T-Test

Introduction

The t-test is used to test the probability that two populations are different. In a two-treatment trial, each population will receive a different treatment (an input or management practice). In R, we can use the `t.test()` function to quickly determine the probability that two treatments are different.

Case Study 1: Wheat Fungicide

Let's start by loading the data and using the `head()` command to examine its structure. Treatment B was a fungicide applied at flag leaf. Treatment A is the untreated control.

```
wheat = read.csv("data/wheat_fungicide.csv")
head(wheat)
```

```
##   Block Treatment Yield
## 1     1         B  95.9
## 2     1         A  93.8
## 3     2         A  92.3
## 4     2         B  97.5
## 5     3         A  98.7
## 6     3         B  98.8
```

T-Test

The `t.test()` function requires at least three arguments. The first is a model statement. In testing whether populations that receive different treatments have different yields, we are modelling yield as a function of treatment.

We can express this in the following statement:

```
model = Yield ~ Treatment
```

We can then plug the model statement into our `t.test`. The second argument to the t.test, `data=wheat`, tells R which data frame to analyze. The third, `paired=TRUE`, tells R the treatments were paired (blocked) in our experiment.

```
wheat_t_test = t.test(model, data=wheat, paired=TRUE)

wheat_t_test
```

```
##
##  Paired t-test
##
## data:  Yield by Treatment
## t = -3.7166, df = 9, p-value = 0.004796
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.871641 -1.428359
```

```
## sample estimates:
## mean of the differences
##                    -3.65
```

What would happen if we left out the `paired=TRUE` argument?

```
wheat_t_test_independent = t.test(model, data=wheat)

wheat_t_test_independent
```

```
##
##   Welch Two Sample t-test
##
## data:  Yield by Treatment
## t = -3.9101, df = 17.91, p-value = 0.001034
## alternative hypothesis: true difference in means between group A and group B is not equal to 0
## 95 percent confidence interval:
##  -5.611876 -1.688124
## sample estimates:
## mean in group A mean in group B
##           94.65           98.30
```

We notice that the t-value, degrees of freedom, and p-value are all different. This is because R analyzes the two treatments as independent populations with their own variances – instead of analyzing the differences between each pair of treatments. The df of 17.91 is not a whole number – this also tells us that R is treating the variances differently and using algorithms it would not use with a paired design.

Lets look at the correct analysis again.

```
wheat_t_test
```

```
##
##   Paired t-test
##
## data:  Yield by Treatment
## t = -3.7166, df = 9, p-value = 0.004796
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.871641 -1.428359
## sample estimates:
## mean of the differences
##                    -3.65
```

Highlights of the output include the **p-value**, and the **95% confidence interval**. The p-value tells us there is about a 0.005, or 0.5% probability that the difference between the treatments is actually zero, give the actual observed difference. The 95% confidence interval is a range around the measured difference, -3.65, that is likely to include the true difference between the populations. That the confidence interval does not include zero is a reminder that the difference between the two populations is significant.

There are a couple of other arguments we may include in the "t.test" function. Remember that the t-test tests whether the measured difference between populations is different from a hypothetical difference. Usually, this hypothetical difference is zero, as we can see in the output above where it says "true difference in means is not equal to 0". But it is good to extra-careful and specify this by including the argument `mu=0` in our t.test function.

```
wheat_t_test_mu_zero = t.test(model, data=wheat, paired=TRUE, mu=0)

wheat_t_test_mu_zero
```

```
##
##  Paired t-test
##
## data:  Yield by Treatment
## t = -3.7166, df = 9, p-value = 0.004796
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.871641 -1.428359
## sample estimates:
## mean of the differences
##                   -3.65
```

Occasionally, we might want to test whether the observed difference between populations is something other than 0. What if in our wheat trial we wanted to test whether the observed difference between populations was different than -1? In that case, we would specify mu=-1 in our code.

```
wheat_t_test_mu_1 = t.test(model, data=wheat, paired=TRUE, mu=-1)
```

```
wheat_t_test_mu_1
```

```
##
##  Paired t-test
##
## data:  Yield by Treatment
## t = -2.6983, df = 9, p-value = 0.02446
## alternative hypothesis: true difference in means is not equal to -1
## 95 percent confidence interval:
##  -5.871641 -1.428359
## sample estimates:
## mean of the differences
##                   -3.65
```

We can see that our p-value has changed, as has the "alternative hypothesis" statement.

Case Study 2: Apple Variety

Lets look at our apple data:

```
apple = read.csv("data/apple_genotype.csv")
head(apple)
```

```
##   rep     gen    yield
## 1  R1  Golden 121.7750
## 2  R1 Redspur 127.5900
## 3  R2  Golden 146.7222
## 4  R2 Redspur 126.7625
## 5  R3  Golden 145.7500
## 6  R3 Redspur 132.9364
```

Our model is now that yield is a function of gen (genotype). Our model statement is therefore:

```
apple_model = yield ~ gen
```

And our t-test is as follows:

```
apple_t_test = t.test(apple_model, data=apple, paired = TRUE, mu=0)
```

```
apple_t_test
```

```
## 
##  Paired t-test
## 
## data:  yield by gen
## t = -0.41501, df = 4, p-value = 0.6994
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -28.58021  21.14724
## sample estimates:
## mean of the differences
##                -3.716484
```

We can see from above that our p-value is about 0.70. Sometimes it snows in April. And sometimes our treatments aren't significantly different. If we look at our confidence interval, we can confirm that it includes zero.

Practice

What is the t-test value for the "soybean_fungicide.csv" trial? Assume the treatments are paired. *Your t-test should reveal a p-value of 0.001048.*

What is the t-test value for the "soybean_manganese.csv" trial? Assume the treatments are paired. *Your t-test should reveal a p-value of 0.005143.*