## Stat 135, Fall 2006 A. Adhikari HOMEWORK 8 SOLUTIONS

1. Running the test is easy enough:  $\chi^2 = 12.1835$  with 6 degrees of freedom, so p is about 0.06, which would not be very convincing for the null hypothesis of "no relation" if I could make sense of p in this case. The test would work if the data were multinomial, e.g. if all psychology students were independent replicas of each other in terms of grades, as were the biology students, and the others. It is particularly hard to imagine everyone in the "other" category as an i.i.d. replicas. That's not even counting the problems that would arise with independence if the grades were assigned on a curve.

So I just don't like this use of the chi-squared test.

2. First turn the table into a contingency table by replacing the "number of patients" column by a "no nausea" column.  $\chi^2 = 24.5$  with 4 degrees of freedom, so the *p*-value is very tiny (no problem with the assumptions here; it's a randomized experiment). So nausea and the drugs are dependent.

I computed the proportion of nausea cases for each drug. It's clear that chlorpromazine and 150 mg of pentobarbital do better than the placebo. Here's a test for the equality of the proportions of nausea cases among placebo and chlorpromazine. Under the null hypothesis the estimate of the common p is (95 + 52)/(165 + 152) = 0.4637 and so the estimated SE of the difference between the proportions is 0.056. The observed difference between the percents is 0.23365, which is more than 4 SEs away from 0. So the proportion of nausea under the placebo is greater than that under chlorpromazine.

Yes, you could run a chi-squared test for the difference between chlorpromazine and the placebo. But it's very heavy machinery for just comparing two percents.

Chlorpromazine and pentobarbital (150 mg) are not significantly different from each other.

**3.** There are 10 distinct pairs from the ranks 1, 2, 3, 4, and 5. Here are the pairs, their sums, and the rank sum of the remaining 3.

(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	$(3,\!4)$	$(3,\!5)$	(4,5)
3	4	5	6	5	6	7	7	8	9
12	11	10	9	10	9	8	8	7	6

The values of  $U_Y$  consist of 6 subtracted from the lower line of rank sums. So the distribution of  $U_Y$  is:

value  $0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6$ probability  $0.1 \ 0.1 \ 0.2 \ 0.2 \ 0.2 \ 0.1 \ 0.1$ 

4. From lecture notes or text page 440, the rank sum of the X's is n(n+1)/2 plus the number of (i, j) pairs for which  $X_i > Y_j$ . That is,

$$T_X = \sum_{i=1}^n \sum_{j=1}^n I(X_i > Y_j) + \frac{n(n+1)}{2}$$

**a.** So  $E(T_X)$  is n(n+1)/2 plus the expected number of (i, j) pairs for which  $X_i > Y_j$ . This expected count is  $n^2 P(X > Y)$  where X and Y are independent, respectively standard normal and

normal (1, 1). Now P(X > Y) = P(Y - X < 0) which is about 0.24; that's easy to find because Y - X has the normal distribution with mean 1 and variance 2. So:

$$E(T_X) = 0.24n^2 + n(n+1)/2$$

**b.** Because n(n+1)/2 is just a constant shift,  $Var(T_X)$  is equal to the variance of the number of (i, j) pairs for which  $X_i > Y_j$ . Now

$$Var \sum_{(i,j)} I(X_i > Y_j) = \sum_{(i,j)} \sum_{(r,s)} Cov(I(X_i > Y_j), I(X_r > Y_s))$$

There are going to be 4 kinds of covariance terms, depending on how many indices are common to the two pairs.

(i) (i, j) = (r, s). There are  $n^2$  such terms. For each one,

$$Cov(I(X_i > Y_j), I(X_i > Y_j)) = Var(I(X_i > Y_j)) = 0.24 \times 0.76 = 0.1824$$

(ii)  $i \neq r$  and  $j \neq s$ . In this case the covariance is 0, because  $(X_i > Y_j)$  is independent of  $(X_r > Y_s)$ .

(iii) i = r and  $j \neq s$ . There are  $n^2(n-1)$  such terms. For each one, the covariance is

$$Cov(I(X_i > Y_j), I(X_i > Y_s)) = E(I(X_i > Y_j \text{ and } X_i > Y_s)) - 0.24 \times 0.24 = 0.114 - 0.24^2 = 0.056$$

The 0.114 is obtained by simulation as discussed in lecture; it's roughly right.

(iv)  $i \neq r$  and j = s. There are  $n^2(n-1)$  such terms.

$$Cov(I(X_i > Y_j), I(X_r > Y_j)) = E(I(X_i > Y_j \text{ and } X_r > Y_j)) - 0.24 \times 0.24 = 0.114 - 0.24^2 = 0.056$$

The 0.114 is obtained by simulation as well. For those of you who understand bivariate normals, I'll explain below why it's the same as the result of the previous simulation. It's a consequence of Var(X) = Var(Y).

So in the end,  $Var(T_X) = 0.1824n^2 + 0.112n^2(n-1)$ 

Note involving bivariate normals. The two probabilities that have been simulated are  $P(X_1 > Y_1 \text{ and } X_1 > Y_2)$  and  $P(X_1 > Y_1 \text{ and } X_2 > Y_1)$ . These can be written respectively as  $P(X_1 - Y_1 > 0 \text{ and } X_1 - Y_2 > 0)$  and as  $P(X_1 - Y_1 > 0 \text{ and } X_2 - Y_1 > 0)$ . The joint distribution of  $X_1 - Y_1$  and  $X_1 - Y_2$  is bivariate normal with means -1 and -1, variances 2 and 2, and covariance equal to

$$Cov(X_1 - Y_1, X_1 - Y_2) = Cov(X_1, X_1) = Var(X_1) = 1$$

because all the other covariances are 0 by independence. The joint distribution of  $X_1 - Y_1$  and  $X_2 - Y_1$  is bivariate normal with means -1 and -1, variances 2 and 2, and exactly the same covariance of 1, because

$$Cov(X_1 - Y_1, X_2 - Y_1) = (-1)^2 Cov(Y_1, Y_1) = Var(Y_1) = 1$$

That is why the two probabilities that are being simulated are the same.

5. By now the R part should be easy (use sample), so I won't give you the code. I hope that you have noticed that it's enough just to draw the observed histogram of  $\bar{X}_B$  from the groups of 8. That's because  $\bar{X}_A - \bar{X}_B$  is a fixed linear transformation of  $\bar{X}_B$ . The total of the 21 values is fixed

(it's 1680.1), so  $\bar{X}_A - \bar{X}_B$  is just  $(1680.1 - 8\bar{X}_B)/13 - \bar{X}_B = 1680.1/13 - (21/13)\bar{X}_B$ . Therefore the difference is large precisely when  $\bar{X}_B$  is small.

**a.** The distribution of the means of the samples of size 8 (and hence the distribution of the differences of the means of the two groups) is a slightly skewed normal.

Among my 1000 samples of size 8, only 6 had means less than 79.97875 which is the mean of the Machine B sample. They are the same 6 for which the difference was greater than 0.04201923, the observed difference of the means of the samples from the two machines. So the null hypothesis of equal distributions is rejected.

This is consistent with the calculation in Example A of Section 11.2.1, where the text calculates a parametric confidence interval (based on normality assumptions and the t distribution) for the difference between the two underlying means. That interval doesn't contain 0. The permutation test is therefore also consistent with the result of the corresponding t-test in Example B.

**b.** The permutation test is similar to the Mann-Whitney test in that under the null hypothesis, all permutations of the 21 values are equally likely. The difference is that Mann-Whitney throws away the values and replaces them by ranks, but the permutation test uses all the values.

6. Any sensible diagram that puts both distributions on the same axes, like boxplots or dotplots, will reveal that the two distributions are quite different.

7. I don't see any pairing here. So I'll use methods for comparing independent samples.

**b.** It's an act of faith to assume normality. Of course the distribution of each set of 10 looks nothing like the normal, but then just 10 replications gives no sense of the shape of the underlying distribution. So assuming both underlying distributions are normal, let's think about the SDs. The two sample SDs are quite different so it's not tempting to use the pooled t techniques. I'll use the unpooled version with 9 as the conservative choice of degrees of freedom. The difference in means is -7.77 (present – absent) with an estimated SE of  $\sqrt{S_X^2/10 + S_Y^2/10} = 2.055$ . The upper 5% point of  $t_9$  is 2.26, so the interval is  $-7.7 \pm 2.26 \times 2.055$  which is about (-12.34, -3.056).

The only thing that would change had we done the bizarre calculation of d.f. given in the text is that the df would have changed to about 14 so that the 2.26 would have changed to 2.14. That doesn't make much of a difference to the interval.

c) The confidence interval doesn't contain 0 so we know that the t test will reject the null hypothesis of equal means, at the 5% level. To do the test anyway, note that t = -7.7/2.05 = -3.75 so that a conservative p-value is 0.0046 using the  $t_9$  curve. Or you can do

> t.test(mice[,1], mice[,2])

and realize that R will do absolutely everything for you, even down to calculating those nasty fractional degrees of freedom. Either way, p is very small so you reject the null hypothesis of equal means.

8. The assumptions are that I have two i.i.d. samples, independent of each other. I am testing the null hypothesis that the two underlying "population" distributions are the same, that is,  $H_0: F = G$  where F and G are the two cdfs. The alternative is that  $F \neq G$ .

Use the Wilcoxon rank sum test (R computes U, not T):

> wilcox.test(mice[,1], mice[,2])

to see that W = 12 (which means that the rank sum was  $12 + 10 \times 11/2 = 67$ ). The *p*-value is about 0.003 So the data go against the null hypothesis of equal distributions. That's consistent with the conclusion of the *t*-test.

You should also try to combine the two samples into one long list and compute the rank-sum

directly:

> micelong  $\leftarrow$  c(mice[,1], mice[,2]) > sum(rank(micelong)[1:10]) That gives 67, consistent with W = 12.

9. There's no pairing here, so I'll use methods for comparing independent samples. "Normal theory" means that you assume the underlying distributions or normal, or at least that the sample size is large enough that the distribution of the difference between the sample means is approximately normal. These assumptions are not bad in this case. I drew the four histograms (two variables for each gender) and they all look roughly bell-shaped though the SDs are different. The sample sizes are both 65 so the usual normal (or t) approximation to the distribution of the difference between the samples mean will be quite good.

**a.** You know that the t with large d.f. is very close to normal, so save yourself some labor and just use R.

> t.test(bodytemp[1:65, 1], bodytemp[66:130,1])

Notice the df of the t - that thing's essentially the same as the normal. And with sample sizes of 65, whether you pool the SDs or not doesn't make a difference either. The interval is about (-0.054, -0.039). This just barely avoids 0; if I'd done a 99% interval, it would have contained 0. **b.** Same thing: the interval is (-3.24, 1.67), which contains 0.

10. In the previous problem we've done the parametric tests to compare the underlying difference in the means; R does the test and the CI simultaneously. You can use the confidence intervals or the *p*-values to make your decision. In the case of heart rates the null is clearly good. In the case of body temperatures the data are pointing to a difference in the underlying means, though the 95% interval is very close to hitting 0 and the 99% interval in fact does contain 0. (That's equivalent to the *p*-value being between 1% and 5%.)

For body temperatures the rank-sum test has almost the same p-value as the normal-theory test (0.027 versus 0.024). The data indicate a difference in the underlying means. For heart rates the data clearly support the null (p about 39%), just as the normal-theory test showed.