

Lab 9 Part I

Classic non-parametric rank sum tests

T-tests and ANOVA belong to the branch of “parametric statistics”, meaning that they rely on the assumption that the residuals are normally distributed and have equal variances. If your data does not meet these assumptions and transformations are unsuccessful (or clearly impossible from the outset), you can use a number of classic non-parametric tests. Of course, assumptions of independence (spatial and temporal) as well as design considerations (sufficient replications, no pseudo-replications, randomization) still apply.

The principle of rank sum tests is to sort your data and assign ranks from 1 to N for each observation (if there is a tie due to two or more observations having the same value, the average rank is assigned to each). Then, the sum of ranks is calculated for each treatment. If there are no treatment effects, the expectation is that the difference among rank sums is zero, but there is usually a small difference if you use a real-world sample.

If you repeatedly took two samples from the same non-normal population, calculated the differences among rank sums, you would get a theoretical distribution of rank sums from which you can derive p-values. This distribution actually looks quite similar to a normal distribution, with a center at zero, but the spread is a function of your sample size (maximum rank). Similar to the t-test situation, you can now ask: what is the probability that I get a difference as big or bigger in my rank sums by random chance.

9.1 Dataset

```
lentils=read.csv("lentil_blocked.csv")
head(lentils)
```

9.2 Wilcoxon rank sum test (t-test alternative if non-normal, but similar distributions)

Let's first compare the Wilcoxon rank sum test to the t-test in a comparison of two samples. Let's subset the dataset to just have one location and two varieties for a t-test to find out whether variety A and B have different yields:

First, let's subset varieties A and B and run a t-test for reference:

```
VarA=lentils[lentils$VARIETY=="A",]
VarB=lentils[lentils$VARIETY=="B",]

t.test(VarA$YIELD, VarB$YIELD)
t.test(VarA$YIELD, VarB$YIELD, alternative ="greater")
```

If your assumptions of normality were violated for either variety, or if you have an outlier that you think represents a true measurement that you don't want to delete, you can use the Wilcoxon rank sum test:

```
wilcox.test(VarA$YIELD, VarB$YIELD)
wilcox.test(VarA$YIELD, VarB$YIELD, alternative="greater", exact=NULL)
```

This tests for differences among two medians (not means!), and gives one-tailed (H_0 : bigger or smaller medians) or two-tailed probabilities (H_0 : equal medians) for the alternative hypothesis being due to random chance.

You can still detect a significant difference with the Wilcoxon rank sum test, but you will get a larger p-value than for the t-test (i.e., a higher probability that the observed difference could be due to random chance). Non-parametric tests are generally less powerful to detect differences than parametric tests. That is why they are a secondary choice.

Note that there is an "exact;" option that you can specify if your sample sizes are not large. If you leave it out, the procedure will take some computational shortcuts. You need to set it to NULL in R for large sample sizes, otherwise the procedure may run for hours or days.

Note that you can use the Wilcoxon test for paired observations, equivalent to a paired t-test, which is quite powerful:

```
dat=read.csv("before_after.csv")
head(dat)
wilcox.test(dat$BEFORE, dat$AFTER, paired=T)
```

9.3 Kolmogorov-Smirnov test (t-test alternative if non-normal & differently shaped distributions)

The Wilcoxon rank sum test is not suitable if you have distributions for groups that do not roughly look similar in a histogram. In that case, you can use the even less-powerful Kolmogorov-Smirnov test:

```
ks.test(VarA$YIELD, VarB$YIELD)
ks.test(VarA$YIELD, VarB$YIELD, alternative="less", exact=NULL)
```

9.4 Kruskal-Wallis test (alternative to one-way ANOVA for non-normal distributions)

The Kruskal-Wallis test is an extension of the Wilcoxon rank sum test for more than two treatment levels (i.e. variety A, B, and C). The Kruskal-Wallis test is equivalent to a one-way ANOVA. If you have factorial experiments, you have to subset your dataset. As in ANOVA, a significant treatment

effect indicates that at least one population median differs from the other, and it has to be followed up by pair-wise Wilcoxon tests.

```
kruskal.test(YIELD~VARIETY,data=lentils)
```

9.5 Friedman test (nonparametric multi-factorial experimental designs with a randomized complete block design)

There is no rank sum based statistical tests for multi-factor experimental designs. However, there is one exception to analyze a randomized complete block design. Here is the code – note that Block and Variety (or whatever you name the variables) can't be stated in reverse order:

```
friedman.test(YIELD~BLOCK|VARIETY,data=lentils)
```