Math 1180:Lab13

Kyle Gaffney

April 9th, 2014

1 Statistical Hypothesis

Arguably the largest goal of science and the scientific method is to test hypotheses. We do this by collecting data related to the question we are trying to answer and then compare these experimental results to what is called the null hypothesis.

1.1 Null Hypothesis

The null hypothesis H_0 is a statement that nothing of interest is happening. In other words if you ran experiments to determine whether a cellular protein is causing the cell to become cancerous the null hypothesis would be that the protein has no role on cancerous development of the cell. If we are collecting data on whether or not a vaccine is causing immunity to a virus we would have the null hypothesis that it isn't. We then say that the significance level is the probability of rejecting our null hypothesis. In other words if the results of our experiment are statistically significant with respect to a certain test then it is highly likely that we should reject our null hypothesis. In other words something interesting happened.

1.2 Alternate Hypothesis

The alternate hypothesis H_a is the scientific hypothesis that we are interested in analyzing with our data. For example with the vaccination problem an alternate hypothesis would be that the vaccine caused people to become immune (or in other words we assume a positive relationship between vaccination and immunity).

1.3 Statistical Test

We would now like to find a way to evaluate whether or not we can reject our null hypothesis. Accepting the null hypothesis is not the same thing as saying that the null hypothesis is correct, just that we do not have sufficient information to reject it. There are 4 situations resulting from

our null hypothesis being either true or false, and the test results coming back as either accepting or rejecting the null hypothesis.

Our test is correct if we accept a true hypothesis or if we reject a false null hypothesis.

We have type 1 error if we our statistical test rejects a true null hypothesis. The probability that this occurs is called the significance level. Alternatively we could accept a false null hypothesis with our test. This is called type 2 error and we call the probability of this happening is one minus the power of the test.

Ideally we would like a single test that minimizes both type 1 and type 2 error. Unfortunately this is impossible, the more powerful a test is at rejecting a false null hypothesis the more likely it is to reject a true null hypothesis. This means that the type 1 error will be large for a powerful test. Alternatively if a test is weaker it is more likely to accept a true null hypothesis, but also more likely to accept a false null hypothesis. This will introduce a large type 2 error.

2 p-value test

Consider the following problem: We have a null hypothesis that the fraction of people with an allele is equal to q=.13 This will be our null hypothesis. First let us consider an alternative hypothesis of q¿.13, this means that we believe that this allele should be found more frequently than 13 percent of the time and this is the hypothesis that we wish to test. Suppose that we collected data on a group of 100 people and found that 20 people have the allele.

To find the p-value we need to add the probabilities of all results as extreme or more extreme than the measured value of 20. Since our alternative hypothesis is q¿.13 we need a one-tailed test. The p-value then is:

$$P = \sum_{k=20}^{100} b(k; 100, 0.13)$$

This is a sum of 81 binomially distributed probabilities, which would be a nightmare to do by hand. Let us use R to do the calculation for us.

Recall that:

$$b(k; n, p) = \binom{n}{k} (p)^k (1-p)^{n-k}$$

Where:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Then:

```
S=0
for (i in 20:100){
S=S+choose(100,i)*(.13)^i*(1-.13)^(100-i)}
S
```

This means that there is a 3.2 percent chance of making a type 1 error, or in other words there is a 3.2 percent probability that the null hypothesis of q=.13 is correct. We would consider such a result to be significant so we can feel fairly confident rejecting our null hypothesis in favor of our alternative hypothesis.

2.1 2-sided p-test

Now suppose that there is a cocky scientist in our group that is convinced that exactly 13 percent of people will express the allele. We no longer care to claim that it should be more, but rather that he is wrong. Then our null hypothesis is q=.13 as before, but our new alternative hypothesis is $q \neq .13$. With this new kind of alternative hypothesis we want to allow extreme events to happen on both sides. Suppose that our lab measures 20 people out of 100 have the allele as before, then the Two tail p-test is:

$$P = \sum_{k=20}^{100} b(k; 100, .13) + \sum_{i=0}^{6} b(i; 100, .13)$$

Note that 13 bisects the endpoints of our summations. This is because we wish to consider extreme events on both sides of the null hypothesis.

Once again this will be easier to calculate in R:

```
S2=0
for (i in 0:6){
S2=S2+choose(100,i)*(.13)^(i)*(1-.13)^(100-i)}
for (i in 20:100){
S2=S2+choose(100,i)*(.13)^(i)*(1-.13)^(100-i)}
S
```

Which type of p-test had a larger p-value, can you explain why this would be?

3 Monte Carlo Method for Finding the P-value

Occasionally you will face problems where the sums required to calculate the p-values cant be written down let alone even calculated. For example suppose that we had a problem where rather then tracking an allele at a single locus instead we were considering it at all possible loci. We know that this destribution is like a sum of binomial distributions and will be roughly normal, but the

function for this distribution is unknown and would be more work than its worth to calculate. In this case you would want to simulate a bunch of data using a computer and then record how many times out of the total the extreme events happen. Let us consider it for this problem of the number of people with an allele out of 100.

We start by considering the null hypothesis of q=.13 (or 13 people out of 100). Now with a Bernoulli random variable of p=.13 we want to simulate 1000 trials of flipping 100 weighted coins. We count the number of heads (with probability p=.13) and this tells us how many people in that trial had the allele.

```
Results=0
for (i in 1:1000){
Results[i]=rbinom(100,p=.13)}
head(Results)
summary(Results)
```

Now we want to count up how many times a trial had 20 or more people with the allele:

```
sum(Results>19)
P1side=sum(Results>19)/1000
P2side=(sum(Results>19)+sum(Results<7))/1000
```

How do these p-values compare to the exact p-values we calculated above?

4 Assignment for the Week

An interesting question pertinent scientifically as well as socially is related to the human sex ratio. Are you equally likely to have your child be a boy or a girl? Suppose that in a certain country there were 1000 children born in a year, with 511 of the children being boys.

1. Using the null hypothesis that q=.5 evaluate the p-value for the p-test associated with the alternative hypothesis that boys are more likely to be born than girls (q¿.5), using the experimental data.

   Hint 511 is the extreme measurement, instead of considering a population of 100 like before we are instead considering a population of 1000. I would not recommend doing this by hand.

2. Now do the 2-sided t-test for the null hypothesis q=0.5, and the alternate hypothesis $q \neq .5$. How does this compare to the p-value you got for the 1-sided test?

3. Were you able to draw any conclusions about whether or not we should accept or reject the null hypothesis for the 1-sided p-test? What about for the 2-sided p-test?

4. Now do a Monte Carlo Simulation when p=.5 to get an estimate for the p-value of the 1-sided test like we did above.

   Hint: rbinom(n,k,p) will do n-trials of k many coin flips with probability of heads =p. The output will be a vector of length n whose entries are results of the trials. For this problem I recommend letting n be at least 1000.

5. In the year 1988 in the United states there were 2,002,424 boys born and 1,907,086 girls born, out of a total of 3,909,510 newborns according to the National Center for Health Statistics. Do a Monte Carlo simulation of the null hypothesis to get the p-value of the 1-sided p-test with the maximal value of 2,002,424. In other words instead of doing 1000 trials like above do 1,000,000 trials, and instead of simulating 100 alleles you will be simulating 3,909,510 births. What is the p-value you get from the Monte Carlo Simulation? What does this tell you about the null hypothesis?