Chapter 1

Introduction to Statistics

Section 1-2, Basic Skills and Concepts

1. Statistical significance is indicated when methods of statistics are used to reach a conclusion that some treatment or finding is effective, but common sense might suggest that the treatment or finding does not make enough of a difference to justify its use or to be practical. Yes, it is possible for a study to have statistical significance but not a practical significance.
2. If the source of the data can benefit from the results of the study, it is possible that an element of bias is introduced so that the results are favorable to the source.
3. A voluntary response sample is a sample in which the subjects themselves decide whether to be included in the study. A voluntary response sample is generally not suitable for a statistical study because the sample may have a bias resulting from participation by those with a special interest in the topic being studied.
4. Even if we conduct a study and find that there is a correlation, or association, between two variables, we cannot conclude that one of the variables is the cause of the other.
5. There does appear to be a potential to create a bias.
6. There does not appear to be a potential to create a bias.
7. There does not appear to be a potential to create a bias.
8. There does appear a potential to create a bias.
9. The sample is a voluntary response sample and is therefore flawed.
10. The sample is a voluntary response sample and is therefore flawed.
11. The sampling method appears to be sound.
12. The sampling method appears to be sound.
13. Because there is a 30% chance of getting such results with a diet that has no effect, it does not appear to have statistical significance, but the average loss of 45 pounds does appear to have practical significance.
14. Because there is only a 1% chance of getting the results by chance, the method appears to have a statistical significance. The result of 540 boys in 1000 births is above the approximately 50% rate expected by chance, but it does not appear to be high enough to have practical significance. Not many couples would bother with a procedure that raises the likelihood of a boy from 50% to 54%.
15. Because there is a 23% chance of getting such results with a program that has no effect, the program does not appear to have statistical significance. Because the success rate of 23% is not much better than the 20% rate that is typically expected with random guessing, the program does not appear to have practical significance.
16. Because there is a 25% chance of getting such results with a program that has no effect, the program does not appear to have statistical significance. Because the average increase is only 3 IQ point, the program does not appear to have practical significance.
17. The male and female pulse rates in the same column are not matched in any meaningful way. It does not make sense to use the difference between any of the pulse rates that are in the same column.
18. Yes, the source of the data is likely to be unbiased.
19. The data can be used to address the issue of whether males and females have pulse rates with the same average (mean) value.
20. The results do not prove that the populations of males and females have the same average (mean) pulse rate. The results are based on a particular sample of five males and five females, and analyzing other samples might lead to a different conclusion. Better results would be obtained with larger samples.

21. Yes, each IQ score is matched with the brain volume in the same column, because they are measurements obtained from the same person. It does not make sense to use the difference between each IQ score and the brain volume in the same column, because IQ scores and brain volumes use different units of measurement. For example, it would make no sense to find the difference between an IQ score of 87 and a brain volume of 1035 cm^3.

22. The issue that can be addressed is whether there is a correlation, or association, between IQ score and brain volume.

23. Given that the researchers do not appear to benefit from the results, they are professionals at prestigious institutions, and funding is from a U.S. government agency, the source of the data appears to be unbiased.

24. No. Correlation does not imply causation, so a statistical correlation between IQ score and brain volume should not be used to conclude that larger brain volumes cause higher IQ scores.

25. It is questionable that the sponsor is the Idaho Potato Commission and the favorite vegetable is potatoes.

26. The sample is a voluntary response sample, so there is a good chance that the results are not valid.

27. The correlation, or association, between two variables does not mean that one of the variables is the cause of the other. Correlation does not imply causation.

28. The correlation, or association, between two variables does not mean that one of the variables is the cause of the other. Correlation does not imply causation.

29. a. The number of people is $(0.39)(1018) = 397.02$.

    b. No. Because the result is a count of people among 1018 who were surveyed, the result must be a whole number.

    c. The actual number is 397 people.

    d. The percentage is $\frac{255}{1018} = 0.25049 = 25.049\%$.

30. a. The number of women is $(0.38)(427) = 162.26$.

    b. No. Because the result is a count of women among 427 who were surveyed, the result must be a whole number.

    b. The actual number is 162 women.

    d. The percentage is $\frac{30}{427} = 0.07026 = 7.026\%$.

31. a. The number of adults is $(0.14)(2302) = 322.28$

    b. No. Because the result is a count of adults among 2302 who were surveyed, the result must be a whole number.

    c. The actual number is 322 adults.

    d. The percentage is $\frac{46}{2302} = 0.01998 = 1.998\%$.

32. a. The number of adults is $(0.76)(2513) = 1909.88$.

    b. No. Because the result is a count of adults among 2513 who were surveyed, the result must be a whole number.

    b. The actual number is 1910 adults.

    d. The percentage is $\frac{327}{2513} = 0.13012 = 13.012\%$.

33. Because a reduction of 100% would eliminate all of the size, it is not possible to reduce the size by 100% or more.

34. If the Club eliminated all car thefts, it would reduce the odds of car theft by 100%, so the 400% figure is impossible.

35. If foreign investment fell by 100% it would be totally eliminated, so it is not possible for it to fall by more than 100%.

36. Because a reduction of 100% would eliminate all plague, it is not possible to reduce it by more than 100%.

Section 1-2, Beyond the Basics

37. Without our knowing anything about the number of ATVs in use, or the number of ATV drivers, or the amount of ATV usage, the number of 740 fatal accidents has no context. Some information should be given so that the reader can understand the rate of ATV fatalities.

38. All percentages of success should be multiples of 5. The given percentage cannot be correct.

39. The wording of the question is biased and tends to encourage negative response. The sample size of 20 is too small. Survey respondents are self-selected instead of being selected by the newspaper. If 20 readers respond, the percentages should be multiples of 5, so 87% and 13% are not possible results.

Section 1-3, Basic Skills and Concepts

1. A parameter is a numerical measurement describing some characteristic of a population, whereas a statistic is a numerical measurement describing some characteristic of a sample.

2. Quantitative data consist of numbers representing counts or measurements, whereas categorical data can be separated into different categories that are distinguished by some characteristic that is not numerical.

3. Parts (a) and (c) describe discrete data.

4. The values of 1010 and 55% are both statistics because they are based on the sample. The population consists of all adults in the United States.

5. Statistic
6. Parameter
7. Parameter
8. Statistic
9. Parameter
10. Parameter
11. Statistic
12. Statistic
13. Continuous
14. Discrete
15. Discrete
16. Continuous
17. Discrete
18. Discrete
19. Continuous
20. Continuous
21. Nominal
22. Ratio
23. Interval
24. Ordinal
25. Ratio
26. Nominal
27. Ordinal
28. Interval

29. The numbers are not counts or measures of anything, so they are at the nominal level of measurement, and it makes no sense to compute the average (mean) of them.

30. The flight numbers do not count or measure anything. They are at the nominal level of measurement, and it does not make sense to compute the average (mean) of them.

31. The numbers are used as substitutes for the categories of low, medium, and high, so the numbers are at the ordinal level of measurement. It does not make sense to compute the average (mean) of such numbers.

32. The numbers are substitutes for names and are not counts or measures of anything. They are at the nominal level of measurement, and it makes no sense to compute the average (mean) of them.

Section 1-3, Beyond the Basics

33. a. Continuous, because the number of possible values is infinite and not countable.

    b. Discrete, because the number of possible values is finite.

    c. Discrete, because the number of possible values is finite.

    d. Discrete, because the number of possible values is finite and countable.

34. Either ordinal or interval is a reasonable answer, but ordinal makes more sense because differences between values are not likely to be meaningful. For example, the difference between a food rated 1 and a food rated 2 is not necessarily the same as a difference between a food rated 9 and a food rated 10.

35. With no natural starting point, temperatures are at the interval level of measurement so ratios such as "twice" are meaningless.

Section 1-4, Basic Skills and Concepts

1. No. Not every sample of the same size has the same chance of being selected. For example, the sample with the first two names has no chance of being selected. A simple random sample of n items is selected in such a way that every sample of the same size has the same chance of being selected.

2. In an observational study, you would examine subjects who consume fruit and those who do not. In the observational study, you run a greater risk of having a lurking variable that affects weight. For example, people who consume more fruit might be more likely to maintain generally better eating habits, and they might be more likely to exercise, so their lower weights might be due to these better eating and exercise habits, and perhaps fruit consumption does not explain lower weights. An experiment would be better, because you can randomly assign subjects to the fruit treatment group and the group that does not get the fruit treatment (control condition), so lurking variables are less likely to affect the results.

3. The population consists of the adult friends on the list. The simple random sample is selected from the population of adult friends on the list, so the results are not likely to be representative of the much larger general population of adults in the United States.

4. Because there is nothing about left-handedness or right-handedness that would affect being in the author's classes, the results are likely to be typical of the population. The results are likely to be good, but, in general, a convenience sample is not a good choice.

5. Because the research participants are subjected to anger and confrontation, they are given a form of treatment, so this is an experiment, not an observational study.

6. Because the subjects were given a treatment consisting of Lipitor, this is an experiment.

7. This is an observational study because the therapists were not given any treatment. Their responses were observed.

8. This is an observational study because the survey respondents were not given any treatment. Their responses were observed.

9. Cluster
10. Convenience
11. Random
12. Systematic
13. Convenience
14. Random
15. Systematic
16. Cluster
17. Random
18. Cluster
19. Convenience
20. Systematic

21. The sample is not a simple random sample. Because every 1000^{th} pill is selected, some samples have no chance of being selected. For example, a sample consisting of two consecutive pills has no chance of being selected, and this violates the requirement of a simple random sample.

22. The sample is not a simple random sample. Not every sample of 1500 adults has the same chance of being selected. For example, a sample of 1500 women has no chance of being selected.
23. The sample is a simple random sample. Every sample of size 500 has the same chance of being selected.
24. The sample is a simple random sample. Every sample of the same size has the same chance of being selected.
25. The sample is not a simple random sample. Not every sample has the same chance of being selected. For example, a sample that includes people who do not appear to be approachable has no chance of being selected.
26. The sample is not a simple random sample. Not all samples of the same size have the same chance of being selected. For example, a sample would not be selected which included people who do not appear to be approachable.

Section 1-4, Beyond the Basics

27. Prospective study
28. Retrospective study
29. Cross-sectional study
30. Prospective study
31. Matched pairs design
32. Randomized block design
33. Completely randomized design
34. Matched pairs design
35. Blinding is a method whereby a subject (or person who evaluates results) in an experiment does not know whether the subject is treated with the DNA vaccine or the adenoviral vector vaccine. It is important to use blinding so that results are not somehow distorted by knowledge of the particular treatment used.
36. **Prospective**: The experiment was begun and results were followed forward in time. **Randomized**: Subjects were assigned to the different groups through the process of random selection, and whereby they had the same chance of belonging to each group. **Double-blind**: The subjects did not know which of the three groups they were in, and the people who evaluated results did not know either. **Placebo-controlled**: There was a group of subjects who were given a placebo, and, by comparing the placebo group to the two treatment groups, the effect of the treatments might be better understood.

Section 1-5, Basic Skills and Concepts

1. A spreadsheet is a collection of data organized in an array of cells arranged in rows and columns, and it is used to summarize, analyze, and perform calculations with the data.
2. Worksheet
3. An Excel file that contains worksheets
4. A single spreadsheet, which is a collection of data organized in an array of cells arranged in rows and columns.
5. The top half of the worksheet is deleted, then restored.
6. The top half of the worksheet is deleted, then restored.
7. The top half of the worksheet is deleted, then restored.
8. The top half of the worksheet is deleted, then restored.
9. The described worksheet is printed.
10. The described worksheet is printed.
11. The described worksheet is printed.
12. The described worksheet is printed.

Section 1-5, Beyond the Basics

13. If using Excel 2013, 2010, or 1007, click the File tab; then select Save As. If using Excel 2007, click on the Office Button; then Select Save As. Enter the new file name and location for the saved file.

14. a-d. Your Descriptive Statistics output will look similar to the output shown below.

	Column1	
2		
3	Mean	77.5
4	Standard Error	1.826795
5	Median	77
6	Mode	78
7	Standard Deviation	11.55367
8	Sample Variance	133.4872
9	Kurtosis	-0.1632
10	Skewness	0.502442
11	Range	48
12	Minimum	56
13	Maximum	104
14	Sum	3100
15	Count	40

Mean = 77.5. The mean is the arithmetic average.
Range = 48. The range is the difference between the highest and lowest pulse rates.
Minimum = 56. The minimum is the lowest pulse rate.
Maximum = 104. The maximum is the highest pulse rate.
Sum = 3100. The sum is the arithmetic sum of the pulse rates included in the sample.
Count = 40. The count is the number of pulse rates included in the sample.

Chapter Quick Quiz

1. No. The numbers do not measure anything.
2. Nominal
3. Continuous
4. Quantitative data
5. Ratio
6. False
7. No
8. Statistic
9. Observational study
10. False

Review Exercises

1. a. Discrete
   b. Ratio
   c. Stratified
   d. Cluster
   e. The mailed responses would be a voluntary response sample, so those with strong opinions are more likely to respond. It is very possible that the results do not reflect the true opinions of the population of all customers.
2. The survey was sponsored by the American Laser Centers, and 24% said that their favorite body part is the face, which happens to be a body part often chosen for some type of laser treatment. The source is therefore questionable.
3. The sample is a voluntary response sample, so the results are questionable.
4. a. It uses a voluntary response sample, and those with special interests are more likely to respond, so it is very possible that the sample is not representative of the population.
   b. Because the statement refers to 72% of all Americans, it is a parameter (but it is probably based on a 72% rate from the sample, and the sample percentage is a statistic).
   c. Observational study

5. a. If they have no fat at all, they have 100% less than any other amount with fat, so the 125% figure cannot be correct.

   b. The exact number is (0.58)(1182) = 685.56. The actual number is 686.

   c. $\frac{331}{1182} = 0.28003 = 28.003\%$

6. The Gallop poll used randomly selected respondents, but the AOL poll used a voluntary response sample. Respondents in the AOL poll are more likely to participate if they have strong feelings about the candidates, and this group is not necessarily representative of the population. The results from the Gallop poll are more likely to reflect the true opinions of American voters.

7. Because there is only a 4% chance of getting the results by chance, the method appears to have statistical significance. The results of 112 girls in 200 births is above the approximately 50% rate expected by chance, but it does not appear to be high enough to have practical significance. Not many couples would bother with a procedure that raises the likelihood of a girls from 50% to 56%.

8. a. Random

   b. Stratified

   c. Nominal

   d. Statistic, because it is based on a sample.

   e. The mailed responses would be a voluntary response sample. Those with strong opinions about the topic would be more likely to respond, so it is very possible that the results would not reflect the true opinions of the population of all adults.

9. a. Systematic

   b. Random

   c. Cluster

   d. Stratified

   e. Convenience

   f. No, although this is a subjective judgment.

10. a. (0.52)(1500) = 780 adults

    b. $\frac{345}{1500} = 0.23 = 23\%$

    c. Men: $\frac{727}{1500} = 0.485 = 48.5\%$

       Women: $\frac{773}{1500} = 0.515 = 51.5\%$

Cumulative Review Exercises

1. You will use Excel's AVERAGE function to solve the problem. Key in the formula **=AVERAGE(A2:A49)** in an Excel worksheet and press [**Enter**]. The mean is equal to 11. Because the flight numbers are not measures or counts of anything, the result does not have meaning.

2. You will use Excel's AVERAGE function to solve the problem. Key in the formula) **=AVERAGE(D2:D21** in an Excel worksheet and press [**Enter**]. The mean is equal to 101, and it is reasonably close to the population mean of 100.

3. You will use Excel's STANDARDIZE function to solve the problem. Key in the formula **=STANDARDIZE(247,176,6)** in an Excel worksheet and press [**Enter**]. The standardized score is equal to 11.83 which is an unusually high value.
4. Key in the formula **=(175-172)/(29/SQRT(20))** in an Excel worksheet and press [**Enter**]. The result is 0.46.
5. Key in the formula **=1.96^2*0.25/0.03^2** in an Excel worksheet and press [**Enter**]. The result is 1067.
6. Key in the formula **=(88-88.57)^2/88.57** in an Excel worksheet and press [**Enter**]. The result is 0.0037.
7. Key in the formula **=((96-100)^2+(106-100)^2+(98-100)^2)/(3-1)** in an Excel worksheet and press [**Enter**]. The result is 28.
8. Key in the formula **=SQRT(((96-100)^2+(106-100)^2+(98-100)^2)/(3-1))** in an Excel worksheet and press [**Enter**]. The result is 5.3.
9. Key in the formula **=0.6^14** in an Excel worksheet and press [**Enter**]. The result is 0.000783642.
10. Key in the formula =8^12 in an Excel worksheet and press [**Enter**]. The result is 68719476736.
11. Key in the formula **=7^14** in an Excel worksheet and press [**Enter**]. The result in scientific notation is 6.78223E+11. The result, not in scientific notation, is 678223072849. Hint: Change the format from General to Number to view the result as an ordinary number, not in scientific notation.
12. Key in the formula **=0.3^10** in an Excel worksheet and press [**Enter**]. The result in scientific notation is 5.9049E-06. The result, not in scientific notation, is 0.00000590493. Hint: Change the format from General to Number and set the number of decimal places at 10 to view the result as an ordinary number, not in scientific notation.