

Name _____

1) One of the reasons that the Monitoring the Future (MTF) project was started was "to study changes in the beliefs, attitudes, and behavior of young people in the United States." Data are collected from 8th, 10th, and 12th graders each year. To get a representative nationwide sample, surveys are given to a randomly selected group of students. In Spring 2004, students were asked about alcohol, illegal drug, and cigarette use. Describe the W's, if the information is given. If the information is not given, state that it is not specified.

- Who:
- What:
- When:
- Where:
- How:
- Why:

2) Consider the following part of a data set:

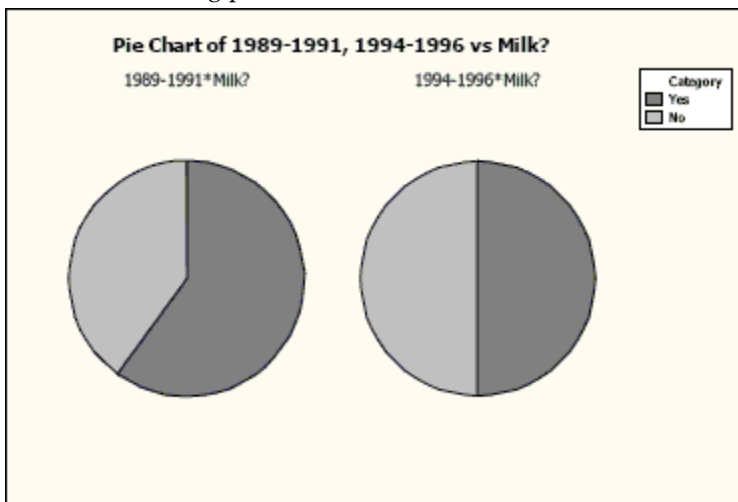
Age (years)	Sex	Only child?	Height (inches)	Weight (pounds)	Credit Hours	GPA	Major
21	Female	Yes	67.00	140.0	16	3.60	animal science
20	Female	No	62.00	130.0	18	3.86	biology
28	Female	No	64.00	188.0	21	3.25	psychology
21	Male	No	65.00	140.0	15	2.95	psychology
24	Female	No	67.00	130.0	20	3.00	anthropology
22	Male	Yes	68.00	135.0	15	2.94	journalism

List the variables in the data set. Indicate whether each variable is treated as categorical or quantitative in this data set. If the variable is quantitative, state the units.

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" (Shanthy 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

- 3) Find the following:
- a. What percent of the young girls reported that they drink milk? 3) a. _____
 - b. What percent of the young girls were in the 1989-1991 survey? b. _____
 - c. What percent of the young girls who reported that they drink milk were in the 1989-1991 survey? c. _____
 - d. What percent of the young girls in 1989-1991 reported that they drink milk? d. _____
- 4) What is the marginal distribution of milk consumption? 4) _____
- 5) Do you think that milk consumption by young girls is independent of the nationwide survey year? Use statistics to justify your reasoning. 5) _____
- 6) Consider the following pie charts of a subset of the data above: 6) _____



Do the pie charts above indicate that milk consumption by young girls is independent of the nationwide survey year? Explain.

7) A brake and muffler shop reported the repair bills, in dollars, for their customers yesterday.

88	283	312	290	172
154	400	381	346	181
203	118	143	252	227
56	192	292	213	422

- Sketch a histogram for these data.
- Find the mean and standard deviation of the repair costs.
- Is it appropriate to use the mean and standard deviation to summarize these data? Explain.
- Describe the association of repair costs.

- 7) a. _____
 b. _____
 c. _____
 d. _____

8) On Monday, a class of students took a big test, and the highest score was 92. The next day, a student who had been absent made up the test, scoring 100. Indicate whether adding that student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:

- mean
- median
- range
- IQR
- standard deviation

- 8) a. _____
 b. _____
 c. _____
 d. _____
 e. _____

9) The body temperature of students is taken each time a student goes to the nurse's office. The five-number summary for the temperatures (in degrees Fahrenheit) of students on a particular day is:

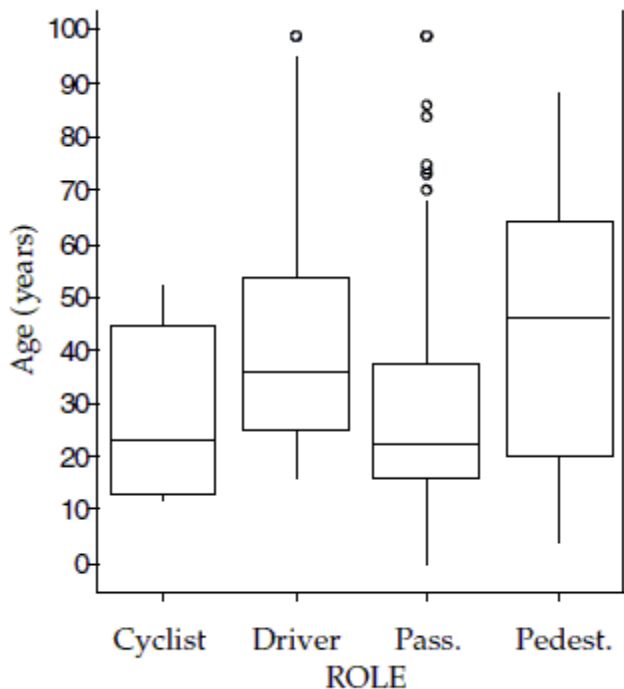
Min	Q1	Median	Q3	Max
96.6°	97.85°	98.25°	98.6°	101.8°

- Would you expect the mean temperature of all students who visited the nurse's office to be higher or lower than the median? Explain.
- After the data were picked up in the afternoon, three more students visited the nurse's office with temperatures of 96.7°, 98.4°, and 99.2°. Were any of these students outliers? Explain.

- 9) a. _____
 b. _____

10) The boxplots show the age of people involved in accidents according to their role in the accident.

10) a. _____



b. _____

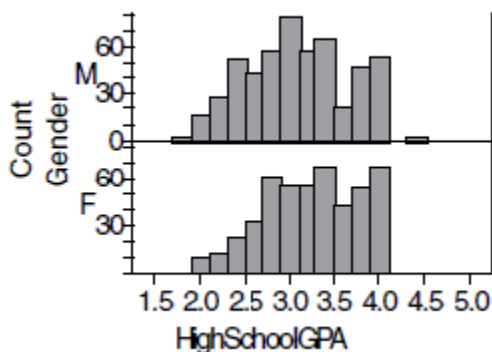
c. _____

d. _____

e. _____

- Which role involved the youngest person, and what is the age?
- Which role had the lowest median age, and what is the age?
- Which role had smallest range of ages, and what is it?
- Which role had the largest IQR of ages, and what is it?
- Which role generally involved the oldest people? Explain.

11) One thousand students from a local university were sampled to gather information such as gender, high school GPA, college GPA, and total SAT scores. The results were used to create histograms displaying high school grade point averages (GPA's) for both males and females. Compare the grade distribution of males and females.



12) Students taking an intro stats class reported the number of credit hours that they were taking that quarter. 12) _____

Summary statistics are shown in the table.

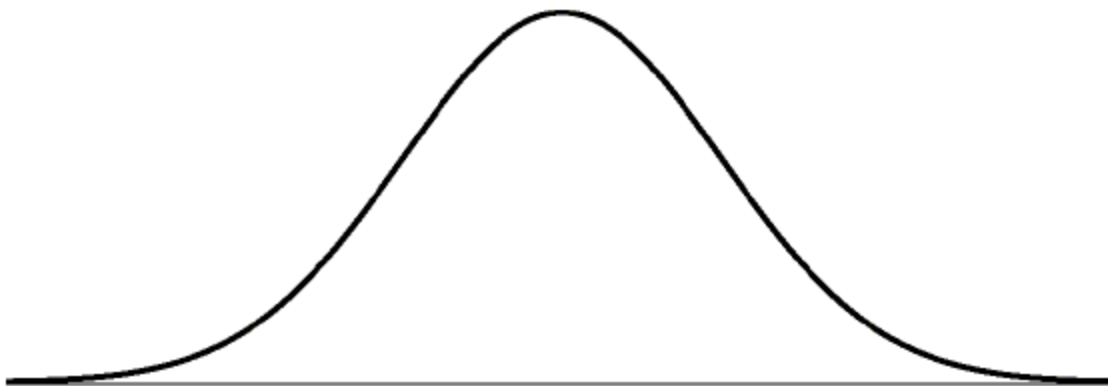
\bar{x}	16.65
s	2.96
min	5
Q1	15
median	16
Q3	19
max	28

a. Suppose that the college charges \$73 per credit hour plus a flat student fee of \$35 per quarter. For example, a student taking 12 credit hours would pay $\$35 + \$73(12) = \$911$ for that quarter.

- i. What is the mean fee paid?
- ii. What is the standard deviation for the fees paid?
- iii. What is the median fee paid?
- iv. What is the IQR for the fees paid?

b. Twenty-eight credit hours seems like a lot. Would you consider 28 credit hours to be unusually high? Explain.

13) The Wechsler Adult Intelligence Scale - Revised (WAIS-R) follow a Normal model with mean 100 and standard deviation 15. Draw and clearly label this model.



14) Adult female Dalmatians weigh an average of 50 pounds with a standard deviation of 3.3 pounds. Adult female Boxers weigh an average of 57.5 pounds with a standard deviation of 1.7 pounds. One statistics teacher owns an underweight Dalmatian and an underweight Boxer. The Dalmatian weighs 45 pounds, and the Boxer weighs 52 pounds. Which dog is more underweight? Explain.

15) Human body temperatures taken through the ear are typically 0.5°F higher than body temperatures taken orally. Making this adjustment and using the 1992 Journal of the American Medical Association article that reports average oral body temperature as 98.2°F , we will assume that a Normal model with an average of 98.7°F and a standard deviation of 0.7°F is appropriate for body temperatures taken through the ear.

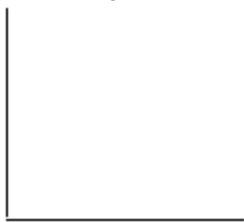
15) a. _____
 b. _____
 c. _____

- a. An ear temperature of 97°F may indicate hypothermia (low body temperature). What percent of people have ear temperatures that may indicate hypothermia?
- b. Find the interquartile range for ear temperatures.
- c. A new thermometer for the ear reports that it is more accurate than the ear thermometers currently on the market. If the average ear temperature reading remains the same and the company reports an IQR of 0.5°F , find the standard deviation for this new ear thermometer.

16) After conducting a survey at a pet store to see what impact having a pet had on the condition of the yard, a news reporter stated "There appears to be a strong correlation between the owning a pet and the condition of the yard." Comment on this observation.

17) On the axes below, sketch a scatterplot describing...
 a. a strong positive association
 b. a weak negative association

17) _____



18) A study by a prominent psychologist found a moderately strong positive association between the number of hours of sleep a person gets and the person's ability to memorize information.

18) a. _____
 b. _____

- a. Explain in the context of this problem what "positive association" means.
- b. Hoping to improve academic performance, the psychologist recommended the school board allow students to take a nap prior to any assessment. Discuss this reasoning.

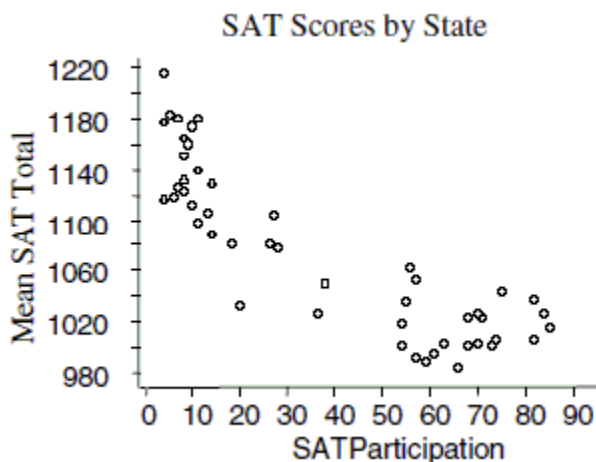
19) A common objective for many school administrators is to increase the number of students taking SAT and ACT tests from their school. The data from each state from 2003 are reflected in the scatterplot below.

19) a. _____

b. _____

c. _____

d. _____



- Write a few sentences describing the association.
- Estimate the correlation. $r =$ _____
- If the point in the top left corner (4, 1215) were removed, would the correlation become stronger, weaker, or remain about the same? Explain briefly.
- If the point in the very middle (38, 1049) were removed, would the correlation become stronger, weaker, or remain about the same? Explain briefly.

An article in the *Journal of Statistics Education* reported the price of diamonds of different sizes in Singapore dollars (SGD). The following table contains a data set that is consistent with this data, adjusted to US dollars in 2004:

2004 US \$	Carat
494.82	0.12
768.03	0.17
1105.03	0.20
1508.88	0.25
1826.18	0.28
2096.89	0.33

2004 US \$	Carat
688.24	0.15
944.90	0.18
1071.75	0.21
1504.44	0.26
1908.28	0.29
2409.76	0.35

2004 US \$	Carat
748.10	0.16
1076.18	0.19
1289.20	0.23
1597.63	0.27
2038.09	0.32

20) Make a scatterplot and describe the association between the size of the diamond (carat) and the cost (in US dollars).

21) Create a model to predict diamond costs from the size of the diamond.

21) _____

22) Do you think a linear model is appropriate here? Explain.

23) Interpret the slope of your model in context.

24) Interpret the intercept of your model in context.

25) What is the correlation between cost and size?

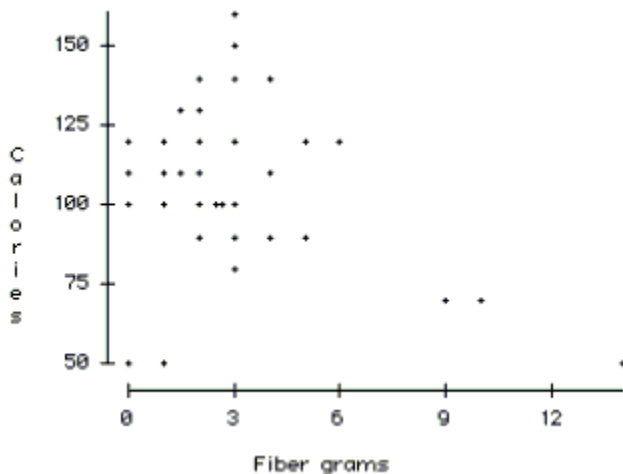
25) _____

26) Explain the meaning of R^2 in the context of this problem.

27) Would it be better for a customer buying a diamond to have a negative residual or a positive residual from this model? Explain.

27) _____

Current research states that a good diet should contain 20–35 grams of dietary fiber. Research also states that each day should start with a healthy breakfast. The nutritional information for 77 breakfast cereals was reviewed to find the grams of fiber and the number of calories per serving. The scatterplot below shows the relationship between fiber and calories for the cereals.



28) Do you think there is a clear pattern? Describe the association between fiber and calories.

29) Comment on any unusual data point or points in the data set. Explain.

30) Do you think a model could accurately predict the number of calories in a serving of cereal that has 22 grams of fiber? Explain.

31) The average movie ticket prices in selected years since 1948 are listed in the table below.

Year	Movie Ticket Price
1948	\$0.36
1954	\$0.49
1958	\$0.68
1963	\$0.86
1967	\$1.22
1971	\$1.65
1974	\$1.89
1974	\$2.03
1976	\$2.13
1977	\$2.23
1978	\$2.34
1979	\$2.47
1980	\$2.69
1981	\$2.78
1982	\$2.94
1983	\$3.15
1984	\$3.36
1985	\$3.55
1986	\$3.71
1987	\$3.91
1988	\$4.11

31) a. _____

b. _____

a. Use re-expressed data to create a model that predicts ticket prices. (Hint: scale the year)

b. Find the movie ticket price this model predicts for 2004.

32) During a chemistry lab, students were asked to study a radioactive element which decays over time. The results are in the table.

Time (in days)	0	2	4	6	8	10
Element (in grams)	320	226	160	115	80	57

32) a. _____

b. _____

a. Model the remaining mass of the element.

b. Find the predicted amount of the element remaining after thirty minutes.

Questions 33–35: As a 4-H project, Billy is raising chickens. He feeds and waters them every day, and collects the eggs every other day, selling them to people in the neighborhood. He has found that each hen's nest will contain from 0 to 2 eggs. Based on past experience he estimates that there will be no eggs in 10% of the nests, one egg in 30% of the nests, and 2 eggs in the other 60%. Conduct a simulation to estimate how many nests Bill will have to visit to collect a dozen eggs.

33) Describe how you will use a random number table to conduct this simulation.

34) Show three trials by clearly labeling the random number table given below. Specify the outcome for each trial.

5 7 5 2 8 7 8 3 0 5 6 3 5 0 8 2 9 4 1 8

9 0 6 7 6 3 1 9 9 3 5 4 6 3 6 1 7 8 7 7

3 1 5 7 4 2 8 0 4 2 7 2 6 2 1 8 4 8 1 8

35) State your conclusion.

36) A statistics teacher wants to know how her students feel about an introductory statistics course. She decides to administer a survey to a random sample of students taking the course. She has several sampling plans to choose from. Name the sampling strategy in each.

36) a. _____

b. _____

a. There are four ranks of students taking the class: freshmen, sophomores, juniors, and seniors. Randomly select 15 students from each class rank.

c. _____

b. Randomly select a class rank (freshmen, sophomores, juniors, and seniors) and survey every student in that class rank.

c. Each student has a nine-digit student number. Randomly choose 60 numbers.

d. _____

d. Using the class roster, select every fifth student from the list.

37) Explain why the second plan suggested above, sampling all students from one class rank, might be biased. Be sure to name the kind(s) of bias you describe.

38) Listed below are the names of 20 students who are juniors. Use the random numbers listed below to select five of them to be in your sample. Clearly explain your method.

Adam	Chris	Dave	Deirdre	Dick
Ellen	Eric	Joan	John	Judi
Joy	Kenny	Laura	Mary	Paul
Peter	Rachel	Rob	Sara	Stacey

04028 55259 81183 40754 60209 06765 27306

28370 82669 83236 77479 90618 43707 78695

39) Name and describe the kind of bias that might be present if the statistics teacher decides that instead of randomly selecting students to survey on how they feel about the course she just...

39) a. _____

b. _____

a. asks students to volunteer for the survey.

b. gives the survey during class one day.

Researchers plan to investigate a new medication that may reduce blood pressure for individuals with higher than average blood pressure. 90 volunteers with higher than average blood pressure are solicited. Volunteers are randomly assign 100 mg of the medicine, 200 mg of the medicine, or a placebo. Blood pressure will be measured at the beginning and at the conclusion of the study.

- 40) Identify the subjects. 40) _____

- 41) Identify the treatments. 41) _____

- 42) Identify the response variable. 42) _____

- 43) Describe an advantage to random assignment of treatment. 43) _____

- 44) Describe an advantage of the placebo. 44) _____

- 45) Describe a disadvantage of using volunteers in this study. 45) _____

- 46) Is this study blind? 46) _____

An article in a local newspaper reported that dogs kept as pets tend to be overweight. Veterinarians say that diet and exercise will help these chubby dogs get in shape. The veterinarians propose two different diets (Diet A and Diet B) and two different exercise programs (Plan 1 and Plan 2). Diet A: owners control the portions of dog food and dog treats; Diet B: a mixture of fresh vegetables with the dog food and substitute regular dog treats with baby carrots. Plan 1: three 30-minute walks a week; Plan 2: 20-minute walks daily. Sixty dog owners volunteer to take part in an experiment to help their chubby dogs lose weight.

- 47) Identify the following: 47) a. _____
 - a. the subjects: b. _____
 - b. the factor(s) and the number of level(s) for each:
 - c. the number of treatments: c. _____
 - d. whether or not the experiment is blind (or double-blind): d. _____
 - e. the response variable: e. _____

- 48) Design an experiment to determine whether the diet and exercise programs are effective in helping dogs to lose weight.

- 49) Five multiple choice questions, each with four possible answers, appear on your history exam. What is the probability that if you just guess, you
- a. get none of the questions correct? _____
- b. get all of the questions correct? _____
- c. get at least one of the questions wrong? _____
- d. get your first incorrect answer on the fourth question? _____

- 50) The Masterfoods company manufactures bags of Peanut Butter M&M's. They report that they make 10% each brown and red candies, and 20% each yellow, blue, and orange candies. The rest of the candies are green.
- a. If you pick a Peanut Butter M&M at random, what is the probability that
- i. it is green? _____
- ii. it is a primary color (red, yellow, or blue)? _____
- iii. it is not orange? _____
- b. If you pick four M&M's in a row, what is the probability that
- i. they are all blue? _____
- ii. none are green? _____
- iii. at least one is red? _____
- iv. the fourth one is the first one that is brown? _____

c. After picking 10 M&M's in a row, you still have not picked a red one. A friend says that you should have a better chance of getting a red candy on your next pick since you have yet to see one. Comment on your friend's statement.

- 51) A survey of families revealed that 58% of all families eat turkey at holiday meals, 44% eat ham, and 16% have both turkey and ham to eat at holiday meals.
- a. What is the probability that a family selected at random had neither turkey nor ham at their holiday meal? _____
- b. What is the probability that a family selected at random had only ham without having turkey at their holiday meal? _____
- c. What is the probability that a randomly selected family having turkey had ham at their holiday meal? _____
- d. Are having turkey and having ham disjoint events? Explain. _____

- 52) Many school administrators watch enrollment numbers for answers to questions parents ask. Some parents wondered if preferring a particular science course is related to the student's preference in foreign language. Students were surveyed to establish their preference in science and foreign language courses. Does it appear that preferences in science and foreign language are independent? Explain.

	Chemistry	Physics	Biology	Total
French	16	10	8	34
Spanish	35	23	44	102
Total	51	33	52	136

- 53) For purposes of making budget plans for staffing, a college reviewed student's year in school and area of study. Of the students, 22.5% are seniors, 25% are juniors, 25% are sophomores, and the rest are freshmen. Also, 40% of the seniors major in the area of humanities, as did 39% of the juniors, 40% of the sophomores, and 36% of the freshmen. What is the probability that a randomly selected humanities major is a junior? Show your work. 53) _____

A small business just leased a new computer and color laser printer for three years. The service contract for the computer offers unlimited repairs for a fee of \$100 a year plus a \$25 service charge for each repair needed. The company's research suggested that during a given year 86% of these computers needed no repairs, 9% needed to be repaired once, 4% twice, 1% three times, and none required more than three repairs.

- 54) Find the expected number of repairs this kind of computer is expected to need each year. Show your work. 54) _____

- 55) Find the standard deviation of the number of repairs each year. 55) _____

- 56) What are the mean and standard deviation of the company's annual expense for the service contract? 56) _____

- 57) How many times should the company expect to have to get this computer repaired over the three-year term of the lease? 57) _____

- 58) What is the standard deviation of the number of repairs that may be required during the three-year lease period? On what assumption does your calculation rest? Do you think this assumption is reasonable? Explain. 58) _____

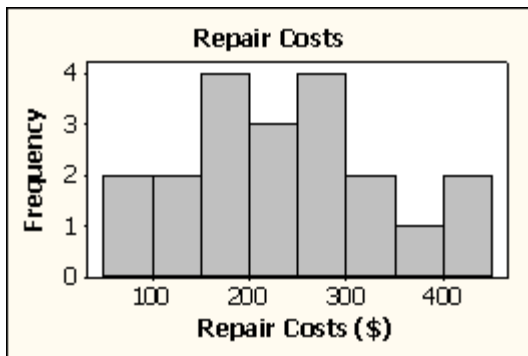
- 59) The service contract for the printer estimates a mean annual cost of \$120 with standard deviation of \$30. What is the expected value and standard deviation of the total cost for the service contracts on computer and printer? 59) _____

- 60) Which service contract should the company expect to cost more each year? How much more? With what standard deviation? 60) _____

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

- 1) Who: 8th, 10th, and 12th graders
What: alcohol, illegal drug, and cigarette use
When: Spring 2004
Where: United States
How: survey
Why: "to study changes in the beliefs, attitudes, and behavior of young people in the United States"
- 2) Categorical: sex, only child?, major
Quantitative: age (years), height (inches), weight (pounds), credit hours, GPA
- 3) a. 56.9%
b. 38.9%
c. 41.1%
d. 60.0%
- 4) Yes: 56.9%; No: 43.1%
- 5) No. 56.9% of all young girls surveyed reported drinking milk, but 60% of the young girls reported drinking milk in the 1989–1991 survey. Since these percentages differ, milk consumption and year are not independent.
- 6) No. It looks like there is some sort of relationship between milk consumption and nationwide survey year, since the percentage of young girls who reported drinking milk is a larger slice of the pie chart for the 1989–1991 survey than the same response for the 1994–1996 survey.
- 7) a.

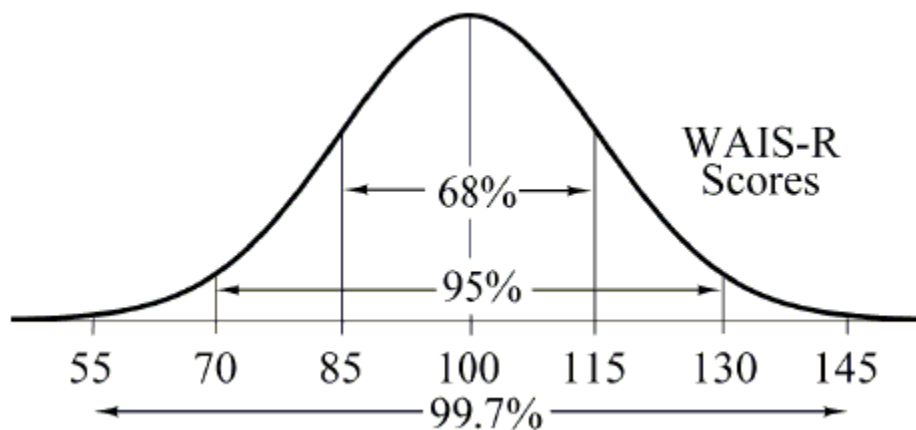


- b. $\bar{x} = \$236.25$; $s = \$103.43$
- c. Yes, the data are roughly unimodal and symmetric with no outliers.
- d. The repair costs averaged \$236.25, ranging from \$56 to \$422 with a standard deviation of \$103.43. The distribution was approximately symmetric, with typical repair costs clustered between \$150 and \$300.
- 8) a. mean: increase
b. median: stay about the same
c. range: increase
d. IQR: stay about the same
e. standard deviation: increase
- 9) a. The mean temperature of all students would probably be higher than the median. Using the five-number summary, it appears the data are skewed to the right.
b. $IQR = 98.6^\circ - 97.85^\circ = 0.75^\circ$. Since $1.5(IQR) = 1.125^\circ$, the fences are $97.85^\circ - 1.125^\circ = 96.725^\circ$ and $98.6^\circ + 1.125^\circ = 99.725^\circ$. The lowest temperature (96.7°) being added to the data set is smaller than the lower fence (96.725°) so it is an outlier on the low end. The highest temperature (99.2°) being added to the data set is not above the upper fence (99.725°) so it is not an outlier on the high end.

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

- 10) a. Passenger, less than 1 year
b. Passenger, 21 years
c. Cyclist, 40 years
d. Pedestrian, 44 years
e. Pedestrian. While the oldest person involved in an accident is not a pedestrian, the median age for pedestrians is almost 45 years, while the median ages in the other groups are between 22 and 35 years old. The oldest 50% of the Pedestrian group, from 45 to 87 years, is generally older than the youngest 75% of two groups – Cyclist and Passenger, and only the Driver group has any of its middle 50% as old. The Driver and Passenger groups have a few people older than the Pedestrian group.
- 11) The distributions of high school GPA for both males and females are skewed to the left, and both distributions appear to be centered at a GPA of about 3.0. The distribution of male GPA appears slightly more spread out than the distribution of female GPA.
- 12) a. i. $\$35 + \$73(16.65) = \$1250.45$
ii. $\$73(2.96) = \216.08
iii. $\$35 + \$73(16) = \$1203$
iv. $IQR = \$73(19-15) = \292
b. $IQR = 19 - 15 = 4$ credit hours
High outliers will lie above $Q3 + 1.5IQR = 19 + 1.5(4) = 25$ credit hours. Since 28 credit hours exceeds 25 credit hours, I would consider 28 credit hours to be unusually high.
- 13)



14) Dalmation: $z_D = \frac{45 - 50}{3.3} = -1.52$

Boxer: $z_B = \frac{52 - 57.5}{1.7} = -3.24$

The Dalmatian is 1.52 standard deviations underweight, while the Boxer is 3.24 standard deviations underweight. So, the Boxer is more underweight.

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

15) a. $z = \frac{97 - 98.7}{0.7} = -2.43$, so $P(z < -2.43) = 0.0075$

About 0.75% of people have ear temperatures that may indicate hypothermia.

b. The z-scores associated with the IQR are $z = -0.67$ and $z = 0.67$. So, we need to solve for y in each of the following equations: $-0.67 = \frac{y - 98.7}{0.7}$ and $0.67 = \frac{y - 98.7}{0.7}$. We get $y = 98.7 - 0.67(0.7) = 98.2$ and $y = 98.7 + 0.67(0.7) = 99.2$. The

interquartile range is $IQR = 99.2^\circ\text{F} - 98.2^\circ\text{F} = 1.0^\circ\text{F}$.

c. The new IQR is 0.5°F , while the old IQR was 1.0°F . So, we want

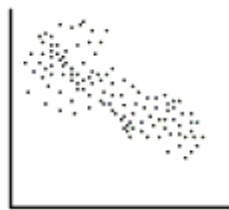
$IQR = [98.7 + 0.67\sigma] - [98.7 - 0.67\sigma] = 0.5$, or $1.34\sigma = 0.5$. Thus, $\sigma = \frac{0.5}{1.34} = 0.37$. Our new standard deviation is 0.37°F .

16) The variables – owning a pet and condition of the yard – are both categorical variables. Correlation cannot be calculated with categorical variables.

17) a.



b.

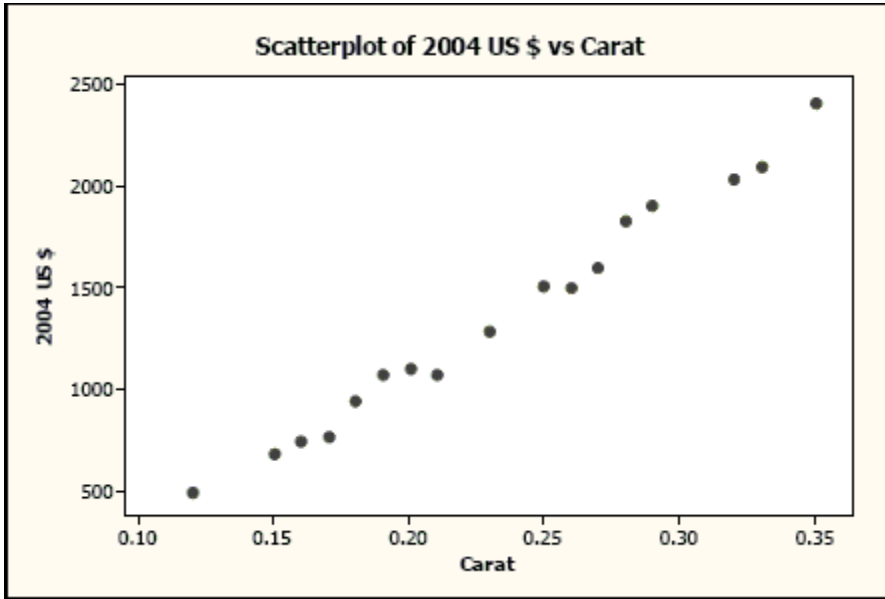


- 18) a. A positive association means in general people who had more sleep were able to memorize more information.
b. The psychologist is attributing association to cause and effect. There is an implication that more sleep will cause better memorization, therefore causing an increase in assessments scores. Perhaps people who had memorized more were able to sleep more restfully, or perhaps differences in brain chemistry allowed some people to memorize more and to sleep more easily.
- 19) a. There is a moderate, negative, linear association between the percent of students taking the SAT test and the total SAT score. It appears that the states with a larger percentage of students taking the SAT test have lower average total scores.
b. $r = -0.76$ (answers between -0.6 and -0.9 are acceptable)
c. If the point in the top left corner (4, 1215) were removed, the correlation would become stronger because the remaining points show a pattern with slightly less scatter.
d. If the point in the very middle (38, 1049) were removed, the correlation would remain about the same; this point does not contribute much to the scatter.

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

20) There is a strong, positive, linear association between the size of the diamond and its cost. The cost of a diamond increases with size.



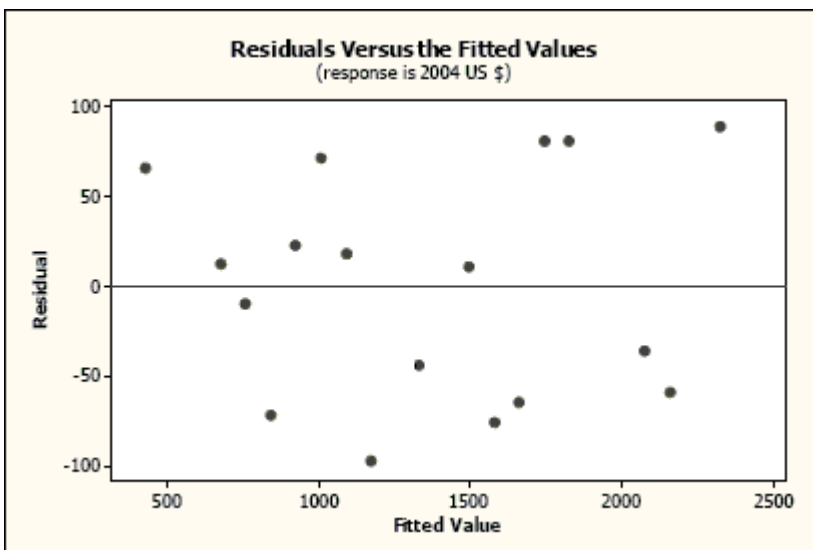
21) The regression equation is
 $2004\text{ US \$} = -559 + 8225\text{ Carat}$

Predictor	Coef	SE Coef	T	P
Constant	-558.52	57.88	-9.65	0.000
Carat	8225.1	239.1	34.40	0.000

$S = 64.9355$ $R\text{-Sq} = 98.7\%$ $R\text{-Sq}(\text{adj}) = 98.7\%$

Predicted cost = $-558.52 + 8225.1(\text{carat})$

22) A linear model is appropriate for this problem. The residual plot shows no obvious pattern.



Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

- 23) The slope of the model is 8225.1. The model predicts that for each additional carat, the cost of the diamond will increase by \$8225.10, on average. This can also be interpreted as for each additional 0.01 carat, the cost of the diamond will increase by \$82.251, on average.
- 24) The intercept of the model is -558.52. The model predicts that a diamond of 0 carats costs -\$558.52. This is not realistic.
- 25) The correlation, r , is $r = \sqrt{0.987} = 0.993$. Since the scatterplot shows a positive relationship, the positive value must be used.
- 26) $R^2 = 0.987$. So 98.7% of the variation in diamond prices can be accounted for by the variation in the size of the diamond.
- 27) It would be better for customers to have a negative residual from this model, since a negative residual would indicate that the actual cost of the diamond was less than the model predicted it to be.
- 28) There is no clear pattern. At first glance, there appears to be a weak, negative association between grams of fiber and the number of calories in the cereals. Yet, the five points at the bottom of the graph are outside the pattern, with extremely low numbers of calories. Additionally, the three points on the right of the scatterplot have an unusually high amount of fiber, making them outliers and influential points.
- 29) The points in the bottom left corner seem to have extremely low calorie content for cereals between zero and six grams of fiber. The points with 9, 10 and 14 grams of fiber appear to have an unusually high amount of fiber for their calorie content, making them outliers and influential points. These three points would also be leverage points, creating the impression that there is a negative association between grams of fiber and the number of calories in the cereals.
- 30) This data contains information about cereals with fiber content between 0 and 14 grams. It would be extrapolation to try to use this data to predict the calorie content of cereals with 22 grams of fiber.
- 31) a. Let explanatory variable be Year - 1900; so, 1948 is input as 48.
Let response variable be $\hat{\log}(\text{Ticket Price})$
Exponential model: $\log(\text{Ticket}) = -1.73 + 0.0269(\text{Year} - 1900)$
b. $\log(\hat{\text{Ticket}}) = -1.73 + 0.0269(104) = 1.0676$
 $\hat{\text{Ticket}} = 10^{1.0676} = \11.68
- 32) a. $\log \text{Element} = 2.505 - 0.0749(\text{time})$
b. Time is measured in days, so 30 minutes, or half an hour, is $\frac{1}{48}$ days.
 $\log \hat{\text{Element}} = 2.505 - 0.0749(\frac{1}{48}) \approx 2.5034$
 $10^{2.5034} \approx 318.74$ grams
- 33) Look at one digit at a time. Let 0 = no eggs; 1, 2, 3 = one egg; 4 - 9 = two eggs. Go across the row of digits one at a time, adding up the number of eggs until there are 12 or more. Count the number of nests visited.
- 34)
- | | | | |
|-----------|------------------------|-----------|-----------|
| 5 7 5 2 8 | 7 8 3 0 5 | 6 3 5 0 8 | 2 9 4 1 8 |
| 2 2 2 1 2 | 2 2 13 eggs, 7 nests | | |
- | | | | |
|-----------|--------------------------|-----------|-----------|
| 9 0 6 7 6 | 3 1 9 9 3 | 5 4 6 3 6 | 1 7 8 7 7 |
| 2 0 2 2 2 | 1 1 2 12 eggs, 8 nests | | |
- | | | | |
|-----------|----------------------------|-----------|-----------|
| 3 1 5 7 4 | 2 8 0 4 2 | 7 2 6 2 1 | 8 4 8 1 8 |
| 1 1 2 2 2 | 1 2 0 2 13 eggs, 9 nests | | |
- 35) According to this simulation, Billy will have to visit an average of 8 nests to collect a dozen eggs.

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

- 36) a. stratified
b. cluster
c. simple
d. systematic
- 37) Undercoverage might occur. The class rank selected may not be representative of all students taking the course. For example, perhaps the senior rank contains students who have failed the course previously, and these students might have a more negative attitude toward the class.
- 38)

01 - Adam	02 - Chris	03 - Dave	04 - Deirdre	05 - Dick
06 - Ellen	07 - Eric	08 - Joan	09 - John	10 - Judi
11 - Joy	12 - Kenny	13 - Laura	14 - Mary	15 - Paul
16 - Peter	17 - Rachel	18 - Rob	19 - Sara	20 - Stacey

04028 55259 81183 40754 60209 06765 27306

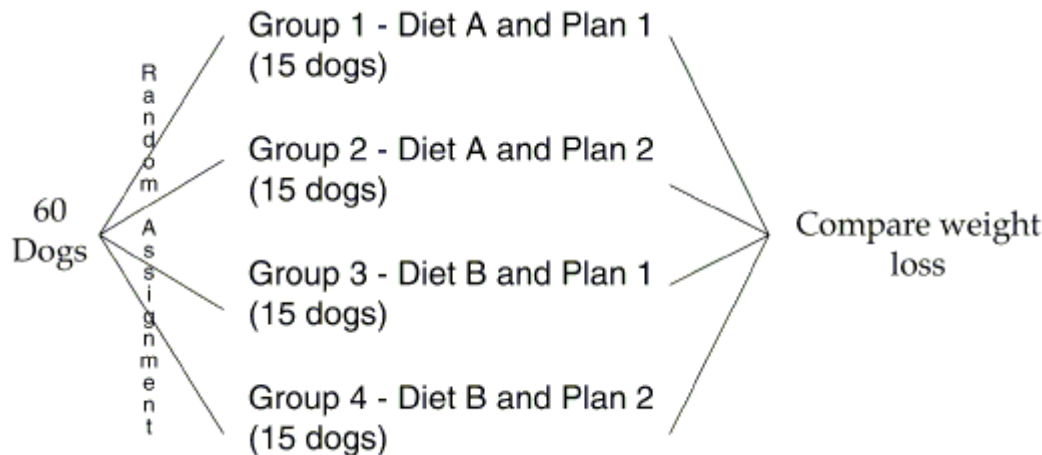
28370 82669 83236 77479 90618 43707 78695

- Assign two digit numbers to each of the juniors, as noted in the table. Use the random digits, in groups of 2, to select the first five people in the sample. Ignore any digits 21-99 and 00, because there are no corresponding people, and ignore repeated numbers.
- 04 - Deirdre, 02 - Chris, 85 - ignore, 52 - ignore, 59 - ignore, 81 - ignore, 18 - Rob, 34 - ignore, 07 - Eric, 54 - ignore, 60 - ignore, 20 - Stacey
- People selected: Deirdre, Chris, Rob, Eric, and Stacey
- 39) a. Voluntary response sample—the bias would probably be toward those students who enjoy the course, and these students would rate the course more favorably.
b. Convenience sample—the bias would probably be toward students who happen to attend class that day. They may enjoy the course more than students who are skipping class!
- 40) The subjects are the 90 volunteers with higher than average blood pressure.
- 41) There are 3 treatments, 100 mg, 200 mg, and the placebo.
- 42) The response variable is the change in blood pressure.
- 43) Randomization will equalize variability for which we cannot control, helping to ensure comparable, homogeneous treatment groups. We may be able to establish causation as opposed to association.
- 44) We have an extra comparison with a control group. Blood pressure may change due to other variables.
- 45) We are not able to generalize the results to a larger population.
- 46) The subjects are blind, assuming the 100 mg, 200 mg and placebo appear to be the same.
- 47) a. 60 chubby dogs
b. diet (two levels) and exercise (two levels)
c. four treatments (Diet A and Plan 1, Diet A and Plan 2, Diet B and Plan 1, Diet B and Plan 2)
d. This design is at best single-blind, since the owners know which diet and exercise plan their dogs are on, but the evaluators do not have to be given this information.
e. weight loss

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

48)

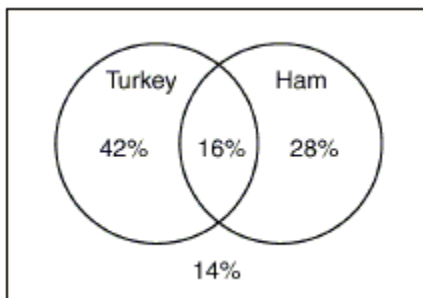


- 49) a. $P(\text{get no questions correct}) = P(\text{get all questions wrong}) = (0.75)^5 = 0.2373$
 b. $P(\text{get all questions correct}) = (0.25)^5 = 0.00098$
 c. $P(\text{get at least 1 question wrong}) = 1 - P(\text{get all questions correct}) = 1 - (0.25)^5 = 0.999$
 d. $P(\text{get first incorrect on fourth question}) = P(\text{correct}) \times P(\text{correct}) \times P(\text{correct}) \times P(\text{wrong})$
 $= (0.25) \times (0.25) \times (0.25) \times (0.75) = 0.0117$

- 50) a. i. $P(\text{green}) = 1 - P(\text{not green}) = 1 - (0.1 + 0.1 + 0.2 + 0.2 + 0.2) = 1 - 0.8 = 0.2$
 ii. $P(\text{red, yellow, or blue}) = P(\text{red}) + P(\text{yellow}) + P(\text{blue}) = 0.1 + 0.2 + 0.2 = 0.5$
 iii. $P(\text{not orange}) = 1 - P(\text{orange}) = 1 - 0.2 = 0.8$
 b. i. $P(\text{all blue}) = (0.2)^4 = 0.0016$
 ii. $P(\text{none green}) = (0.8)^4 = 0.4096$
 iii. $P(\text{at least one red}) = 1 - P(\text{none red}) = 1 - (0.9)^4 = 1 - 0.6561 = 0.3439$
 iv. $P(\text{fourth is first brown}) = P(\text{not brown}) \times P(\text{not brown}) \times P(\text{not brown}) \times P(\text{brown})$
 $= (0.9) \times (0.9) \times (0.9) \times (0.1) = 0.0729$
 c. Assuming a very large number of M&Ms are manufactured, selections of candies are essentially independent, so the colors of the M&M's picked so far have no effect on the color of the next M&M selected.

- 51) a. $P(\text{neither ham nor turkey}) = 1 - P(\text{ham} \cup \text{turkey})$
 $= 1 - [P(\text{ham}) + P(\text{turkey}) - P(\text{ham} \cap \text{turkey})]$
 $= 1 - [0.44 + 0.58 - 0.16] = 1 - 0.86 = 0.14$

Or, using the Venn diagram below, 14%



- b. $P(\text{ham only}) = P(\text{ham}) - P(\text{ham} \cap \text{turkey}) = 0.44 - 0.16 = 0.28$

Or, using the Venn diagram above, 28%

- c. $P(\text{ham} | \text{turkey}) = \frac{P(\text{ham} \cap \text{turkey})}{P(\text{turkey})} = \frac{0.16}{0.58} = 0.2759$

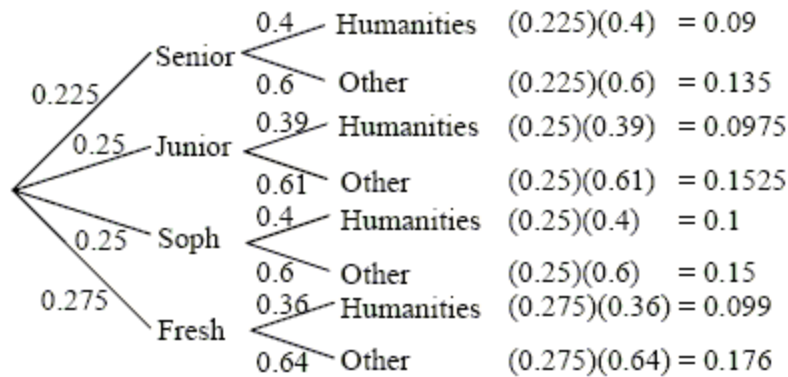
- d. No, the events are not disjoint, since some families (16%) have both ham and turkey at their holiday meals.

Answer Key

Testname: AP STATS 1ST SEMESTER REVIEW CH 1-15

52) Overall, 102 of 136, or 75%, preferred Spanish. 35 of 51, or 68.6%, of students in Chemistry had Spanish. 23 of 33, or 69.6%, of students in Physics had Spanish, and 44 of 52, or 84.6% of students in Biology had Spanish. Chemistry and Physics students were much less likely to take Spanish than Biology students. It does appear that there is an association between preference in science and foreign language.

53)



$$P(\text{humanities major}) = 0.09 + 0.0975 + 0.1 + 0.099 = 0.3865$$

$$P(\text{junior} \mid \text{humanities major}) = \frac{P(J \cap H)}{P(H)} = \frac{0.0975}{0.3865} = 0.2523$$

54) $E(X) = 0(0.86) + 1(0.09) + 2(0.04) + 3(0.01) = 0.20$ repairs

55) $Var(X) = (0 - 0.2)^2(0.86) + (1 - 0.2)^2(0.09) + (2 - 0.2)^2(0.04) + (3 - 0.2)^2(0.01)$

$$Var(X) = 0.30 \quad SD(X) = 0.55 \text{ repairs}$$

56) Let $C = 100 + 25X$; $E(C) = 100 + 25(0.20) = \105 ; $SD(C) = 25(0.55) = \$13.69$

57) $E(X_1 + X_2 + X_3) = 0.20 + 0.20 + 0.20 = 0.60$ repairs.

58) We must assume that the number of repairs is independent from year to year. This may be risky as some computers may be more trouble-prone than others.

$$Var(X_1 + X_2 + X_3) = 0.30 + 0.30 + 0.30 = 0.90, \text{ so } SD(X) = 0.95 \text{ repairs.}$$

59) $E(C + P) = 105 + 120 = \$225$; $Var(C + P) = (13.69)^2 + 30^2 = 1087.42$, so $SD(C + P) = \$32.98$

60) The printer; $E(P - C) = 120 - 105 = \$15$ more; $SD(P - C) = \$32.98$ (same as the sum)