Analysis of Root Growth Experiment

T. Neeman

01/05/2019

## Reproducible Research

Experimental setup: The researcher would like to compare the effect of a point mutation on root growth in *Arabidopsis*. She set up 5 plates; in each plate, she placed 10 newly germinated seeds – 5 of each genotype. She measured root length on Days 1, 2, 7 and 9.

### Making your analysis transparent

R is more than just a fun package for data analysis. It is also an essential tool for every biologist. In order to use R, you need to organise your data so that it can be easily interpreted. Once data are in the correct format, it’s easy to visualise the patterns in your data. In R, you can easily model your data, taking into account data structure, and obtain mean estimates, standard errors, and correct inference. Also, a set of code (suitably annotated) leaves a clear audit trail of exactly what you did to reach your conclusions. You can include the R markdown files in your appendices, and you can send it (together with your beautifully formatted data) to the reviewers. They will easily be able to replicate your analysis.

## Sample analysis of root growth using R

As part of a standard R protocol, we set up the libraries we plan to use for the analysis. In this case, we need (1) a library to read excel files; (2) a graphics library for exploratory analysis and summarising our findings; (3) a library for mixed effects linear models; (4) a library for reporting means, SE and posthoc p-values.

library(readxl)  
library(ggplot2)  
library(lmerTest)

library(emmeans)

## Set up an R project, import data

Having a separate directory for each of your research projects ensures that you and your lab team can keep track of all of your plots and analyses. Imagine you were writing up your research to submit to a journal. Three months after you submit, you receive comments back from the referees asking for changes to your analysis. Now you’ll have all the relevant documents stored in clearly marked folders and you’ll be able to re-visit all of your work, and revise accordingly.

root<-read\_excel("Root growth measurements.xlsx",sheet=1)  
head(root)

## # A tibble: 6 x 5  
## Day Plate Rep Genotype Rootlength  
## 1 1 Plate 1 1 Wild Type 11  
## 2 1 Plate 1 2 Wild Type 12  
## 3 1 Plate 1 3 Wild Type 9  
## 4 1 Plate 1 4 Wild Type 11  
## 5 1 Plate 1 5 Wild Type 12  
## 6 1 Plate 1 6 Mutant 7

str(root)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 200 obs. of 5 variables:  
## $ Day : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Plate : chr "Plate 1" "Plate 1" "Plate 1" "Plate 1" ...  
## $ Rep : num 1 2 3 4 5 6 7 8 9 10 ...  
## $ Genotype : chr "Wild Type" "Wild Type" "Wild Type" "Wild Type" ...  
## $ Rootlength: num 11 12 9 11 12 7 9 6 10 6 ...

## Changing Data types

R recognises (more or less) three types of data: numbers, factors, and text fields. In our case, we want Rep and Day to be factors, although we may also want to consider Day as a numbers, if we want to model trends in root growth over time.

root$DayF<-factor(root$Day)  
root$Rep<-factor(root$Rep)  
root$Plate<-factor(root$Plate)  
root$Genotype<-factor(root$Genotype)  
str(root)

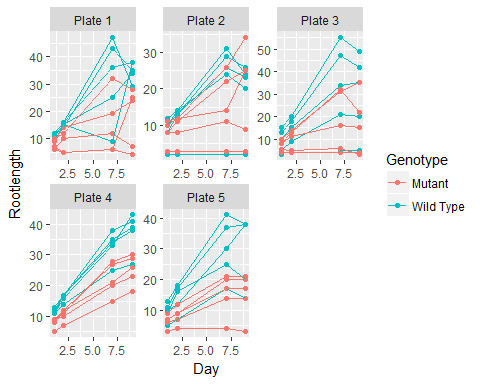
## Classes 'tbl\_df', 'tbl' and 'data.frame': 200 obs. of 6 variables:  
## $ Day : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Plate : Factor w/ 5 levels "Plate 1","Plate 2",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Rep : Factor w/ 10 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Genotype : Factor w/ 2 levels "Mutant","Wild Type": 2 2 2 2 2 1 1 1 1 1 ...  
## $ Rootlength: num 11 12 9 11 12 7 9 6 10 6 ...  
## $ DayF : Factor w/ 4 levels "1","2","7","9": 1 1 1 1 1 1 1 1 1 1 ...

## Visualising your data

This is the most important part of data analysis. You have an opportunity to understand your data, note patterns and variation. Looking at your data will allow you to spot outliers quickly. You can also assess what kind of model to use to compare Genotypes.

ggplot(root,aes(x=Day, y=Rootlength,colour=Genotype,group=Rep))+  
 geom\_point()+  
 geom\_line()+  
 facet\_wrap(~Plate, scales="free")

## Warning: Removed 1 rows containing missing values (geom\_point).



**Modelling your data**

A statistical model of the data should capture all the structure in your data. In this case, we have 5 plates, and comparisons can be made within plates (Notice that root growth differs between plates). We also have measured the same root over time.

Time is an important factor, and when comparing root growth, we want to know if the difference between WT and mutant is changing over time. We call this a Time by Genotype interaction.

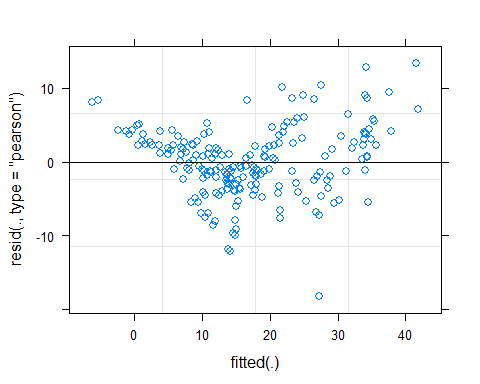
The anova function applied to our model object shows the Analysis of Variance Table. It indicates the main patterns of the data. We see immediately that there is a Time by Genotype interaction, indicating that the pattern we observed above is “real”.

Model assessment: The model fits means to every Day and Genotype combination. If we subtract off the estimated means from the actual data, what remains (“residuals”) should be just noise. We graph this noise against the fitted values.

mod1<-lmer(Rootlength~DayF\*Genotype+(1|Plate)+(1|Plate:Rep), data=root)  
anova(mod1)

## Analysis of Variance Table of type III with Satterthwaite   
## approximation for degrees of freedom  
## Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)   
## DayF 9784.1 3261.4 3 142.438 113.564 < 2.2e-16 \*\*\*  
## Genotype 487.7 487.7 1 43.361 16.983 0.0001669 \*\*\*  
## DayF:Genotype 992.0 330.7 3 142.438 11.514 8.364e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(mod1)



emmeans(mod1, revpairwise~Genotype|DayF)

## Loading required namespace: pbkrtest

## $emmeans  
## DayF = 1:  
## Genotype emmean SE df lower.CL upper.CL  
## Mutant 7.42000 1.706412 24.45 3.901553 10.93845  
## Wild Type 9.96000 1.706412 24.45 6.441553 13.47845  
##   
## DayF = 2:  
## Genotype emmean SE df lower.CL upper.CL  
## Mutant 9.28000 1.706412 24.45 5.761553 12.79845  
## Wild Type 13.73846 1.724256 25.41 10.190180 17.28674  
##   
## DayF = 7:  
## Genotype emmean SE df lower.CL upper.CL  
## Mutant 17.08000 1.706412 24.45 13.561553 20.59845  
## Wild Type 29.96000 1.706412 24.45 26.441553 33.47845  
##   
## DayF = 9:  
## Genotype emmean SE df lower.CL upper.CL  
## Mutant 18.56000 1.706412 24.45 15.041553 22.07845  
## Wild Type 30.24000 1.706412 24.45 26.721553 33.75845  
##   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95   
##   
## $contrasts  
## DayF = 1:  
## contrast estimate SE df t.ratio p.value  
## Wild Type - Mutant 2.540000 2.320457 88.94 1.095 0.2766  
##   
## DayF = 2:  
## contrast estimate SE df t.ratio p.value  
## Wild Type - Mutant 4.458459 2.333610 90.44 1.911 0.0592  
##   
## DayF = 7:  
## contrast estimate SE df t.ratio p.value  
## Wild Type - Mutant 12.880000 2.320457 88.94 5.551 <.0001  
##   
## DayF = 9:  
## contrast estimate SE df t.ratio p.value  
## Wild Type - Mutant 11.680000 2.320457 88.94 5.033 <.0001

## Reporting the results of your analysis

results1<-summary(emmeans(mod1, ~Genotype\*DayF))  
head(results1)

## Genotype DayF emmean SE df lower.CL upper.CL  
## Mutant 1 7.42000 1.706412 24.45 3.901553 10.93845  
## Wild Type 1 9.96000 1.706412 24.45 6.441553 13.47845  
## Mutant 2 9.28000 1.706412 24.45 5.761553 12.79845  
## Wild Type 2 13.73846 1.724256 25.41 10.190180 17.28674  
## Mutant 7 17.08000 1.706412 24.45 13.561553 20.59845  
## Wild Type 7 29.96000 1.706412 24.45 26.441553 33.47845  
##   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95

ggplot(results1,aes(DayF,emmean, fill=Genotype))+  
 geom\_bar(stat="identity",width=.5, position=position\_dodge(width=0.5))+  
 geom\_errorbar(aes(ymax=emmean+SE, ymin=emmean-SE),width=0.2,  
 position=position\_dodge(width=0.5))+  
 ylab("Mean Root Length (mm)")+  
 xlab("Day")+  
 annotate("text",x=c(2.87,3.87),y=c(20,22),label="\*\*\*")

