The Pretest! Pretest! Pretest! Assignment (Example 1)

May 20, 2003

1 Statement of Purpose and Description of Pretest Procedure

Statement of Purpose: there is a belief among many people that they can tell difference between Pepsi and Coke. My gut feeling is that people tend to greatly exaggerate their ability to do accomplish this feat. I will try and develop an experiment to help resolve whether my suspicion is founded or not. In the next four sub-sections I will describe various aspects of a pretest that I devised in order to study this question. In section **??**, I will describe how I would use what I learned from this pretest in order to implement a second, more comprehensive, pretest.

1.1 Quantifying Belief

In this section, I will describe a protocol for determining whether an individual **believes** they "really can" tell the difference between Pepsi and Coke. This protocol will consist of a pair of questions.

Question 1: Imagine, while blindfolded, you were presented with a soft drink which is either a either a Coke or a Pepsi. If you were asked to determine by sipping the beverage whether this beverage was a Coke or Pepsi, would you be able to do so?

If they answer yesish, then ask:

Question 2: Assuming you are allowed to clean your palette between sips, would you estimate that you could correctly identify the soft drink in the above situation 90 percent of the time or better?

If they answer yes to this second question then I will claim that the subject truly believes they can tell the difference between Pepsi and Coke.

I asked 31 students in our math 10 class this question and found that about 45 percent, $\frac{14}{31}$ of them, answered yes to both these questions. **If** I view this class as a random selection of Dartmouth students (which is admittedly a highly dubious thing to do!), then I can conclude that with about 90 percent confidence the real fraction, p_{real}, of the Dartmouth student body that will answer yes to these question, will satisfy that

$$\hat{p}_{real} - z_{0.05}\frac{1}{2\sqrt{n}} \leq p_{real} \leq \hat{p}_{real} + z_{0.05}\frac{1}{2\sqrt{n}}$$

$$\frac{14}{31} - 1.645\frac{1}{2\sqrt{31}} \leq p_{real} \leq \frac{14}{31} + 1.645\frac{1}{2\sqrt{31}}$$

$$.304 \leq p_{real} \leq .599.$$

Notice I have used the fact that the standard deviation of a success failure experiment is always less than $\frac{1}{2}$ to avoid exaggerating my chances here, and also have utilized the normal approximation to the binomial.

1.2 The Null Hypothesis

From the previous section I find that a reasonable null hypothesis will be that people **can** correctly estimate their ability to distinguish Coke and Pepsi. We see that if we restrict our attention to subjects who answer yes to question 2 , then we can quantify this null hypothesis via a binomial distribution with $p = .9$ being the success rate. With this assumption and using the normal approximation to the binomial, we find that the we can run a highly significant hypothesis test with a cut off at

$$p_{cut} = .75$$

if we use

$$n = 22.$$

In other words, we find that if our population truly can distinguish between these beverages 90 percent of the time, then there is **about** a 1 in a 100 chance that when we sample the population 22 times that we find $\hat{p}$ to be

$$p - z_{.01}\frac{\sqrt{p(1-p)}}{\sqrt{n}} = .9 - \frac{\sqrt{0.9(0.1)}}{\sqrt{22}} = .751$$

or less.

In summary our test statistic is

$$\hat{p} = \frac{Correct}{Total},$$

where $Correct$ is the number of correct soda identifications and $Total$ is the total number of attempts. We decided that we will perform 22 attempts, and that we reject the null hypothesis if the number correct is less than or equal to 16, and cannot reject the null hypothesis if if the number correct is greater than or equal to 17, since

$$\frac{16}{22} = .727 < .751 < .772 = \frac{17}{22}.$$

1.3 Implementation and Results

I will take for my pretest population randomly selected student in our Math 10 class. Since the math 10 class is small I can put there names in a random name generator and select them randomly with replacement until Ive tested my needed 22 students. In particular with respect to our class, I truly can implement the binomial assumption utilized in the previous section. .

The test involves two type of administrators: visible and hidden. The hidden administrators are hidden completely form the subject and each time the test is performed are asked to flip a coin and, if turns up heads, pour a sip or two of Coke into an clean empty cup. If the flip turns up tails they pour a sip or two of Pepsi into an clean empty cup. They then jot down on piece of paper what they poured. Then a visible administrator retrieves the cup **attempting to interact as little as possible with the hidden administrators**. At this point the subject is allowed to sample the beverage and attempt to determine if the soda is Coke or Pepsi. After doing so the visible administrator tell the subject's response to the hidden administrators, who then record this response next to the actual guess on a sheet of paper. In order not to bias the results, if a subject is tested more than once, they are offered a cup of water so that they may clean there palette.

My administrators and I found the following responses:

Actual	*Response*
Pepsi	*Pepsi*
Coke	*Coke*
Coke	*Pepsi*
Pepsi	*Coke*
Coke	*Coke*
Pepsi	*Pepsi*
Pepsi	*Pepsi*
Pepsi	*Pepsi*
Coke	*Pepsi*
Coke	*Pepsi*
Pepsi	*Pepsi*
Pepsi	*Coke*
Pepsi	*Coke*
Pepsi	*Pepsi*
Pepsi	*Pepsi*
Pepsi	*Pepsi*
Coke	*Coke*
Coke	*Coke*
Pepsi	*Pepsi*
Coke	*Coke*
Coke	*Coke*
Coke	*Coke*

With two hidden and three visible administrators this experiment took nearly 30 seconds per trial. In total a third liter of each the coke and Pepsi were needed, 30 plastic cups were needed, and a fourth liter of water.

Notice, by its design the test is very close to double blind, in the sense that the visible administrator who interacts with the subject does not know the content of the cup.

1.4 Analysis of Results

Notice 16 of our 22 trial were successful, hence we compute

$$\frac{Correct}{Total} = \frac{16}{22} = 0.727,$$

and find it is $< .75$. Hence, based on our analysis in the section 1.2, we will reject our null hypothesis. Notice there is a rule for using the normal distribution to approximate the binomial distribution in a "socially acceptable" way, namely, we need that $\hat{p} * 22 = 16$ and $(1 - \hat{p})22 = 6$ to each be greater than 5. Hence our approximation is "socially acceptable".

Notice our 90 percent confidence intervals do no not include the $p = .9$ claim. Namely, using the normal approximation, we would say that we are 90 percent confident that the true p value satisfies

$$\hat{p} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

with $\hat{p} = .727$, $\hat{\sigma} = \sqrt{.727(1 - .727)}$, $\alpha = 0.05$ and $n = 22$, in other rather that

$$.571 \leq p \leq .88.$$

Notice that we still might feel that we are in fishy territory since $(1-p)22 = 2.2 < 5$. Hence it might be nice to view the actual binomial distribution governing this experiment. We have the following table:

k	$P(X = k)$	$P(X \leq k)$	$P(X \geq k)$
12.	.00183	.00207	100.
13.	.0126	.0147	100.
14.	.0732	.0879	100.
15.	.351	.439	99.9
16.	1.38	1.82	99.6
17.	4.39	6.21	98.2
18.	11.0	17.2	93.8
19.	20.8	38.0	82.8
20.	28.1	66.1	62.0
21.	24.1	90.2	33.9
22.	9.85	100.	9.85

We see that in the real distribution there is actually a probability of 1.82 percent associated to our 16. This is best articulated by the notion that the actual P value associated to our experiment is the 1.82 percent form the table, while the normal approximation would associate a P value according the z value

$$z = \frac{\frac{16}{22} - .9}{\frac{\sqrt{0.9(0.1)}}{\sqrt{22}}} = -2.7$$

or rather a 0.3 percent! For larger n such a discrepancies will of course become smaller, but we can certainly sense the "approximation". Perhaps the "socially acceptable" rule of thumb should also include the notion that pn and $(1-p)n$ should also each be bigger than 5 (since $.9 * 22 = 2.2 < 5$). To be safe in any test I would want to use at least 50 trials 5 (since $0.1 * 5 = 5$), or resort to using the actual binomial distribution.

Notice: if we used the table to set up our original hypothesis test, then we would reject the Null hypothesis if the number correct is less than or equal to 15, and cannot reject the Null hypothesis if the number correct is greater than or equal to 16. Notice if we had set things up this way then we would not have rejected the Null hypothesis!

2 The Next Pretest

My conclusion is that this test went fairly well. Clearly the most important step in the next pretest is to determine ways of sampling a larger population. Perhaps the Dartmouth student body. Setting up a table at the Food Court and asking passer bye to participate might work. Clearly we need to pick a spot where our hidden administrators can be properly hidden. Such a method for sampling random students will clearly have a bias toward the students who frequent food court and have free time. Perhaps it is best to be honest and acknowledge this via "several hundred randomly selectstudents frequenting the Food Court, were asked the following...".If we dont want to sell our results this way, then it would be prudent to test whether we are capturing at least a somewhat random sample of the student population under these conditions. To do so it would be wise to, at the very minimum, collect some demographics on the students who respond to the initial survey. For example we could jot down their years and genders, and then later compare the break down of our sample demographics to the true break down of the Dartmouth Student population. We then could compare these figures and detect any obvious bias in the selection method. For example comparing the true Dartmouth gender ratio with our sample ratio at the 90 percent confidence level, and a χ^2 test on the year distribution would at the 90 percent confidence level would be a pair useful test. Assuming I can get the actual gender and year break downs, which I will need to look into.

Note: an analysis of this demographic data might be interesting with respect to our question as well. Namely, does gender or year affect ones perception and/or true ability to distinguish these beverages? Clearly to perform this sort of analysis many more samples will be needed.

Important Note: During this second pretest our administrators should make sure that question s1 and 2 are understandable. Our administrators will ask these questions out loud to the subjects, and should take notes if there is any consistent confusion among our subjects. If there is some confusion, then perhaps simplifying our two questions into one would be wise. First let us see if there is any confusion.

Important Note: Due to the nature of our selection process, we may introduce an interviewer bias (meaning that our visible administrators may naturally tend to ask a non-random group of particularly non-intimidating or easily accessible passersby). We can avoid this by asking the administrators to always attempt to utilize the **first** available and willing passer. If possible, our visible administrator should attempt to keep track of how many passersby refuse to even answer the questions. This might help us estimate the extend of the non-response bias . Once again the above demographic data may help identify the extent of this problem.

Our numbers: It is my belief that at the Food Court there will be a continual stream of people, and hence asking question 1 and 2 to 300 people seems reasonable. From section 1.1, I can expect that at least 100 of these people will answer yes to question 2. Hence, if a had similar administrator resources as I had during pretest 1, then based on the data described in section 1.3, it seems reasonable that in a couple hours together with a liter of water, 200 cups and 2 liters each of Coke and Pepsi, that we could perform our test on 100 subjects. Notice in this case that

$$p_{cut} = p - z_{.01}\frac{\sqrt{p(1-p)}}{\sqrt{n}} = .9 - \frac{\sqrt{0.9(0.1)}}{\sqrt{100}} = 0.83$$

would be a good cut off for our test statistic, $\hat{p} = \frac{Correct}{Total}$. Hence if we find that fewer than 83 subjects respond correctly, then we can safely reject the null hypothesis, and call the result highly significant.

Notice, as discussed in section in section **??**, that our chosen n is bigger than 50. Just to be safe let us look at the binomial distribution for $p = .9$ and $n = 100$ and compare.

k	$P(X = k)$	$P(X \leq k)$	$P(X \geq k)$
77.	.00745	.0114	100.
78.	.0198	.0312	100.
79.	.0496	.0808	100.
80.	.117	.198	99.9
81.	.260	.458	99.8
82.	.543	1.00	99.5
83.	1.06	2.06	99.0
84.	1.93	3.99	97.9
85.	3.27	7.26	96.0
86.	5.13	12.4	92.7
87.	7.43	19.8	87.6
88.	9.88	29.7	80.2

Wow! Notice there is **exactly** a 99 percent chance that we see 83 or more correct guess under the null hypothesis!