Bio 483M
Wilcoxon Rank Sum Test
Adam Sasson John Wolters

The Wilcoxon Rank Sum Test, also known as the Mann-Whitney Test, is the non-parametric analogue to the two sample t-test. This test answers the same type of question as the t-test, however, it releases the assumption of a normal distribution. Like the t-test, the Wilcoxon Rank Sum test allows you to test if different samples are significantly different from each other.

As this is a non-parametric test, the assumption that the data follows a normal distribution is relaxed. At times the question of normality can be ambiguous; therefore it is useful to compare the results of the Wilcoxon Rank Sum Test to those of a two-sample t-test

Assumptions for the Wilcoxon Rank Sum Test:
- The observed values represent a random sampling of the population
- The two variables are independent
- The underlying distributions are continuous
- The measurements have an ordinal scale, in other words the data can be ranked from lowest to highest

Hypotheses:
- Ho: No ordinal difference in median for the population
- H1: There is an ordinal difference in the medians for the populations

#For this example we will use a dataset of plant growth under different treatments that comes pre-loaded with R
 PlantGrowth

#There are three different groups, a control and two different treatment groups: ctrl, trt1, trt2
#As such we would like to compare the control group to each of the treatment groups
attach(PlantGrowth)

control=weight[group=="ctrl"]
treatment1=weight[group=="trt1"]
treatment2=weight[group=="trt2"]

#For a Wilcoxon Rank Sum test the decision rule is very similar to most tests of significance
#You must set an alpha value and compare that to your p-value
#For this test we will use a=.05, if p < a then the results are significant

#Running a Wilcoxon Rank Sum test in R is very simple using the wilcox.test function.
wilcox.test(control,treatment1,exact=T,alternative="two.sided",paried=F)

    Wilcoxon rank sum test with continuity correction

data: control and treatment1
W = 67.5, p-value = 0.1986
alternative hypothesis: true location shift is not equal to 0

#Since the default settings are for a two sided test for unpaired data it is not necessary to include these
#exact=T specifies that you want the test to compute exact p-values

#We obtained a p-value greater than .05 so we cannot reject the null that treatment 1 is significantly different from the control
wilcox.test(control,treatment2,exact=T)

	Wilcoxon rank sum test with continuity correction

data: control and treatment2
W = 25, p-value = 0.06402
alternative hypothesis: true location shift is not equal to 0

#The p-value is greater than alpha, however it is so close that it is worth further investigation
wilcox.test(treatment1,treatment2,exact=T)

	Wilcoxon rank sum test with continuity correction

data: treatment1 and treatment2
W = 16, p-value = 0.01133
alternative hypothesis: true location shift is not equal to 0

#The p-value is less than alpha, treatment 1 is significantly different from treatment 2

#The results for the wilcox.test can be very similar to that for the two-sample t.test, but different in very significant ways.
#To compare lets run two sample t.tests for the same data, using the same value of a=.05 and a very similar null and alternative hypothesis.
t.test(control,treatment1,alternative="two.sided",conf.level=.95)

	Welch Two Sample t-test

data: control and treatment1
t = 1.1913, df = 16.524, p-value = 0.2504
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2875162 1.0295162
sample estimates:
mean of x mean of y
   5.032   4.661

#The p-value is greater than alpha so we reject the null for this test as well

t.test(control,treatment2,alternative="two.sided",conf.level=.95)

    Welch Two Sample t-test

data: control and treatment2
t = -2.134, df = 16.786, p-value = 0.0479
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.98287213 -0.00512787
sample estimates:
mean of x mean of y
   5.032   5.526

 #The p-value is less than alpha so we cannot reject the null. Treatment 2 is significantly different from the control.
 #This conflicts directly with the results of the wilcox.test. Notice that both are very close the the alpha value;
surrounding it on both above and below.
 #At times like this the question of normality becomes the principal concern. To determine which test we should use it is
necessary to investigate how closely the data follows a normal distribution.
 t.test(treatment1,treatment2,alternative="two.sided",conf.level=.95)

    Welch Two Sample t-test

data: treatment1 and treatment2
t = -3.0101, df = 14.104, p-value = 0.009298
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.4809144 -0.2490856
sample estimates:
mean of x mean of y
   4.661   5.526

 #The p-value is less than alpha so we reject the null. This agrees with the wilcox.test.

In this particular example we have encountered an example in which the decision to use either the Wilcoxon Rank Sum test or the t-test had drastic repercussions. For an alpha value of .05 the t-test gave us a p value that allowed us to reject the null hypothesis, however the Wilcoxon Rank Sum test, at the same alpha value, did not allow us to reject the null. In this case it is very important to reexamine the data carefully and determine whether the data follows a normal

distribution or not.

> qqnorm(control,main="Control Group")
> qqnorm(treatment2,main="Treatment Group Two")
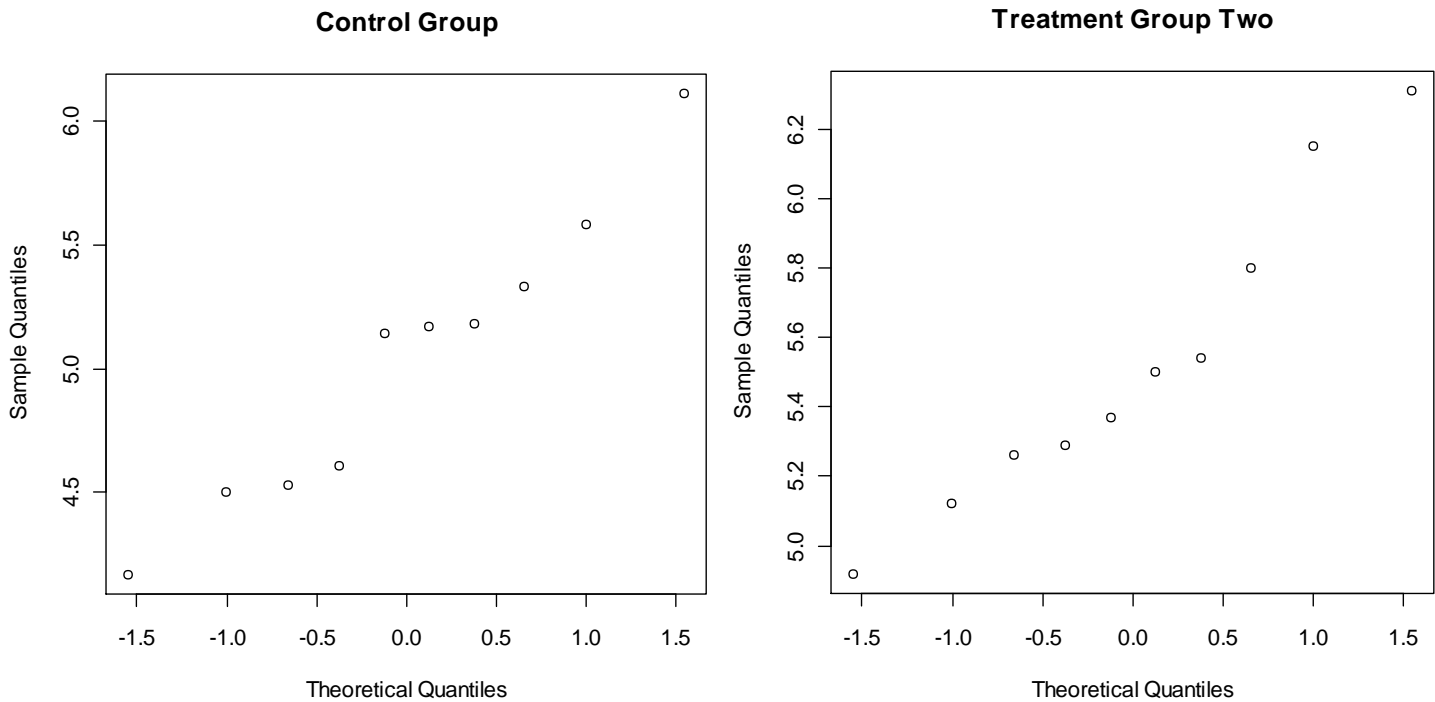
Upon examining the qq-plots, it is still not evident as to whether the control group is Gaussian or not. At this point the decision is left for the statistician to make. If you want to be careful and conservative, then you should use the Wilcoxon Rank Sum test. If you feel comfortable assuming the data is normal, either because it has a linear resembling pattern (which it does) or because you have additional biological information about the sample, then you should use the t-test as it is a more powerful test.

As we have demonstrated, this process can be very precarious at times. For this reason it is important to know how to use statistics as tools of a trade rather than blindly following an algorithmic approach.