

Statistical comparisons course

Marcus Schmidt, University of Goettingen

PART 1

R preperation

Please do the following 3 things before the acutal work in R:

1 – Download R

You can download R for free at:

<https://cran.r-project.org/bin/windows/base/>

2 – Download some R-packages

Download the R-packages (holding certain functions) which we may use. We will only be able to work properly if this is done before the seminar.

Open R, go to File > New Script and type in:

```
install.packages("multcomp")  
install.packages("lme4") install.packages("car")  
install.packages("pgirmess")  
install.packages("MASS")  
install.packages("agricolae")
```

Then click anywhere into the first line and enter Contol + R to run this line. You need to be online to do that. Choose a mirror (you can take Germany (Münster)) and the package will be downloaded. Continue with the other 4 lines accordingly.

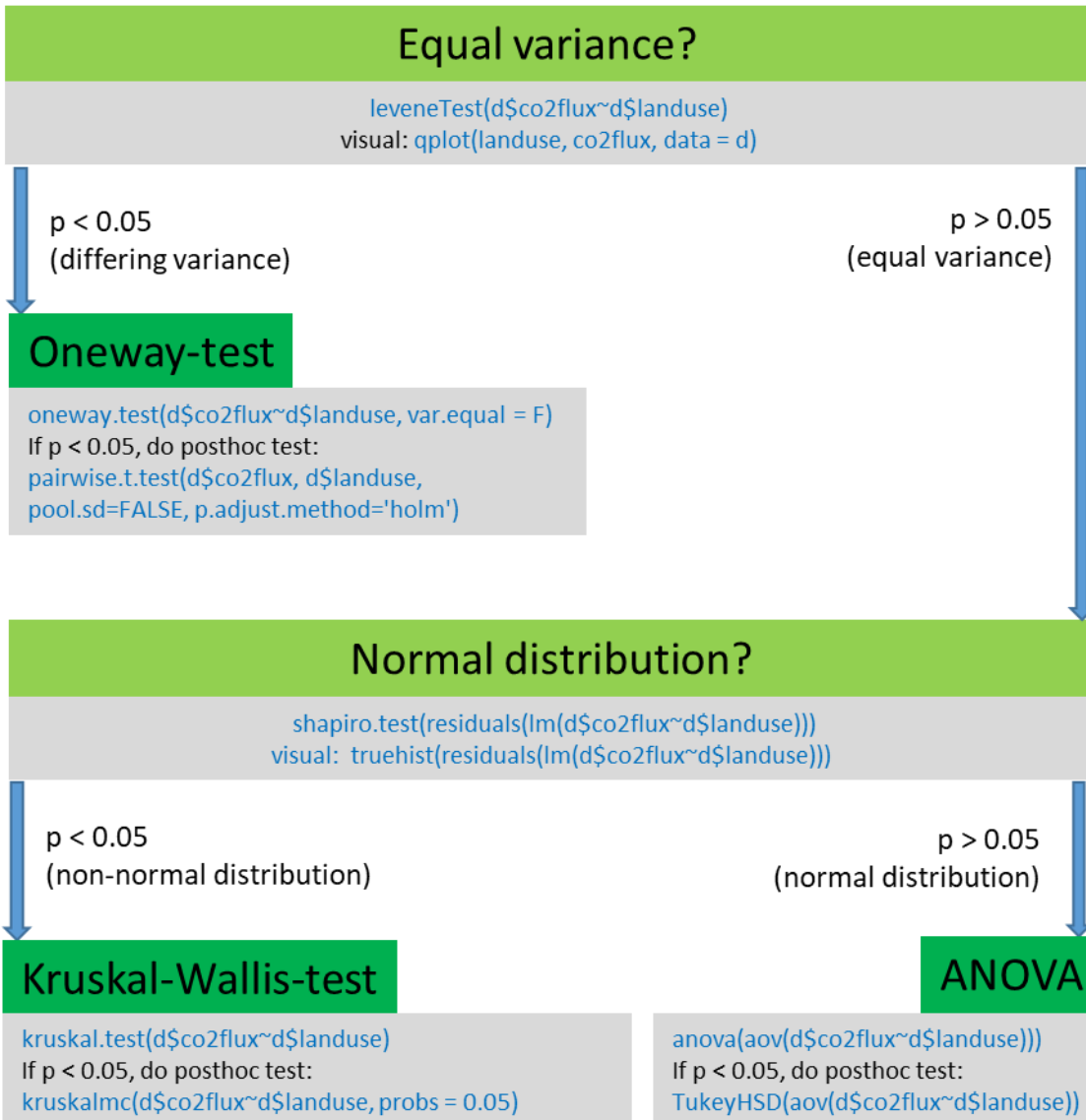
3 – Format your data ans save as .csv file

Please do so, for example in Excel. Each observation should be one row and each variable one column. There should be no blank spaces or signs like % or ? etc. Empty fields will become NA in R.

PART 2

Comparing multiple means in R

Note: This following comparisons of means for multiple groups have been checked with the Centre of Statistics Göttingen (thanks to Thomas Kneib and Peter Pütz)



Footnote: pool.sd in pairwise.t.test would take a common standard deviation for all groups. Since we are working with a differing variance at this point, we don't want that, therefore use pool.sd = F

There are several methods of p-adjustment you can choose. The point is that when you compare many groups, you would always find some differences just by chance. So the p-adjustment takes care of this.

What should stick:

- Know the assumptions and how to test them
- Significance tests need to have a p smaller than 0.05 to talk about a significant difference
- tests fulfilling more assumptions are generally more likely to detect differences

Correlation

When to use:

Use this to correlate two parameters.

Assumptions and testing:

Use method="pearson" if all parameters are normally distributed (Shapiro-Wilk test $p > 0.05$). Use method="spearman" if any parameters are not normally distributed (Shapiro-Wilk test $p \leq 0.05$).

```
shapiro.test(d$SOC)
```

```
shapiro.test(d$N)
```

How to use:

[Example: correlation of clay and cation exchange capacity] `df`

```
<- data.frame(d$SOC, d$N)
```

```
cor(d, use="pairwise.complete.obs", method="spearman")
```

```
# result: Spearman r = 0.78
```

```
par(mfrow=c(1,1)) plot(d$SOC,
```

```
d$N)
```

```
plot(df)
```

```
cor.test(d$SOC, d$N, method="spearman")$p.value
```

```
# result: p-value < 0.000008 # getting  $R^2$  value:
```

```
0.7762557^2 #
```

```
 $R^2 = 0.60$ 
```