

TEACHER

# STAT 113

## Work Sheet 10: Chapter 24

- Here are counts from a study of all firearm-related deaths in Wisconsin between 2000 and 2002 for children and youth under the age of 25 where the type of firearm used was known. We want to compare the types of firearms used in homicides and in suicides.

	Handgun	Long gun	
Homicides	106	17	123
Suicides	59	166	225
	165	183	348

- What percent of deaths were suicides and used handguns?

$$\frac{59}{348} = 16.95\%$$

- What percent of homicides used long guns?

$$\frac{17}{123} = 13.8\%$$

- Long guns were used in what percent of deaths?

$$\frac{183}{348} = 52.59\%$$

- Here is a table that breaks down Purdue's 1697 faculty members by both gender and academic rank:

	Female	Male	
Assistant professors	151	254	405
Associate professors	154	397	551
Professors	99	642	741
	404	1293	1697

- What percent of the faculty members are female?

$$\frac{404}{1697} = 23.81\%$$

- Among male faculty members, what is the percent who are professors?

1293

$$\frac{642}{1293} = 49.65\%$$

- What percent of the faculty members are female associate professors?

$$\frac{154}{1697} = 9.07\%$$

- What percent of female faculty members were associate professors or professors?

404

$$\frac{253}{404} = 62.6\%$$

3. How is the hatching of water python eggs influenced by the temperature of the snake's nest? Researchers assigned newly laid eggs to one of three temperature: hot, neutral, or cold. Hot duplicated the extra warmth provided by the mother python, and cold duplicated the absence of the mother. Here are the data on the number of eggs and the number that hatched:

	Number of eggs	Hatched
Cold	27	16 $\frac{16}{27} = 12.4\%$
Neutral	56	38 $\frac{38}{56} = 29.4\%$
Hot	104	75 $\frac{75}{104} = 58.14\%$
	187	129

- (a) Make a two-way table of temperature by outcome (hatched or not)

	Not hatched	Hatched
Cold	11	16
Neutral	18	38
Hot	29	75

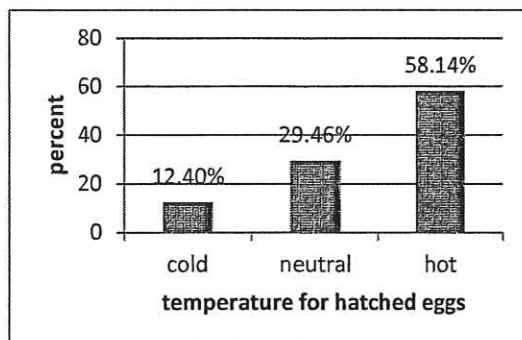
- (b) What percent of newly laid eggs were assigned to cold?

$$\frac{27}{187} = 14.4\%$$

- (c) What percent of newly laid eggs were assigned to neutral and hatched?

$$\frac{38}{187} = 20.3\%$$

- (d) Find the conditional distribution of temperature for hatched eggs and draw a labeled bar graph to display this distribution.



$$\frac{16}{129} = 12.4$$

$$\frac{38}{129} = 29.4$$

$$\frac{75}{129} = 58.14$$

4. Below are the numbers of flights on time and delayed for Alaska Airlines at five airports in one month.

	On time	Delayed
Los Angeles	497	62
Phoenix	221	12
San Diego	212	20
San Francisco	503	102
Seattle	1841	305

3214

501

559

233

3775

Among the flights to Los Angeles, what percent of flights were delayed?

$$\frac{62}{559} = 11.1\%$$

- (a) What percent of delayed flights were **not** flying to Los Angeles?

$$\frac{439}{501} = 87.6\%$$

- (b) What percent of flights were on time?

$$\frac{3214}{3775} = 86.7\%$$

- (c) What percent of flights were on time and the destination is Phoenix or San Diego?

$$\frac{221 + 212}{3775} = \frac{433}{3775} = 11.4\%$$



A researcher suspected a relationship between people's preference in movies and preference in pizza. A random sample of 100 people produced the following two-way table:

Favorite Movie	Ground			
	Pepperoni	beef	Mushrooms	
Jurassic Park	20	5	10	35
Lethal Weapon (I)	8	15	12	35
Gone with the Wind	15	2	13	30
	43	22	35	100

1. Calculate the marginal distributions in percents for each variable.

JP 35%

Pepp 43%

LW 35%

GB = 22%

GW 30%

Mush = 35%

2. Compute (in percents) the conditional distribution of favorite movie among those who prefer ground beef topping. Show the distribution in a table.

JP  $\frac{5}{22} \approx 22.7\%$

LW  $\frac{15}{22} \approx 68.2\%$

GW  $\frac{2}{22} \approx 9.1\%$

3. Briefly describe your findings in words.

Among people who prefer Ground Beef as pizza topping, Lethal Weapon was most popular at 68%, compared to Jurassic Park at 23% and Gone w/wind at 9.1%.

4. Compute (in percents) the conditional distribution of favorite pizza topping among those who chose *Gone with the Wind* as their favorite movie. Show the distribution in a table.

Pepp  $\frac{15}{30} = 50\%$

ground beef  $\frac{2}{30} = 6.7\%$

mush  $\frac{13}{30} = 43.3\%$

5. Briefly describe your findings in words.

Among people whose favorite movie is *Gone w/wind* half prefer pepperoni in pizza, followed by mushrooms + then ground Beef at 6.7%



In 2006, an electronic replay system debuted in both men's and women's professional tennis. Each player is allowed two unsuccessful challenges per set. Here are some data on the results of challenges made during the first few months of the new system.

	Successful	Unsuccessful	
Men	201	288	489
Women	126	224	350
	327	512	839

1. Calculate the marginal distribution (in percents) for each of the two variables.

$$\text{men } \frac{489}{839} = 58.3\% \quad \text{succ. } \frac{327}{839} = 39.97\%$$

$$\text{wom } \frac{350}{839} = 41.7\% \quad \text{unsucc. } \frac{512}{839} = 61.03\%$$

2. Write a sentence describing what each marginal distribution tells you.

more men made challenges

most challenges unsuccessful

3. Calculate the two conditional distributions of challenge results, given gender. Display your results in a table.

	Successful	Unsucc.
Men	$\frac{201}{489} = 41.1\%$	$\frac{288}{489} = 58.9$
women	$\frac{126}{350} = 36\%$	$\frac{224}{350} = 64\%$

4. Write a few sentences describing the relationship between gender and challenge results.

men make more challenges. They are more successful than women

**Smoking and Education**

A

200 adults shopping at a supermarket were asked about the highest level of education they had completed and whether or not they smoke cigarettes. Results are summarized in the table.

	Smoker	Non-smoker	Total
High school	32	61	93
2 yr college	5	17	22
4+ yr college	13	72	85
Total	50	150	200

1. Discuss the W's.

2. Identify the variables.

3. a) What percent of the shoppers were smokers with only high school educations? \_\_\_\_\_  
 b) What percent of the shoppers with only high school educations were smokers? \_\_\_\_\_  
 c) What percent of the smokers had only high school educations? \_\_\_\_\_

$$\frac{32}{200} = 16\%$$

$$\frac{32}{93} = 34.4\%$$

$$\frac{32}{50} = 64\%$$

B

Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from "Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes" (Shanthi A. Bowman, *Journal of the American Dietetic Association*, 102(9), pp. 1234-1239):

		Nationwide Food Survey Years			Total
		1987-1988	1989-1991	1994-1996	
Drinks Fluid Milk	Yes	354	502	366	1222
	No	226	335	366	927
	Total	580	837	732	2149

1. Find the following:
- What percent of the young girls reported that they drink milk? \_\_\_\_\_
  - What percent of the young girls were in the 1989-1991 survey? \_\_\_\_\_
  - What percent of the young girls who reported that they drink milk were in the 1989-1991 survey? \_\_\_\_\_
  - What percent of the young girls in 1989-1991 reported that they drink milk? \_\_\_\_\_
2. What is the marginal distribution of milk consumption?

$$\frac{1222}{2149} = 56.9\%$$

$$\frac{837}{2149} = 38.9\%$$

$$\frac{502}{1222} = 41.1\%$$

$$\frac{502}{837} = 59.9\%$$

To determine if people's preference in dogs had changed in the recent years, organizers of a local dog show asked people who attended the show to indicate which breed was their favorite. This information was compiled by dog breed and gender of the people who responded. The table summarizes the responses.

1. Identify the variables and tell whether each is categorical or quantitative.

	Female	Male	Total
Yorkshire Terrier	73	59	132
Dachshund	49	47	96
Golden Retriever	58	33	91
Labrador	37	41	78
Dalmatian	45	28	73
Other breeds	86	67	153
Total	348	275	623

2. Which of the W's are unknown for these data?

3. Find each percent.

- What percent of the responses were from males who favor Labradors?
- What percent of the male responses favor Labradors?
- What percent of the people who choose Labradors were males?

$$41/623 = 6.5\%$$

$$41/275 = 14.9\%$$

$$41/78 = 52.5\%$$

4. What is the marginal distribution of breeds?

$$\frac{132}{623}$$

$$\frac{96}{623}$$

$$\frac{91}{623}$$



Name \_\_\_\_\_

Date \_\_\_\_\_

## Worksheet – Marginal and Conditional distributions

Following are statistics from exit polls from the 2004 presidential election.

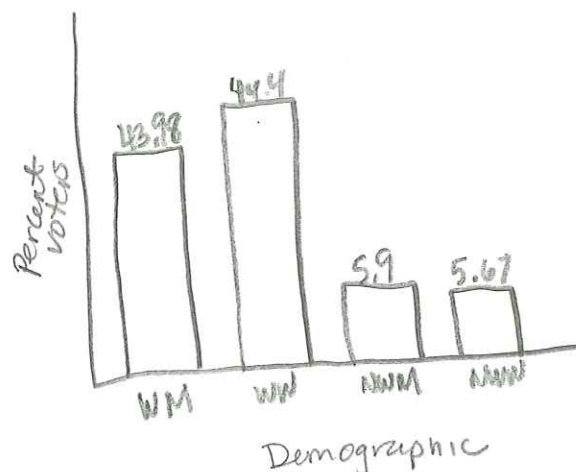
	Bush	Kerry	
White Men	3049	1820	4869
White Women	3080	2464	5544
Non White Men	410	915	1325
Non-White Women	393	1229	1622
	6932	4608	11540

1. Complete the chart by finding the total values.
2. Determine the marginal distribution of votes for each candidate.

3. Determine the conditional distribution of gender categories among Bush voters.

WM	$\frac{3049}{6932} = 43.98\%$
WW	$\frac{3080}{6932} = 44.4\%$
NWM	$\frac{410}{6932} = 5.9\%$
NWW	$\frac{393}{6932} = 5.67\%$

4. Create a bar graph to represent #3 and summarize your findings about Bush voters.



	Bush	Kerry	
No High School	268	273	541
HS Grad	1563	1412	2975
Some college	2360	2011	4371
College Grad	1847	1634	3481
Post Grad	962	1202	2164
	7000	6532	13532

5. Complete the chart by finding the total values.

6. Find the marginal distribution of education.

7. Determine the conditional distribution of education among Kerry voters.

$$\begin{aligned}
 \text{NHS} & \quad \frac{273}{6532} = 4.19\% \\
 \text{HS} & \quad \frac{1412}{6532} = 21.6\% \\
 \text{Some} & \quad \frac{2011}{6532} = 30.79\% \\
 \text{Coll} & \quad \frac{1634}{6532} = 25.01\% \\
 \text{Post} & \quad \frac{1202}{6532} = 18.4\%
 \end{aligned}$$

8. Determine the conditional distribution of education among Bush voters.

## 4.3 Worksheet

Name: \_\_\_\_\_

**PROBLEM 1**

In a 1980 study, researchers looked at the relationship between the type of college (public or private) attended by 3265 members of the class of 1960 who went into industry and the level of job each member had in 1980. The results were:

<b>Management Level</b>	<b>Public</b>	<b>Private</b>
High	75	107
Middle	962	794
Low	732	595

1. Compute the marginal counts.
2. Compute the conditional distributions of management level given college type (in percents).  
[Write the numbers next to the counts in the above table.]

(Graph not available for transfer.)

3. A **segmented bar graph** allows a graphical comparison of distributions. Each bar describes one group, and the bar is divided into segments to show the distribution for that group. Each bar has height 100%. Show the conditional distributions from (2) as a segmented bar graph. Be sure to label both axes and provide a key to identify the segments. One group has been done for you.
4. Comment on the observed relationship.

## 4.3 Worksheet

Name: \_\_\_\_\_

**PROBLEM 2**

A researcher suspected a relationship between people's preferences in movies and preference in pizza. A random sample of 100 people produced the following two-way table:

Favorite Movie		Ground		
		Pepperoni	Beef	Mushrooms
Jurrasic Park		20	5	10
Lethal Weapon (I)		8	15	12
Gone With the Wind		15	2	13

1. Enter the overall (marginal) distributions on the table.
2. Compute (in percents) the conditional distribution of favorite movie among those who prefer ground beef topping. Show the distribution in a table.

SKIP

3. Briefly describe your finding in words.
4. Compute (in percents) the conditional distribution of favorite pizza topping among those who chose Gone With the Wind as their favorite movie. Show the distribution in a table.

5. Briefly describe your finding in words.



## Quiz 4.3C

## AP Statistics

Name: \_\_\_\_\_

Two candy factories, A and B, produce 3 flavors of gum: spearmint, cinnamon, and wintergreen. Here is a two-way table of factory (x) by gum flavor (y), with each entry in the table representing the number of packs of gum of a certain flavor produced during a given day.

	Spearmint	Cinnamon	Wintergreen	
A	250	150	50	450
B	175	180	25	380
	425	330	75	830

1. Compute the marginal distribution of gum flavor (expressing your results in percents). Work in the space below, or write in the table.

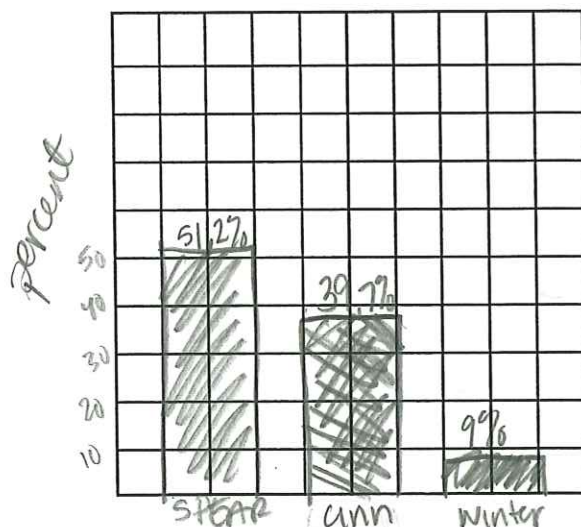
Spearmint  $425/830 = 51.2\%$

Cinn  $330/830 = 39.7\%$

Wintergreen  $75/830 = 9\%$  B:  $380/830 = 45.7\%$

A:  $450/830 = 54.2\%$

2. Construct a bar graph to show your results.



3. Compute the conditional distribution of gum flavor given factory. Show the distribution in a table.

$250/450 = 55.6\%$   $150/450 = 33.3\%$   $50/450 = 11.1\%$

$175/380 = 46.1\%$   $180/380 = 47.4\%$   $25/380 = 6.5\%$



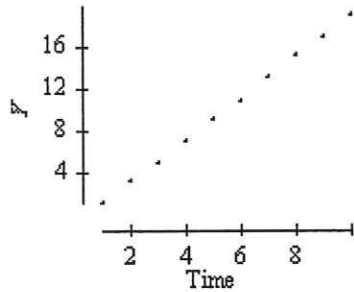


## Chapter 4: More about Relationships between Two-Variables

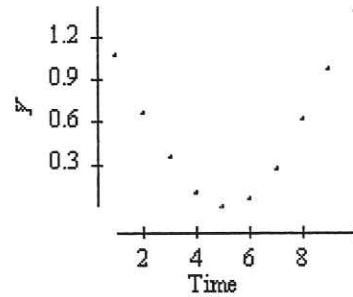
5. Suppose we measure a response variable  $Y$  at each of several times. A scatterplot of  $\log Y$  versus time of measurement looks approximately like a positively sloping straight line. We may conclude that
- A) the correlation between time of measurement and  $Y$  is negative, since logarithms of positive fractions (such as correlations) are negative.
  - B) the rate of growth of  $Y$  is positive, but slowing down over time.
  - C) a logarithmic growth model would approximately describe the relationship between  $Y$  and the time of measurement.
  - D) a mistake has been made. It would have been better to plot  $Y$  versus the logarithm of the time of measurement.
  - ☒ E) an exponential growth model would approximately describe the relationship between  $Y$  and time of measurement.
6. Using least-squares regression, I determine that the (base 10) logarithm of the population of a country is approximately described by the equation
- $$[\hat{y}] \log(\text{population}) = -13.5 + 0.01 \times (\text{year})$$
- Based on this equation, the population of the country in the year 2010 should be about
- A) 6.6.
  - B) 735.
  - C) 2,000,000.
  - ☒ D) 3,981,072.
  - E) 33,000,000.
7. Which of the following would provide evidence that a power law model describes the relationship between a response variable  $y$  and an explanatory variable  $x$ ?
- A) A scatterplot of  $y$  versus  $x$  looks approximately linear.
  - B) A scatterplot of  $\log y$  versus  $x$  looks approximately linear.
  - C) A scatterplot of  $y$  versus  $\log x$  looks approximately linear.
  - ☒ D) A scatterplot of  $\log y$  versus  $\log x$  looks approximately linear.
  - E) A scatterplot of the square root of  $y$  versus  $x$  looks approximately linear.

8. Which of the following scatterplots would indicate that  $Y$  is growing exponentially over time?

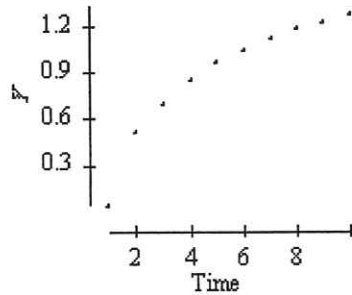
A)



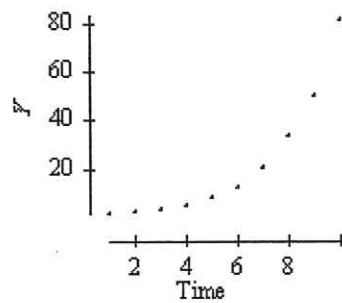
C)



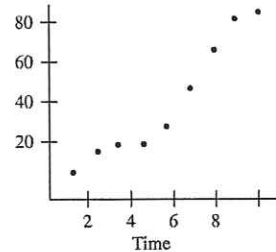
B)



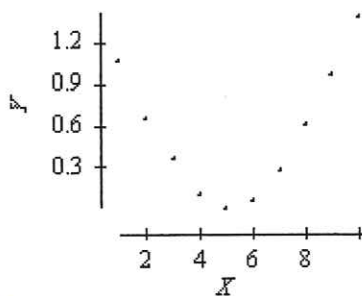
D)



E)



9. A scatterplot of a response variable  $Y$  versus an explanatory variable  $X$  is given below.

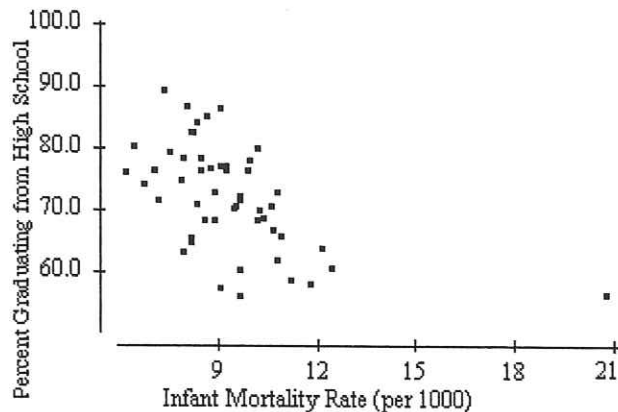


Which of the following is true?

- A) There is a nonlinear relationship between  $Y$  and  $X$ .
- B) There is a very strong positive correlation between  $Y$  and  $X$  because there is an obvious relationship between these variables.
- C) There is a monotonic relationship between  $Y$  and  $X$ .
- D) There is a strong quadratic relationship between  $Y$  and  $X$ .
- E) All of the above.

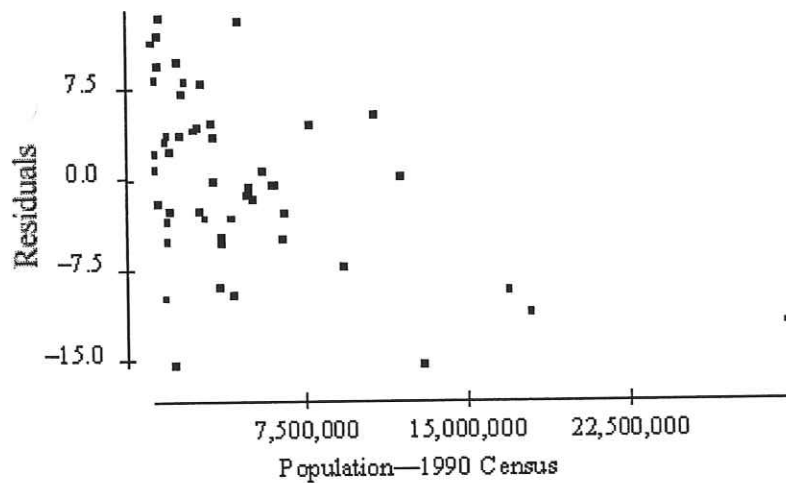
Use the following to answer questions 16 and 17:

The scatterplot below plots, for each of the 50 states, the percent of 18-year-olds in the state  $Y$  in 1990 that graduated from high school versus the state's infant mortality rate (deaths per 1,000 births)  $X$  in 1990.



16. For the data above, the correlation between  $X$  and  $Y$  is  $r = -0.54$ . If instead of plotting these variables for each of the 50 states we plotted the values of these variables for each county in the United States, we would expect the value of the correlation  $r$  to be
- A) exactly the same.
  - ☒ B) smaller.
  - C) 0.54 (the magnitude is the same, but the sign changes).
  - D) much higher and probably near 1 since there are many more counties than states.
  - E) much smaller and probably near 0 since there are many more counties than states.

17. Referring to the information above, the least-squares regression line was fitted to the data in the scatterplot and the residuals were computed. A plot of the residuals versus the 1990 population in the state is given below.



This plot suggests

- A) that states with larger populations have lower infant mortality rates due to superior hospital facilities.
  - B) that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with small populations.
  - ☒ C) that population may be a lurking variable in understanding the association between infant mortality rate and percent graduating from high school.
  - D) that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with large populations.
  - E) none of the above.
18. Two variables, an explanatory variable  $x$  and a response variable  $y$ , are measured on each of several individuals. The correlation between these variables is found to be 0.88. To help us interpret this correlation, we should do which of the following?
- A) Compute the least-squares regression line of  $y$  on  $x$  and consider whether the slope is positive or negative.
  - B) Interchange the roles of  $x$  and  $y$  (i.e., treat  $x$  as the response variable and  $y$  as the explanatory variable) and recompute the correlation.
  - ☒ C) Plot the data.
  - D) Determine whether  $x$  or  $y$  has larger values before computing the residuals.
  - E) All of the above.



28. A researcher computed the average SAT math score of all high school seniors who took the SAT exam for each of the 50 states. The researcher also computed the average salary of high school teachers in each of these states and plotted these average salaries against the average SAT math scores for each state. The plot showed a distinct negative association between average SAT math scores and teacher salaries. The researcher may legitimately conclude which of the following?
- A) Increasing the average salary of teachers will cause the average of SAT math scores to decrease, but it is not correct to conclude that increasing the salaries of individual teachers causes the SAT math scores of individual students to increase.
  - B) States that pay teachers high salaries tend to do a poor job of teaching mathematics, on average.
  - C) As the pay for an individual teacher increases, the teacher's students are more likely to do poorly on the SAT math.
  - ☒ D) The data used by the researcher do not provide evidence that increasing the salaries of teachers will cause the performance of students on the SAT math to get worse.
  - E) States in which students tend to perform poorly in mathematics probably have a higher proportion of problem students and thus need to pay teachers higher salaries in order to attract them to teach in those states.
29. The average number of home runs hit by major league baseball players is greater now than it was three decades ago. A researcher suspects that the reason may be that baseballs are livelier now than they were 30 years ago. To check this he tested two baseballs, one that was manufactured 30 years ago (but never used) and one that was new. He noticed that the new baseball bounced higher than the older ball when both were dropped from the same height; that is, the new baseball was livelier than the old one. The researcher can legitimately conclude
- ☒ A) that there is a positive association between the liveliness of the balls tested and the average number of home runs hit in the year that the ball was manufactured.
  - B) that newer baseballs are livelier than older baseballs.
  - C) that there is good evidence that the increase in the liveliness of baseballs has caused the increase in home runs. This is because there is a positive association between liveliness of baseballs and average number of home runs hit and because there is a plausible theory for the observed association.
  - D) that baseballs have been gradually getting livelier over the last three decades.
  - E) all of the above.
30. A researcher notices that in a sample of adults, those that take larger amounts of vitamin C have fewer illnesses. However, those that take larger amounts of vitamin C also tend to exercise more. As explanations for having fewer illnesses, the variables "amount of vitamin C taken" and "amount of exercise" are
- A) skewed.
  - ☒ B) confounded.
  - C) common responses.
  - D) symmetric.
  - E) linked.

31. In 1982 Kennesaw, Georgia, passed a law requiring all citizens to own at least one gun. Although the law was never enforced, six months after the law was passed the number of burglaries in that month was less than in the month prior to passage of the law. We may conclude which of the following?
- A) Gun ownership and burglary rates are negatively associated.
  - B) Gun ownership causes a reduction in crime. This is because there is a negative association between gun ownership and burglary rates and because there is a plausible explanation for this association (gun ownership acts as a deterrent to crime).
  - C) Criminals are more likely to avoid homes in towns where guns are more prevalent.
  - D) All of the above.
  - E) None of the above.
32. A study of the salaries of full professors at Upper Wabash Tech shows that the median salary for female professors is considerably less than the median male salary. However, further investigation shows that the median salaries for male and female full professors are about the same in every department (English, physics, etc.) of the university. This apparent contradiction is an example of
- A) extrapolation.
  - B) Simpson's paradox.
  - C) confounded responses.
  - D) correlation.
  - E) causation.
33. The reversal of the direction of an association when lurking variables are taken into account is called
- A) Simpson's paradox.
  - B) least-squares regression.
  - C) confounding.
  - D) a residual plot.
  - E) negative association.