***Exercises 13A***

**1.** *Sample*: Two Hinckley cars

*Population*: All Hinckley cars

*Sample size*: Two cars are a very small sample when we are discussing potentially millions of cars. This reduces the likelihood that the sample is representative of the population.

*Potential bias*: The sample excludes any cars that other owners might praise. It doesn’t allow for the possibility of evidence that would go against its claims. This reduces the likelihood that the sample is representative of the population.

*Randomness*: This is not a random sample because not every Hinckley car had an equal chance of getting into the sample. This reduces the likelihood that the sample is representative of the population.

**2.** *Sample*: 132 dogs brought to one veterinarian

*Population*: All breeds of dogs

*Sample size*: The sample size is not large enough to capture accurately all the hundreds of millions of dogs of various breeds. This reduces the likelihood that the sample is representative of the population.

*Potential bias*: We are not told that the sample contains the correct proportions of all possible dog breeds. We don’t know if the dogs brought to this particular veterinarian are truly representative of the various breeds. This reduces the likelihood that the sample is representative of the population.

*Randomness*: This is not a random sample because it just included dogs that were brought to one place. Therefore, not every dog had an equal chance of getting into the sample. This reduces the likelihood that the sample is representative of the population.

**3.** *Sample*: 1600 psychiatric outpatients at a Chicago hospital

*Population*: The study does not specify, but it seems that the researchers might be willing to generalize (at least in an approximate way) to similar psychiatric outpatients.

*Sample size*: The size is large enough to be acceptable. However, it would be better if the study was conducted in more than one city, because it would have bolstered the case that the sample might be representative of the population.

*Potential bias*: The sample is taken from one hospital in one city. In addition, we are not told how the patients were divided into the two groups. It is possible that those placed into the drug therapy group were different from those in the other group; perhaps their problems were not as severe at the beginning of the study. This potential bias affects the likelihood that the sample is representative of the population.

*Randomness*: The patients might have been placed randomly into the two groups, but we are not told this. If they were not randomly separated, then not every patient had an equal chance of getting into the two groups. Additionally, since the sample was in only one hospital, the 1600 patients were not randomly taken from the total group of psychiatric outpatients. This reduces the likelihood that the sample is representative of the population.

**4.** *Sample*: 4000 UFO sightings around the world

*Population*: All UFOs

*Sample size*: The size is large enough, but we are not told the source of this data. Specifically, we are not told what constitutes a UFO sighting. Are the researchers using newspaper accounts, TV reports, government reports, or just word-of-mouth accounts?

*Potential bias*: We are not given any information on how the researchers analyze the data. For example, what do they mean by “have never been adequately explained by any government agency”? The criteria they use for this dismissal may bias their interpretation of the data. The same problem occurs for their claim that the number of “hoaxes” is insignificant. It does not allow for the possibility of evidence, or a different interpretation of the data, that would go against their claims. These potential biases reduce the likelihood that the sample is representative of the population.

*Randomness*: We are not told that the sample was randomly gathered. The researchers may have purposely chosen examples to support their contention, and disregarded any negative evidence. This reduces the likelihood that the sample is representative of the population.

**5.** *Sample*: 6000 urban public high school seniors throughout the United States

*Population*: All U.S. high school seniors

*Sample size*: The sample is large and taken from throughout the United States. This raises the likelihood that the sample is representative of the population.

*Potential bias*: The sample excludes private high schools. This reduces the likelihood that the sample is representative of the population, because there may be a significant difference between the two groups’ test scores. Their exclusion from the study may bias the results in one direction or another.

*Randomness*: This is a random sample. This raises the likelihood that the sample is representative of the population. However, it would be better for the researchers to generalize to public school seniors because that is what they studied.

**6.** *Sample*: Ninety-three wars that took place in the last 200 hundred years

*Population*: All future wars

*Sample size*: The size is adequate, if it in fact represents a substantial number of “wars.” It would help if we knew more about the researcher’s inclusion or exclusion of conflicts.

*Potential bias*: The sample is based on the researcher’s own definition of war. For all we, know the researcher may have included cases of a citizen of one country killing a citizen of another country. He or she may have counted as a civil war any riots or public displays of protest that led to a killing. It is always possible to find data to support any claim if you purposely include data for your claim and exclude data that goes against your claim. These concerns reduce the likelihood that the sample is representative of the population.

*Randomness*: We are not given any indication that this is a random sample. This reduces the likelihood that the sample is representative of the population.

**7.** *Sample*: People chosen for a soft drink survey

*Population*: All Americans

*Sample size*: We are not given the sample size. Eight out of 10 might refer to just 10 people in the survey, or 1000. If the sample is small (which seems likely), then we are discussing a very small sample when we are generalizing to millions of people. This reduces the likelihood that the sample is representative of the population.

*Potential bias*: The sample may be one of a series of surveys done. It is not unusual for a company to do many such surveys until they find one that gives the results they are seeking. The published results may have excluded any evidence against the claim. This potential bias reduces the likelihood that the sample is representative of the population.

*Randomness*: This is not a random sample. This reduces the likelihood that the sample is representative of the population.

**8.** *Sample*: The dreams of thirty volunteers for 1 year

*Population*: All humans’ dreams

*Sample size*: Although the amount of dreams is quite large, the sample of people used to get those dreams is quite small. This reduces the likelihood that the sample is representative of the population.

*Potential bias*: The sample includes only the dreams that the subjects either remembered or felt like revealing to the researchers. Since the data represents only what the subjects decide to reveal, it has the potential of being biased. In addition, the researchers are dependent on the accuracy and honesty of the subjects, both of which can be a source of bias. This reduces the likelihood that the sample is representative of the population.

*Randomness*: This is not a random sample. This reduces the likelihood that the sample is representative of the population.

**9.** *Sample*: The results of the World Series from 1903 to 2008 and the correlation to sales

*Population*: Future World Series winners

*Sample size*: The sample includes the results from 105 years. The sample is certainly large enough to be representative of the past winners, since it includes nearly the entire past population. However, since the researchers are projecting into the future, there may be reasons to think that future society may not be the same as that represented in the sample.

*Potential bias*: The sample clearly shows a past trend. However, since we know that cigarette and liquor sales are affected by many social factors, the extended trend may be a simple correlation and not an indication of any real connection (the issue of *correlation* will be explored further in the next chapter). This allows us to question the likelihood that the sample is representative of the population.

*Randomness*: In a sense, randomness is not an issue here. If all the World Series results are included, then no data are missing regarding other World Series.

**10.** *Sample*: 20,000 students who were accepted to law and medical schools in the United States for the past 20 years

*Population*: All students planning to apply to law or medical school

*Sample size*: The sample is very large and seems to include students from all areas of the United States. This increases the likelihood that the sample is representative of the population.

*Potential bias*: The sample does not check for other possible contributing factors, such as the number of students having completed law and medical school preparation classes outside the university setting. We would also need to be provided with evidence that philosophy majors “tend to score higher” on the logical reasoning section. The researchers seem to explain the acceptance rate based on the results of one section of the tests.

*Randomness*: We are not told if the sample is random or not. Although the sample is large, a random sample would reduce the likelihood of getting a non-representative sample.

***Exercises 13B***

**I.**

**1.** Mean: 5.17; Median: 4; Mode: 4

**2.** Mean: 8.14; Median: 9; Mode: 11

**3.** Mean: 6.44; Median: 5; Modes: 3, 13

**4.** Mean: 25; Median: 25; Mode: No mode

**5.** Mean: 186.7; Median: 200; Mode: 200

**II.**

**1.** Mean: $52,500; Median: $42,000; Mode: $42,000

**2.** Mean: $1,660,000; Median: $100,000; Mode: $100,000

**3.** Mean: $10; Median: $10; Mode: $10

**4.** Mean: $2,159.50; Median: $427; Mode: No mode

**5.** Mean: $1,200.30; Median: $1,000; Mode: $1,000

**III.**

**1.** Mean: 3.15; Median: 3.26; Mode: 3.26

**2.** Mean: 2.82; Median: 2.73; Mode: No mode

**3.** Mean: 3.45; Median: 3.25; Mode: 3.25

**4.** Mean: 2.50; Median: 2.50; Mode: No mode

**5.** Mean: 3.13; Median: 3.16; Mode: 3.16

**IV.**

**1.** Mean: 61.14"; Median: 67"; Mode: 67"

**2.** Mean: 44.67"; Median: 44.5"; Mode: No mode

**3.** Mean: 27"; Median: 27"; Mode: 27"

**4.** Mean: 33"; Median: 33"; Mode: No mode

**5.** Mean: 55.29"; Median: 74"; Modes: 74", 80"

***Exercises 13C***

**I.**

**1.** The standard deviation is 3.13.

Step 1: 5.17

Step 2: -3.17; -2.17; -1.17; -1.17; 2.83; 4.83

Step 3: 10.05, 4.71, 1.37, 1.37, 8.01, 23.33

Step 4: 48.84

Step 5: 9.77

Step 6: 3.13

**2.** The standard deviation is 3.24

Step 1: 8.14

Step 2: -5.14; -3.14; -1.14; 0.86; 2.86; 2.86; 2.86

Step 3: 26.42; 9.86; 1.3; 0.74; 8.18; 8.18; 8.18

Step 4: 62.86

Step 5: 10.48

Step 6: 3.24

**3.** The standard deviation is 4.77.

Step 1: 6.44

Step 2: -5.44; -4.44; -3.44; -3.44; -1.44; 0.56; 4.56; 6.56; 6.56

Step 3: 29.6; 19.71; 11.83; 11.83; 2.07; 0.31; 20.8; 43.03; 43.03

Step 4: 182.21

Step 5: 22.78

Step 6: 4.77

**4.** The standard deviation is 12.91.

Step 1: 25

Step 2: -15; -5; 5; 15

Step 3: 225; 25; 25; 225

Step 4: 500

Step 5: 166.7

Step 6: 12.91

**5.** The standard deviation is 73.67

Step 1: 186.7

Step 2: -86.7; -76.7; 13.3; 13.3; 23.3; 113.3

Step 3: 7,516.9; 5,882.9; 176.9; 176.9; 542.9; 12,836.9

Step 4: 27,133.1

Step 5: 5,426.62

Step 6: 73.67

**II.**

**1.** The standard deviation is 48,093.66

Step 1: 52500

Step 2: -46,500; -10,500; -10,500; 67,500

Step 3: 2,162,250,000; 110,250,000; 110,250,000; 4,556,250,000

Step 4: 6,939,000,000

Step 5: 2,313,000,000

Step 6: 48,093.66

**2.** The standard deviation is 56,901.67

Step 1: 1,660,000

Step 2: -1,560,000, -1,560,000, -1,560,000, 1,340,000; 3,340,000

Step 3: 2,433,600,000; 2,433,600,000; 2,433,600,000; 1,795,600,000; 11,155,600,000

Step 4: 12,951,200,000

Step 5: 3,237,800,000

Step 6: 5,6901.67

**3.** The standard deviation is 0.

Step 1: 10

Step 2: 0, 0, 0, 0

Step 3: 0, 0, 0, 0

Step 4: 0

Step 5: 0

Step 6: 0

**4.** The standard deviation is 3,761.10.

Step 1: 2,159.50

Step 2: -2,152.50; -2,082.50; -1,382.50; 5,617.50

Step 3: 4,633,256.25; 4,336,806.25; 1,911,306.25; 31,556,306.25

Step 4: 42,437,675.00

Step 5: 14,145,891.67

Step 6: 3,761.10

**5.** The standard deviation is 1,642.90.

Step 1: 1,200.30

Step 2: -1199.80; -1199.30; -200.30; -200.30; 2799.70

Step 3: 1,439,520.04; 1,438,320.49; 40,120.09; 40,120.09; 7,838,320.09

Step 4: 10,796,400.80

Step 5: 2,699,100.20

Step 6: 1,642.90

**III.**

**1.** The standard deviation is 0.72.

Step 1: 3.15

Step 2: -1.01; -0.66; 0.11; 0.11; 0.63; 0.84

Step 3: 1.02; 0.44; 0.01; 0.01; 0.40; 0.71

Step 4: 2.59

Step 5: 0.52

Step 6: 0.72

**2.** The standard deviation is 0.82.

Step 1: 2.82

Step 2: -0.94; -0.79; -0.26; 0.01; 0.82; 1.07

Step 3: 0.88; 0.62; 0.07; 0.01; 0.67; 1.14

Step 4: 3.39

Step 5: 0.68

Step 6: 0.82

**3.** The standard deviation is 0.28.

Step 1: 3.45

Step 2: -0.20; -0.20; -0.20; 0.30; 0.30

Step 3: 0.04; 0.04; 0.04; 0.09; 0.09

Step 4: 0.30

Step 5: 0.08

Step 6: 0.28

**4.** The standard deviation is 1.22.

Step 1: 2.50

Step 2: -1.5; -0.5; 0.05; 1.5

Step 3: 2.25; 0.0025; 0.0025; 2.25

Step 4: 4.51

Step 5: 1.50

Step 6: 1.22

**5.** The standard deviation is 0.17.

Step 1: 3.13

Step 2: -0.27; -0.17; 0.03; 0.03; 0.13; 0.23

Step 3: 0.07; 0.03; 0.00; 0.00; 0.02; 0.05

Step 4: 0.17

Step 5: 0.03

Step 6: 0.17

**IV.**

**1.** The standard deviation is 18.20.

Step 1: 61.14

Step 2: -24.14; -16.14; -13.14; 5.86; 5.86; 16.86; 24.86

Step 3: 582.74; 260.50; 172.66; 34.34; 34.34; 284.26; 618.02

Step 4: 1986.86

Step 5: 331.14

Step 6: 18.20

**2.** The standard deviation is 5.28.

Step 1: 44.67

Step 2: -6.67; -4.67; -0.67; 0.33; 4.33; 7.33

Step 3: 44.49; 21.81; 0.45; 0.11; 18.75; 53.73

Step 4: 139.34

Step 5: 27.87

Step 6: 5.28

**3.** The standard deviation is 2.19.

Step 1: 27

Step 2: -4; 0; 0; 0; 2; 2

Step 3: 16; 0; 0; 0; 4; 4

Step 4: 24

Step 5: 4.8

Step 6: 2.19

**4.** The standard deviation is 2.16.

Step 1: 33

Step 2: -3; -2; -1; 0; 1; 2; 3

Step 3: 9; 4; 1; 0; 1; 4; 9

Step 4: 28

Step 5: 4.67

Step 6: 2.16

**5.** The standard deviation is 27.22.

Step 1: 55.29

Step 2: -31.29; -28.29; -27.29; 18.71; 18.71; 24.71; 24.71

Step 3: 979.06; 800.32; 744.74; 350.06; 350.06; 610.58; 610.58

Step 4: 4445.40

Step 5: 740.90

Step 6: 27.22

***Exercises 13E***

**1.** Since the Bureau of Labor Statistics has access to large amounts of data, we can assume that the size of the data set is adequate. The *median* weekly income for persons with a bachelor’s degree was $1,025 (in 2009); therefore, we know that 50% of that data set made more than that amount and 50% made less. For those with only a high school diploma, 50% of that data set made more than $626 a week and 50% made less.

On the surface, the difference between $1,025 and $626 is substantial. However, we are not told either the *mean* or the *standard deviation* in the two sets of data. Without that information, we cannot determine the amount of diversity in the sets. In addition, people with bachelor’s degrees in engineering and computer science average three to four times the yearly salary of many social science majors. The article does not distinguish between majors. It therefore may misleadingly suggest that any kind of bachelor’s degree puts you in a position to make substantially more than a person with only a high school diploma.

**2.** The group of Best Actress Oscar winners is quite small, and the group of nonwinners is at least four times as large. In addition, the two groups probably differ in many ways other than being Oscar winners. Also, we are not told the mean value and the standard deviation; therefore, without that information, we cannot determine the amount of diversity in the sets.

**3.** It would be nice to know the number of reported decisions in which judges granted a new trial,

denied a request for a new trial, or overturned a verdict, in whole or in part, because of non-Internet actions to see how large the problem is when compared to the total number of decisions. We are also told that “In three-quarters of the cases in which judges declined to declare mistrials, they nevertheless found Internet-related misconduct on the part of jurors.” So, for 75% of those cases the “misconduct” did not reach a level that would cause a judge to act. Finally, the claim that “These figures do not include the many incidents that escape judicial notice” does not help the analysis. Obviously, if an incident escapes judicial notice it cannot be included in the study. The statement seems to hint that the undetected cases are substantial, but no evidence is offered to support this.

**4.** Some background information is helpful here: A picogram is one thousand times smaller than a nanogram. If Contador was found to have 50 picograms of a banned substance, then that would be equivalent to 0.05 nanograms. Christiane Ayotte claims that tainted beef contains at most 1 nanogram of Clenbuterol. Given this, if Contador ate the beef, then he would have had far more than 50 picograms in his sample. Therefore, the Clenbuterol must have come from some other source.

**5.** First, we are told that the “hundreds of millions of dollars” are wasted, presumably because of late penalties incurred with the IRS. Second, “workers have forgone huge amounts of money in matching 401(k) contributions because they never got around to signing up for a retirement plan.” However, no accurate figures are given to support this claim. Third, we are told that “Seventy percent of patients

suffering from glaucoma risk blindness because they don’t use their eyedrops regularly.” However, no information is given to show how this figure was arrived at; we are not told the kind of study, the sample size, or whether it was random. Fourth, the claim that “Procrastination also inflicts major costs on businesses and governments” has no supporting evidence. Also, the term “major costs” is vague. Finally, the claim that “the bankruptcy of General Motors was due in part to executives’ penchant for delaying tough decisions” has no supporting evidence.

**6.** The comparisons of the 2004 Smarty Jones record crowd of 120,139 at Belmont, as compared with the 72,739 at Churchill Downs to see Zenyatta requires some background information. Belmont Park is in New York, while Churchill Downs is in Kentucky. Attendance for the Belmont Stakes has generally been higher than the Breeders’ Cup. Also, when a Triple Crown is on the line attendance at Belmont is at its highest. The attendance for Zenyatta’s Breeders’ Cup year was the highest of any Breeders’ Cup year that preceded it.

The number of articles comparison may not be the best guide. For example, any year in which a Triple Crown is on the line generates a lot of interest. Given this, perhaps it is better not to compare a potential Triple Crown horse with Zenyatta’s, but instead compare Zenyatta’s Breeder’s Cup with other years.

**7.** The report shows how statistics were taken out of context and used misleadingly. The comparison of New York to New Jersey was flawed in that the study did not consider the high cost of running a retail store in New York as compared with New Jersey. Therefore, the 10% difference should not have been blamed on price tags. This mistake was compounded when the Coalition for Retail Pricing Modernization applied the unsubstantiated claim regarding New York/New Jersey to Michigan stores. Their claim that price tags cost Michigan stores $2.2 billion annually is not supported.