

Statistics 411/511

Lab 2

Lab Instructions: The structure of the labs allows you to work along with the TA in lab, or work at your own pace, either in the lab or not. 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 2: t-tests

Objectives for this Lab

- Do a paired t-test.
 - Calculate the t-statistic and p-value by hand, with R's assistance.
 - Calculate a confidence interval by hand, with R's assistance.
 - Do a two-sample t-test.
1. As in Lab 1, start up RStudio. (Click the window icon at the bottom left of the screen. Type RStudio in the "Search programs and files box," then click the RStudio icon.) Download and open Lab2.r.
 2. Load the Sleuth3 R package. If you are working in Bexell, you'll have to install the package again as described in item 5(a) of Lab 1.

```
> library(Sleuth3)
```

3. Produce a histogram of the differences in the twin study (case0202).

```
> with(case0202,hist(Unaffected-Affected))
```

4. Next, we'll do a paired t-test to test if the population of differences in left hippocampus volumes have non-zero mean. A paired t-test is just a one-sample t-test on the differences between the pairs.

- (a) Create the differences, stored in a vector called `diffs`. Either of the following commands will accomplish this.

```
> diffs<-case0202$Unaffected-case0202$Affected  
> diffs<-with(case0202,Unaffected-Affected)
```

You can view the differences in the R Console by typing the name of the vector.

```
> diffs
```

If you create a histogram of `diffs`, you'll get the same graph as in item 2.

```
> hist(diffs)
```

- (b) Do the t-test.

```
> t.test(diffs)
```

- (c) Examine the output from `t.test()`. Find the t-statistic, degrees of freedom, p-value, and 95% confidence interval.

- (d) As noted in item 7(c) of Lab 1, there are usually multiple ways to accomplish the same thing in R. Another way to do the paired t-test is to give both groups to `t.test()` but specify you want a paired test.

```
> with(case0202,t.test(Unaffected,Affected,paired=TRUE))
```

Except for some slight differences in formatting, the output should be exactly the same as for the previous `t.test()` command. The `paired=TRUE` syntax in the above command is an example of an optional argument. The “default value” of this argument is `FALSE`, so if you give `t.test()` two groups and don’t specify `paired=TRUE`, you will get a two-sample t-test.

```
> with(case0202,t.test(Unaffected,Affected))
```

Caution: You get very different results if you do a two-sample t-test. The twin data are *paired* and therefore the two-sample t-test is inappropriate. In Chapter 3, we will focus on the assumptions of t-tests. You may recall that one of the assumptions of the two-sample t-test is that the two samples are *independently* drawn from two populations.

5. As we encounter more complicated data analyses, it will be useful to be able to calculate t-statistics, p-values, and confidence intervals “by hand” because R may not calculate the items we need.

- (a) Obtain the elements needed to calculate the t-statistic by hand. Refer to the equations on page 35 of the *Sleuth* for $SE(\bar{Y})$ and the t-ratio. In the twin study, Y refers to the difference in response between the two twins, and \bar{Y} refers to the sample mean of these differences. This sample mean is the “estimate” of the population mean in the t-ratio. The “parameter” is the null hypothesized value for the population mean, and is almost always 0, as it is here (why?). The notation s denotes the sample standard deviation, and n is always the sample size. Below is R code to calculate each of these quantities.

```
> sample.mean<-mean(diffs) #Calculate the sample mean of the differences
> sample.sd<-sd(diffs)     #Calculate the sample standard deviation
> n<-length(diffs)         #Find the sample size
```

You should get $\bar{Y} = 0.1986667$, $s = 0.2382935$, and $n = 15$. View the values by typing the names of the R variables.

```
> sample.mean
> sample.sd
> n
```

- (b) Calculate the standard error of the sample mean $SE(\bar{Y})$ according to the formula at the top of page 35 of the *Sleuth*.

```
> se.mean<-sample.sd/sqrt(n) #Calculate the standard error of the sample mean.
```

You should get $SE(\bar{Y}) = 0.06152713$.

- (c) Now calculate the t-ratio according to the formula at the bottom of page 35. Since the null hypothesized value of the population mean is 0, this is simply

$$t\text{-ratio} = \frac{\text{Estimate}}{\text{SE}(\text{Estimate})}$$

which is $\bar{Y}/\text{SE}(\bar{Y})$:

```
> sample.mean/se.mean
```

Your calculated t-ratio should be the same as in the output of `t.test()` obtained in item 4(d) above.

- (d) Calculate the p-value of the two-tailed test. This is the area under the t distribution to the right of 3.228928 and to the left of -3.228928 . Refer to Display 2.5 on page 37 and the discussion at the bottom of page 36 of the *Sleuth*. R's `pt()` command calculates probabilities from t distributions. For example,

```
> pt(3.228928,14)
```

calculates the area to the left of 3.228928 under the t distribution with 14 degrees of freedom. (Recall for a one-sample t -test, the degrees of freedom are $n - 1$.)

We want the area to the *right* of 3.228928, and since the total area is 1, the area to the left is

```
> 1-pt(3.228928,14)
```

And actually, we need twice this number to account for the area to the left of -3.228928 , so calculate the p-value as

```
> 2*(1-pt(3.228928,14))
```

Compare this to the output from the t -test in item 4(d).

- (e) Similarly, we can calculate the 95% confidence interval “by hand.” Refer to the formula in Display 2.6 on page 38 of the *Sleuth*. We have already calculated \bar{Y} and $\text{SE}(\bar{Y})$, so all we need is t -quantile $t_{n-1}(0.975)$. R will calculate this quantile for you:

```
> qt(.975,14)
```

You should get 2.144787. For a 95% confidence interval, this number should always be near 2 as long as the degree of freedom are not too small.

Calculate the confidence interval:

```
> sample.mean-qt(.975,14)*se.mean
```

```
> sample.mean+qt(.975,14)*se.mean
```

Compare to the output from `t.test` in item 4(d).

6. Do a two-sample t -test to analyze the finch data of Case Study 2.1.1.

- (a) Start by plotting the data. If you're in Bexell, you'll have to install the `ggplot2` package.

```
> library(ggplot2)
```

```
> qplot(factor(Year),Depth,data=case0201,geom="boxplot")
```

R thinks that the data in `Year` are numbers, but we are thinking of `Year` as a “factor” or grouping variable. If you omit `factor()`, you'll get a warning and a strange-looking plot.

- (b) Think about the data and what we want to test. Are the samples paired, or are they independent?

As usual, the null hypothesis is “no difference.” Is the alternative one-sided or two-sided? Which side?

- (c) Though we plan to do a one-sided t-test, first do a two-sided test. The output will give a two-sided confidence interval, which is what we usually want.

```
> t.test(Depth~Year,data=case0201,var.equal=TRUE)
```

The optional argument `var.equal` has default `FALSE` which means `t.test` usually does not assume the two populations have equal variance. We will explore this issue in Chapter 3.

Find the 95% confidence interval in the R output and compare to Display 2.9 on page 43 and to the Statistical Conclusion on pages 29 and 30 of the *Sleuth*.

- (d) To do a one-sided test, we need to determine how R orders the groups. That’s apparent from the output of the two-sided test above. The last line of the output gives the group means in R’s order. This is usually a sensible ordering—numeric or alphabetical. Here, the 1976 group is first.

If we let μ_1 denote the population mean depth for 1976 and μ_2 the population mean depth for 1978, then the null hypothesis is $H_0 : \mu_1 - \mu_2 = 0$. What’s the alternative hypothesis?

- (e) Do the one-sided two-sample t-test.

```
> t.test(Depth~Year,data=case0201,var.equal=TRUE,alternative="less")
```

In the R output, identify the t-statistic and the p-value. How would you report the conclusion of this test?