Welch's t-Test

In this activity we will show you how to use Welch's t-test to compare two independent samples. First, here are the assumptions for Welch's t-test:

 1. The first sample of size n_1 is drawn from a normal population with mean μ_1 and variance σ_1^2.

 2. The second sample of size n_2 is also drawn from a normal population with a mean μ_2 and variance σ_2^2.

 3. The two samples are independent.

Unlike the independent samples in the activity Pooled t-Test (http://msemac.redwoods.edu/~darnold/math15/spring2013/R/Activities/PooledTTest.html), we do **not** assume that the variances of the two parent distributions are equal. That is, we do **not** assume that $\sigma_1 = \sigma_2$.

We learned in the activity Distribution of the Differences of the Sample Means for Two Random Samples (http://msemac.redwoods.edu/~darnold/math15/spring2013/R/Activities/DistributionForTwoIndependentSamples.html that if two samples of size n_1 and n_2 are drawn from two normal distributions, one having mean μ_1 and variance σ_1^2, the second having mean μ_2 and variance σ_2^2, the distribution of $\overline{X}_1 - \overline{X}_2$ is a normal distribution having mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$. That is:

$$\overline{X}_1 - \overline{X}_2 \text{ is } N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

If we replace σ_1^2 and σ_2^2 with the sample variances s_1^2 and s_2^2, respectively, then the t-statistic becomes:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_2 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Historically, the difficulty with this model is determining what to use for the degrees of freedom. In 1947, B. L. Welch wrote an article entitled *The Generalization of a "Student's" Problem when Several Different Population Variances are Involved.* Welch established the following formula for the degrees of freedom for this current t-test:

$$\text{df} = \frac{\left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{\sigma_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{\sigma_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Reading Activities of Third Grade Children

The following example comes from David Moore and George McCabe's *Introduction to the Practice of Statistics, Second Edition.*

An educator believes that new directed reading activities in the classroom will help elementary school pupils improve some aspects of their reading ability. She arranges for a third grade class of 21 students to take part in these activities for an 8-week period. A control classroom of 23 third graders follows the same curriculum without the activities. At the end of 8 weeks, all students are given a Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve.

Here are the reading scores for both the treatment group and the control group.

Treatement	Group	Control	Group
24	56	42	46
43	59	43	10
58	52	55	17
71	62	26	60
43	54	62	53
49	57	37	42
61	33	33	37
44	46	41	42
67	43	19	55
49	57	54	28
53		20	48
		85	

Next, let's state the hypotheses for the test.

- H_0: There is no difference between the mean reading scores of the treatment and control groups.
- H_1: The mean reading score of the treatment group is higher than the mean reading score of the control group.

Now, let's let μ_1 represent the mean reading score of the treatment group and μ_2 represent the mean reading score of the control group and write the hypotheses in symbolic format.

- H_0: $\mu_1 = \mu_2$
- H_1: $\mu_1 > \mu_2$

If we subtract μ_2 from both sides of the symbolic hypotheses, then the hypotheses are prepared for Welch's t-test.

- H_0: $\mu_1 - \mu_2 = 0$
- H_1: $\mu_1 - \mu_2 > 0$

Note that this is a one-tailed test whose direction of extreme is to the right.

Testing Assumptions of the Test

We begin by testing whether the treatment group and the control group were drawn from normal populations. First, enter the data.

```
treatment=c(24,43,58,71,43,49,61,44,67,49,53,56,59,52,62,54,57,33,46,43,57)
control=c(42,43,55,26,62,37,33,41,19,54,20,85,46,10,17,60,53,42,37,42,55,28,48)
```
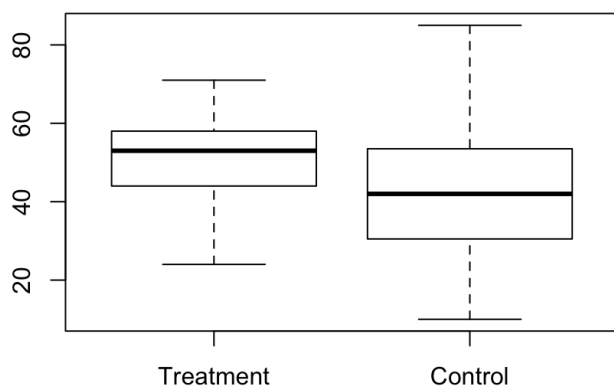
Let's draw some **qqnorm** plots.

```
par(mfrow=c(1,2))
qqnorm(treatment,main="Treatment Group")
qqline(treatment)
qqnorm(control,main="Control Group")
qqline(control)
```
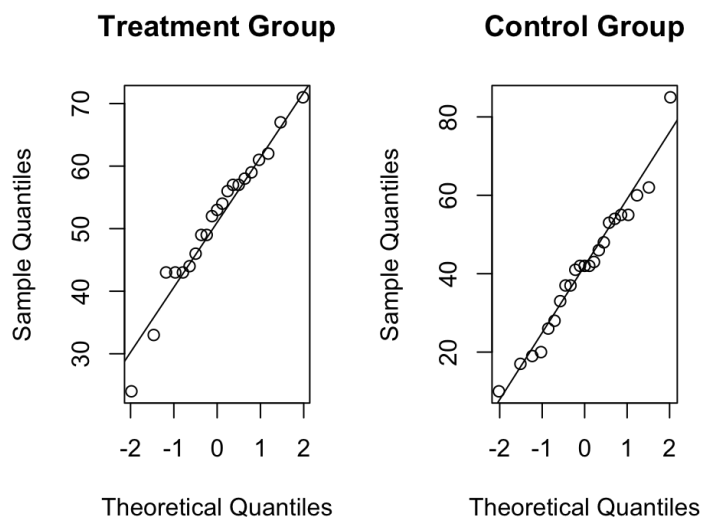
```
par(mfrow=c(1,1))
```

The evidence looks good. Both groups seem to be drawn from normal populations. Now, let's draw a boxplot to do some more comparison.

```
boxplot(treatment,control,
        names=c("Treatment","Control"))
```


Note that the median score of the Treatment Group seems higher than the median score of the Control Group. Furthermore, the inner quartile range (IQR) of the Treatment Group seems less than the inner quartile range of the Control Group. Also, the whiskers of the control group extend further in both directions than the whiskers of the Treatment Group, leading one to believe that the sample deviation of the Control Group is larger than the sample deviation of the Treatment Group.

Let's load the **psych** package again and use its **describe** command to compare the sample deviations.

```
library(psych)
describe(treatment)
```

```
##    vars  n  mean    sd median trimmed   mad min max range  skew kurtosis
## 1    1 21 51.48 11.01    53  52.12 10.38  24  71    47 -0.54     0.04
##      se
## 1 2.4
```

```
describe(control)
```

```
##    vars  n  mean    sd median trimmed   mad min max range skew kurtosis
## 1    1 23 41.52 17.15    42  41.11 17.79  10  85    75 0.27    -0.04
##      se
## 1 3.58
```

Sure enough, the standard deviation of the Treatment Group is 11.01 and the standard deviation of the Control Group is 17.15. They are different. Because the pooled t-test depends on these being the same, in this example we will use Welch's t-test to make a decision on our hypotheses.

Performing Welch's t-Test

First, we store the data we need for the t-statistic and Welch's degrees of freedom formula.

```
n1=length(treatment)
xbar1=mean(treatment)
s1=sd(treatment)
n2=length(control)
xbar2=mean(control)
s2=sd(control)
```

Because the null hypothesis assumes that there is no difference in the population means, $\mu_1 - \mu_2 = 0$. Thus, the t-statistic formula becomes:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_2 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

We can now use R to compute the t-statistic as follows.

```
t.stat=(xbar1-xbar2)/sqrt(s1^2/n1+s2^2/n2)
t.stat
```

```
## [1] 2.310889
```

Next, we use Welch's formula to calculate the degrees of freedom for the t-distribution. Note that the formula

$$df = \dfrac{\left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{\sigma_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{\sigma_2^2}{n_2}\right)^2}{n_2 - 1}}$$

will be extremely difficult to enter. However, if we let

$$A = \frac{s_1^2}{n1} \quad \text{and} \quad B = \frac{s_2^2}{n_2},$$

then Welch's formula becomes

$$df = \dfrac{(A + B)^2}{\dfrac{A^2}{n_1 - 1} + \dfrac{B^2}{n_2 - 1}},$$

which will be a bit easier to enter. So, let's begin by entering A and B in R.

```
A=s1^2/n1
B=s2^2/n2
```

Now enter the degrees of freedom formula that has been adjusted with the variables A and B.

```
df=(A+B)^2/(A^2/(n1-1)+B^2/(n2-1))
df
```

```
## [1] 37.8554
```

Notice that the degrees of freedom is not a whole number.

Calculating the p-Value

We now state the definition of the p-value, then translate it symbolically using the t-statistic. The alternative hypothesis was $H_1\colon \mu_1 - \mu_2 > 0$, so the direction of extreme is to the right.

$$p\text{-value} = P(\text{Observed Value or more extreme under } H_0)$$
$$= P(t \geq 2.311)$$

Next, we draw the distribution and shade the region representing the t-value.

```
tt=seq(-4,4,length=200)
yy=dt(tt,df)
plot(tt,yy,type="l",col="blue")
xx=seq(2.311,4,length=100)
zz=dt(xx,df)
polygon(c(2.311,xx,4),c(0,zz,0),col="gray")
arrows(2.311,.15,2.311,.05,lwd=2,angle=20,col="red")
text(2.311,.15,"2.311",pos=3)
text(2.311,.25,"df = 37.86")
```
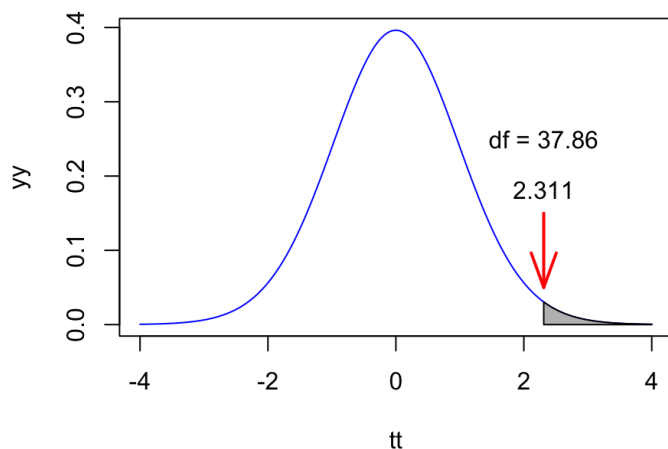
We can now use R's **pt(t,df)** command to compute the p-

value. Remember that the command returns the area to the left, but we want the area shaded to the right. Hence, we'll have to subtract our answer from 1.

```
p=1-pt(t.stat,df)
p
```

```
## [1] 0.01319121
```

Hence, the p-value is 0.01319.

Technology Helps Us Avoid Cumbersome Computations

The formula for the t-statistic and Welch's formula for the degrees of freedom are difficult to remember and they are hard to enter and arrive at the correct answer. As the statistics tests become more difficult, statisticians begin to depend more and more on technology. In this case, it is far easier to use R's **t.test** command to determine the p-value for the test of the hypotheses.

Typing **?t.test** in the console window brings up the help file for the **t.test** command. Here is the important piece from the help file.

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

We're not using a paired t-test, so we won't change the default value of the argument **paired = FALSE**. We are not assuming the variance of our two groups are equal, so we won't change the default value of the argument **var.equal = FALSE**. Our null hypotheses is $H_0: \mu_1 - \mu_2 = 0$, so we won't change the default value of the argument **mu = 0**. We're not doing a confidence interval, so we will also ignore that argument. However, the direction of extreme is to the right, so we will enter the argument **alternative = "greater"**.

```
t.test(treatment,control,alternative="greater")
```

```
##
##   Welch Two Sample t-test
##
## data:   treatment and control
## t = 2.3109, df = 37.855, p-value = 0.01319
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##   2.691293      Inf
## sample estimates:
## mean of x mean of y
##   51.47619  41.52174
```

Note that the title of the result is "Welch Two Sample t-test," which indicates that R's **t.test** command is using the formulas developed in this activity. Also, note that the output of R's **t.test** command produces the same t-statistic, nearly the same degrees of freedom (roundoff error?), and the same p-value with a minimal amount of work.

Enjoy!