Math 180, Fall 2014: Assignment 10 Solutions

November 22, 2014

1 Bombay plague

Part of this was straightforward. Many of you went online and found the p values for χ^2 both with and without Yates' correction as well as for Fisher's exact test, and you gave me the values for both one-tailed and two tailed tests, which I accepted. However, you probably found the question of whether one should use a one-tailed or two-tailed test confusing because I did not explain fully the following in class: In general, a χ^2 test with two or more degrees of freedom must be considered as a two-tailed test since it is measuring the deviation of an observed distribution from an expected one but it is generally meaningless to speak of the direction of the deviation. However, the case of one degree of freedom, e.g., in analyzing a 2 x 2 table, is different. To understand why, observe that *if the numbers in the boxes are large enough*, then as a first approximation we can view the question as one of whether the mean of one binomial distribution differs significantly from that of another. (One needs an expectation value of at least 5 in each box; in the first table they are both over 6.) In this case, we can ask if it differs significantly upward (or downward, but pick one), in which case we are dealing with a one-tailed test, or if overall there is a significant difference – a two-tailed test. (One might expect that the probability associated with the two-tailed test should then be twice that associated with the one-tailed, and this will be the case if the numbers involved are large enough, but here Fisher's exact test will already reveal a slight difference. The reason for this difference will appear when we consider the second table in the homework.) To illustrate the approximation, let's consider the first table in the homework.

	Infected	Not Infected	
Uninoculated	10	117	127
Inoculated	3	144	147
	13	261	N = 274

Table 1: Infection Rates, Inoculated v. Uninoculated. Fisher Exact one-tailed $p = 0.022993820$, two-tailed $p = 0.042148613$

For the approximation, view this by rows as giving the data for two populations, "Uninoculated" and "Inoculated". In the first, 10 out of 127 people have been infected, in the second, 3 out of 147. In the first, the probability of infection is $p_1 = 10/127$; in the second it is $p_2 = 3/147$. Consider these as producing two binomial distributions. Recall now that if we make n tosses of a biased coin with $p(head) = p$ and $p(tail) = 1 - p = q$, then the mean number of heads in these n tosses will be np and the variance in the number of heads will be npq. In the first distribution, the number of "tosses" is $n_1 = 127$, in the second it is $n_2 = 144$. Suppose now that we watch two people, the first tossing a biased coin with $p(head) = 10/127$ a total of 127 times, the second tossing a biased coin with $p(head) = 3/144$. For each of these trials (the first tossing $n_1 = 127$ times, the second $n_2 = 144$), we record the difference between the number of heads obtained by the first person and that obtained by the second. Doing this repeatedly is a process which produces numbers (the differences) whose variance will be the sum of the variances of the two individual binomial distributions, namely

$$n_1p_1q_1 + n_2p_2q_2 = 127 \times \frac{10}{127} \times (1 - \frac{10}{127}) + 147 \times \frac{3}{147} \times (1 - \frac{3}{147}) = 12.15.$$

As we really have only one trial, from which the two individual variances are being estimated, we should really correct the first term in the sum by multiplying by 127/126 and the second by multiplying by 147/146, but this will make only a small difference which for simplicity we will overlook. (The precise formula is actually somewhat more complicated and must be used with smaller values of n_1 and n_2, but for those we have here our simple approximation is quite adequate.) The standard deviation of the process is therefore (without any corrections) $\sqrt{12.15} = 3.49$. The trial has produced a difference of $10 - 3 = 7$ or a z-score of 7/3.49 or approximately 2. The area under one tail for this value of z is approximately .0228.

What does this mean? Notice that the expected value of the difference is 7; if we got a value of 0 it would be 2 standard deviations below the expected value. Therefore, in the "trials", the probability of seeing at least as many infected individuals in the inoculated population as in the inoculated is .0228. (Of course, there could never be exactly equal numbers because we are dealing with integers, so we might want to consider a correction for continuity, but we will also neglect that for simplicity.) Likewise, with this method of computation the probability of seeing a difference greater than 7+7 (two standard deviations up from 7) or 14 would be .0228, so a two-tailed test would give a p double that of .0456, approximately. The approximation was very useful before modern computers and online computation became available because one could do it by hand and get the result just using a z-table. Moreover, in this case it still does better than either χ^2 test!

Going online, one finds the value of χ^2 without the Yates correction to be 5.130, giving $p = 0.0118$ for a one-tailed test or $p = 0.0235$ for a two tailed test. These are almost useless. (The second should be exactly twice the first; there is round-off error.) With Yates' correction the values are $p = 0.0239$ for the one-tailed test and 0.0477 for the two-tailed, which is better. Fisher's exact test, the best choice, gives, for a one-tailed test, $p = 0.0230$.

Notice that the approximation gives a better result (closer to that of Fisher's exact test) than the χ^2 test even with Yate's correction. For Fisher's exact two-tailed test, however, going online you will find that the value of p is not double that for a one-tailed test. The reason will become clear and the difference more evident in following, where the data are sparse and the calculations sufficiently simple.

When the data are sparse, as with the mortality table for those infected, one *must* use Fisher's exact test. To understand the calculation of a one-tailed and two-tailed test here we must introduce the important concept of the *odds ratio.* The second table in the homework compares the mortality rates amongst the infected for those uninoculated and those inoculated.

	Died	Recovered	
Uninoculated	6	4	10
Inoculated	0	3	3
	6	7	N = 13

Table 2: Hypothetical Death Rates, Inoculated v. Uninoculated. One-tailed $p = 0.122377622$; two-tailed $p = 0.1923076923$

The probability given by Fisher's exact test is

$$\frac{\binom{6}{0} \times \binom{7}{3}}{\binom{13}{3}} = \frac{1 \times \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3}}{\frac{13 \cdot 12 \cdot 11}{1 \cdot 2 \cdot 3}} = \frac{210}{1716} = 0.122377622.$$

(We did this in class, where I said, Suppose that Death now reaches its hand into the jar of 13 infected and takes a sample of 6, in which all happen to be uninoculated. What is the probability of such a sample?) This is the probability of getting exactly the distribution in the table, keeping the marginals fixed. However, if you go online some of the calculators will just give the result 0.192308 without explanation. The better ones will explain that the answer is 0.1224 for a one-tailed and 0.192308 for a two tailed test. To understand what this is about, let's look at the rows of the table. (We would get the same results looking at columns.) In the first row we see that amongst those who were infected but uninoculated, 6 died and 4 recovered. For these the odds of dying were therefore 6:4 or 1.5. For those who were inoculated, it was 0:3 or just 0. The odds ratio is the ratio between these two numbers, namely, 1.5/0 which here happens to be ∞. Often the odds ratio is given as the measure of effectiveness of a treatment. By itself, however, it conveys little information since there is no measure of significance attached (no p value). However, it is a measure of how different the distributions are in the two rows. It is important to notice that if the rows had been interchanged, that is, if the data for the inoculated had been listed before those for the uninoculated, then the odds ratio would be replaced by its reciprocal, which in this case would be 0. The main point is the following. If a table has odds ratio greater than or equal to 1, then a table with the same marginals but higher odds ratio will have lower individual probability. If the odds ratio of the first table is less than one then a table with the same marginals and lower odds ratio will have lower probability. In general, the farther the odds ratio from 1, up or down, the lower the probability.

Because here we have an odds ration of ∞, let's consider first a hypothetical outcome, given in the following table, where *it is important that the marginals have been kept the same*. If all that had been reported was the number uninoculated, the number inoculated, the number that had died, and the number that had recovered, then the marginals would have been fixed but the actual entries in the table unknown. We are considering one of the possible tables, had this been the case.

	Died	Recovered	
Uninoculated	5	5	10
Inoculated	1	2	3
	6	7	N = 13

Table 3: Hypothtical Death Rates, Inoculated v. Uninoculated. One-tailed $p = 0.5629370629$, two tailed $p = 1$

Here the odds ratio i (5:5)/(1:2) = 2, which is between 1 and ∞ and therefore has, as expected, higher probability than Table 2. (Note that *both* rows have changed in order to preserve the marginals; in the bottom row we shifted 1 from right to left, in the top row 1 from left to right.) When such a 2×2 table is given, for Fisher's one-tailed test do the following. If the odds ratio is greater than one, define the more extreme tables to be those *with the same marginals* with odds ratios at least equal to the given one and having lesser individual probabilities. Adding the probabilities of all of these to that of the original table gives the value of p for Fisher's one-tailed test. If the odds ratio of the given table is less than one then take the tables with odds ratio less than or equal to the given one and therefore having individual probabilities less than or equal to that of the given one. (When the odds ratio is exactly equal to one then the entries in the table are distributed exactly as predicted by the marginals; any test will give $p = 1$.) For a two-tailed test, take the inverse of the odds ratio of the original table. If that reciprocal is greater than 1, consider also all tables with that and larger odds ratios whose individual probabilities are less than or equal to that of the original table; if it is less than 1 examine tables with smaller odds ratios. Now add to the value of the one-tailed test the sum of the probabilities of all these tables; this is the p for the two-tailed test. Notice that this generally will not give you twice the value of p for the one-tailed test. The result will be 1 if all tables are more extreme than the given one since then you will be summing over all tables, which is what happens here. Of course, the value of p for the two-tailed test could not be double that for the one-tailed test since it would then exceed 1.

For our hypothetical Table 3, the only more extreme table is the original Table 2, so the value of p for the one-tailed Fisher's test would be its probability plus that of the original Table 2. (As an exercise, calculate the individual probabilities and compare with the given values.) No possible table is more extreme in the same direction (has greater odds ratio) than then original Table 2, so for it, the one-tailed test is just its individual probability. *We should accept this in this particular instance since the probabilities derived from the Table 1 show that inoculation is highly likely to beneficial.*

As an illustration of how to perform two tailed test when Table 2 is the given table, first take the reciprocal of its odds ratio, which is 0. Now we must look for all tables with odds ratio 0 and with probability less than or equal to that of Table 2; the direction has been reversed. There is exactly one possible table with odds ratio zero (where, as it happens, the denominator in the quotient is ∞):

	Died	Recovered	
Uninoculated	3	7	10
Inoculated	3	0	3
	6	7	N = 13

Table 4: The additional table for a two-tailed test. One-tailed $p = 0.0699300699$, two-tailed $p = 0.0699300699$

Its probability is less than or equal to that of the original mortality Table 2 (in fact, strictly less), so to get p for the two-tailed test we must add its individual probability to that of Table 2, getting $p = 0.1923076923$, as indicated. (If you enter Table 2 in some of the online calculators for Fisher's exact test, you will get just the result 0.192308 without explanation. When you know, as here, that a one-tailed test is appropriate, be sure to use a calculator that will give it, e.g., Vassarstats.) In this example we also encounter a certain anomaly. Suppose now that Table 4 had been the one originally given. Since there is no more extreme table, its individual probability would be the p for Fisher's one-tailed test. *But if we ask for the two-tailed test, the result remains the same, since the only table one could consider with the inverse odds ratio is Table 2, but its probability is greater than that of Table 4, so it must be excluded.* In this case the one-tailed and two tailed tests give the same answer; there is no difference!

2 Atlantic City Roulette

There are 38 pockets; if the wheel is well-oiled and balanced, the probability of landing on any one of them is 1/38. On 16 of them (reds) our gambler wins \$1, on 16 he loses \$1, and on each of the other two he loses \$.5. Therefore, the expectation for a single bet is $-\frac{1}{38} = .02632$, or, on the average, the gambler will lose 2.63 on 100 bets.

What is the standard deviation of this expectation? (After all, he could lose more or less on the next 100 bets.) This part is tricky, but think of it in the following way. Suppose that we spin the roulette wheel only once. Then we would want the standard deviation of 38 numbers, 17 of which are +1, 17 are -1, and 2 are -.5, and whose mean is -.02632. Noting that subtracting the mean means adding .02632, their variance is

$$\frac{1}{38} \times [18 \times (1 + .02632)^2 + 18 \times (-1 + .02632)^2 + 2 \times (-.5 + .02632)^2] = .910$$

and their standard deviation is $\sqrt{.910} = .954$. If we spin the wheel n times then the standard deviation ('standard error') of the sum is will be $\sqrt{n} \times .954$. Here $n = 100$, so the s.d. of 100 bets is \$9.54. Therefore, z = \$2.63/\$9.54 = .276, which gives a probability of about .39 that the gambler breaks even or comes out ahead.