Lab #12

In lab #12, we are going to analyze two-group studies in which the outcome is continuous. Today's lab will focus on the use of the two-sample t-test, confidence intervals for the difference between two means, and transformations and rank-based methods.

A note about permutation tests: It is possible to perform permutation tests in R, but as these tests are computer intensive and not particularly common, we won't be going over them.

1 Infant Diarrhea Study

Diarrhea is a major public health problem in underdeveloped countries, especially for babies. Diarrhea leads to dehydration, which results in millions of deaths each year worldwide. Bismuth salicylate (the active ingredient in Pepto Bismol) has been shown to reduce diarrhea in adults. Researchers in Peru conducted a double-blind randomized controlled trial, published in The New England Journal of Medicine, to determine whether it would do so in infants suffering from diarrhea as well.

In their study, all infants received the standard therapy for diarrhea: oral rehydration. In addition to the rehydration, 85 babies received bismuth salicylate, while 84 babies received a placebo. The total stool volumes for all infants over the course of their illness was measured. To adjust for body size, the researchers divided by body weight to obtain their outcome of interest: stool output per kilogram of body weight. The results of their study are available on the course website.

2 Exploration

Download and attach the data set *diarrhea.txt*. With continuous data, it is always a good idea to explore and plot your data before diving into the analysis. Box plots are useful:

```
boxplot(Stool~Group, col = "grey", main = "Effect of bismuth salicylate")
```

So are histograms:

```
require(lattice)
histogram(~Stool|Group)
```

So it seems as though bismuth salicylate does help. Now we must ask – how do we know whether or not this is a statistically significant difference? What test should we use?

So far we've used many methods for testing the differences between groups, but this is the first one where we have 1) a continuous outcome and 2) two independent groups. So we turn to our new tool: 2-sample t-tests.

Note: Since our data aren't all that normal looking, we may actually not want to use the 2-sample t-test here, and we could either account for this by doing a permutation test or we will learn how to do a better "nonparametric" test soon. For now though, we will consider only the 2-sample t-test.

3 Two-Sample t-tests

Student's 2-sample t-test

"By hand" – Recall from the "Student's t-test: procedure" slide from the 4-9 class notes these three equations that we use to perform the test by hand:

- $SD_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$
    - (This one is not from the notes, but it's the formula for the pooled variance. You likely won't be asked to compute it, but you'll need to know how to use it in the following 2 equations.)
- $SE_d = SD_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
- $t = \frac{\bar{x}_1-\bar{x}_2}{SE_d}$ ~ T with $n_1 + n_2 - 2$ degrees of freedom

So we need to compute several statistics from the data in order to do this t-test (consult code). Running the code we come up with the following (I've made it resemble how this may be presented on a test):

Test question format:

> There are 85 infants in the treatment group, and 84 infants in the control group. The sample mean for the treatment group is 181.8706, the sample mean for the control group is 260.2976. The pooled standard deviation is 227.07. Perform a 2-sample t-test testing whether or not bismuth salicylate significantly improves the condition of infants.

Do this by hand, then (as usual) we'll see there was an easy way to do it in R all along.

"Using R" – R's `t.test()` function should look familiar – we used it a couple weeks ago when doing 1-sample tests and paired t-tests. The same function can also be used to compute this two-sample test, using the following syntax:

```
# Student's 2-sample t-test
t.test(Stool~Group, var.equal = TRUE)
```

The `var.equal = TRUE` statement is what tells R to do the Student's t-test vs. the Welch t-test. (Do you remember the difference between these two? We'll take a look at the difference in a second, but first let's check out the output from the Student's test.

```
Two Sample t-test

data:  Stool by Group
t = 2.245, df = 167, p-value = 0.02608
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
9.457362 147.396700
sample estimates:
mean in group Control mean in group Treatment
260.2976               181.8706
```

This looks similar to the output we saw from a one-sample t-test, but this output gives us 2 means (these are the same values we calculated for the "by hand" version. R omits the steps it took calculating the pooled standard deviation.

<u>Conclusions</u>: Infants in the treatment group had significantly less diarrhea than infants in the control group (p = .0261). [Infants in the control group had significantly more diarrhea than infants in the treatment group].

Welch's 2-sample t-test

In calculating the pooled variance, I found that the standard deviation in the treatment group was 197, and the standard deviation in the control group was 253.6. These are different by a little bit, so if I was worried about my assumption of equal variance that we made in doing the Student's t-test (recall we told R `var.equal = TRUE`), we can just omit that line from the code and R will by default run the Welch t-test.

```
# Welch 2-sample t-test
t.test(Stool~Group)
```

Note: it is possible to do the Welch t-test by hand, but the calculation is a little bit complex so we'll only do this one in R.

Notice our p-value is very close to the one we obtained using the Student's test. This will generally be the case when the variances are so close.

4 Practice Problem

Some infants are born with congenital heart defects and require surgery very early in life. One approach to performing this surgery is known as “circulatory arrest." A downside of this procedure, however, is that it cuts off the flow of blood to the brain, possibly resulting in brain damage. An alternative procedure, “low-flow bypass" maintains circulation to the brain, but does so with an external pump that potentially causes other sorts of injuries to the brain.

To investigate the treatments, surgeons at Harvard Medical School conducted a randomized controlled trial. In the trial, 70 infants received low-flow bypass surgery and 73 received the circulatory arrest approach. The researchers looked at two outcomes: the Psychomotor Development Index (PDI), which measures physiological development, and the Mental Development Index (MDI), which measures mental development. For both indices, higher scores indicate greater levels of development. The results of their study are on the course website.

a) Calculate the standard deviations of PDI and MDI for each group. Based on these values, should we do a Students test or a Welch test?

b) Conduct a t-test (whichever one you decided was most appropriate) to determine whether the difference in *physiological* development between the infants in the circulatory arrest group and the low-flow bypass group is statistically significant.

c) Conduct a t-test (whichever one you decided was most appropriate) to determine whether the difference in *mental* development between the infants in the circulatory arrest group and the low-flow bypass group is statistically significant.

d) Calculate the 95% confidence intervals for these two tests. Interpret these in terms of the context of the study. What does the confidence interval mean? Which group did better?

e) If your child had to have open-heart surgery as an infant, which treatment option would you prefer? Why?