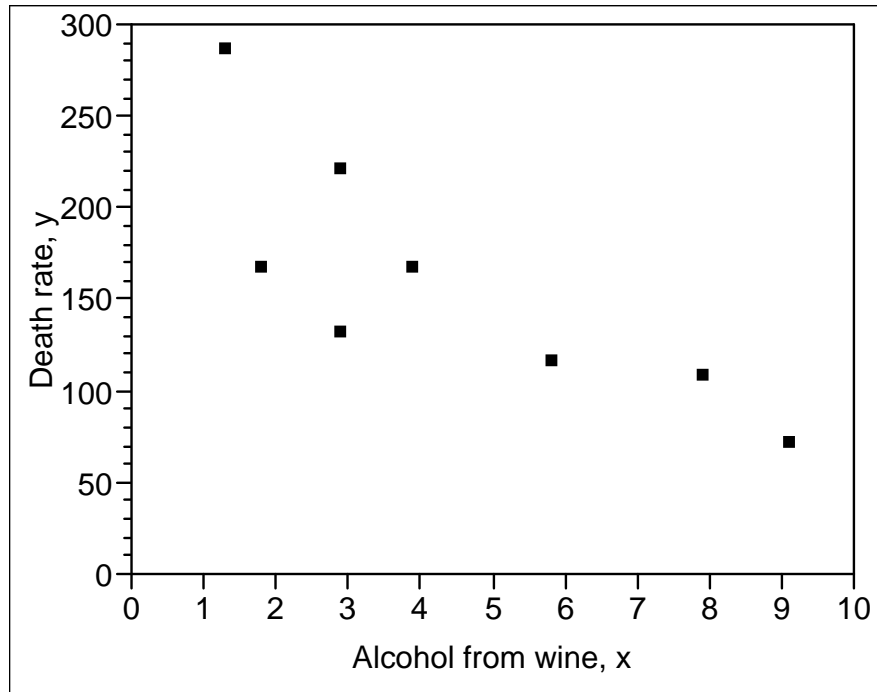
- a) [2] What is the population?

Below are summary statistics.

Alcohol from wine, x		Heart death rate, y			
Mean	4.45	Mean	158.0		
Std Dev	2.864	Std Dev	68.804	Correlation	-0.8156
SS(x)	57.4	SS(y)	33138.0	SS(xy)	-1124.8

- b) [4] Compute the least squares estimate of the slope, b_1 .
- c) [5] Give an interpretation of the slope within the context of the problem.
- d) [4] Compute the least squares estimate of the y-intercept, b_0 .
- e) [3] Give an interpretation of the y-intercept within the context of the problem.
- f) [2] Give the equation of the least squares regression line and use it to predict the death rate from heart disease for another European country, Spain, which has a wine consumption of 6.5 liters per person.
- g) [2] The actual death rate from heart disease in Spain is 86. What is the residual for Spain?

- h) [3] Graph the least squares regression line on the plot below. In order to get full credit it must be obvious to me that you are using the equation to draw the line.



Formulas

$$\bar{x} = \frac{\sum x}{n} \quad s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} \quad z_x = \frac{x - \bar{x}}{s_x}$$

$$\bar{y} = \frac{\sum y}{n} \quad s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}} \quad z_y = \frac{y - \bar{y}}{s_y}$$

$$r = \frac{\sum z_x z_y}{n-1} = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y}$$

$$b_1 = \frac{SS(xy)}{SS(x)} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad \text{or} \quad b_1 = r \frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$\hat{y} = b_0 + b_1 x \quad \text{residual} = y - \hat{y}$$