

Tap Water Hypothesis Test Proposal

Math 5 Crew

February 6, 2004

Goal:

I go out of my way not to drink Hanover tap water, and, when given the choice, I always choose to drink bottled spring water rather than Hanover tap water. For the most part, I do this because I've had rather unpleasant experiences with Hanover tap water's flavor, especially during the summer. Also, I have found that many other Hanover residents agree with me that Hanover's tap water is not pleasant. However, it seems possible that for the most part this is all in our minds. Namely, during most of the year, and especially during winter, is there really a difference? I've designed the following hypothesis test to help me determine whether or not in winter there really is an inherent preference for the flavor of bottled spring water over the flavor of Hanover's tap water among Hanover residents.

Null Hypothesis

To articulate the null hypothesis will require us to determine whether a subject is a taster¹ or at least in a tasting mood. To accomplish this, each time I test a subject I will first decide whether they are an Opinionated Taster, where an *Opinionated Taster* is a subject that answers yes to the following Question 1 and either right or left to the following Question 2.

Question 1: "Can you detect a difference between these two samples of water?" (Answers: Yes, No)

Question 2: "Which of the two samples do you prefer the taste of?" (Answers: left, right, they are different but I have no preference.)

Null Hypothesis: Opinionated Tasters will be equally likely to prefer Hanover tap water and bottled spring water.

Note: In order to make sense of these results, it is important that I report the percent of **all** people who are "Opinionated Tasters". The following percent is

¹ It is well known fact that for genetic reasons people really do have very different taste sensitivities. I feel that I'm probably a taster and wish to compare my experiences with people who share this idiosyncrasy.

crucial if one wants to understand the problem as a whole:

$$p_t = \frac{|\text{Opinionated Taster that prefer spring water}|}{|\text{Total Surveyed}|}.$$

Even if I decide that Opinionated Tasters prefer spring water, if p_t is small, I would be much less inclined to report these results at a Hanover town meeting. Clearly, in my final report I will report p_t (and its margin of error) and discuss this situation more carefully.

Alternate Hypothesis

Alternate Hypothesis: Opinionated Tasters will prefer bottled spring water to Hanover tap water.

Parameter

If you choose a subject at random from the Hanover community who qualifies as an Opinionated Taster, then there is some probability that they will prefer the bottled spring water to Hanover tap water. Call this probability p_{true} . This is the *parameter* that I will be testing. Under the null hypothesis, the tasters would be equally likely to prefer tap water over bottled water; hence the null hypothesis is that $p_{true} = p_{null} = 0.5$, while the alternate hypothesis is that $p_{true} > p_{null} = 0.5$.

Note: While this is a reasonable alternate hypothesis, I'd really like to find that more than a mere majority of the Opinionated Tasters prefer bottled spring water! Hence, when I report my results I will certainly include an estimate for p_{true} and its margin of error.

Test Statistic

Suppose I have collected the data associated to N randomly sampled Opinionated Tasters. Let K be the number of these Tasters who preferred the bottled spring water. My *test statistic* will be

$$P = K/N.$$

Under the Null hypothesis, K will be distributed via a binomial distribution with probability $p_{null} = .5$. In particular, I can estimate all the probabilities associated to P using the actual binomial distribution or, provided $N \geq 30$, the normal approximation to the binomial distribution.

Significance Level

Since I will want to be able to report these results as statistically significant, I will choose the *significance level* to be 0.05. In other words, I will leave only a five percent chance of a *type 1 error*. In this setting, a type 1 error corresponds to concluding that my Opinionated Tasters prefer spring water, when there is in fact no true preference.

Critical region

Let us assume $N \geq 30$. This test is right sided since my alternative hypothesis is that $p_{true} > p_{null}$. Therefore, to find the critical region I will need to find z_0 with the area to the right of z_0 under the standard normal curve equal to the significance level of 0.05. Hence, using a table or Excel, I find that $z_0 = 1.65$, and I will reject the null hypothesis provided

$$P \geq p_{null} + z_0 \sqrt{\frac{p_{null}(1 - p_{null})}{N}},$$

which is therefore P 's *critical region*. Denote the right hand side of this inequality as

$$p_{crit} = p_{null} + z_0 \sqrt{\frac{p_{null}(1 - p_{null})}{N}}$$

and call this quantity the *critical value*. In my Excel workbook (in the sheet titled "Normal Approx"), I find that for any $N \geq 30$ (where N is the value under "Opinionated Tasters") I can find p_{crit} (where p_{crit} is the value under "Critical Value").

Power Hypothesis

My feeling is that at least 80 percent of opinionated tasters will prefer spring water. I will now compute the *power* of this test with respect to the belief that $p_{true} = p_{pow} = .8$. If $N \geq 30$ I can use the normal approximation, and the area to the right of the standardized p_{crit} under the standard normal curve will be the probability that I correctly conclude the alternate hypothesis assuming $p_{pow} = p_{true}$. The standardized p_{crit} equals

$$\frac{p_{crit} - p_{pow}}{\sigma} = \frac{p_{crit} - p_{pow}}{\sqrt{\frac{p_{pow}(1 - p_{pow})}{N}}}$$

In my Excel workbook (in the sheet titled "Normal Approx") I have demonstrated how to find this power (where the power is the value under "Power") for any $N \geq 30$. Notice, if I believe that at least 80 percent of opinionated tasters will prefer spring water, then the chance of a *type 2 error* is at most $1 - power$. In this setting, a type 2 error corresponds to the chance that $p_{true} \geq .8$, but I report no preference.

Pre-test Population

For the pretest I will test all the students that show up on Friday, February 6th, minus the two or three volunteer administrators. Hence my N will depend on the number of opinionated tasters in this pretest population. For this pretest I hope N will be at least 30. This assumption is an expression of the fact that I suspect p_t is bigger than $3/4$. This pretest will help decide if my feelings regarding

p_t are accurate. Using the formulas developed above, with $N = 30$, I would have $p_{crit} = .65$ with the power of test equaling 98 percent with respect to my $p_{pow} = .8$ hypothesis. In particular, the chance of a type 2 error is less than 2 percent. If we fail to have $N \geq 30$ then I will be forced to determine the above critical value and power via the binomial distribution. Attached is an example of this done with Excel for $N = 25$ (which I could easily modify to handle the N I actually obtain). This can be found in the sheet titled "Binomial." With $N = 25$, I have $p_{crit} = .72$ with the power of the test equaling 89 percent with respect to my $p_{pow} = .8$ hypothesis. Notice, in this case I was forced to have a significance level of only 2.16 percent due to the lack of opinions. This helps explain the (relatively) small power of the test.

Note: Clearly this is **not** a random sample, and in a bigger test I would attempt to sample a randomly chosen group of Hanover residents, and keep testing them until I found (say) 100 opinionated tasters. With $N = 100$, I would have $p_{crit} = .582$ while the power would be nearly 100 percent with regard to my $p_{pow} = .8$ hypothesis. In my final report I will carefully develop how I would implement this sampling.

Equipment

For this experiment I would like (a least) four administrators: the *Pourer*, the *Distributor*, and 2 *Testers*. Since there are 44 students in the class, to be safe I should bring (at least) 88 opaque colored cups, $1/4 * 44 = 11$ cups of Spring Water (In the pretest I will use Poland Spring Water from Gallon containers that have been brought to room temperature), $1/4 * 44 = 11$ cups of tap water (In the pretest I will use tap water collect from Bradley Hall on February 4t, 2004 that have been brought to room temperature), a Data Chart 2 and watch for each Tester, a Data Chart 1 for the Pourer and 44 crackers. Also I should also allow my $(44/2) * 2 = 44$ minutes.

Protocol

Room The Pourer will be in an isolated area easily accessible by the Distributor, while the Testers should be set up in somewhat isolated areas with easy access to the subjects and the Distributor. We will attempt to test the subjects away from factors that might influence them (like the other subjects). Testers and subjects must have no view of the Pourer.

To begin I will announce what will take place: "You will be asked to sample 2 cups of water, one containing spring water and one Hanover tap water. You will have 20 seconds to taste them, and we ask that you **avoid** looking at them carefully, since we want to understand your impression about how these two samples of water taste, not look. A cracker will be available if you need a pallette cleanser during

this period. At the end of your twenty seconds, you will be asked to answer the following questions: (State Question 1,2, and 3 and the possible answers).”

Pourer: The Pourer will have a copy of Data Chart 1 going from 1 to 44. Starting at row 1, the Pourer will fill the left and right cups as indicated in the row, tear off the appropriate number slip from the right hand side of Data Chart 1, and place this slip next to the cups corresponding to this row.

Distributor: Facing the same direction as the Pourer, the Distributor will collect the number slip and cups (in the appropriate hands!), bring the cups to an available Tester, and place the number slip and the cups in front of the available Tester. The Distributor need to make sure their left and the Tester’s left agree! The Distributor must avoid watching the filling procedure to insure double blindness of the experiment.

Tester: The Tester will fill out a row of Data Chart 2 for each subject. First they will copy down the number off the slip, and then they will ask the subject for their year and gender and mark down this information on the sheet. Then they will have the tasters taste the two samples making sure that left and right are not switched. Crackers will be available and Subjects may be reminded of what is going on if they seem confused. They will ask questions 1, 2, and

Question 3: Which one do you believe is tap water? (Answers: left, right, they are different but I cannot tell which is which.)

The Tester will mark the answers in the appropriate row.

Note: An explanation of this third question and the reason why we are collecting demographic data will be explained carefully in my final report.