Tufte Handout

June 27, 2017

Lab 1 Data-analysis tips

In lab one you will be collecting measurements on several dependent variables, in each of two manipulated conditions (the independent variable). For each dependent variable you will want to determine whether the manipulation had an effect. That is, did the independent variable cause a change in the dependent variable. We know that differences can sometimes be observed by chance alone, so we want to conduct an inferential statistical test to determine the probability that our observed difference could have been produced by chance alone. To do this we will be conducting several t-tests. This is a short primer on the process. You can conduct t-tests in the software of your choice, or by hand using a calculator (or in excel). Here, we will use the free and open-source statistical package called R, to illustrate the process.

Let's imagine we have two groups of 10 subjects each. Group A receives condition 1 of the independent variable, and Group B recieves condition 2 of the independent variable. We then measure some behavior for all of the subject in all of the groups. To make this more concrete, let's say 10 subjects drink coffee, and the the 10 subjects drink tea. Then we present all of the subjects with a piece of art and ask them rate how beautiful they think it is on a scale from 1 to 7.

When we collect all the data we should have 20 total ratings, one for each subject in each group.

For example, if you put the data in a table it might look something like the following. Note, the grey text box shows the R code used to simulate the data. For, each group, we sample 10 numbers from a normal distribution with a mean of 4, and a standard deviation of .5. Then we put the numbers in a table.

```r
coffee<-round(rnorm(10,4,.5))
tea<-round(rnorm(10,4,.5))
all_data<-data.frame(coffee,tea)
kable(all_data,format="latex")
```

coffee	tea
4	5
3	4
3	4
3	3
4	3
4	5
4	4
4	4
4	4
3	4

We can do some quick descriptive statistics, for example, we might want to know the means of the beauty ratings for the coffee and tea groups.

```
mean(coffee)
```

```
## [1] 3.6
```

```
mean(tea)
```

```
## [1] 4
```

The means aren't very different, and of course we should expect they should be similar. After all, we sampled these means from the exact same distribution. So, we should expect that on average, the means should both be close to 4. However, they won't necessarily be exactly 4, because of variability introduced by random sampling.

The t-test

What we want to do next is conduct an independent samples t-test. We want to determine whether any possible difference between the coffee and tea groups could have been produced by chance alone. We can conduct a t-test in R very easily using the t.test function.

```
t.test(coffee,tea,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  coffee and tea
## t = -1.5, df = 18, p-value = 0.151
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9602459  0.1602459
```

```
## sample estimates:
## mean of x mean of y
##       3.6       4.0
```

R gives us back the t values, the degrees of freedom (df), and the associated p-value. The p-value tells us the likelihood that our difference, or a difference greater than the one we observed could have been produced by chance.

One more time

Let's try this whole process again, but this time we will simulate data with an actual difference between the groups. For example, let's say we want to simulate the idea that drinking coffee makes people think the art is less beautiful by at least 2 points, and then reconduct the t-test with the new simulated data. We will sample numbers from a normal distribution with mean 3 for the coffee group, and mean 5 for the tea group (for an average expected difference of 2).

```
coffee<-round(rnorm(10,3,.5))
tea<-round(rnorm(10,5,.5))
all_data<-data.frame(coffee,tea)
kable(all_data,format="latex")
```

coffee	tea
3	5
3	5
3	4
3	6
2	5
4	5
3	6
3	5
3	5
3	5

```
mean(coffee)
```

```
## [1] 3
```

```
mean(tea)
```

```
## [1] 5.1
```

```
t.test(coffee,tea,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  coffee and tea
## t = -9, df = 18, p-value = 4.404e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.590215 -1.609785
## sample estimates:
## mean of x mean of y
##       3.0       5.1
```

Writing up the results of a t-test

We've now conducted two different t-tests, and received different results on each them. You will likely find different results for all of the t-tests that you conduct for the lab experiment. However, you will use the basic sentence structure to report all of the results. When you report the results of your experiment along with statistical tests there are two important features to include, the pattern of the results, and the inferential statistic. In this situation, we would simply report the means and the t-test information. Here are is an with made-up numbers.

The coffee group gave a lower mean beauty rating (M = 3.4) than the tea group (M = 5.6), and the difference was significant, t (18) = 5.4, p < .001.

So, just in one sentence we tell the reader what the means were in both conditions, as whether the result was significant. APA style recommends reporting exact p-values when they are greater than .001 (for example p = .047). If the p-value is less than .001, then you just need to report p < .001.