Stat 140: Taste Test

A Taste Test

Suppose your friend claims to be able to tell the difference between the tastes of Coke and Pepsi. Within your groups, design an experiment to test this claim. You should answer the following:

1. What materials will we need to conduct this experiment?

2. What procedure will we follow? As part of your response, address

- What are the experimental units?
- What aspects of the experiment (if any) will you control?
- What aspects of the experiment (if any) will you randomize?
- What aspects of the experiment (if any) will you replicate?
- What aspects of the experiment (if any) will you block?
- Will you use any sort of blinding? (Why?)

3. What data do we record?

4. What is the population parameter you are trying to estimate? (For that matter, what is the population?)

5. What is the sample statistic?

6. How do we reach a conclusion based on the data? Is there a certain cutoff for the number of correct guesses required, or something similar?

A Simple Experiment

Let's suppose that you decided to run a fairly simple experiment, as follows: You prepare 5 cups of soda; each cup is an experimental unit. For each cup, you randomly decide whether that cup has Coke or Pepsi in it by flipping a coin. You control the temperature of the sodas so that both Coke and Pepsi are the same temprature; you control the cups, using the same cups for both sodas; and so on. You give the cups to your friend one at a time and ask her to determine whether each cup has Coke or Pepsi in it. You record the number of cups where your friend made a correct determination of the soda type. (I don't claim this is the best experimental design – you probably came up with something better on the previous page. But this one is simple enough for us to analyze without too much complication.)

Suppose your friend correctly identifies the soda on four out of five trials. That's pretty good, but is it enough to conclude that she can tell the difference between Coke and Pepsi?

Let's examine this using a kind of reverse logic: What if your friend was **not** able to distinguish between Coke and Pepsi? How likely would she be to guess correctly at least four times out of five? We will say that the experimental results provide strong evidence that she can distinguish between Coke and Pepsi if she was unlikely to get at least 4 correct by random guessing.

Our exploration will have three phases. The first two will be familiar to most of you already, since we did something very similar on the first day of class: 1. A physical simulation flipping coins 2. A simulation in R 3. A more formal mathematical analysis

Phase 1: A Simulation Study

The key question here is to determine what results would occur in the long run under the assumption that your friend is **not** able to tell the difference between Coke and Pepsi. (We will call this assumption the null model or null hypothesis.) We will answer this question by simulating the process of determining soda types in five trials over and over, assuming that each guess is random and has a 50/50 chance of being correct.

a) Flip a coin 5 times, and record whether you got heads or tails on each flip. Record the total number of heads that you obtain, which represents the number of sodas guessed correctly. (Note that this is a sample statistic based on your data set, which is the recorded result for each of the five guesses.)

b) Repeat step a) four more times. You will have five sample statistics based on five different samples obtained under the null hypothesis that your friend is just guessing. Combine your simulation results with your classmates on the board.

c) The plot on the board is an approximation of the sampling distribution of the total number of correct guesses, assuming the null hypothesis is true. The sample statistic obtained from the actual data is 4 (since your friend guessed correctly 4 times). Based on the combined simulation results on the board, does it seem that guessing correctly 4 out of 5 times would be surprising if in fact your friend was not able to tell the difference between Coke and Pepsi?

Phase 2: Automating the Simulation using `R`

The coin-flipping process got us an **approximation** to the sampling distribution of the total number of correct guesses if the null hypothesis is true, but our approximation would be improved by repeating the simulation process thousands of times. Let's use R to do that; lab 8 on Gryd will guide you through this process. Then, continue with the questions below:

d) Find the proportion of the 1000 simulated repetitions of the experiment that you performed in R that produced 4 or more correct guesses.

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 the null hypothesis 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 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.

e) Is the proportion you got in part d) small enough to consider the actual result obtained in the experiment surprising, assuming the null hypothesis that your friend was just guessing?

f) In light of your answer to part e), would you say that the experimental data provide strong evidence that your friend was not just guessing? Can we reject the null hypothesis as being inconsistent with the data?

g) Would your answer to part f) change if your friend had guessed correctly all five times? What would the approximate p-value be?

Phase 3: Mathematical Calculations

Repeating the simulation 1000 times is better than just a few times, but the approximation to the sampling distribution we got above by doing the simulation 1000 times is still not perfect. We can do better. Our ultimate goal here is to calculate the exact p-value. That is, if the null hypothesis was true and our friend could not distinguish between Coke and Pepsi, what are the chances that she would guess correctly at least 4 times? The next few questions will lead you through an exact calculation of this probability.

The first step is to think about all possible **outcomes** of the experiment. In our case, we can represent a single outcome by a sequence of five letters of the form "CCIIC", where a C means that the guess was Correct, and an I means the guess was Incorrect. For example, the sequence "CCIIC" means that the first two guesses were correct, the third and fourth guesses were incorrect, and the fifth guess was correct.

The **sample space** is the set of all possible **outcomes** of a random phenomenon. We use the capital letter S to denote the sample space.

h) Write down all possible outcomes (i.e., all sequences of five letters made up of C's and I's). This is the sample space. You should have 32 possible outcomes. I realize this is tedious; bear with me.

i) An *event* is a set of outcomes. We use capital letters near the beginning of the alphabet to denote events. One example of an event is that your friend guesses correctly four times out of five; let's use the letter A to denote that event. This event A is the set of all outcomes with four C's and one I. Another event that we are interested in is the event that your friend guesses correctly five times out of five; let's call that event B. Write down the outcomes in the event A and the outcomes in the event B.

A =

B =

j) If all of the possible outcomes are equally likely, the probability of an event A is calculated as follows:

$$P(A) = \frac{\text{Number of outcomes in the event A}}{\text{Number of outcomes in the sample space, S}}$$

Note that in our example, if your friend is actually just guessing randomly then all of the outcomes you wrote down in part h) are equally likely. Use this to find the probabilities of the events A and B we defined in part i).

k) For the purpose of our hypothesis test, we're interested in the probability that your friend would guess correctly at least four times, if she was guessing randomly. Two events are said to be *disjoint* if they don't contain any outcomes in common. Verify that the events A and B we defined above are disjoint by checking that in your answer to part i), there are no outcomes that are part of both A and B.

l) If A and B are disjoint events, you can calculate the probability of either event occurring as the sum of the probability that A occurs and the probability that B occurs:

$$P(A \text{ or } B) = P(A) + P(B)$$

Use this formula, in combination with you answers from part j), to calculate the probability that your friend would guess correctly at least 4 times out of five. This is the exact p-value.

m) Compare the exact p-value from part l) to the approximate p-value from part d). How close was the approximate p-value to the exact p-value?

Wrap Up

The process we followed in this lab to analyze the strength of evidence from the experiment had four steps:

1. **Collect data** (ask your friend to identify the soda in 5 trials)
2. **Calculate a sample statistic** based on the data (total number correct = 4)
3. **Obtain the sampling distribution** for the sample statistic, in a hypothetical universe where the **null hypothesis** holds. In general, the null hypothesis says that "there is no effect". The exact meaning of that sentence changes depending on the context; in our application, the null hypothesis is that our friend was just guessing and was equally likely to be right or wrong on each guess. We have looked at two ways of finding the sampling distribution assuming the null hypothesis is true:
   a. Simulation
   b. Calculating probabilities of events
4. **Calculate a p-value**: the probability of observing a sample statistic at least as extreme as the one we got from our actual data in step 2, if the null hypothesis were true.
5. Draw a **conclusion**.

All of these steps are important, but the conceptual foundation of this approach is in step 3. As we've seen here, calculating probabilities of events directly can be tedious – and our experimental setup was pretty simple! We'll spend the next week or two looking at rules for how to do these calculations and make them less tedious.