Statistics 411/511
Lab 4

Lab Instructions: If you want to work along with the TA in lab, please sit near the front. If you prefer to go to lab but work at your own pace, please sit near the back, and wait for the appropriate time to ask any questions.

Lab 4: Rank-Sum Test, Sign Test, Wilcoxon Signed-Rank Test, Welch's t-Test

Objectives for this Lab

- Perform a Wilcoxon rank-sum test and obtain a confidence interval.
- Perform a sign test and a Wilcoxon signed-rank test on the twin hippocampus data.
- Do a Welch's t-test, which does not require the equal-variance assumptions

1. As usual, start up RStudio. Load the Sleuth3 and ggplot2 R packages. If you are working in Bexell, you'll have to install the packages again as described in item 5(a) of Lab 1.

   ```
   > library(Sleuth3)
   > library(ggplot2)
   ```

2. Download Lab4.r to your Z: drive, then open Lab4.r in RStudio (File>Open file...).

3. R's `wilcox.test()` function performs Rank-Sum tests for two-sample comparisons when the normality assumption is violated (Section 4.2 in the *Sleuth*). Let's do a rank-sum test on the cognitive load data from Section 4.1.2.

   (a) Look at the first few lines of the data frame.

   ```
   > head(case0402)
   ```

   This data frame contains three columns named `Time, Treatment,` and `Censored`. The last column contains a 0 for the *uncensored* observations (those times less than 300) and a 1 for the *censored* observations. Censoring is common in studies that observe times-to-event, such as survival times for transplant patients.

   (b) Check side-by-side boxplots.

   ```
   > qplot(Treatment,Time,data=case0402,geom="boxplot")
   ```

   These don't look so bad, until you notice that the Conventional boxplot lacks an upper whisker, due to the censoring. Checking histograms reveals very skewed distributions.

   ```
   > qplot(Time,data=case0402,geom="histogram",xlim=c(0,300)) +
   +    facet_grid(Treatment ~ .)
   ```

   (c) Perform a rank-sum test of the hypothesis that there is no difference between the groups. Since the research question is whether the "Modified" group had lower times than the "Conventional" group, this should be a one-sided test, so first check a summary of `Treatment` to discover how R orders the groups.

   ```
   > summary(case0402$Treatment)
   ```

   Since the "Conventional" group is first, the alternative is that times from the first group are more likely to be *greater* than times from the second group.

```
> wilcox.test(Time~Treatment,data=case0402,alternative="greater",
+            exact=FALSE,correct=FALSE)
```

Setting `exact=FALSE` gives a p-value based on the normal approximation (Section 4.2.3; Display 4.7). Setting `correct=FALSE` means a continuity correction is not applied. (The default is `correct=TRUE`, so if you omit `correct=FALSE`, you will get the continuity correction.)

Compare your output with Display 4.7 on page 93. You should have almost the same p-value. Why is it not the same?

Also, R's W statistic doesn't match $T = 137$, calculated in Display 4.5. The relationship between W and T is $W = T - n_1(n_1 + 1)/2$ where n_1 is the number of observations in the first sample ("Conventional").

(d) To obtain a confidence interval for the difference δ between treatments (i.e. if Y is a student's time after studying the conventional materials, then $Y - \delta$ is the time if the student had instead studied using the modified materials), do a two-sided test and specify `conf.int=TRUE`.

```
> wilcox.test(Time~Treatment,data=case0402,exact=FALSE,correct=FALSE,
+            conf.int=TRUE)
```

The resulting interval is not quite as reported in Display 4.8, but it is very close. The *Sleuth* only considers integer δ.

4. A sign test is a non-parametric (i.e. no distributional assumptions) alternative to a paired t-test. As discussed at the beginning of Section 4.4.1, the test statistic is the number of positive differences K between the paired responses. If the null hypothesis that the median difference is 0 is true, this statistic has a binomial distribution with probability of success 0.5, that is, we'd expect about half of the differences to be positive and half to be negative. Therefore, to perform this test in R, we use `binom.test()`.

(a) First, calculate the differences.

```
> difference<-with(case0202,Unaffected-Affected)
```

(b) Count how the number of pairs altogether as well as the number of positive differences.

```
> length(difference)
> length(which(difference>0))
```

This tells you there are 15 pairs and 14 of the differences are positive. That's a lot more than half. Calculate a p-value to ascertain the strength of evidence against the null hypothesis using `binom.test()`. We will do a one-sided test because that's what's done on page 100 of the text. However, when we first met this case study, the research question seemed to be two-sided ("is there a difference?" not "is there a positive difference?").

```
> binom.test(14,15,alternative="greater")
```

You should get an exact p-value of 0.0004883. "Exact" means no normal approximation was used (see page 100 in the *Sleuth*).

5. Perform a Wilcoxon signed-rank test on the twin data to do a one-sided test of H_0: population of differences is symmetric about 0.

```
> wilcox.test(difference,exact=FALSE,alternative="greater")
```

This is the same function we used in item 2(b) above, but it is doing a different test for us now. In item 3(c), the first argument was the formula `Time~Treatment`. There were two categories in the grouping variable `Treatmt`, and R understood that we were comparing two (unpaired) groups. This time, the first argument to `wilcox.test()` is a single column of data `difference`, and R understands this to be a single sample of the differences within the pairs.

The signed-rank statistic is $V = 111$ and the one-sided p-value is 0.002064, almost the same as reported on page 102 of the *Sleuth*. The conclusion is the same in either case: there is strong evidence against the null.

6. Perform a Welch's t-test to test if two population means are different when the assumption of equal variance is not met.

   We will use the finch data from Chapter 2 for illustration, even though we have no reason to doubt the equal-variance assumption. Perform Welch's t-test. This uses our old friend `t.test()` but without the `var.equal=TRUE` option.

   ```
   > t.test(Depth~Year,data=case0201) # Not assuming equal variance
   ```

   Notice that the two-sided p-value is 8.739×10^{-6} whereas *with* `var.equal=TRUE`, it is a bit smaller. This illustrates a general principle: when the assumptions *are* met, the t-tools make better use of the information in the sample. Unfortunately, we never know for sure if the assumptions are met.