NUMB3RS Activity: The Odds-On Favorite
Episode: "Longshot"

Topics: Probability, odds, and odds ratios **Grade Level:** 9 - 12
Objective: To understand the relationships between and among probabilities, odds, odds ratios, and statistical significance
Materials: Scientific calculator; TI-83 Plus/TI-84 Plus graphing calculator (for the extensions only)
Time: 15 - 20 minutes

Introduction

In "Longshot," a man is murdered at the racetrack after he wins the "Pick 6" (selecting the winner of six consecutive horse races). Don brings Charlie a notebook that was found on the body. It contains pages of horse racing data and equations. Charlie determines that the equations were designed to pick the *second* place winner, not first place, possibly to be less conspicuous. Parts of these equations use the "logit" (LŌ – jit) function, a specific probability function that uses logarithms and odds ratios. Because the logit function can get pretty complicated, this activity lays its foundations, namely the relationship between probability, odds, and odds ratios.

Discuss with Students

Odds ratios are used to measure the *effect size* between two variables. This is used to determine whether an experiment is *statistically significant*. Examples of experiments include the difference in preferences between males and females, or between two or more age groups, or even between weight loss programs. *Odds ratios* are a way to demonstrate the degree of differences between variables; more formal methods for measuring significance include *hypothesis tests* and *confidence intervals*.

- *effect size*: the size of the relationship between two variables
- *statistically significant*: a difference that is greater than what might be expected to happen by chance alone
- *odds ratio*: one measure that can be used to determine the degree of association between two variables
- *hypothesis tests*: inferential procedures that use sample data to evaluate the credibility of a hypothesis about a population
- *confidence intervals*: an estimated range of values that is likely to include an unknown population parameter

Although the essential concepts and definitions are included in the activity, the teacher should be mindful that the definition of probability is the ratio of the number of *favorable* (desired) outcomes to the total number of *possible* outcomes. Odds, in turn, are the ratio of the number of favorable outcomes to the number of *unfavorable* outcomes (these last two categories are complementary; see "Extensions" for more on this). Finally, the *odds ratio* is the ratio of the odds of an event occurring in one group to the odds of the same event occurring in another group, as in the examples in the paragraph above.

L. Charles (Chuck) Biehl, The Charter School of Wilmington, Wilmington, DE

Student Page Answers:

1.

	Soft Drink	Fruit Drink
Boys	0.39 (46/117)	0.61 (71/117)
Girls	0.31 (37/120)	0.69 (83/120)

2. *The sum of each pair of probabilities is 1, because there are only two choices, and one of them has to be chosen.*

3.

	Soft Drink	Fruit Drink
Boys	0.65 (46/71)	1.54 (71/46)
Girls	0.45 (37/83)	2.24 (83/37)

4. *A ratio of 1 means that the preference is equally likely in both groups.*

5.

	Soft Drink	Fruit Drink
Boys/Girls	1.44 (0.65/0.45)	0.69 (1.54/2.24)
Girls/Boys	0.69 (0.45/0.65)	1.45 (2.24/1.54)

6. *Boys are 1.44 times as likely to prefer soft drinks to fruit drinks as girls, and/or girls are 1.45 times as likely to prefer fruit drinks to soft drinks as boys.* **7.** *No. The odds ratio is less than 1.8, so the difference we think we see might be due to chance.*

Extensions:

1. *Odds for boys is p/(1-p). Odds for girls is q/(1-q). The ratio of these is (p/(1-p))/(q/(1-q)) = p(1-q)/(q(1-p)).* **2.** *A logistic curve has an S-shape. From left to right, the value of the dependent variable grows slowly at first, and then begins to rise sharply (like exponential growth), but then "levels off."* **3.** *The shape of a logistic curve reveals the inherent flaw with an exponential model for spread of disease. If the dependent variable is the number of new cases reported, the spread can begin like exponential growth, but is bound to level off. Especially in the worst-case scenario – when every single member of the population gets the disease – the number of new cases would fall to zero.*

L. Charles (Chuck) Biehl, The Charter School of Wilmington, Wilmington, DE

Name: ______________________________ Date: _______________

NUMB3RS Activity: The Odds-On Favorite

In “Longshot”, a man is murdered at the racetrack after he wins the “Pick 6” (selecting the winner of six consecutive races). Don brings Charlie a notebook that was found on the body. It contains pages of horse racing data and equations. Charlie determines that the equations were designed to pick the *second* place winner, not first place, possibly to be less conspicuous. Parts of these equations use the “logit” (LŌ – jit) function, a specific probability function that uses logarithms and odds ratios. Because the logit function can get pretty complicated, this activity lays its foundations, namely the relationship between probability, odds, and odds ratios.

Consider the following data, collected from a study to determine high school students’ preference in beverages. Each of the students was asked whether he or she prefers soft drinks or fruit drinks. Use this table for the questions below. Express all answers to the questions in this activity as decimals to the nearest hundredth.

	Soft Drink	**Fruit Drink**	**Total**
Boys	46	71	117
Girls	37	83	120

Probability is defined as the ratio of the number of *favorable outcomes* (the ones you are counting, whether they are actually good things or not) to the number of *total outcomes* (all possible outcomes). Suppose that the data above are considered good enough to make predictions about the preferences of high school students in general.

1. Complete this table with the probability for each event. For example, the probability that a boy will prefer soft drinks is the number of boys who prefer soft drinks divided by the total number of boys.

	Soft Drink	**Fruit Drink**
Boys		
Girls		

2. What is the sum of the probabilities for boys’ preferences? What is the sum of the probabilities for girls’ preferences? Why is this?

Notice that there is a difference between boys’ and girls’ preferences. Is this difference significant, or could it just be due to chance? There are many ways to answer this question; a relatively simple one that relates to the math in “Longshot” is called the *odds ratio*. This is exactly what it says it is; namely the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. Odds are another way of expressing probability. Odds are the ratio of the number of *favorable* outcomes to

L. Charles (Chuck) Biehl, The Charter School of Wilmington, Wilmington, DE

the number of *unfavorable* outcomes for each group. For example, the odds in favor of girls preferring soft drinks is the ratio of the number choosing soft drinks to those who do not. Similarly, the odds for girls preferring fruit drinks are found by taking the reciprocal.

3. Complete the same table using odds (e.g., number of boys preferring one/number of boys preferring the other).

	Soft Drink	**Fruit Drink**
Boys		
Girls		

Finally, the measure for actually measuring the difference between boys’ and girls’ preferences comes in the form of an *odds ratio*, the ratio of boys’ preference to girls’.

4. What does it mean if the ratio of boys’ odds to girls’ odds is exactly 1?

5. Complete this table using the ratios of the respective odds (e.g., odds for boys/odds for girls for each choice).

	Soft Drink	**Fruit Drink**
Boys/Girls		
Girls/Boys		

Suppose that one of the ratios of girls’ preference to boys’ was 2.35 (which it is actually not). This means that the girls would be 2.35 times as likely to have the preference as the boys.

6. Make a similar statement about the results in the table for #5.

Scientists try to distinguish between differences that occur by chance and differences that indicate that the preferences of boys and girls really are different. We say that the difference is “statistically significant” if it is very unlikely that we could have seen this difference by chance. In this situation, we need an odds ratio of at least 1.8 before we can conclude that the difference between the preferences of boys and girls is statistically significant. (See the “Extensions” for more on where the odds ratio of 1.8 comes from.)

7. Can we be confident that there is a real difference between the preferences of boys and girls?

L. Charles (Chuck) Biehl, The Charter School of Wilmington, Wilmington, DE

The goal of this activity is to give your students a short and simple snapshot into a very extensive mathematical topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

Introduction

This activity has laid the foundations for the logit function. Among many other tests and calculations, statisticians use the logit function to determine the strength of the relationship between two variables.

For the Student

1. Show that if the percentage of boys that prefer soft drinks is p and the percentage of girls that prefer soft drinks is q, then the odds ratio is $\dfrac{p(1-q)}{q(1-p)}$.

2. The logit function, for any probability p between 0 and 1, is defined as $\text{logit}(p) = \log\left(\dfrac{p}{1-p}\right)$. If the first group has probability p and the second group has probability q, then the logarithm of odds ratio, $\dfrac{p(1-q)}{q(1-p)}$, is equal to the difference logit(p) – logit(q). The logit function is the algebraic inverse of a *logistic function*, or logistic curve. Logistic curves are used in many kinds of statistical models, including population growth and the spread of disease or rumors. Find an example of a logistic function and graph it. Describe how its properties are appropriate for the applications listed here.

3. The media often reports the spread of such things as disease (like HIV, SARS, or avian flu) as potentially being exponential. Why is a logistic curve more realistic for a mathematical model for the spread of disease than an exponential one?

Related Topics

The odds ratio is considered to be statistically significant if a 95% confidence interval does not contain the value 1 (a ratio of 1:1). In other words, there is 95% confidence that the true ratio is larger (or smaller) than 1. There is a calculator to do this that is based on the logit function. For an explanation of how it works, see the article “The odds ratio” by J. Martin Bland and Douglas G. Altman, *British Medical Journal, volume 320, 27 May 2000, page 1468*. For the calculator to find a 95% confidence interval for the odds ratio, see: **http://www.hutchon.net/ConfidOR.htm**

For a related activity on the probability of complementary events, see the *NUMB3RS* activity “Birthday Surprise” from the episode “Traffic.” To download this activity, go to **http://education.ti.com/exchange** and search for “7463.”

Additional Resources

For those interested in studying the logit function (or “logit transformation,” as it is called in “Longshot”) and its applications, see:
http://mathworld.wolfram.com/LogitTransformation.html

L. Charles (Chuck) Biehl, The Charter School of Wilmington, Wilmington, DE