Print this page

4.1 INTRODUCING HYPOTHESIS TESTS

DATA 4.1 Do Dogs Resemble their Owners?

© Kevin Klöpper/iStockphoto

You may have seen dogs that look like their owners, but is this just a coincidence or do dogs really tend to resemble their owners? To investigate this question statistically, we need data. Roy and Christenfeld[1]Roy, M. and Christenfeld, N., "*Do Dogs Resemble their Owners?*," *Psychological Science*, 2004; 15(5): 361-363. conducted a study testing people's ability to pair a dog with its owner. Pictures were taken of 25 owners and their purebred dogs, selected at random from dog parks. Study participants were shown a picture of an owner together with pictures of two dogs (the owner's dog and another random dog from the study) and asked to choose which dog most resembled the owner. Of the 25 owners, 16 were paired with the correct dog.[2]Each dog-owner pair was viewed by 28 naive undergraduate judges, and the pairing was deemed "correct" if the majority of judges (more than 14) chose the correct dog to go with the owner. Is this convincing evidence that dogs tend to resemble their owners?

To address this question, let's think about what might happen if a dog's looks are completely unrelated to its owner. In this case, the participants' choices would be no better than random guesses for each pair of dogs. Since there are two possible choices, we'd expect people to choose correctly about half the time. Of course, even guessing randomly, people will not always be correct exactly half the time; sometimes they will get slightly more than half correct and sometimes slightly less. While 16 out of 25 is more than 50% correct, how do we know if this is because dogs really resemble their owners or just due to random chance?

Example 4.1

Consider each of the following numbers of hypothetical matches. For each scenario, does the evidence convince you that dogs resemble their owners? Why or why not?

(a) 25 out of 25 correct matches

(b) 10 out of 25 correct matches

(c) 13 out of 25 correct matches

Solution

(a) Observing 25 out of 25 correct matches would provide very convincing evidence that dogs resemble their owners, since this would be very unlikely to happen if participants were just guessing at random.
(b) Observing 10 out of 25 correct matches does not provide any evidence that dogs resemble their owners, since less than 50% were paired correctly.
(c) Observing 13 out of 25 correct matches (52%) is greater than 50%, but this is not convincing evidence because this could easily happen under random guessing.

In the actual study, the conclusion is not obvious. If participants are guessing randomly, how "lucky" would they have to be to get at least 64% (16/25) correct? Is this a commonplace occurrence or an extreme event? How extreme does a result have to be in order to rule out random chance? These are the types of questions we'll be discussing in this chapter.

In Data 4.1, we're using data from the sample (16 out of 25) to assess a claim about a population (do dogs really resemble their owners.) This is the essence of all statistical tests: determining whether results from a sample are convincing enough to allow us to conclude something about the population.

Statistical Tests

A **statistical test** uses data from a sample to assess a claim about a population.

Null and Alternative Hypotheses

In Chapter 3, we use data from a sample to create a confidence interval for a population parameter. In this chapter, we use data from a sample to help us decide between two competing *hypotheses* about a population. In Data 4.1, one hypothesis is that dogs really do tend to look like their owners, and the competing hypothesis is that there is no dog-owner resemblance. We make these hypotheses more concrete by specifying them in terms of a *population parameter* of interest. In this case, we are interested in the population parameter p, the proportion of all purebred dogs that can be correctly matched with their owners. If there is no resemblance, we have $p=0.5$ since guessers would be choosing randomly between two options. However, if dogs do resemble their owners, we have $p>0.5$. Which is correct: $p=0.5$ or $p>0.5$? We use the data in the sample (16 correct out of 25, giving $\hat{p}=0.64$) to try to answer this question.

We refer to the competing claims about the population as the *null hypothesis*, denoted by H_0, and the *alternative hypothesis*, denoted by H_a. The roles of these two hypotheses are *not* interchangeable. The claim for which we seek significant evidence ($p>0.5$ in the dog-owner example) is assigned to the alternative hypothesis. Usually, the null hypothesis is a claim that there really is “no effect” or “no difference.” For the test of dogs resembling owners, where p is the true proportion of correct dog/owner matches, the hypotheses are

$$\begin{aligned} H_0&: \quad p=0.5 \\ H_a&: \quad p>0.5 \end{aligned}$$

In many cases, the null hypothesis represents the status quo or that nothing interesting is happening. The alternative is usually what the experimenter or researcher wants to establish or find evidence for. We assess the strength of evidence by assuming the null hypothesis is true and determining how unlikely it would be to see sample results as extreme as those in the original sample.

Null and Alternative Hypotheses

Null Hypothesis (H_0): Claim that there is no effect or no difference.
Alternative Hypothesis (H_a): Claim for which we seek significant evidence.

The alternative hypothesis is established by observing evidence (data) that contradicts the null hypothesis and supports the alternative hypothesis.

Note that the hypotheses are written in terms of the population parameter p, not in terms of the sample statistic $\hat{p}$. We *know* that the proportion for the sample of 25 owners, $\hat{p}=0.64$, is greater than 0.5. The key question is whether that statistic provides convincing evidence that the proportion of correct matches for all owners is more than 0.5.

In Example 4.1, matching all 25 dogs correctly with their owners would be very unlikely if the null hypothesis ($p=0.5$) were true. (Think of the chance of getting 25 heads in consecutive coin flips.) Observing 25 out of 25 correct matches would be very convincing evidence against the null hypothesis and in favor of the alternative hypothesis, supporting the claim that dogs resemble their owners. If we were to observe 10 out of 25 correct guesses ($\hat{p}=0.40$), we would have no evidence for an alternative hypothesis of $p>0.5$ since the sample statistic is less than 0.5. Seeing 13 out of 25 correct matches would support the alternative hypothesis (since $\hat{p}>0.5$), but the result would not be surprising if the null hypothesis were true. If we can't rule out the null hypothesis of $p=0.5$, we don't have enough evidence to conclude that dogs really resemble their owners.

DATA 4.2 Smiles and Leniency

© Cameron Whitman/iStockphoto

A neutral expression and a smiling expression: Which student gets the harsher punishment?

Can a simple smile have an effect on punishment assigned following an infraction? LeFrance and Hecht[3]LeFrance, M. and Hecht, M. A., "*Why Smiles Generate Leniency*," *Personality and Social Psychology Bulletin*, 1995; 21: 207-214. conducted a study examining the effect of a smile on the leniency of disciplinary action for wrongdoers. Participants in the experiment took on the role of members of a college disciplinary panel judging students accused of cheating. For each suspect, along with a description of the offense, a picture was provided with either a smile or neutral facial expression. A leniency score was calculated based on the disciplinary decisions made by the participants. The full data can be found in **Smiles**. The experimenters have prior knowledge that smiling has a positive influence on people, and they are testing to see if the average lenience score is higher for smiling students than it is for students with a neutral facial expression (or, in other words, that smiling students are given more leniency and milder punishments.)

Example 4.2

In testing whether smiling increases leniency, define the relevant parameter(s) and state the null and alternative hypotheses.

Solution ▶

We are comparing two means in this test, so the relevant parameters are μ_s, the true mean score for smiling students, and μ_n, the true mean score for neutral students. We are testing to see if there is evidence that the average leniency score is higher for smiling students, so the alternative hypothesis is $\mu_s > \mu_n$. The null hypothesis is that facial expression has no effect on the punishment given, so the two means are equal:

$$H_0: \quad \mu_s = \mu_n$$
$$H_a: \quad \mu_s > \mu_n$$

Example 4.3

In Example 4.2, we are testing to see if the leniency score is higher for smiling students. For the two other scenarios described below, state the null and alternative hypotheses.

(a) The experimenters have no prior beliefs about the effect of smiling on leniency and are testing to see if facial expression has any effect.

(b) The experimenters believe that during a hearing for an offense such as cheating, a disciplinary panel will view smiling as arrogant and disrespectful. They are testing to see if there is evidence that smiling will cause harsher punishments (less leniency).

Solution

(a) We are testing to see if there is evidence that the average score for smiling students is different (in either direction) from the average score for neutral students, so the alternative hypothesis is $\mu_s \neq \mu_n$. The null hypothesis is still "no effect." We have

$$\begin{aligned} H_0&: \quad \mu_s = \mu_n \\ H_a&: \quad \mu_s \neq \mu_n \end{aligned}$$

(b) We are testing to see if there is evidence that the average score for smiling students is less than the average score for neutral students, so the alternative hypothesis is $\mu_s < \mu_n$. The null hypothesis is still "no effect." We have

$$\begin{aligned} H_0&: \quad \mu_s = \mu_n \\ H_a&: \quad \mu_s < \mu_n \end{aligned}$$

Notice that, in general, the null hypothesis is a statement of equality, while the alternative hypothesis contains a range of values, using notation indicating greater than, not equal to, or less than. It is relatively straightforward to assess evidence against a statement of equality. In a hypothesis test, we measure evidence against the null hypothesis and for the alternative hypothesis.

In each case in Examples 4.2 and 4.3, the choice of hypotheses is made prior to the analysis of data. While the null hypothesis of "no difference" is the same in each case, the alternative hypothesis depends on the question of interest. In general, the question of interest, and therefore the null and alternative hypotheses, should be determined before any data are examined. In analyzing this study about cheating (or in any situation), we would be cheating in the statistical analysis if we used our sample data to determine our hypotheses!

In any of these examples, we could also phrase the null hypothesis as simply "Smiling has no effect on leniency scores" rather than the more specific claim that the means are equal. For the sake of simplicity in this book, we will generally choose to express hypotheses in terms of parameters, even when the hypothesis is actually more general, such as "no effect."

In Example 4.2, we describe a hypothesis test comparing two means. In Data 4.1 about dogs resembling their owners, we describe a test for whether a single proportion is greater than 0.5. Just as we discussed confidence intervals for any population parameter in Chapter 3, statistical tests can apply to any population parameter(s). In the next example, we consider a hypothesis test for a correlation.

DATA 4.3 Do Teams with Malevolent Uniforms Get More Penalties?

Garrett Ellwood/GettyImages, Inc.

Tom Hauck/Getty Images, Inc.

Most and least malevolent NFL team logos

Frank and Gilovich[4]Frank, M.G. and Gilovich, T., "*The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports*," *Journal of Personality and Social Psychology*, 1988; 54(1): 74-85. describe a study of relationships between the type of uniforms worn by professional sports teams and the aggressiveness of the team. They consider teams from the National Football League (NFL) and National Hockey League (NHL). Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean, and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent (evil-looking) uniforms. To measure aggressiveness, the authors used the amount of penalties (yards for football and minutes for hockey) converted to z-scores and averaged for each team over the seasons from 1970 to 1986. The data are shown in Table 4.1 and stored in **MalevolentUniformsNFL** and **MalevolentUniformsNHL**.

Table 4.1 ***Malevolence rating of uniforms and z-scores for penalties***

NFLTeam	Malevolence	ZPenYds	NHLTeam	Malevolence	ZPenMin
LA Raiders	5.10	1.19	Vancouver	5.33	0.88
Pittsburgh	5.00	0.48	Philadelphia	5.17	2.01
Cincinnati	4.97	0.27	Boston	5.13	0.42
New Orleans	4.83	0.10	New Jersey	4.45	−0.78
Chicago	4.68	0.29	Pittsburgh	4.27	0.64

Kansas City	4.58	−0.19	Chicago	4.18	−0.02
Washington	4.40	−0.07	Montreal	4.18	−0.70
St. Louis	4.27	−0.01	Detroit	4.15	0.44
NY Jets	4.12	0.01	Edmonton	4.15	0.58
LA Rams	4.10	−0.09	Calgary	4.13	−0.40
Cleveland	4.05	0.44	LA Kings	4.05	−0.20
San Diego	4.05	0.27	Buffalo	4.00	−0.68
Green Bay	4.00	−0.73	Minnesota	4.00	−0.11
Philadelphia	3.97	−0.49	NY Rangers	3.90	−0.31
Minnesota	3.90	−0.81	NY Islanders	3.80	−0.35
Atlanta	3.87	0.30	Winnipeg	3.78	−0.30
Indianapolis	3.83	−0.19	St. Louis	3.75	−0.09
San Francisco	3.83	0.09	Washington	3.73	−0.07
Seattle	3.82	0.02	Toronto	3.58	0.34
Denver	3.80	0.24	Quebec	3.33	0.41
Tampa Bay	3.77	−0.41	Hartford	3.32	−0.34
New England	3.60	−0.18			
Buffalo	3.53	0.63			
Detroit	3.38	0.04			
NY Giants	3.27	−0.32			
Dallas	3.15	0.23			
Houston	2.88	0.38			
Miami	2.80	−1.60			

Figure 4.1 shows a scatterplot with regression line of the malevolence ratings vs z-scores of the penalty yardage for the $n=28$ NFL teams in this dataset. The graph shows a somewhat positive association: Teams with more malevolent uniforms tend to have more penalty yards. In fact the most penalized team (LA Raiders, now in Oakland) had the most malevolent uniform, and the least penalized team (Miami Dolphins) had the least malevolent uniform. The sample correlation between malevolence and penalties for the 28 teams is $r=0.43$. Does this provide evidence to conclude that the true correlation is really

positive?

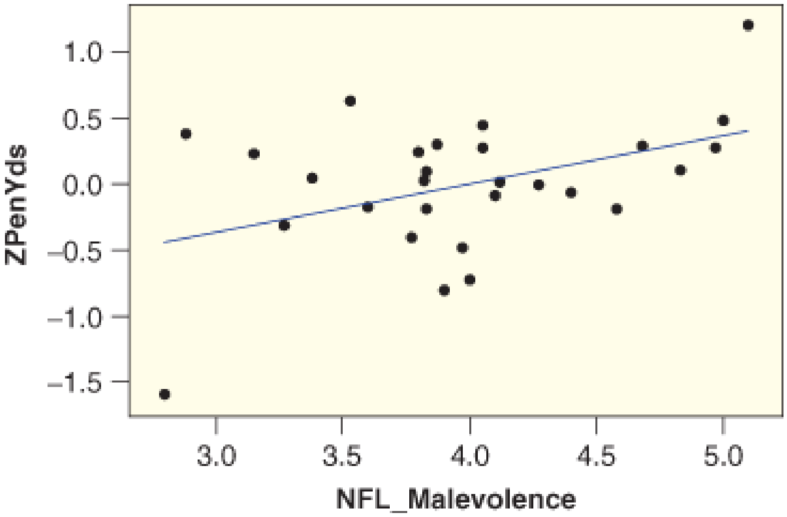

Figure 4.1 ***Relationship between penalties and uniform malevolence for NFL teams***

Example 4.4

Define the parameter of interest and state the null and alternative hypotheses.

Solution

The parameter of interest is the correlation ρ between malevolence of uniforms and number of penalty yards. We are testing to see if the correlation is positive, so the hypotheses are

$$H_0\colon \ \rho=0$$
$$H_a\colon \ \rho>0$$

Even if there were no relationship between the types of jerseys and penalties for the teams, we would not expect the correlation for any sample of teams and seasons to be *exactly* zero. Once again, the key question is whether the statistic for this sample (in this case the sample correlation r) is farther away from zero than we would reasonably expect to see by random chance alone. In other words, is it unusual to see a sample correlation as high as $r=0.43$ if the null hypothesis of $\rho=0$ is really true?

Practice Problems 4.1A

Statistical Significance

Practice Problems 4.1B

This idea, whether the sample results are more extreme than we would reasonably expect to see by random chance if the null hypothesis were true, is the fundamental idea behind statistical hypothesis tests. If data as extreme would be very unlikely if the null hypothesis were true, we say the data are *statistically significant*.[5]Statistical significance will be made more rigorous in Section 4.3. Statistically significant data provide convincing evidence against the null hypothesis in favor of the alternative, and allow us to generalize our sample results to the claim about the population.

Statistical Significance

When results as extreme as the observed sample statistic are unlikely to occur by random chance alone (assuming the null hypothesis is true), we say the sample results are **statistically significant**.

If our sample is statistically significant, we have convincing evidence against H_0 and in favor of H_a.

Example 4.5

If the sample correlation of $r=0.43$ is statistically significant, what does that mean?

Solution

If the sample data are statistically significant, it means that we have convincing evidence against H_0 and for H_a. This means we have convincing evidence that the true correlation is positive, indicating that teams with more malevolent uniforms tend to be more heavily penalized. It also means that we are unlikely to get a sample correlation as high as $r=0.43$ just by random chance if the true correlation ρ is really zero.

DATA 4.4 Divorce Opinions by Gender

© Sawayasu Tsuji/iStockphoto

Is divorce morally acceptable?

Do men and women have different views on divorce? A May 2010 Gallup poll of US citizens over the age of 18 asked participants if they view divorce as “morally acceptable.” Of the 1029 adults surveyed, 71% of men and 67% of women responded “yes.”[6]http://www.gallup.com/poll/117328/marriage.aspx.

Example 4.6

In Data 4.4, what are the population, sample, response variable, and statistical question of interest?

Solution

The population of interest is all US adults. The sample is the 1029 adults surveyed. The response variable is whether or not the respondent views divorce as morally acceptable. We observe that the sample proportions for men and women are not the same; the statistical question is whether this same phenomenon is likely to hold for the population.

Example 4.7

Define the population parameter(s) of interest and state the null and alternative hypotheses for testing whether there is a gender difference in opinions about divorce.

Solution

The parameters of interest are p_m and p_w, the proportions of men and women, respectively, who view divorce as morally acceptable. We are testing to see if there is a difference between the two proportions, so the hypotheses are

$$\begin{aligned} H_0 &: \quad p_m = p_w \\ H_a &: \quad p_m \neq p_w \end{aligned}$$

Example 4.8

Assume the 1029 adults in the Gallup survey were selected by an unbiased method that produced a random sample from the population of all US citizens. Does the fact that a higher proportion of men in the sample view divorce as morally acceptable allow us to conclude that such a difference must exist in the entire population?

Solution

No. Even if the polling methods used in Data 4.4 are perfect (i.e., participants are truly a random sample from the population), the data are still subject to sampling variability. It is possible that the difference we see in the sample proportions between men and women is just a result of random chance. (We'll examine this further in Section 4.3.) If the sample difference can be explained by random chance, then the true difference could be larger, smaller, or even in the other direction.

Example 4.9

If the 1029 adults were randomly selected from all US adults and we find that the results are statistically significant, can we conclude that males and females have different opinions about the moral acceptability of divorce?

Solution

We can never know for sure without surveying the entire population, but if the results are statistically significant, we will have strong evidence that males and females have different opinions about the moral acceptability of divorce.

Practice Problems 4.1C

DATA 4.5 Arsenic Levels in Chicken Meat

© Lauri Patterson/iStockphoto

How much arsenic is in this chicken?

Arsenic-based additives in chicken feed have been banned by the European Union but are mixed in the diet of about 70% of the 9 billion broiler chickens produced annually in the US.[7]"*Arsenic in Chicken Production*," *Chemical and Engineering News: Government and Policy*, 2007; 85(15): 34-35. Many restaurant and supermarket chains are working to reduce the amount of arsenic in the chicken they sell. To accomplish this, one chain plans to measure, for each supplier, the amount of arsenic in a random sample of chickens. The chain will cancel its relationship with a supplier if the sample provides sufficient evidence that the average amount of arsenic in chicken provided by that supplier is greater than 80 ppb (parts per billion).

Example 4.10

For the situation in Data 4.5, define the population parameter(s) and state the null and alternative hypotheses.

Solution ▶

The parameter of interest is μ, the mean arsenic level in all chickens from a supplier. We are testing to see if the mean is greater than 80, so the hypotheses are

$$H_0: \quad \mu = 80$$
$$H_a: \quad \mu > 80$$

Since we are testing to see if there is evidence that the mean is greater than 80, it is clear that the alternative hypothesis is $H_a{:}\mu{>}80$. For the null hypothesis, writing $H_0{:}\mu{\le}80$ makes intuitive sense, as any arsenic level less than 80 is satisfactory. However, it is easier to assess the extremity of our data for a single, specific value ($H_0{:}\mu{=}80$). This is a conservative choice; if the sample mean is large enough to be statistically significant when $\mu{=}80$, it would be even more significant when compared to $\mu{=}78$ or $\mu{=}75$. Thus, for convenience, we generally choose to write the null hypothesis as an equality.

Example 4.11

Suppose the chain measures arsenic levels in chickens sampled randomly from three different suppliers, with data given in Figure 4.2.

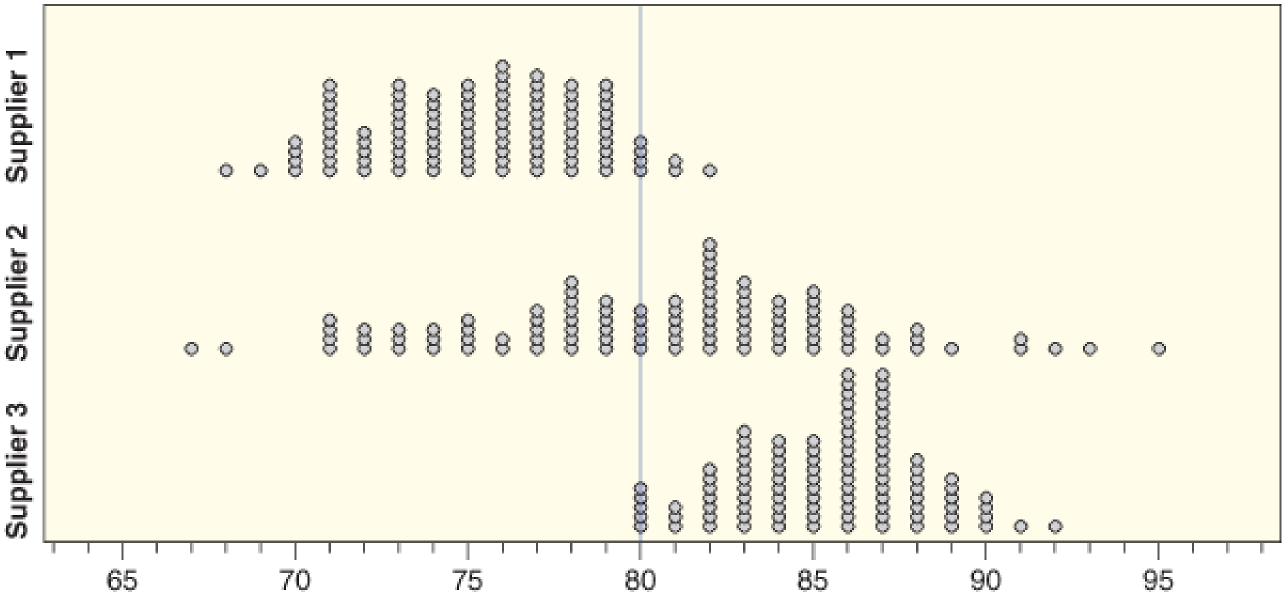

Figure 4.2 *Arsenic levels in chicken samples from three different suppliers*

(a) Which of the samples shows the strongest evidence for the alternative hypothesis?

(b) Which of the samples shows no evidence in support of the alternative hypothesis?

Solution

(a) The sample from Supplier 3 shows the strongest evidence of an average arsenic amount greater than 80, because it has the highest sample mean and all of the sampled chickens have arsenic levels at least 80.

(b) The sample from Supplier 1 shows no evidence of an average arsenic amount greater than 80, since the mean of that sample is less than 80.

Example 4.12

Under what conditions will the chain cancel its relationship with a supplier?

Solution

The chain will cancel its relationship with a supplier if there is evidence that the true mean arsenic level is greater than 80. This evidence is established if a value as high as the observed sample mean is unlikely if the true mean is 80. In other words, the chain will cancel its relationship with a supplier if the data are *statistically significant*.

Practice Problems 4.1D

In this section, we've learned that evidence for a claim about a population can be assessed using data from a sample. If the sample data are unlikely to occur just by random chance when the null hypothesis (usually "no effect") is true, then we have evidence that there is some effect and that the alternative hypothesis is true. We understand that you don't yet know how to determine what is "likely" to occur by random chance when the null hypothesis is true, and that you are probably eager to learn. That is the topic of the next section. By the end of the chapter, we'll return to the examples in this section as well as the situations described in the exercises and find out which of them are statistically significant and which aren't.

SECTION LEARNING GOALS

You should now have the understanding and skills to:

- Recognize when and why statistical tests are needed
- Specify null and alternative hypotheses based on a question of interest, defining relevant parameters
- Recognize that the strength of evidence against the null hypothesis depends on how unlikely it would be to get a sample as extreme just by random chance, if the null hypothesis were true
- Demonstrate an understanding of the concept of statistical significance

Exercises for Section 4.1

SKILL BUILDER 1

In Exercises 4.1 to 4.4, a situation is described for a statistical test and some hypothetical sample results are given. In each case:

(a) State which of the possible sample results provides the most significant evidence for the claim.
(b) State which (if any) of the possible results provide no evidence for the claim.

4.1 Testing to see if there is evidence that the population mean for mathematics placement exam scores is greater than 25. Use Figure 4.3.

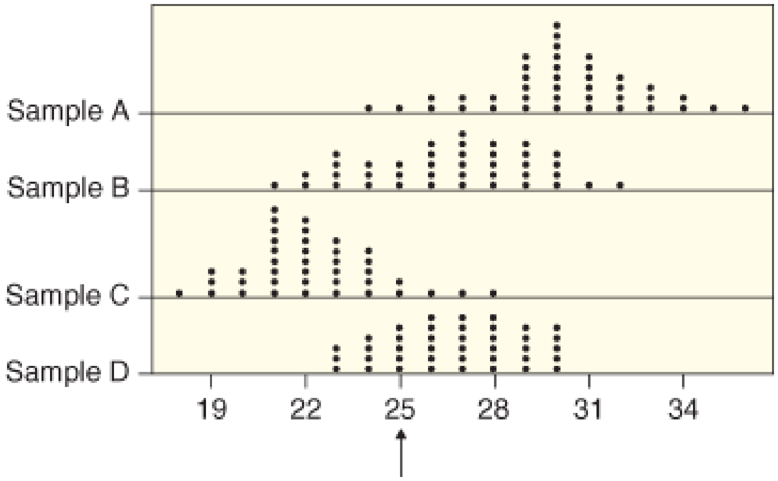

Figure 4.3 ***Samples for Exercise 4.1***

ANSWER ⊕

WORKED SOLUTION ⊕

4.2 Testing to see if there is evidence that the mean service time at Restaurant #1 is less than the mean service time at Restaurant #2. Use Figure 4.4 and assume that the sample sizes are all the same. Sample means are shown with circles on the boxplots.

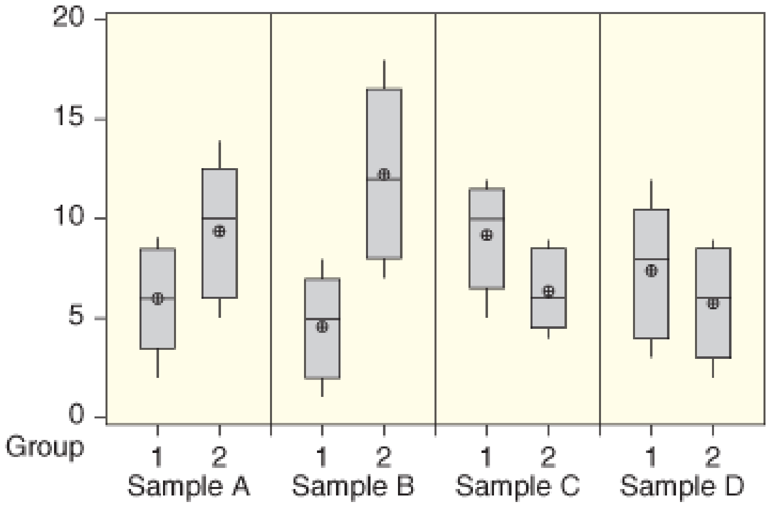

Figure 4.4 ***Samples for Exercise 4.2***

4.3 Testing to see if there is evidence that the correlation between exam grades and hours playing video games is negative for a population of students. Use Figure 4.5.

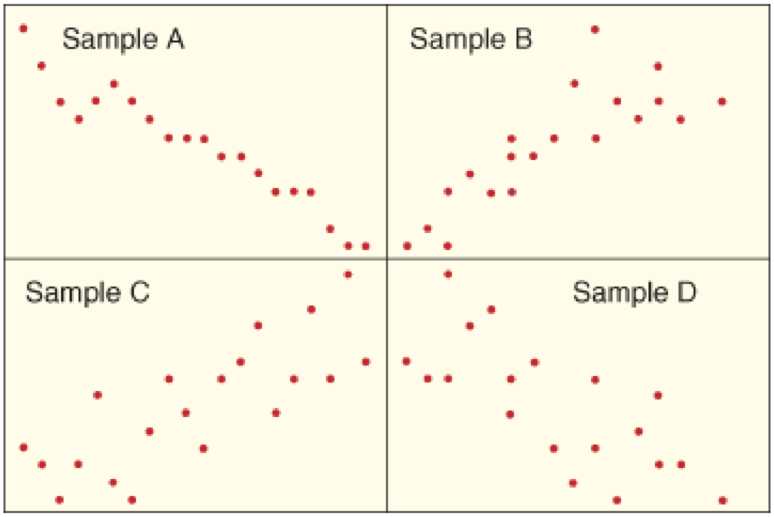

Figure 4.5 ***Samples for Exercise 4.3***

ANSWER ⊕

WORKED SOLUTION ⊕

4.4 Testing to see if there is evidence that the proportion of US citizens who can name the capital city of Canada is greater than 0.75. Use the following possible sample results:

Sample A: 31 successes out of 40

Sample B: 34 successes out of 40

Sample C: 27 successes out of 40

Sample D: 38 successes out of 40

SKILL BUILDER 2

In Exercises 4.5 to 4.8, state the null and alternative hypotheses for the statistical test described.

4.5 Testing to see if there is evidence that the mean of group A is not the same as the mean of group B

ANSWER ⊕

WORKED SOLUTION ⊕

4.6 Testing to see if there is evidence that a proportion is greater than 0.3

4.7 Testing to see if there is evidence that a mean is less than 50

ANSWER ⊕

WORKED SOLUTION ⊕

4.8 Testing to see if there is evidence that the correlation between two variables is negative

SKILL BUILDER 3

In Exercises 4.9 to 4.13, a situation is described for a statistical test. In each case, define the relevant parameter(s) and state the null and alternative hypotheses.

4.9 Testing to see if there is evidence that the proportion of people who smoke is greater for males than for females

ANSWER ⊕

WORKED SOLUTION ⊕

4.10 Testing to see if there is evidence that a correlation between height and salary is significant (that is, different than zero)

4.11 Testing to see if there is evidence that the percentage of a population who watch the Home Shopping Network is less than 20%

ANSWER ⊕

WORKED SOLUTION ⊕

4.12 Testing to see if average sales are higher in stores where customers are approached by salespeople than in stores where they aren't

4.13 Testing to see if there is evidence that the mean time spent studying per week is different between first-year students and upperclass students

ANSWER ⊕

WORKED SOLUTION ⊕

SKILL BUILDER 4

In Exercises 4.14 and 4.15, determine whether the sets of hypotheses given are valid hypotheses.

4.14 State whether each set of hypotheses is valid for a statistical test. If not valid, explain why not.

(a) $H_0:\mu=15$ vs $H_a:\mu\neq 15$

(b) $H_0:p\neq 0.5$ vs $H_a:p=0.5$

(c) $H_0:p_1<p_2$ vs $H_a:p_1>p_2$

(d) $H_0:\bar{x}_1=\bar{x}_2$ vs $H_a:\bar{x}_1\neq\bar{x}_2$

4.15 State whether each set of hypotheses is valid for a statistical test. If not valid, explain why not.

(a) $H_0:\rho=0$ vs $H_a:\rho<0$

ANSWER ⊕

WORKED SOLUTION ⊕

(b) $H_0:\hat{p}=0.3$ vs $H_a:\hat{p}\neq 0.3$

ANSWER ⊕

WORKED SOLUTION ⊕

(c) $H_0:\mu_1\neq\mu_2$ vs $H_a:\mu_1=\mu_2$

ANSWER ⊕

WORKED SOLUTION ⊕

(d) $H_0{:}p{=}25$ vs $H_a{:}p{\neq}25$

ANSWER ⊕

WORKED SOLUTION ⊕

4.16 Pesticides and ADHD

Are children with higher exposure to pesticides more likely to develop ADHD (attention-deficit/hyperactivity disorder)? In a recent study, authors measured levels of urinary dialkyl phosphate (DAP, a common pesticide) concentrations and ascertained ADHD diagnostic status (Yes/No) for 1139 children who were representative of the general US population.[8]Bouchard, M., Bellinger, D., Wright, R., and Weisskopf, M., "*Attention-Deficit/Hyperactivity Disorder and Urinary Meta-bolites of Organophosphate Pesticides*," *Pediatrics*, 2010; 125: e1270-e1277. The subjects were divided into two groups based on high or low pesticide concentrations, and we compare the proportion with ADHD in each group.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

(b) In the sample, children with high pesticide levels were more likely to be diagnosed with ADHD. Can we necessarily conclude that, in the population, children with high pesticide levels are more likely to be diagnosed with ADHD? (Whether or not we can make this generalization is, in fact, the statistical question of interest.)

(c) To assess statistical significance, we assume the null hypothesis is true. What does that mean in this case? State your answer in terms of pesticides and ADHD.

(d) The study found the results to be statistically significant. Which of the hypotheses, H_0 or H_a, is no longer a very plausible possibility?

(e) What do the statistically significant results imply about pesticide exposure and ADHD?

4.17 Beer and Mosquitoes

Does consuming beer attract mosquitoes? A study done in Burkino Faso, Africa, about the spread of malaria investigated the connection between beer consumption and mosquito attraction.[9]Lefvre, T., et al., "*Beer Consumption Increases Human Attractiveness to Malaria Mosquitoes*," *PLoS ONE*, 2010; 5(3): e9546. In the experiment, 25 volunteers consumed a liter of beer while 18 volunteers consumed a liter of water. The volunteers were assigned to the two groups randomly. The attractiveness to mosquitoes of each volunteer was tested twice: before the beer or water and after. Mosquitoes were released and caught in traps as they approached the volunteers. For the beer group, the total number of mosquitoes caught in the traps before consumption was 434 and the total was 590 after consumption. For the water group, the total was 337 before and 345 after.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses for a test to see if, after consumption, the average number of mosquitoes is higher for the volunteers who drank beer.

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Compute the average number of mosquitoes per volunteer before consumption for each group and compare the results. Are the two sample means different? Do you expect that this difference is just the result of random chance?

ANSWER ⊕

WORKED SOLUTION ⊕

(c) Compute the average number of mosquitoes per volunteer after consumption for each group and compare the results. Are the two sample means different? Do you expect that this difference is just the result of random chance?

ANSWER ⊕

WORKED SOLUTION ⊕

(d) If the difference in part (c) is unlikely to happen by random chance, what can we conclude about beer consumption and mosquitoes?

ANSWER ⊕

WORKED SOLUTION ⊕

(e) If the difference in part (c) is statistically significant, do we have evidence that beer consumption increases mosquito attraction? Why or why not?

ANSWER ⊕

WORKED SOLUTION ⊕

4.18 Guilty Verdicts in Court Cases

A reporter on cnn.com stated in July 2010 that 95% of all court cases that go to trial result in a guilty verdict. To test the accuracy of this claim, we collect a random sample of 2000 court cases that went to trial and record the proportion that resulted in a guilty verdict.

(a) What is/are the relevant parameter(s)? What sample statistic(s) is/are used to conduct the test?

(b) State the null and alternative hypotheses.

(c) We assess evidence by considering how likely our sample results are *when* H_0 *is true*. What does that mean in this case?

4.19 Exercise and the Brain

It is well established that exercise is beneficial for our bodies. Recent studies appear to indicate that exercise can also do wonders for our brains, or, at least, the brains of mice. In a randomized experiment, one group of mice was given access to a running wheel while a second group of mice was kept sedentary. According to an article describing the study, “The brains of mice and rats that were allowed to run on wheels pulsed with vigorous, newly born neurons, and those animals then breezed through mazes and other tests of rodent IQ”[10]Reynolds, G., “*Phys Ed: Your Brain on Exercise*,” *The New York Times*, July 7, 2010. compared to the sedentary mice. Studies are examining the reasons for these beneficial effects of exercise on rodent (and perhaps human) intelligence. High levels of BMP (bonemorphogenetic protein) in the brain seem to make stem cells less active, which makes the brain slower and less nimble. Exercise seems to reduce the level of BMP in the brain. Additionally, exercise increases a brain protein called noggin, which improves the brain's ability. Indeed, large doses of noggin turned mice into “little mouse geniuses,” according to Dr. Kessler, one of the lead authors of the study. While research is ongoing in determining which effects are significant, all evidence points to the fact that exercise is good for the brain. Several tests involving these studies are described. In each case, define the relevant parameters and state the null and alternative hypotheses.

(a) Testing to see if there is evidence that mice allowed to exercise have lower levels of BMP in the brain on average than sedentary mice

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Testing to see if there is evidence that mice allowed to exercise have higher levels of noggin in the brain on average than sedentary mice

ANSWER ⊕

WORKED SOLUTION ⊕

(c) Testing to see if there is evidence of a negative correlation between the level of BMP and the level of noggin in the brains of mice

ANSWER ⊕

WORKED SOLUTION ⊕

4.20 Taste Test

A taste test is conducted between two brands of diet cola, Brand A and Brand B, to determine if there is evidence that more people prefer Brand A. A total of 100 people participate in the taste test.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

(b) Give an example of possible sample results that would provide strong evidence that more people prefer Brand A. (Give your results as number choosing Brand A and number choosing Brand B.)

(c) Give an example of possible sample results that would provide no evidence to support the claim that more people prefer Brand A.

(d) Give an example of possible sample results for which the results would be inconclusive: The sample provides some evidence that Brand A is preferred but the evidence is not strong.

INTENSIVE CARE UNIT (ICU) ADMISSIONS

Exercises 4.21 to 4.25 describe tests we might conduct based on Data 2.3. This dataset, stored in **ICUAdmissions**, contains information about a sample of patients admitted to a hospital Intensive Care Unit (ICU). For each of the research questions below, define any relevant parameters and state the appropriate null and alternative hypotheses.

4.21 Is there evidence that mean heart rate is higher in male ICU patients than in female ICU patients?

ANSWER ⊕

WORKED SOLUTION ⊕

4.22 Is there a difference in the proportion who receive CPR based on whether the patient's race is white or black?

4.23 Is there a positive linear association between systolic blood pressure and heart rate?

ANSWER ⊕

WORKED SOLUTION ⊕

4.24 Is either gender over-represented in patients to the ICU or is the gender breakdown about equal?

4.25 Is the average age of ICU patients at this hospital greater than 50?

ANSWER ⊕

WORKED SOLUTION ⊕

4.26 Income East and West of the Mississippi

For a random sample of households in the US, we record annual household income, whether the location is east or west of the Mississippi River, and number of children. We are interested in determining whether there is a difference in average household income between those east of the Mississippi and those west of the Mississippi.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

(b) What statistic(s) from the sample would we use to estimate the difference?

4.27 Relationship between Income and Number of Children

Exercise 4.26 discusses a sample of households in the US. We are interested in determining whether or not there is a linear relationship between household income and number of children.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Which sample correlation shows more evidence of a relationship, $r=0.25$ or $r=0.75$?

ANSWER ⊕

WORKED SOLUTION ⊕

(c) Which sample correlation shows more evidence of a relationship, $r=0.50$ or $r=-0.50$?

ANSWER ⊕

WORKED SOLUTION ⊕

4.28 Red Wine and Weight Loss

Resveratrol, a compound in grapes and red wine, has been shown to promote weight loss in rodents and now in a primate.[11]BioMed Central, "*Lemurs Lose Weight with 'Life-Extending' Supplement Resveratrol*," *ScienceDaily*, June 22, 2010. Lemurs fed a resveratrol supplement for four weeks had decreased food intake, increased metabolic rate, and a reduction in seasonal body mass gain compared to a control group. Suppose a hypothetical study is done for a different primate species, with one group given a resveratrol supplement and the other group given a placebo. We wish to see if there is evidence that resveratrol increases the mean metabolism rate for this species. (This exercise presents hypothetical data. We will see the results from the actual study later in this chapter.)

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

(b) Possible sample results for Species A are shown in Figure 4.6(a) with the mean indicated by a circle on the boxplots. In the sample, is the mean greater for the resveratrol group? Can we necessarily conclude that resveratrol increases the metabolism rate for this species?

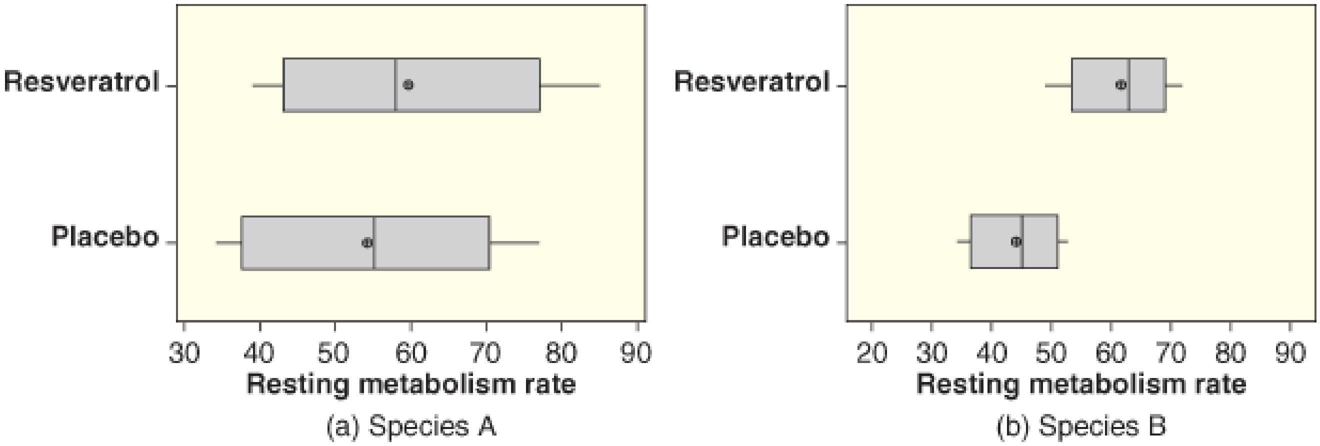

Figure 4.6 ***Does red wine boost metabolism rates?***

(c) Possible sample results for Species B are shown in Figure 4.6(b) and the sample sizes are the same as for Species A. For which of the two species, A or B, is the evidence stronger that resveratrol increases the metabolism rate for this species? Explain your reasoning.

4.29 Flaxseed and Omega-3

Studies have shown that omega-3 fatty acids have a wide variety of health benefits. Omega-3 oils can be found in foods such as fish, walnuts, and flaxseed. A company selling milled flaxseed advertises that one tablespoon of the product contains, on average, at least 3800 mg of ALNA, the primary omega-3.

(a) The company plans to conduct a test to ensure that there is sufficient evidence that its claim is correct. To be safe, the company wants to make sure that evidence shows the average is higher than 3800 mg. What are the null and alternative hypotheses?

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Suppose, instead, that a consumer organization plans to conduct a test to see if there is evidence *against* the claim that the product contains an average of 3800 mg per tablespoon. The consumer organization will only take action if it finds evidence that the claim made by the company is false and the actual average amount of omega-3 is less than 3800 mg. What are the null and alternative hypotheses?

ANSWER ⊕

WORKED SOLUTION ⊕

STATISTICAL TESTS?

In Exercises 4.30 to 4.36, indicate whether the analysis involves a statistical test. If it does involve a statistical test, state the population parameter(s) of interest and the null and alternative hypotheses.

4.30 Polling 1000 people in a large community to determine the average number of hours a day people watch television

4.31 Polling 1000 people in a large community to determine if there is evidence for the claim that the percentage of people in the community living in a mobile home is greater than 10%

ANSWER ⊕

WORKED SOLUTION ⊕

4.32 Utilizing the census of a community, which includes information about all residents of the community, to determine if there is evidence for the claim that the percentage of people in the community living in a mobile home is greater than 10%

4.33 Testing 100 right-handed participants on the reaction time of their left and right hands to determine if there is evidence for the claim that the right hand reacts faster than the left

ANSWER ⊕

WORKED SOLUTION ⊕

4.34 Testing 50 people in a driving simulator to find the average reaction time to hit the brakes when an object is seen in the view ahead

4.35 Giving a Coke/Pepsi taste test to random people in New York City to determine if there is evidence for the claim that Pepsi is preferred

ANSWER ⊕

WORKED SOLUTION ⊕

4.36 Using the complete voting records of a county to see if there is evidence that more than 50% of the eligible voters in the county voted in the last election

4.37 Influencing Voters

When getting voters to support a candidate in an election, is there a difference between a recorded phone call from the candidate or a flyer about the candidate sent through the mail? A sample of 500 voters is randomly divided into two groups of 250 each, with one group getting the phone call and one group getting the flyer. The voters are then contacted to see if they plan to vote for the candidate in question. We wish to see if there is evidence that the proportions of support are different between the two methods of campaigning.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Possible sample results are shown in Table 4.2. Compute the two sample proportions: $\hat{p}_c$, the proportion of voters getting the phone call who say they will vote for the candidate, and $\hat{p}_f$, the proportion of voters getting the flyer who say they will vote for the candidate. Is there a difference in the sample proportions?

Table 4.2 ***Sample A: Is a phone call or a flyer more effective?***

	Will Vote	**Will Not Vote**
Sample A	**for Candidate**	**for Candidate**
Phone call	152	98
Flyer	145	105

ANSWER ⊕

WORKED SOLUTION ⊕

(c) A different set of possible sample results are shown in Table 4.3. Compute the same two sample proportions for this table.

Table 4.3 ***Sample B: Is a phone call or a flyer more effective?***

	Will Vote	**Will Not Vote**
Sample B	**for Candidate**	**for Candidate**
Phone call	188	62
Flyer	120	130

ANSWER ⊕

WORKED SOLUTION ⊕

(d) Which of the two samples seems to offer stronger evidence of a difference in effectiveness between the two campaign methods? Explain your reasoning.

ANSWER ⊕

WORKED SOLUTION ⊕

4.38 Influencing Voters: Is a Phone Call More Effective?

Suppose, as in Exercise 4.37, that we wish to compare methods of influencing voters to support a particular candidate, but in this case we are specifically interested in testing whether a phone call is more effective than a flyer. Suppose also that our random sample consists of only 200 voters, with 100 chosen at random to get the flyer and the rest getting a phone call.

(a) State the null and alternative hypotheses in this situation.

(b) Display in a two-way table possible sample results that would offer clear evidence that the phone call is more effective.

(c) Display in a two-way table possible sample results that offer no evidence at all that the phone call is more effective.

(d) Display in a two-way table possible sample results for which the outcome is not clear: There is some evidence in the sample that the phone call is more effective but it is possibly only due to random chance and likely not strong enough to generalize to the population.

4.39 Mice and Pain

Can you tell if a mouse is in pain by looking at its facial expression? A new study believes you can. The study[12]"*Of Mice and Pain*," *The Week*, May 28, 2010, p. 21. created a "mouse grimace scale" and tested to see if there was a positive correlation between scores on that scale and the degree and duration of pain (based on injections of a weak and mildly painful solution). The study's authors believe that if the scale applies to other mammals as well, it could help veterinarians test how well painkillers and other medications work in animals.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses.

ANSWER ⊕

WORKED SOLUTION ⊕

(b) Since the study authors report that you can tell if a mouse is in pain by looking at its facial expression, do you think the data were found to be statistically significant? Explain.

ANSWER ⊕

WORKED SOLUTION ⊕

(c) If another study were conducted testing the correlation between scores on the "mouse grimace scale" and a placebo (non-painful) solution, should we expect to see a sample correlation as extreme as that found in the original study? Explain. (For simplicity, assume we use a placebo that has no effect on the facial expressions of mice. Of course, in real life, you can never automatically assume that a placebo has no effect!)

ANSWER ⊕

WORKED SOLUTION ⊕

(d) How would your answer to part (c) change if the original study results showed no evidence of a relationship between mouse grimaces and pain?

ANSWER ⊕

WORKED SOLUTION ⊕

4.40 Euchre

One of the authors and some statistician friends have an ongoing series of Euchre games that will stop when one of the two teams is deemed to be *statistically significantly* better than the other team. Euchre is a card game and each game results in a win for one team and a loss for the other. Only two teams are competing in this series, which we'll call Team A and Team B.

(a) Define the parameter(s) of interest.

(b) What are the null and alternative hypotheses if the goal is to determine if either team is statistically significantly better than the other at winning Euchre?

(c) What sample statistic(s) would they need to measure as the games go on?

(d) Could the winner be determined after one or two games? Why or why not?