Example 2-1: Naughty or Nice?
Do children less than a year old recognize the difference between nice, friendly behavior as opposed to mean, unhelpful behavior? Do they make choices based on such behavior? In a study reported in the November 2007 issue of Nature, researchers investigated whether infants take into account an individual’s actions towards others in evaluating that individual as appealing or aversive (Hamlin, Wynn, and Bloom, 2007). In one component of the study, 10-month-old infants were shown a “climber” character (a piece of wood with “google” eyes glued onto it) that could not make it up a hill in two tries. Then they were alternately shown two scenarios for the climber’s next try, one where the climber was pushed to the top of the hill by another character (“friend”) and one where the climber was pushed back down the hill by another character (“foe”). The infant was alternately shown these two scenarios several times. Then the child was presented with both pieces of wood (representing the friend and the foe) and asked to pick one to play with. Videos of this study are available at: www.yale.edu/infantlab/socialevaluation/Helper-Hinderer.html

a) Identify the observational units and variable in this study. Is the variable categorical or quantitative?

Researchers found that 14 of the 16 infants in the study selected the nice toy.

b) Determine the proportion of infants who selected the nice toy.

c) Suggest two possible explanations for this result that the researchers observed.

d) Suppose for the moment that the researchers’ conjecture is wrong, and infants actually have no preference for either type of toy. Would it be possible to have obtained a result as extreme as the researchers found?

e) Again suppose that infants have no preference. How many of the 16 would you expect to select the nice toy? Would you always expect to see that many of the 16 infants select the nice toy? Explain briefly.

f) In your judgment, how many infants, out of the 16, would have to select the friend toy in order for you to fairly well convinced that the researchers’ conjecture is correct, that infants really do have a tendency to prefer the friend toy? Explain
The key question here is to determine what results would occur in the long run under the assumption that infants actually have no preference. (We will call this assumption of no genuine preference the **null model** or **null hypothesis**.) We will answer this question with by simulating (artificially re-creating) the selection process of 16 infants over and over, assuming that infants actually have no genuine preference.

**g)** Describe how we could use a common device to simulate the infants’ selection process.

**h)** Flip a coin 16 times. Record the number of heads that you obtain, which represents the number of your 16 hypothetical infants who choose the nice toy.

**i)** Combine your simulation results with your classmates. Produce a well-labeled dotplot. Identify the observational units and variable in the resulting dotplot.

<table>
<thead>
<tr>
<th>Observational units:</th>
<th>Variable:</th>
</tr>
</thead>
</table>

**j)** Where is the distribution of number of heads in 16 tosses centered? Explain why this makes sense.

**k)** Looking at this dotplot, does it seem that the result obtained by the researchers would have been **surprising** if in fact the infants had no preference? What does this suggest about whether the researchers’ result provides much evidence that the infants do genuinely prefer the nice toy? Explain.

We really need to simulate this random selection process hundreds, preferably thousands of times. This would be very tedious and time-consuming with coins, so we’ll turn to technology.

**l)** Use the “One Proportion” applet, available from our course webpage, to simulate the random process of 16 infants making this toy choice, still assuming the null model that infants have no real preference and so are equally likely to choose either toy. (Start with 1 repetition, and click the “16 tosses” button. Repeat this a total of 5 times. Then ask for 20 repetitions and watch the results. Finally, change the number of repetitions to 975, which will produce a total of 1000 repetitions.) Describe the shape of the resulting dotplot, and comment on whether it is centered where you expected.
m) Based on your simulation results, would you say that it would be very surprising, if infants actually have no genuine preference, that 14 out of 16 infants in the study would have chosen the friend toy just by chance? Explain.

n) Report how many of your 1000 repetitions produced 14 or more infants choosing the friend toy. Also determine the proportion of these 1000 repetitions that produced such an extreme result.

- This proportion is called an approximate p-value. A p-value is the probability of obtaining a result as extreme as the one observed, assuming that there is no genuine preference/difference.
- A small p-value casts doubt on the null model/hypothesis used to perform the calculation (in this case, that infants have no genuine preference).
  - A p-value of .10 or less is generally considered to be some evidence against the null model/hypothesis.
  - A p-value of .05 or less is generally considered to be fairly strong evidence against the null model/hypothesis.
  - A p-value of .01 or less is generally considered to be very strong evidence against the null model/hypothesis.
  - A p-value of .001 or less is generally considered to be extremely strong evidence against the null model/hypothesis.

o) Is this proportion small enough to consider the actual result obtained by the researchers surprising, assuming the null model that infants have no preference and so choose blindly between the two toys?

p) In light of your answers to the previous two questions, would you say that the experimental data obtained by the researchers provide strong evidence that infants in general have a genuine preference for the friend toy over the foe toy? Explain the reasoning process behind your answer.

q) Now suppose that 10 of the 16 infants had chosen the friend toy. Report the approximate p-value, and summarize the conclusion that you would have drawn from this result.
Example 2-2: Lady Tasting Tea
A famous (in statistical circles) study involves a woman who claimed to be able to tell whether tea or milk was poured first into a cup. She was presented with 8 cups containing a mixture of tea and milk, and she correctly identified which had been poured first for all 8 cups.

a) Identify the observational units and variable in this study.

b) Is it possible that the woman could get all 8 correct if she were randomly guessing with each cup?

c) Describe how you could investigate whether it is unlikely that the woman would get all 8 correct if she were randomly guessing with each cup.

d) Use the applet to conduct an appropriate simulation analysis to investigate the woman’s claim based on the empirical result that she correctly identified which was poured first for all 8 cups. Specify the input values, and report the approximate p-value.

e) Would you conclude that the woman’s result produces strong evidence that she does have the ability to distinguish whether milk or tea was poured first? Explain the reasoning process behind your answer.

f) Suppose that I try this for 8 cups and make the correct identification 5 times. I say: “5 out of 8 is more than half, so you have to conclude that I’m doing better than random guessing.” How would you respond?
Example 2-3: Facial Prototyping

A study in *Psychonomic Bulletin and Review* (Lea, Thomas, Lamkin, & Bell, 2007) presented evidence that “people use facial prototypes when they encounter different names.” Similar to one of the experiments they conducted, you will be asked to match photos of two faces to the names Tim and Bob. The researchers wrote that their participants “overwhelmingly agreed” on which face belonged to Tim. We will repeat this study by collecting data from students in our class, and then investigate whether our class data provide strong evidence that people have a natural tendency to associate the name Tim with the face that the researchers proposed.

a) Record the number of students who matched the name Tim with the face that the researchers anticipated, and also the number who did not. What proportion of students put Tim with the “correct” face?

b) Describe (in words) the null model/hypothesis to be investigated with this study.

c) Describe how you could (in principle) use a coin to produce a simulation analysis of whether these data provide strong evidence that Cal Poly students would correctly match the names to faces more than half the time.

d) Use the One-Proportion Inference applet to conduct a simulation (using 1000 repetitions), addressing the question of whether our class results provide strong evidence in support of the conjecture that Cal Poly students would correctly match the names to faces more than half the time. Where does our observed class result falls in that distribution?

e) Report the approximate p-value from this simulation analysis. Also write a sentence or two describing what this approximate p-value means.

f) Summarize what your simulation analysis reveals about whether our class data provide strong evidence in support of the conjecture that Cal Poly students would correctly match the names to faces more than half the time. Also explain the reasoning process behind your conclusion.