Example 4-1: Naughty vs. Nice

We all recognize the difference between naughty and nice, right? What about children less than a year old: Do they recognize the difference and show a preference for nice over naughty? 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 (“helper”) and one where the climber was pushed back down the hill by another character (“hinderer”). The infant was alternately shown these two scenarios several times. Then the child was presented with both pieces of wood (the helper and the hinderer) and asked to pick one to play with. The researchers found that the 14 of the 16 infants chose the helper over the hinderer.

Suppose for the moment that the researchers’ conjecture is wrong, and infants actually have no preference for either type of toy.

a) Is it possible, even if infants actually have no genuine preference, that 14 out of 16 infants in the study would have chosen the helper toy just by chance?

b) Do you think 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 helper toy just by chance?

The key question here is to determine whether the observed result is very surprising under the assumption that infants actually have no preference. (We will call this assumption of no genuine preference the null model.) We could answer this question in two ways:

- first with simulation, and then
- with an exact probability calculation.

c) Describe how you could simulate the infants’ selections, assuming that infants have no genuine preference for either toy.

d) Use the “One Proportion inference” applet to simulate 1 repetition of flipping a fair coin 16 times. Then repeat this four more times. Did you get the same number of heads each time? Did you ever get 14 or more heads?
e) Now enter 995 for the number of repetitions, to produce a total of 1000 repetitions of flipping a fair coin 16 times. Describe the resulting distribution of the number of heads achieved in 16 tosses of a fair coin. (Comment on shape, center, and variability.)

f) Based on the simulated distribution of number of heads, would you say that it would be very surprising to obtain 14 or more heads in 16 tosses of a fair coin? Explain.

g) Describe what your answer to the previous question reveals about whether the results of the infant toy study would be surprising if infants actually had no genuine preference between the two toys.

h) Now use the applet to produce a total of 10,000 repetitions of tossing a fair coin 16 times. Then use the applet to count how many, and what proportion, of the 10,000 repetitions produced 14 or more heads.

i) Is this proportion small enough to conclude that the infant toy study results provide strong evidence that infants actually prefer the nice toy to the naughty toy? Explain your reasoning process.

- This proportion is called an approximate p-value. A p-value is the probability of obtaining a result at least as extreme as the one observed, assuming that there is no genuine preference/difference (i.e., assuming the null model is true).
- 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 provide some evidence against the null model/hypothesis.
  - A p-value of .05 or less is generally considered to provide fairly strong evidence against the null model/hypothesis.
  - A p-value of .01 or less is generally considered to provide very strong evidence against the null model/hypothesis.
- A p-value can be approximated by simulation or with a probability distribution such as normal or t.
- For inference about a proportion, a p-value can be calculated exactly from the binomial distribution.
j) Now let the random variable $X$ represent the number of infants who would choose the helper toy, if in fact infants have no genuine preference for either toy. Describe the probability distribution of $X$.

k) Use this probability distribution to calculate the (exact) $p$-value for the infant toy study.

l) Would you conclude that the experimental data obtained by the researchers provide strong evidence that infants in general have a genuine preference for the helper toy over the hinderer toy? Explain the reasoning process behind your answer.

Example 4-2: Nice vs. Naughty (cont.)
Now suppose that 32 infants had participated in the toy study, with 28 selecting the nice toy.

a) What proportion of the infants selected the nice toy? Is this a parameter or a statistic? What symbol would we use for this?

b) Describe (in words) the parameter of interest in this study. Also identify the symbol that we would use for this.

c) How does the sample proportion who selected the nice toy compare to the original study?

d) Do you think this study provides stronger evidence that infants have a genuine preference for the nice toy, or less evidence to that effect, or the same strength of evidence? Explain.

e) Use an applet simulation to approximate the $p$-value. Then use the applet to calculate the exact (binomial) $p$-value. Is this $p$-value smaller, larger, or about the same as before?
f) Summarize what this example reveals about the role of sample size in assessing strength of evidence.

Example 4-3: Flat Tire?
A legendary story on college campuses concerns two students who miss a chemistry exam because of excessive partying but blame their absence on a flat tire. The professor allowed them to take a make-up exam, and he sent them to separate rooms to take it. The first question, worth five points, was quite easy. The second question, worth ninety-five points, asked: Which tire was it? I will ask each of you to indicate which tire you would pick. Do not confer with anyone else before answering.

a) Which tire would you pick?

b) Name the tire that I predict to be the most popular choice.

c) Record the counts for the class data below.

<table>
<thead>
<tr>
<th>Left front</th>
<th>Left rear</th>
<th>Right front</th>
<th>Right rear</th>
</tr>
</thead>
</table>


d) What proportion of students selected the tire that I predicted? Also indicate the correct symbol for this.

e) The null model asserts that my conjecture is wrong and there is nothing special about the right front tire, so it is equally likely to be picked as any other tire. Let the random variable $X$ be the number of students in our class who would choose the right front tire. Under the null model of “nothing special” about this tire, describe the probability distribution of $X$.

f) Indicate how to calculate the p-value, and then use the applet (or Excel) to calculate it. Also interpret this p-value.

g) Is this p-value small enough that the class data provide fairly strong or very strong evidence against the null model? Summarize your conclusion.