Moore, IPS 6e Chapter 02

Quizzes prepared by Dr. Patricia Humphrey, Georgia Southern University

A major study examined the relationship between cause of death (heart attack, cancer, stroke, accident, etc.) and age. A good way to graphically represent the relationship is with
- ○ a. side-by-side boxplots.
- ○ b. back-to-back stemplots.
- ○ c. a scatterplot.

A researcher examines data from all cities with populations over 100,000 in the United States. He notices that those cities that have a major league baseball team tend to have a greater number of divorces than other cities. One can reasonably conclude that
- ○ a. the presence of a major league baseball team contributes to divorce. Men spend time at the ballpark at the expense of their marriage.
- ○ b. this correlation cannot be explained and is probably accidental. Cities with major league baseball teams should have no more divorces than other cities.
- ○ c. none of the above.

A researcher measures a response variable Y and explanatory variable X on each of several objects. A scatterplot of the measurements is as follows.

The researcher notices that there is a distinct curved pattern in the plot. It would be appropriate to conclude
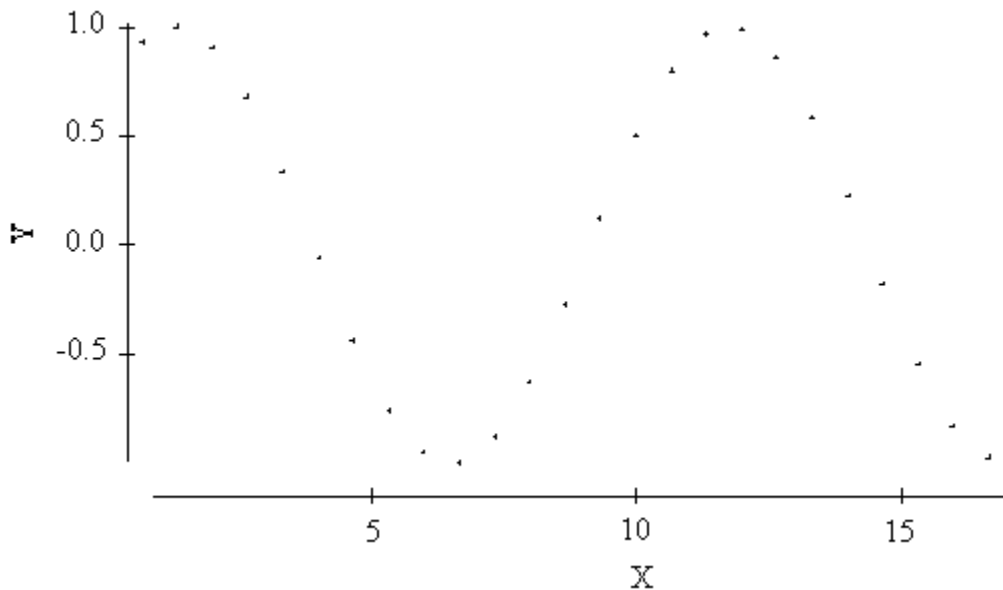
- ○ a. r is small.
- ○ b. r is approximately -2/3 since Y decreases as X increases in approximately 2/3 of the plot.
- ○ c. r is meaningless here.

For each of the 50 states and Washington, D.C., a number of statistics related to education are available for 1992. Among these are the *average* salary of all high school teachers and the percentage of high school seniors who took the SAT exam. The following is a scatterplot of the average teacher salary and the percentage of students taking the SAT exam for all 50 states and Washington, D.C. in 1992.
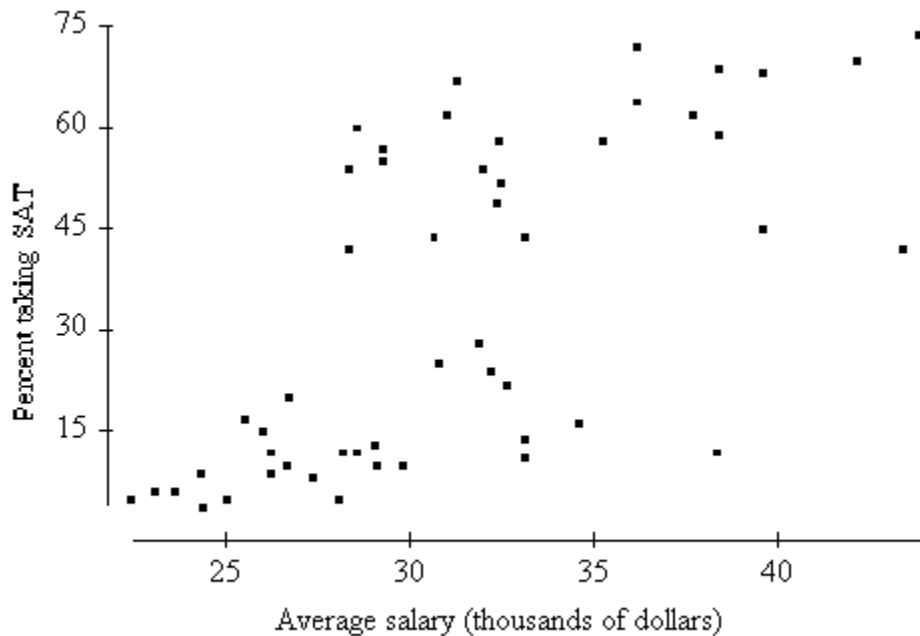
The value of the correlation for these data is 0.66. We may most reasonably conclude that

- ○ a. there is a surprisingly high association between the salaries of individual teachers and the percent of their students who took the SAT.
- ○ b. the plot suggests that there are two clusters of teachers-those teachers who tend to inspire their students to take the SAT, and those who don't.
- ○ c. none of the above.

Which of the following statements is consistent with the scatterplot?
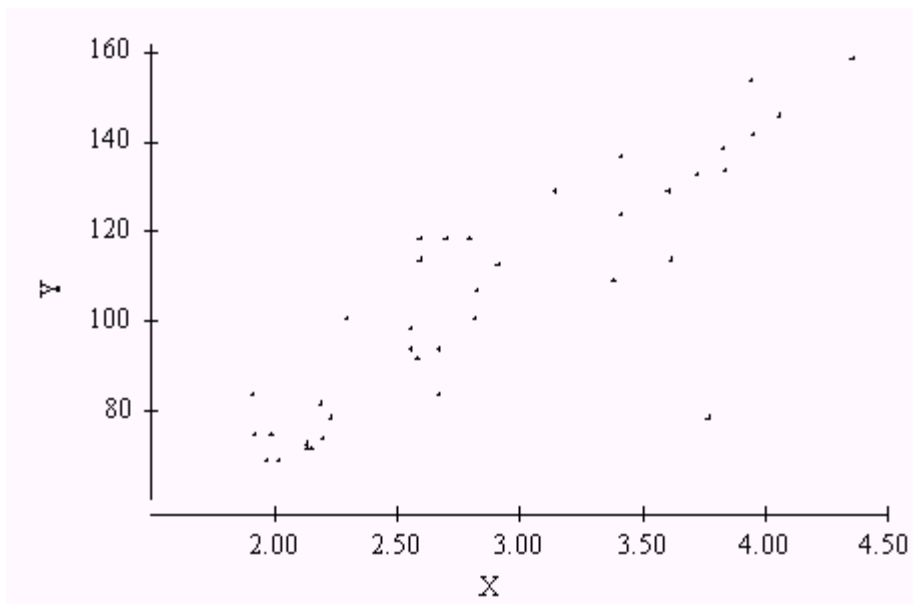
○ a. There is an outlier in the plot and, if the least-squares line were fit to the data, it would have a positive residual.

○ b. There is an outlier in the plot and, if the least-squares line were fit to the data, it would have a negative residual.

○ c. There is evidence that a lurking variable is present.

At a large department store, the amount a shopper spent and the shopper's gender (male or female) were recorded. To determine if gender is useful in explaining the amount of money a shopper spends at the store we could

○ a. make side-by-side boxplots of the distribution of the amount spent by males and the distribution of the amount spent by females.

○ b. compute the correlation between the amount spent and gender.

○ c. compute the least-squares regression line of amount spent on gender.

Below is a plot of the Olympic gold medal winning performance in the high jump (in inches) for the years 1900 to 1996.
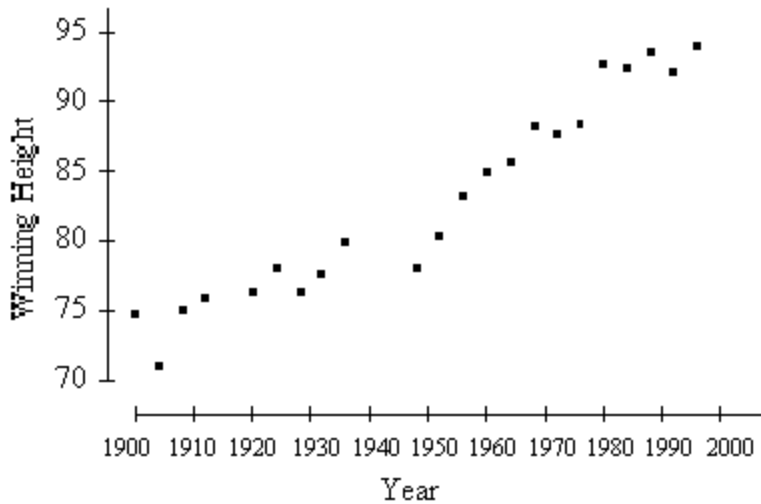
From this plot, the correlation between the winning height and year of the jump is
- ○ a. about 0.95.
- ○ b. about 0.10.
- ○ c. about -0.50.

Below is a plot of the Olympic gold medal winning performance in the high jump (in inches) for the years 1900 to 1996.
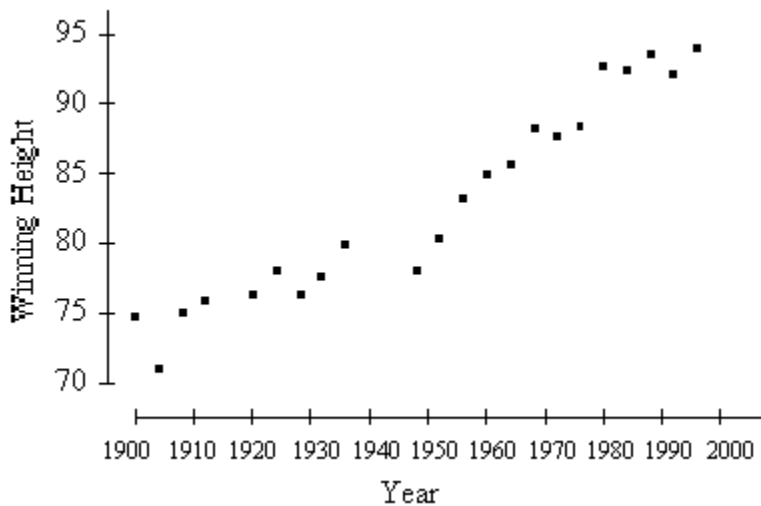
The equation of the least-squares regression line of Winning Height (in inches) on Year is

Winning Height = -364.90 + 0.23 Year

In another millennium (the year 3000) if the Olympics continue to be held, we can expect the Winning Height to be about
- ○ a. 325 inches.
- ○ b. 690 inches.

○ c. none of the above.

A survey of 1000 adults ages 30 to 35 is conducted. The number of years of schooling and the annual salary for each person in the survey is recorded. The correlation between years of schooling and annual salary is found to be 0.27. Suppose instead the average salary of all individuals in the survey with the same number of years of schooling was calculated and the correlation between these averages and years of schooling was computed. This correlation would most likely be

○ a. equal to 0.27.
○ b. larger than 0.27.
○ c. less than 0.27.

The regression line to predict average exam grade from hours of study is y = 15 + 5.6*x. The slope of the regression line indicates

○ a. for any student, an extra hour of study increases the grade 5.6 points.
○ b. on average, an extra hour of study will increase the grade 5.6 points.
○ c. an extra hour of study will increase the grade 15 points.

A business has two types of employees: managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The number of male and female managers at each salary level and the number of male and female workers at each salary level are given in the two tables below.

	Managers	
	Male	Female
$100,000	80	20
$200,000	20	30

	Workers	
	Male	Female
$10,000	30	20
$20,000	20	80

We may conclude

○ a. the mean salary of female managers is $40,000 greater than that of male managers.
○ b. there are more females employed by this company than males, which will lower their salaries.
○ c. the mean salary of female managers is greater than that of male managers and the mean salary of female workers is higher than that of male workers, indicating the mean salary for female employees is higher than the mean salary for male employees.

A researcher wants to study the effect of regular exercise on cholesterol levels. The researcher compares the cholesterol levels of 50 people who belong to a local gym and exercise regularly with

the cholesterol levels of 50 people from the community who did not exercise regularly. The cholesterol levels of the members of the gym were substantially lower. The researcher can conclude that

- ○ a. belonging to a gym reduces cholesterol level.
- ○ b. exercising regularly reduces cholesterol level.
- ○ c. members of a local gym who exercised regularly had lower cholesterol levels than those in the community who did not exercise regularly.


All but one of the following statements contain a blunder. Which one does not contain a blunder?

- ○ a. "There is a correlation of $r = 0.54$ between the position a football player plays and his or her weight."
- ○ b. "The correlation between amount of fertilizer and yield of tomatoes was found to be $r = 0.33$."
- ○ c. "The correlation between the gas mileage of a car and its weight is $r = 0.71$ gallon-pounds."

A review of voter registration records in a small town yielded the following table of the number of males and females registered as Democrat, Republican, or some other affiliation.

	Male	Female
Democrat	300	600
Republican	500	300
Other	200	100

Which of the following bar graphs represents the distribution of Democrats, Republicans, and other affiliations in this town?

○ a.

○ b.

c.
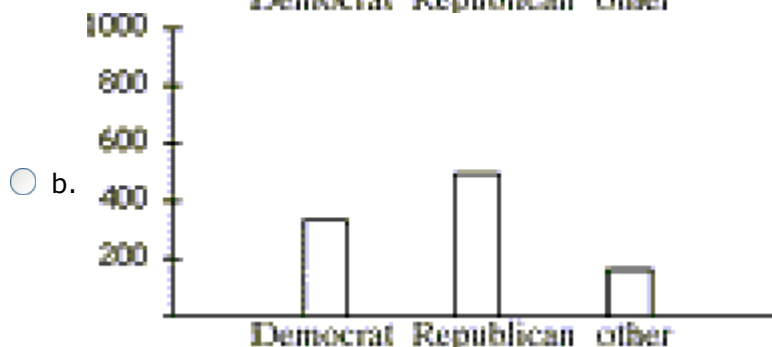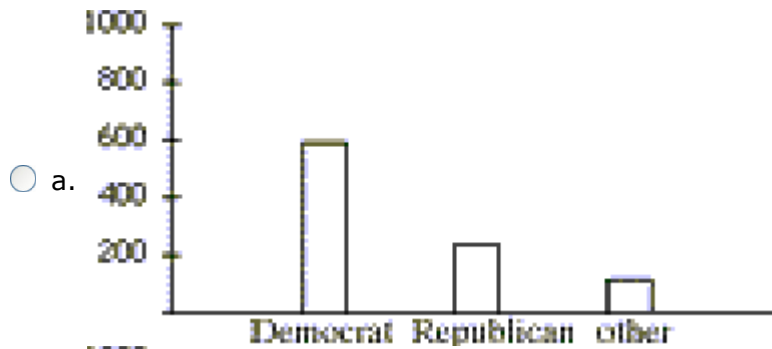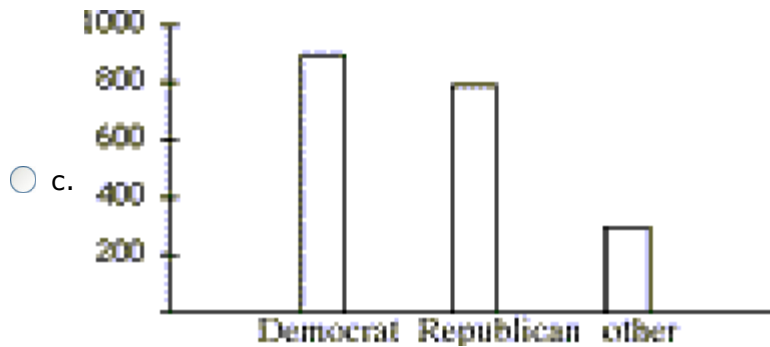
Foresters use regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let y be the volume of timber in cubic feet and x be the diameter in feet (measured at 3 feet above ground level). One set of data gives

$y = -30 + 60x$

The predicted volume for a tree of 18 inches is
- a. 1050 cubic feet.
- b. 90 cubic feet.
- c. 60 cubic feet.

I wish to determine the correlation between the height (in inches) and weight (in pounds) of 21-year-old males. To do this I measure the height and weight of two 21-year-old men. The measured values are

	Male #1	Male #2
Height	70	69
Weight	160	164

The correlation r computed from the measurements on these males is
- a. 1.0.
- b. -1.0.
- c. near 0, since the heights and weights of the men are similar.

In the graph below, the circled point is
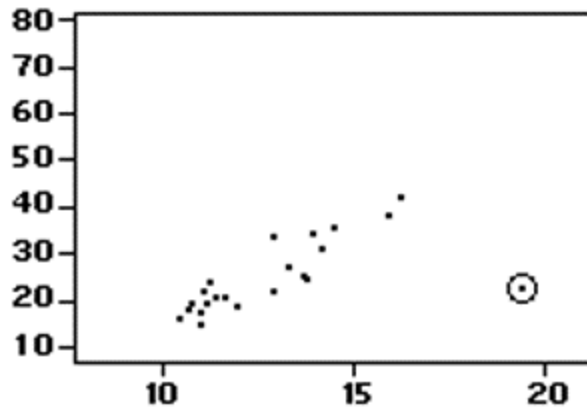
○ a. an influential point. Deleting it should reduce the correlation and improve the fit.
○ b. not an influential point since its y value is not unusually large or small.
○ c. an influential point. Deleting it should increase the estimate of the slope.

The data in the scatterplot below are an individual's weight and the time it takes (in seconds) on a treadmill to raise his or her pulse rate to 140 beats per minute. The o's correspond to females and the +'s to males.
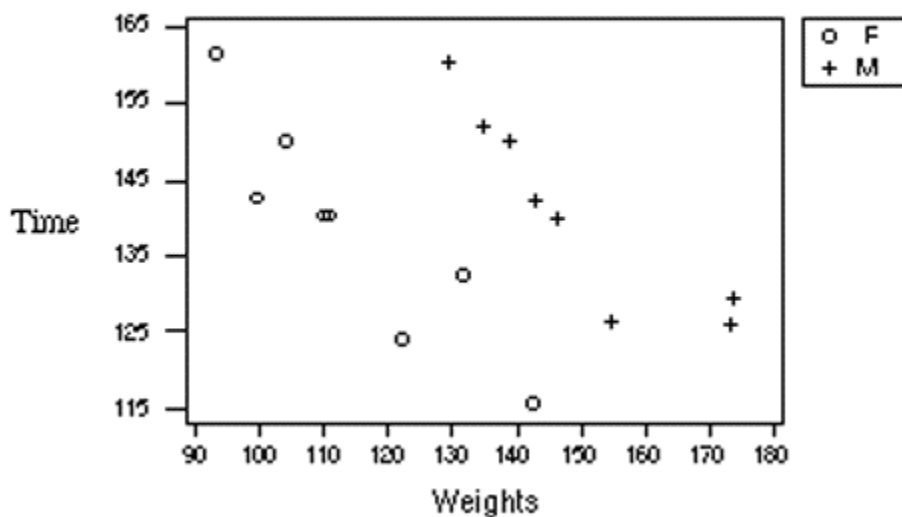
Based on the scatterplot we conclude
○ a. there is a positive correlation r between gender and weight, since men tend to weigh more.
○ b. there is a negative correlation between time and weight for males and for females.
○ c. both (a) and (b) are correct.

The world record time (in minutes) in the marathon is plotted versus the year in which the record was set for men and women separately. The plotting symbol o is used for men and x for women. The data include only records set between 1908 and 1988.

Which of the following statements would be a valid conclusion from this plot?
○ a. The world record times for women show a greater rate of improvement (decrease more rapidly) than the world record times for men.
○ b. We can expect the world record time for women to be lower than that for men sometime before the year 2010.
○ c. By the year 2010, the world record time for men will reach a plateau beyond which no improvement will be possible.

Which of the following is true of the slope of the least-squares regression line?
○ a. It has the same sign as the correlation.
○ b. The square of the slope equals the fraction of the variation in the response variable that is explained by the explanatory variable.
○ c. It is unitless.

Which of the following measures only the strength of a relationship?
○ a. The Correlation coefficient.
○ b. The slope.
○ c. The Coefficient of Determination.

A student has obtained the following computer output from a regression examining the relationship between BTUs input to a furnace and the BTU output.

Predictor	Coef	SE Coef	T	P
Constant	0.1868	0.2153	0.87	0.388
BTU.Out	0.91109	0.01915	47.57	0.000

S = 0.5580   R-Sq = 96.3%

We can conclude from this information
- ○ a. The regression is useful because R-squared is large.
- ○ b. The regression model is not useful because the slope is small.
- ○ c. We don't have enough information to conclude whether or not the regression is useful.

A student has obtained the following computer output from a regression examining the relationship between BTUs input to a furnace and the BTU output.

Predictor	Coef	SE Coef	T	P
Constant	0.1868	0.2153	0.87	0.388
BTU.Out	0.91109	0.01915	47.57	0.000

S = 0.5580   R-Sq = 96.3%

From this we conclude
- ○ a. for every BTU input to a furnace, the output increases by 0.911, on average.
- ○ b. for every 0.187 BTU input to a furnace, the output increases by 0.911, on average.
- ○ c. for every 0.911 BTU input to a furnace, the output increases by 1, on average.

Which of the following does not have an error?
- ○ a. The correlation between students' majors and their GPA is 0.67.
- ○ b. The correlation between students' IQ scores and their foot size is 0.67.
- ○ c. The correlation between students' incomes and their GPA is 1.12.

Statistician William Hammack examined the relationship between the number of public schools in each county of Florida and the crime rate for the county. The data showed a very strong linear relationship with r = 0.970. From this we conclude
- ○ a. public schools cause crime.
- ○ b. there is probably a lurking variable at work.
- ○ c. this is totally meaningless.

A researcher is investigating the relationship between the diameter of a tree and the usable wood volume. He obtains the following plot of the data and regression output.
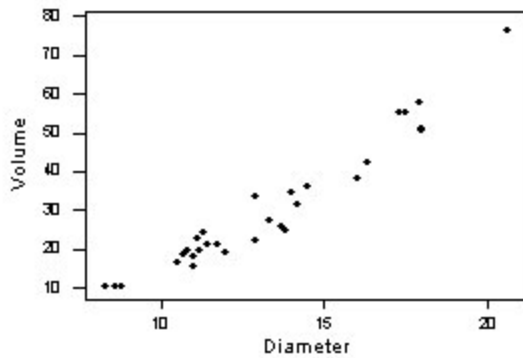
○ a. The point at the upper right of the plot may be influential. We should recalculate the regression without it.

○ b. If a residual plot is random, we will be satisfied with the model.

○ c. First determine if the point at the upper right is influential; if not and a residual plot is random, we will be satisfied with the model.

I want to examine whether or not there is a relationship between a student's grade point average and after-college plans. For a visual display of the data I should choose

○ a. a scatterplot.

○ b. side-by-side boxplots.

○ c. a back-to-back stemplot.

I want to examine the relationship between gas mileage of cars and the engine size (displacement in cubic inches). The explanatory variable is

○ a. engine size.

○ b. gas mileage.

○ c. I could use either one.

The ABC Company has been evaluating the performance of two advertising agencies it deals with. They produce the following scatterplot of sales against advertising expenditures. The two agencies are indicated on the plot with different symbols.

Scatterplot of Sales vs Advertis
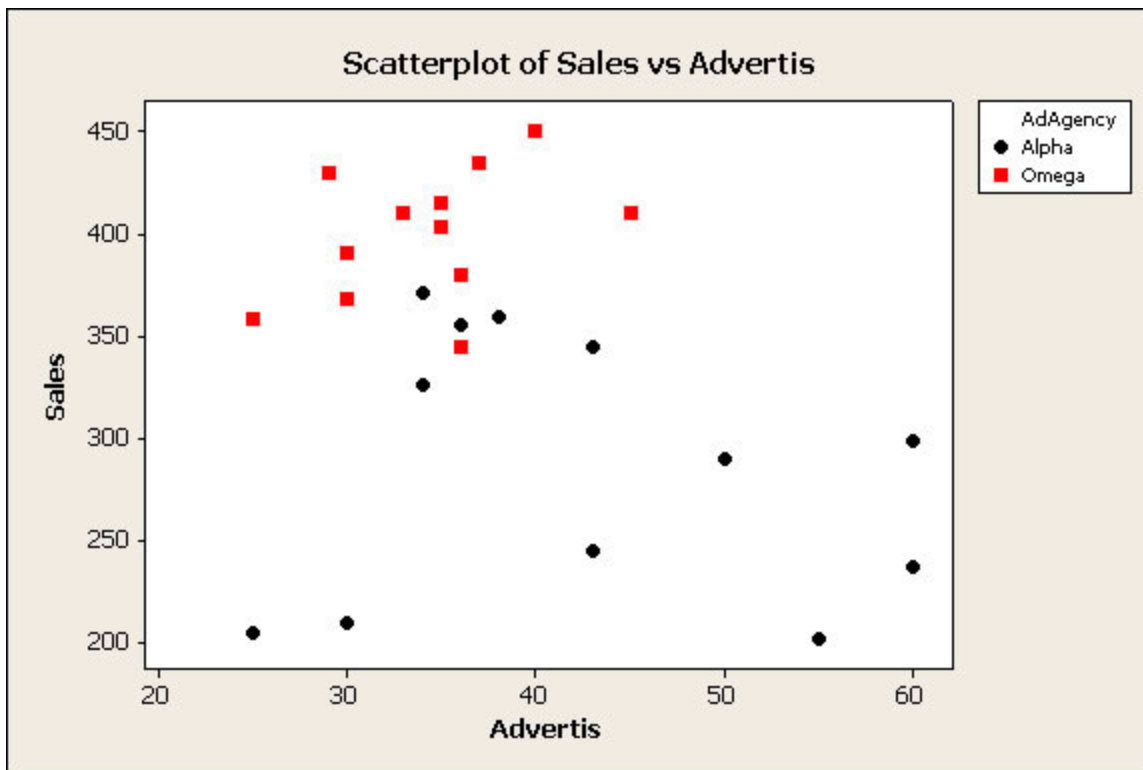
From the plot, the ABC Company should decide
- ○ a. sales don't depend on the advertising company.
- ○ b. sales don't depend on the amount of advertising.
- ○ c. the Omega Company is better than Alpha-sales are generally higher even with less expenditure on advertising.

A recent article in an educational research journal reports a correlation of +0.8 between math achievement and overall math aptitude. It also reports a correlation of -0.8 between math achievement and a math anxiety test. Which of the following interpretations is the most correct?
- ○ a. The correlation of +0.8 indicates a stronger a relationship than the correlation of -0.8.
- ○ b. The correlation of +0.8 is just as strong as the correlation of -0.8.
- ○ c. It is impossible to tell which correlation is stronger.

If females of a certain species of lizard always mate with males that are .75 years younger than they are, what would the correlation between the ages of these male and female lizards be?
- ○ a. 1
- ○ b. 0.75
- ○ c. -1

Suppose you want to know how salaries (in thousands of dollars per year) are related to years of experience for your chosen career. You collect data for a random sample of 100 people with this type of job who have had from 0 to 20 years of experience. The relationship is linear and you determine the regression equation to be: salary = 45.59 + .798 years. Predict the salary for an

individual who has been working 15 years.
- ○ a. $57.56
- ○ b. $57,560
- ○ c. $45,590

Suppose you want to know how salaries (in thousands of dollars per year) are related to years of experience for your chosen career. You collect data for a random sample of 100 people with this type of job who have had from 0 to 20 years of experience. The relationship is linear and you determine the regression equation to be: salary = 45.59 + .798 years. For this equation, the number 45.59 indicates
- ○ a. the y-intercept is meaningless because of extrapolation.
- ○ b. an individual with no experience in this career will earn $45.59 per hour.
- ○ c. an individual with no experience in this career will earn $45,590 per year.

A group of 7th grade teachers want to use IQ scores to predict GPA. Data on 78 students in this 7th grade class give us the following statistics regarding these two variables. (IQ is on a scale of 0-140; GPA is on a scale of 1-12.)

IQ	GPA
$\bar{x} = 108.9$	$\bar{x} = 7.447$
$s = 13.17$	$s = 2.10$
$r = 0.633$	

Find the equation of the least-squares line for this data.
- ○ a. GPA = -3.44 + 0.1*IQ
- ○ b. GPA = 79.34 + 3.97*IQ
- ○ c. IQ = 79.34 + 3.97*GPA

A student group investigates the relationship between IQ and GPA, measured on a 12-point scale. They find the equation to be GPA = -6 + 0.15*IQ. Along comes Marilyn Vos Savant with an IQ of 200. What does this regression say her GPA should be?
- ○ a. 24
- ○ b. Using the line is meaningless for her.
- ○ c. 12

Do companies that invest in education and training perform better in the stock market than those that do not? Companies were characterized as to whether or not they invested in education or training for their employees. Of those characterized as investing in education and training, a sample of 100 was selected. Of those characterized as not investing in education and training, a sample of 200 was selected. The performance of each of the 300 firms selected was rated as above average, average, or below average compared to the stock market as a whole. The results are presented in the following table.

	Market Performance		
Invest in Education/Training	Above Avg.	Avg.	Below Avg.
Yes	35	44	21
No	51	87	62

The marginal distribution of the Market Performance of all firms in the study is

a.
Market Performance		
Above Avg.	Avg.	Below Avg.
86	131	83

b.
Market Performance		
Above Avg.	Avg.	Below Avg.
0.407	0.336	0.253

c.
Market Performance		
Above Avg.	Avg.	Below Avg.
0.287	0.437	0.277

A study was conducted to explore the relationship between alcohol consumption and hypertension. 7400 nurses were surveyed and information about the amount of alcohol they consumed per day (on average) and whether or not they suffered from hypertension was collected. The following table summarizes the information that was obtained.

	Avg. Number of Alcoholic Drinks Per Day			
Hypertension	0	0.01 – 1.00	1.01 – 2.00	> 2.00
Yes	175	200	15	10
No	2600	4000	280	120

Based on the information in the table, we may conclude
- a. the more a nurse drank, on average, the less likely the nurse was to have hypertension.
- b. moderate drinkers (those drinking 0.01 - 1.00 glasses per day on average) had the lowest proportion of cases of hypertension.
- c. nurses who drank no alcohol were less likely to have hypertension than nurses who drank alcohol.

In the table below we examine the relationship between final grade and the reported hours per week each student said they studied for the course.

	A	B	C	D	F	Total
<5	2	3	7	4	5	21
5 – 10	3	5	5	4	6	23
>10	6	5	3	4	2	20
Total	11	13	15	12	13	64

The probability a student earned an A for the course is
- a. 11/64.

○ b. 2/21.
○ c. 21/64.

In the table below we examine the relationship between final grade and the reported hours per week each student said they studied for the course.

	A	B	C	D	F	Total
<5	2	3	7	4	5	21
5 – 10	3	5	5	4	6	23
>10	6	5	3	4	2	20
Total	11	13	15	12	13	64

The probability that a student earned an A if they studied more than 10 hours per week is
○ a. 11/64.
○ b. 6/20.
○ c. 20/64.

In the table below we examine the relationship between final grade and the reported hours per week each student said they studied for the course.

	A	B	C	D	F	Total
<5	2	3	7	4	5	21
5 – 10	3	5	5	4	6	23
>10	6	5	3	4	2	20
Total	11	13	15	12	13	64

According to the table, the probability that a random student both earned a C and studied between 5 and 10 hours per week is
○ a. 15/64.
○ b. 5/64.
○ c. 23/64.

A large company has been sued for sex discrimination. The case brought by the female managers said they were underrepresented in management. However, further analysis of the company by division found that females were actually more likely than males to be managers in each division. This is an example of
○ a. aggregating data.
○ b. Simpson's Paradox.
○ c. biased data collection.

Colleges conduct student evaluations of instruction to determine if there are patterns of either excellence or the lack thereof by their instructors. These can have a large impact on tenure and promotion decisions at many institutions. One particular instructor had ratings in 3 courses in one semester as given by the table below (5 is agree strongly to 1 is disagree strongly) for the statement "The instructor is an excellent teacher."

	1	2	3	4	5	Total
Class A	3	4	8	19	10	44
Class B	8	6	3	10	8	35
Class C	2	4	4	8	10	28
Totals	13	14	15	37	28	107

The chance a student in Class B was neutral on the professor is
- a. 3/35.
- b. 3/15.
- c. 3/107.

A study was performed to examine the personal goals of children in grades 4, 5, and 6. A random sample of students was selected for each of the grades from schools in Georgia. The students received a questionnaire regarding personal goals. They were asked what they would most like to do at school: make good grades, be popular, or be good at sports. Results are presented in the table below by the sex of the child.

	Make good grades	Be popular	Be good in sports
Boys	96	32	94
Girls	295	45	40

The proportion of boys who chose the goal "be good in sports" and the proportion of girls who chose the goal "be good in sports" are
- a. proportion of boys = 0.42, proportion of girls = 0.11.
- b. proportion of boys = 0.70, proportion of girls = 0.30.
- c. proportion of boys = 0.16, proportion of girls = 0.07.

A study was conducted to assess the effect of the drug Imipramine on the treatment of patients with agoraphobia and panic disorder. The three treatment groups included a control group treated with placebo, a low dose of 1.5 mg per kilogram of body weight, and a high dose of 3 milligrams per kilogram of body weight. There were 70 subjects, with 30 assigned at random to the control and 20 to each of the treatment groups. At the end of 24 weeks of treatment, subjects were classified as responders or nonresponders to treatment based on a battery of psychological tests.

	Responder	
	Yes	No
Control	6	24
1.5 mg/kg	9	11
3 mg/kg	9	11

The proportion of responders in the experiment is

- a. 0.3428.
- b. 0.3667.
- c. 0.4500.

Submit

Perception licensed to Bedford, Freeman & Worth Publishing Group