Lecture 9 Models with multiple factors Part II

T. Neeman

26th August 2021

Review from Lecture 1

Let's go back to talk about a couple of the examples from the first lecture. Experimental set-up: Are temperature mechanisms modified in a genetically-engineered tomato plant?

- Genotypes: WT or mutant
- watering conditions: normal or drought
- Outcome: leaf temperature at 7 days post-treatment

If the genetically-engineered plants are drought-resistant, we would expect that leaf temperatures would less affected by drought in the mutant plants. Let's import the data and explore these relationships.

```
library(tidyverse)
library(ggforce)
library(lmerTest)
library(ggResidpanel)
library(emmeans)
```

Import Data set: drought data.csv

```
 drought <- read_csv("../Data/drought data.csv")
```

```
## Rows: 32 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Genotype, WaterCondition
## dbl (2): plant, Temperature
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
glimpse(drought)
```

```
## Rows: 32
## Columns: 4
## $ plant          <dbl> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5~
## $ Genotype       <chr> "WT", "WT", "WT", "WT", "WT", "WT", "WT", "WT", "mutant~
## $ WaterCondition <chr> "Normal", "Normal", "Normal", "Normal", "Normal", "Norm~
## $ Temperature    <dbl> 30.0, 31.0, 31.6, 31.9, 32.0, 32.1, 32.7, 33.5, 30.0, 3~
```

Notice that there are 32 measurements, but there are multiple measurements per plant. We can get more detail about the data structure using group_by() %>% summarise(). There were a total of 8 plants: 4 WT and 4 mutant. Two plants of each genotype received normal watering conditions, and two plants of each genotype were exposed to drought conditions.

```
drought %>%
  group_by(WaterCondition,Genotype, plant) %>%
  summarise(count = n())
```

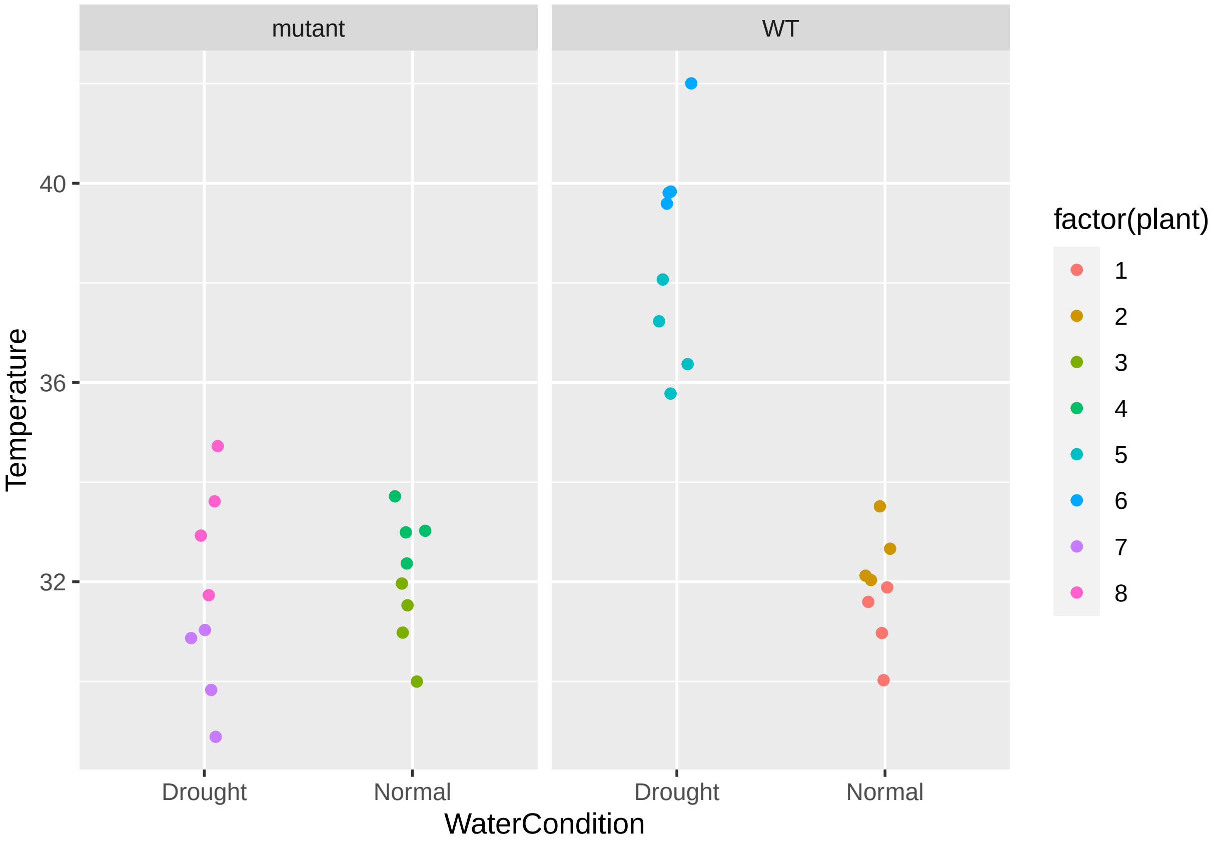
```
## ‘summarise()‘ has grouped output by ’WaterCondition’, ’Genotype’. You can
## override using the ‘.groups‘ argument.
```

```
## # A tibble: 8 x 4
## # Groups:   WaterCondition, Genotype [4]
##   WaterCondition Genotype plant count
##   <chr>          <chr>    <dbl> <int>
## 1 Drought        mutant       7     4
## 2 Drought        mutant       8     4
## 3 Drought        WT           5     4
## 4 Drought        WT           6     4
## 5 Normal         mutant       3     4
## 6 Normal         mutant       4     4
## 7 Normal         WT           1     4
## 8 Normal         WT           2     4
```

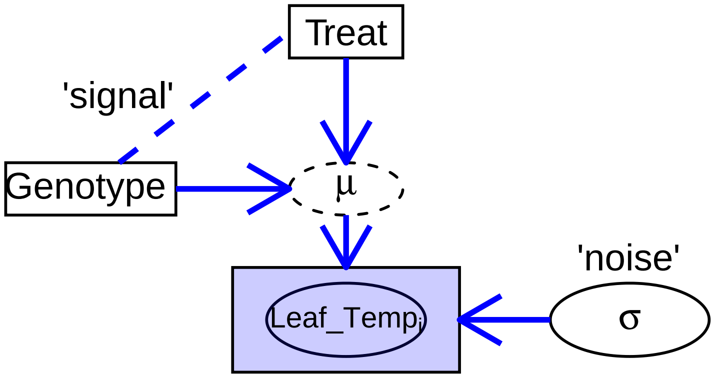
Data exploration

When we explore the data, it’s good also include the information about the data structure (plant).

```
ggplot(drought, aes(x=WaterCondition, y= Temperature, col=factor(plant)))+
  geom_jitter(width=0.1)+
  facet_wrap(~Genotype)
```



There is evidence that the mutant plants are more drought-resistant, but the strength of the evidence is limited by the small number of plants. Recall from an earlier lecture the schematic diagram of this model:



The dotted line indicates that Genotype may affect the Treatment effect, and also Treatment may affect the Genotype effect. Where would we add plant(ID) to this diagram?

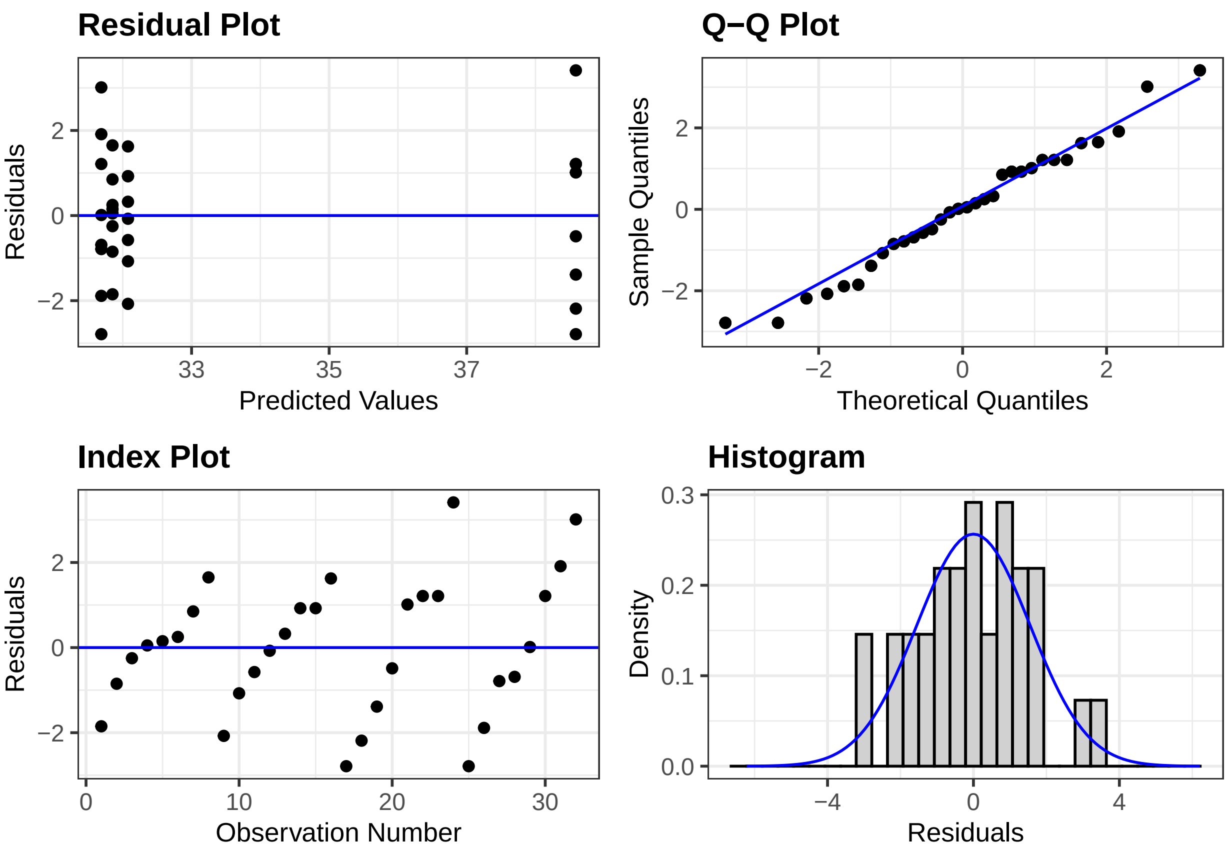
How important is it to include plant(ID) in the model? Let's fit a model with/ without plant as a random effect.

```
#Ignore plant as a factor
mod1<-lm(Temperature~Genotype*WaterCondition, data=drought)
anova(mod1)
```

```
## Analysis of Variance Table
```

```
##
## Response: Temperature
##                            Df  Sum Sq Mean Sq F value    Pr(>F)
## Genotype                    1  89.111  89.111  33.289 3.407e-06 ***
## WaterCondition              1  80.645  80.645  30.127 7.304e-06 ***
## Genotype:WaterCondition     1 101.531 101.531  37.929 1.195e-06 ***
## Residuals                  28  74.953   2.677
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

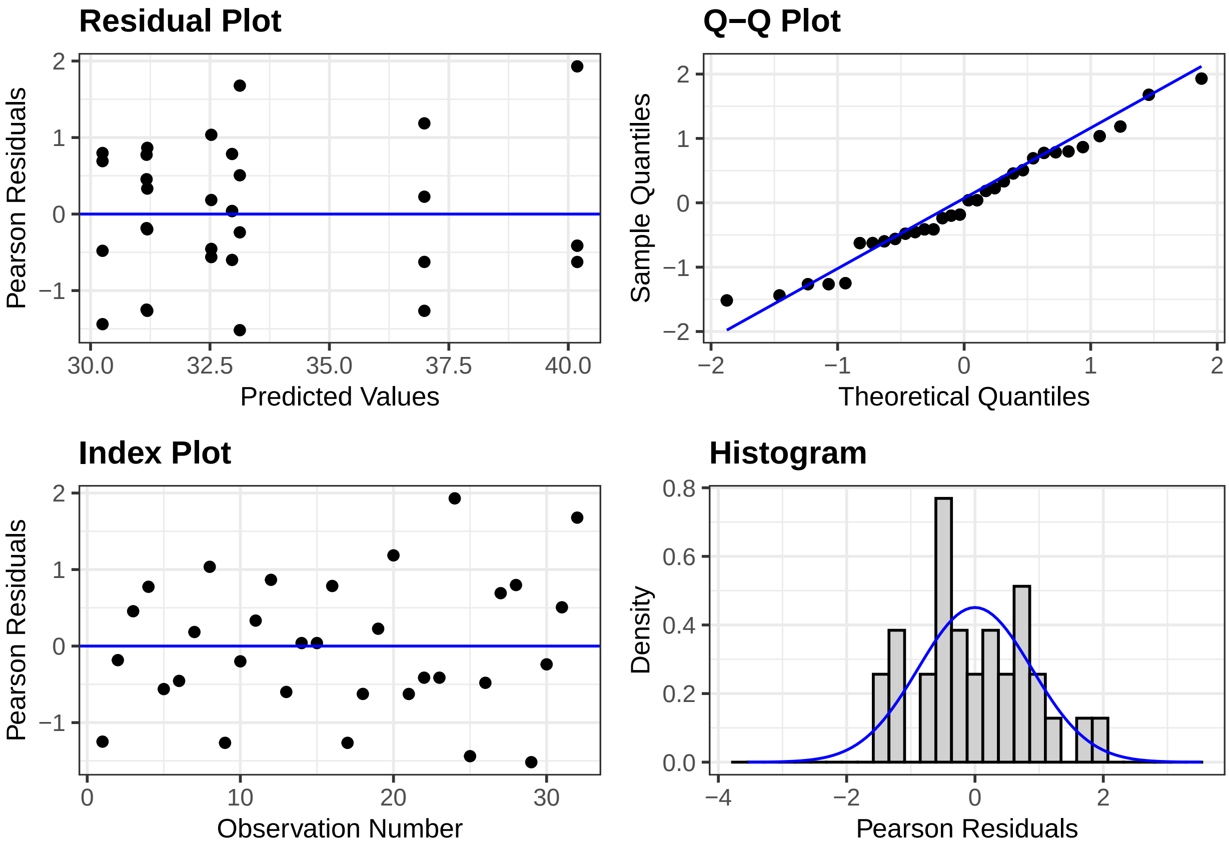
```
resid_panel(mod1)
```



```
#Include plant as a factor
mod2<-lmer(Temperature~Genotype*WaterCondition + (1|plant), data=drought)
anova(mod2)
```

```
## Type III Analysis of Variance Table with Satterthwaite’s method
##                         Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## Genotype                5.8403  5.8403     1     4  6.6257 0.06171 .
## WaterCondition          5.2854  5.2854     1     4  5.9962 0.07054 .
## Genotype:WaterCondition 6.6543  6.6543     1     4  7.5491 0.05150 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
resid_panel(mod2)
```


Residual Plot
Pearson Residuals
Predicted Values
Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Index Plot
Pearson Residuals
Observation Number
Histogram
Density
Pearson Residuals