



Sharon Nielsen Statistical Consulting & Training

---

## SC1-2016 Pot Trial PNG

Prepared for Robert Magarey

---

Sharon Nielsen Statistical Consulting & Training  
1 Patamba Street  
Koorinal 2650  
Phone: 0411017865

ABN: 24 679 927 061



# Contents

- 1 Background and aims 2
- 2 Design 2
- 3 Statistical Methods 2
  - 3.1 Measurements . . . . . 2
- 4 Results 4
  - 4.1 Dead Heart . . . . . 4
    - 4.1.1 Data summary . . . . . 4
    - 4.1.2 Analysis . . . . . 4
  - 4.2 Bored Internodes . . . . . 6
    - 4.2.1 Data summary . . . . . 6
    - 4.2.2 Analysis . . . . . 6
  - 4.3 Tunnel length and Variety . . . . . 9
    - 4.3.1 Data summary . . . . . 9
    - 4.3.2 Analysis . . . . . 10
  - 4.4 Tunnel length and Bored Internodes . . . . . 12
    - 4.4.1 Analysis . . . . . 12
- References 14

# 1 Background and aims

A glasshouse pot experiment was established in PNG on 10/10/2016. Six sugarcane varieties were tested to examine their resistance to the sugarcane top borer (*Scirpophaga excerptalis*).

## 2 Design

The experimental design was a randomised complete block design (RCBD) consisting of 16 replicate blocks. The six varieties were randomly allocated to 6 pots surrounding a central pot which was used for the application of 2 batches of egg infestations. The same variety (RQ117) was planted to the central pot. Each pot contained (mostly) 5 cane stalks. Each of the 5 stalks in each plant were assessed, however the main interest is on the infestation that occurs on the varieties planted to the peripheral pots.

## 3 Statistical Methods

### 3.1 Measurements

Three measurements of interest were:

1. Dead heart: A bernoulli (0, 1) measurement that indicates whether the growing point is dead (1) or alive (0)
2. Number of Bored Internodes.
3. Tunnel length (cm), measuring the distance from the borer's point of entry to where it finally pupated/exited the stalk as an adult moth. A longer tunnel indicates greater damage.

Note: Number of Bored Internodes and Tunnel length can only be measured on plants where a dead heart was formed (i.e. where successful infestation occurred).

The analysis will be done in four parts (one each for Dead Heart and Number of Bored Internodes and two for Tunnel Length).

1. The first analysis will be based on a generalised linear mixed model (GLMM) was fitted to the Dead heart bernoulli (0, 1) data with the logit function as the canonical link for this model. Varieties were fitted as a fixed effect, to test for differences between the six varieties. The random terms included blocks, pots within blocks and one indexed the observational units of stalks within pots. To fit these data to the GLMM we assume the responses from the individual stalk to be independently and identically distributed. Because of this assumption, the scale parameter was fixed to be 1. This model was fitted using the `asrem1` function in the `ASRem1-R` package (Butler, 2009) in R (R Core Team, 2016).
2. The second analysis involves the bored internode variable, this data is count data that does not follow the usual Poisson distribution partly due to an over abundance of zeros. Due to this the non-zero data will firstly be analysed using a linear mixed model. As the model assumptions are not met a second analysis is done using a generalized linear model with a quasi-Poisson distribution. This distribution attempts to describe the variance in the data that cannot be explained by a Poisson distribution. This model was fitted using the `glm` function in R (R Core Team, 2016), therefore, replicate and pot within replicate were

fitted as fixed terms in the model. Also, for completeness, the data is examined using a Fisher's exact test of independence, to determine whether there is a relationship between the number of bored internodes and Variety (ignoring Replicate and pot within Replicate).

3. The last two analyses rely on linear mixed models using ASReml-R (Butler, 2009). These are used to examine the relationship between Tunnel Length and Variety and then between Tunnel Length and the Number of Bored Internodes. The first model can symbolically be written as:

$$\text{Tunnel Length} \sim \text{mean} + \text{Variety} + \textit{Rep} + \textit{Rep:Pot}$$

The second model can symbolically be written as:

$$\text{Tunnel Length} \sim \text{mean} + \text{Bored Internodes} + \text{Variety} + \text{Bored Internodes:Variety} + \textit{Rep} + \textit{Rep:Pot}$$

**Notation:** Terms fitted in the model as random are italicised; all other terms are fitted as fixed terms.

The model assumptions are that the residuals are normally distributed, they have a constant variance and are independent. Also, that the factor level variances are equal for the treatments, this is tested using the Brown-Forsythe Test.

To determine whether the random term in the model are statistically significant, we use a log-likelihood ratio test, to test

$H_0$  : The reduced model is true

$H_1$  : The current model is true

Here to test the null hypothesis that an arbitrary group of  $k$  coefficients from the model is set equal to zero (e.g. no relationship with the response), we need to fit two nested models, the:

1. **reduced model** which omits the  $k$  predictors in question, and
2. **current model** which includes them.

The likelihood-ratio test is

$$\Delta = -2 \times \log L \text{ from reduced model} - (-2 \times \log L \text{ from current model})$$

and the degrees of freedom is  $k$  (the number of coefficients in question). The  $p$ -value is  $P(\chi_k^2 \geq \Delta)$ .

Following the analysis of data with an appropriate linear model or linear mixed model with qualitative explanatory variables, if there is significance evidence from the model that the explanatory variable means differ, a multiple comparison test is used to determine which of the means are different. In this report a Tukey's multiple comparison test is used to determine which means among a set of means differ from the rest at a family significance level of 5%.

NOTE: The data contained results for the central pot. These were removed for this analysis.

## 4 Results

### 4.1 Dead Heart

#### 4.1.1 Data summary

The overall raw proportion of canes with dead hearts is 0.14, with a standard deviation of 0.35.

#### 4.1.2 Analysis

The ANOVA table follows:

|         | Df | denDF | F.inc | Pr    |
|---------|----|-------|-------|-------|
| Variety | 5  | 87.8  | 3.95  | 0.003 |

From the ANOVA table we see there is a significant effect due to Variety on Dead Heart,  $p$ -value = 0.003.

From the log-likelihood ratio test we can see there is significant variance associated with Replicate,  $p$ -value < 0.001. Also, we can see there is significant variance associated with Pot within Replicate,  $p$ -value < 0.001.

The predicted values for Variety follow:

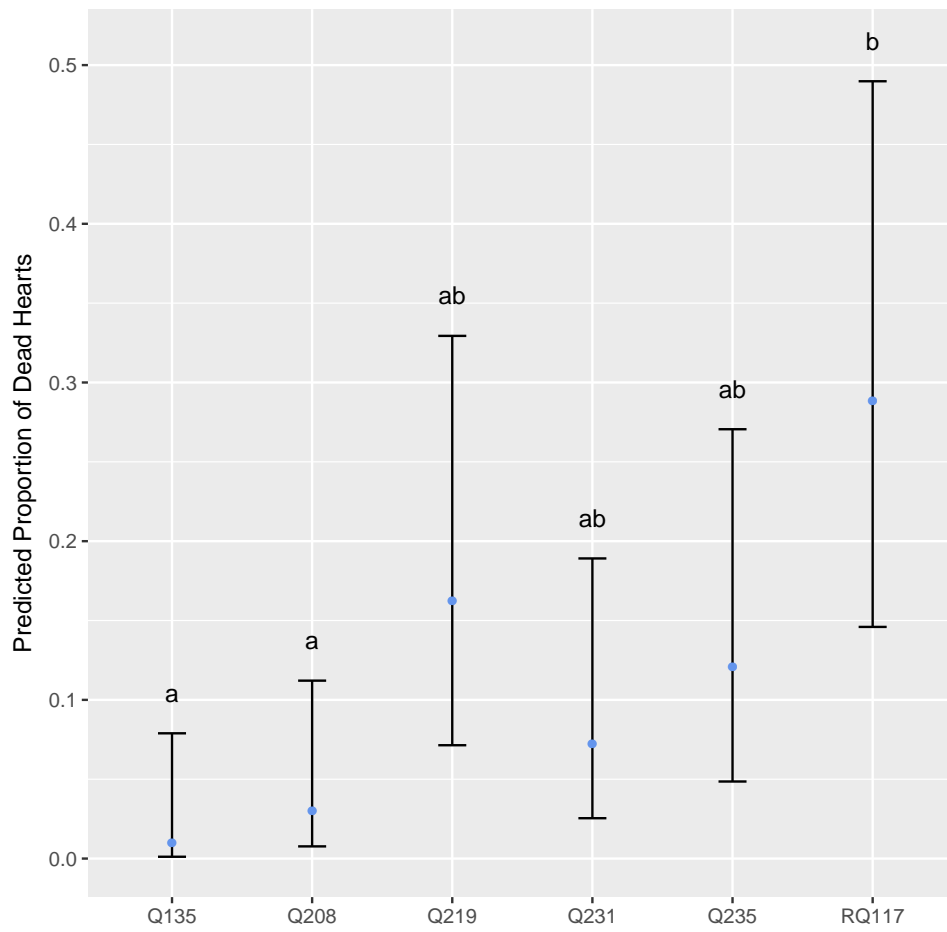
| Variety | Predicted Value | Standard Error | Ranking |
|---------|-----------------|----------------|---------|
| RQ117   | -0.90           | 0.440          | b       |
| Q219    | -1.64           | 0.473          | ab      |
| Q235    | -1.98           | 0.506          | ab      |
| Q231    | -2.55           | 0.559          | ab      |
| Q208    | -3.47           | 0.712          | a       |
| Q135    | -4.62           | 1.101          | a       |

These results indicate that Varieties Q135 and Q208 have significantly less Dead Hearts than RQ117, at the 5% family significance level.

The back-transformed predicted values follow:

| Variety | transformed.value | approx.se |
|---------|-------------------|-----------|
| RQ117   | 0.29              | 0.08      |
| Q219    | 0.16              | 0.05      |
| Q235    | 0.12              | 0.04      |
| Q231    | 0.07              | 0.03      |
| Q208    | 0.03              | 0.02      |
| Q135    | 0.01              | 0.01      |

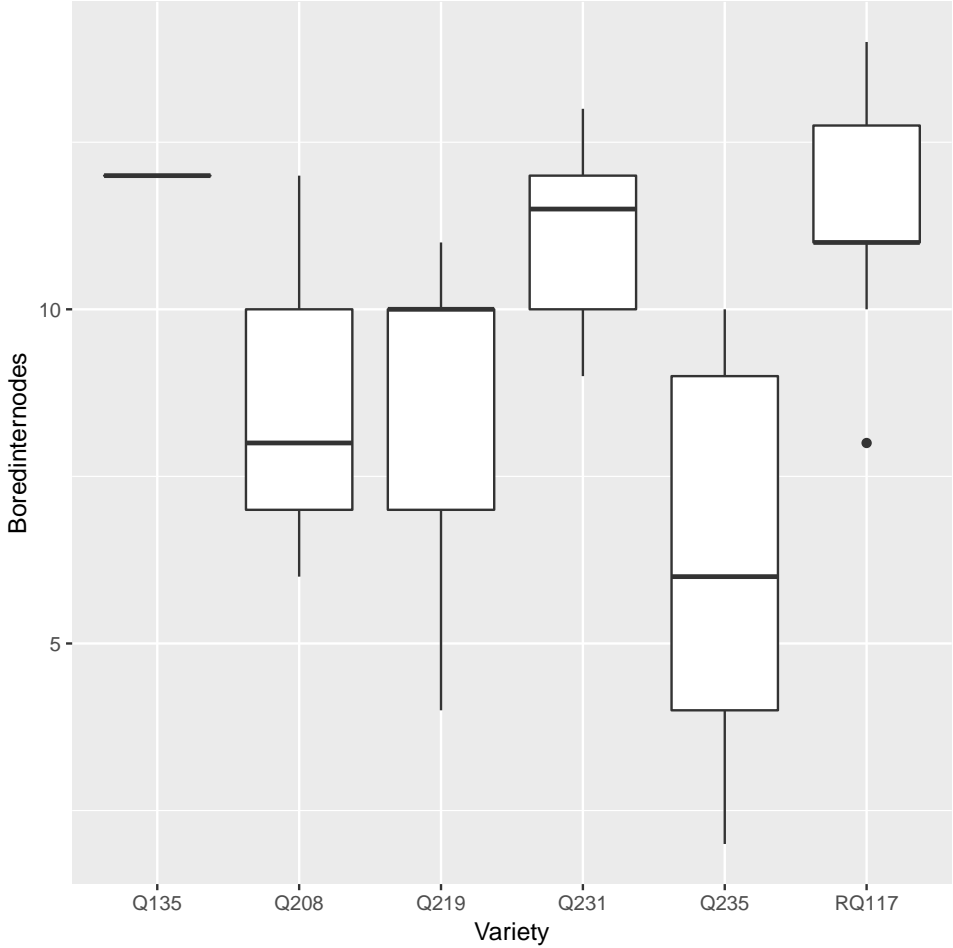
A graph of the predicted means follows:



Note: The error bars represent the 95% confidence interval.

## 4.2 Bored Internodes

### 4.2.1 Data summary



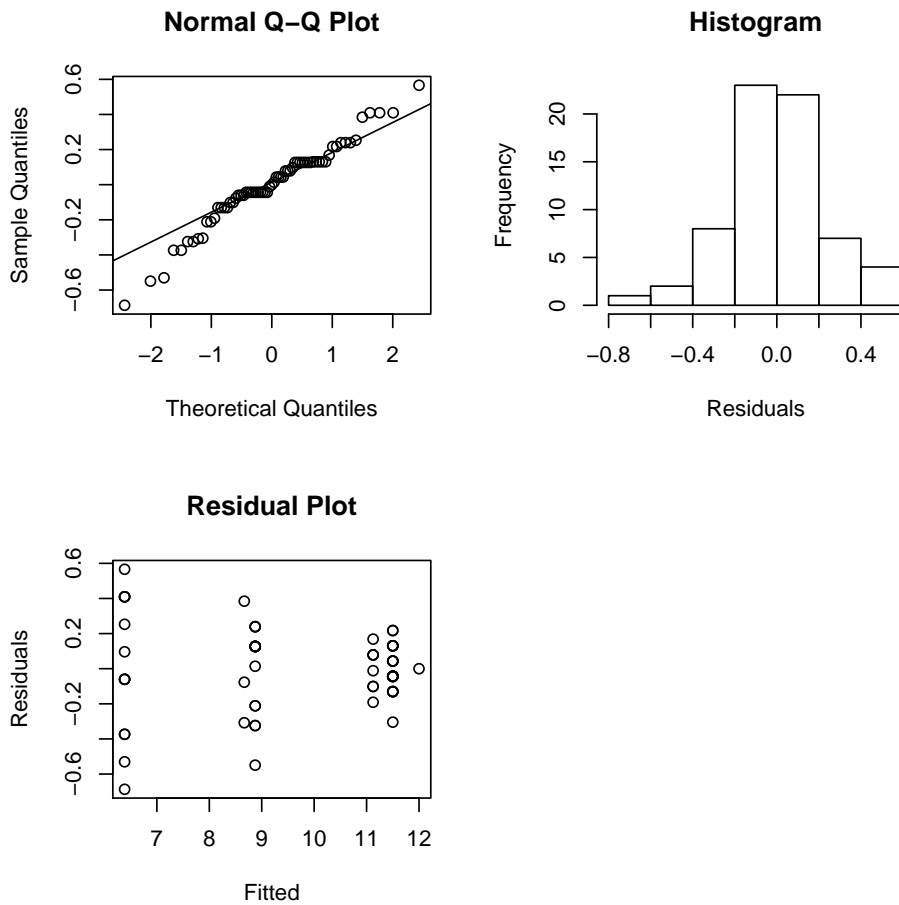
### 4.2.2 Analysis

The variances of the Varieties are the same (as the Brown-Forsythe Test  $p$ -value = 0.251).



The model assumptions can be checked using the plots below:

### Residual Plots



The Shapiro-Wilk test of normality is used to determine if the residuals are normally distributed. In this case we can conclude that the residuals are derived from a population that is normally distributed (as the Shapiro-Wilk  $p$ -value = 0.072). We can see in the residual plot that the residuals have a constant variance and we assume that the residuals are independent.

The ANOVA table follows:

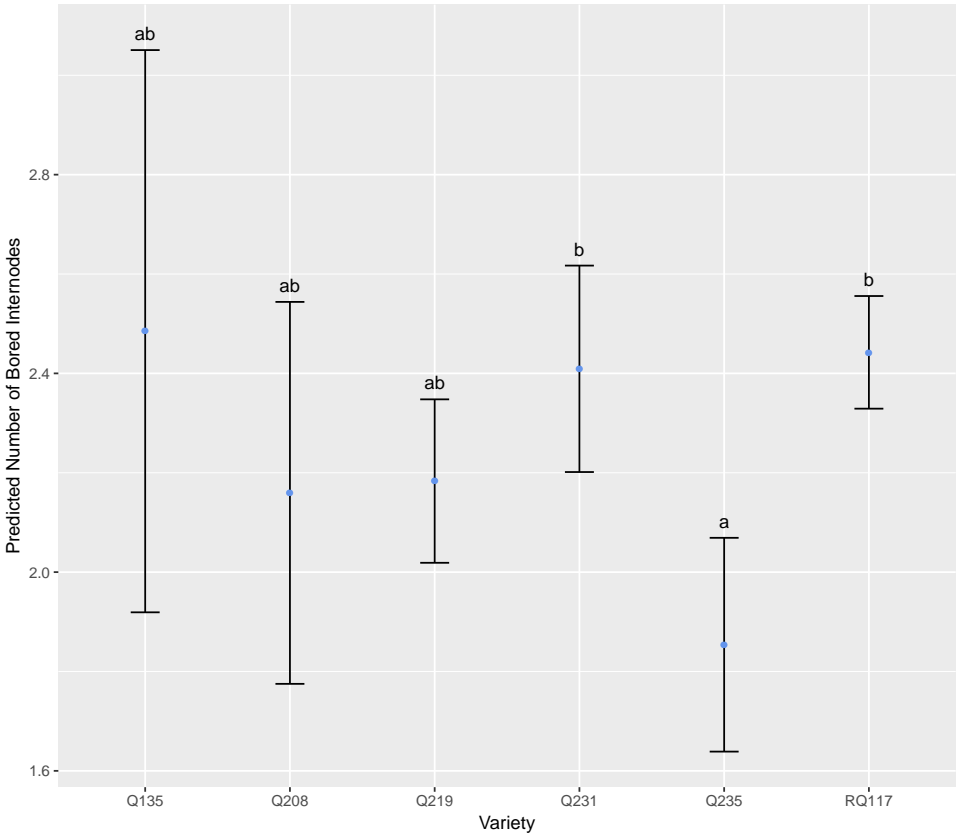
|         | Df | denDF | F.inc | Pr    |
|---------|----|-------|-------|-------|
| Variety | 5  | 58.0  | 5.30  | 0.000 |

From the ANOVA table we see there is a significant effect due to Variety on Bored Internodes,  $p$ -value < 0.001.

The predicted values for the main effect of Variety follow:

| Variety | Predicted Value | Standard Error | Ranking |
|---------|-----------------|----------------|---------|
| Q235    | 1.9             | 0.1            | a       |
| Q208    | 2.2             | 0.2            | ab      |
| Q219    | 2.2             | 0.1            | ab      |
| Q231    | 2.4             | 0.1            | b       |
| RQ117   | 2.4             | 0.1            | b       |
| Q135    | 2.5             | 0.3            | ab      |

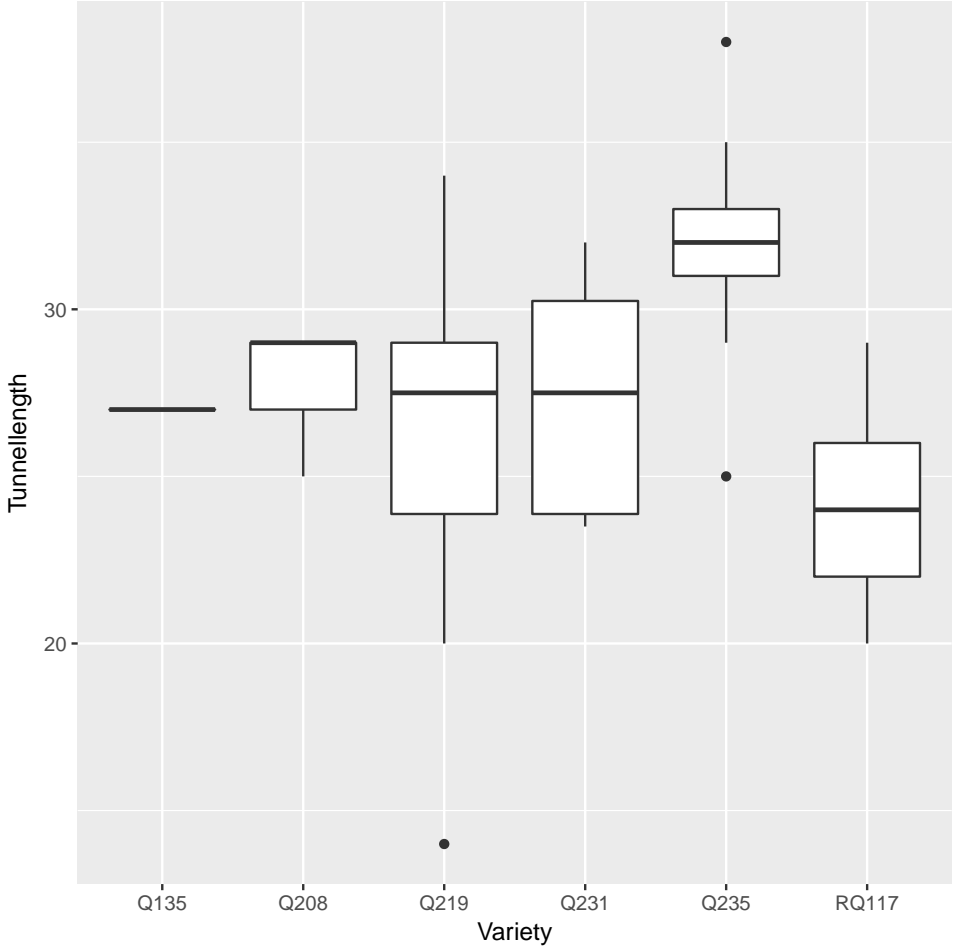
This indicates that Q235 has significantly less bored internodes than Q231 and RQ117, at the family 5% significance level.



Note: The error bars represent the 95% confidence interval.

### 4.3 Tunnel length and Variety

#### 4.3.1 Data summary

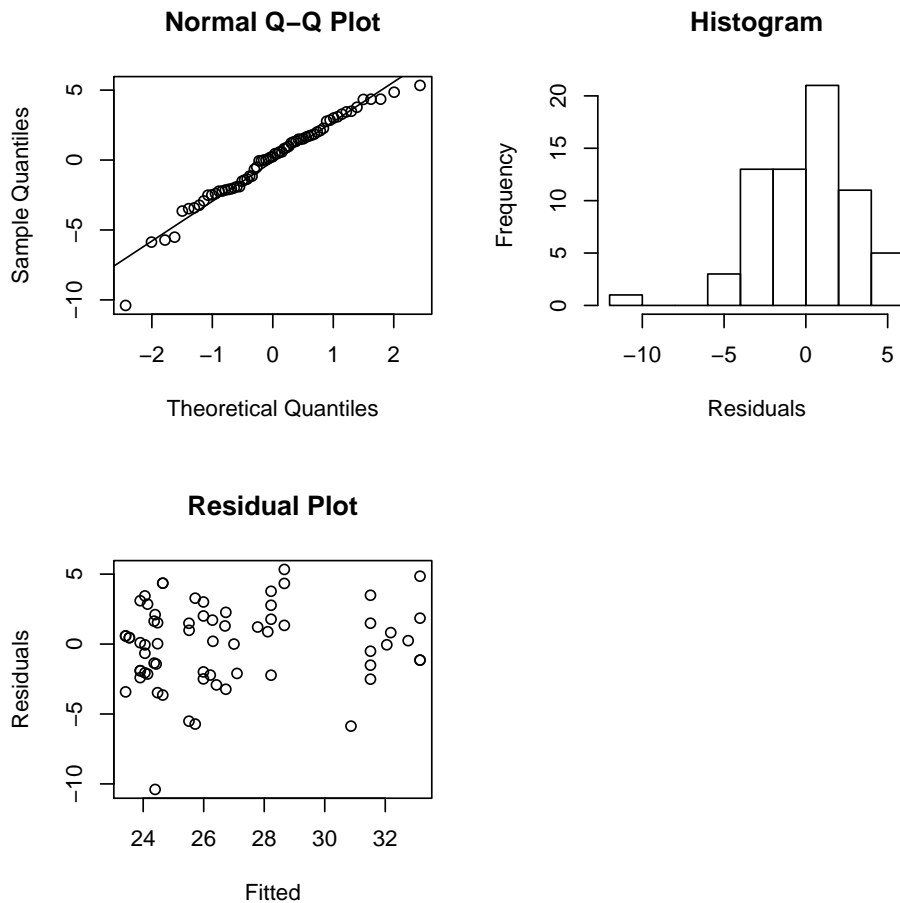


### 4.3.2 Analysis

The variances of the Varieties are the same (as the Brown-Forsythe Test  $p$ -value = 0.19).

The model assumptions can be checked using the plots below:

#### Residual Plots



The Shapiro-Wilk test of normality is used to determine if the residuals are normally distributed. In this case we can conclude that the residuals are derived from a population that is normally distributed (as the Shapiro-Wilk  $p$ -value = 0.051). We can see in the residual plot that the residuals have a constant variance and we assume that the residuals are independent.

The ANOVA table follows:

|         | Df | denDF | F.inc | Pr    |
|---------|----|-------|-------|-------|
| Variety | 5  | 22.4  | 7.28  | 0.000 |

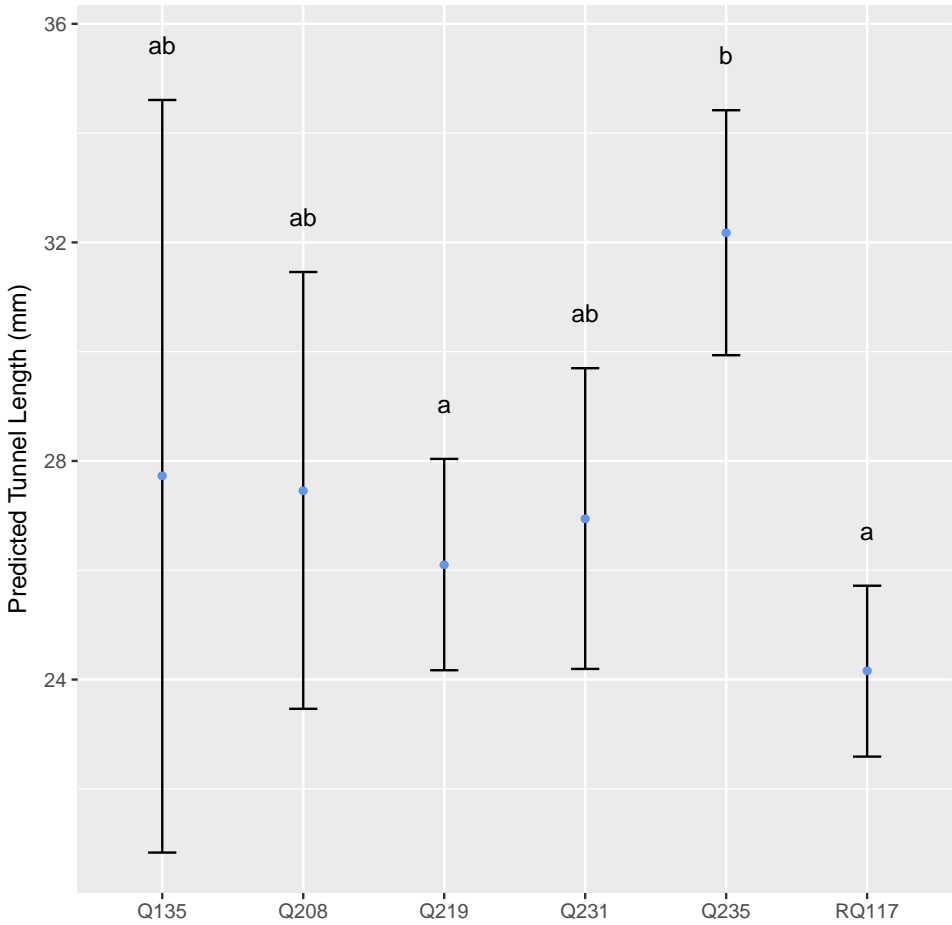
From the ANOVA table we see there is a significant effect due to Variety on Tunnel Length,  $p$ -value < 0.001.

The predicted values for Variety follow:

| Variety | Predicted Value | Standard Error | Ranking |
|---------|-----------------|----------------|---------|
| RQ117   | 24.15           | 0.798          | a       |
| Q219    | 26.10           | 0.987          | a       |
| Q231    | 26.95           | 1.404          | ab      |
| Q208    | 27.46           | 2.039          | ab      |
| Q135    | 27.72           | 3.513          | ab      |
| Q235    | 32.18           | 1.144          | b       |

These results indicate that Varieties RQ117 and Q219 have significantly shorter tunnel lengths compared to variety Q235 at the 5% family significance level.

A graph of the predicted means follows:



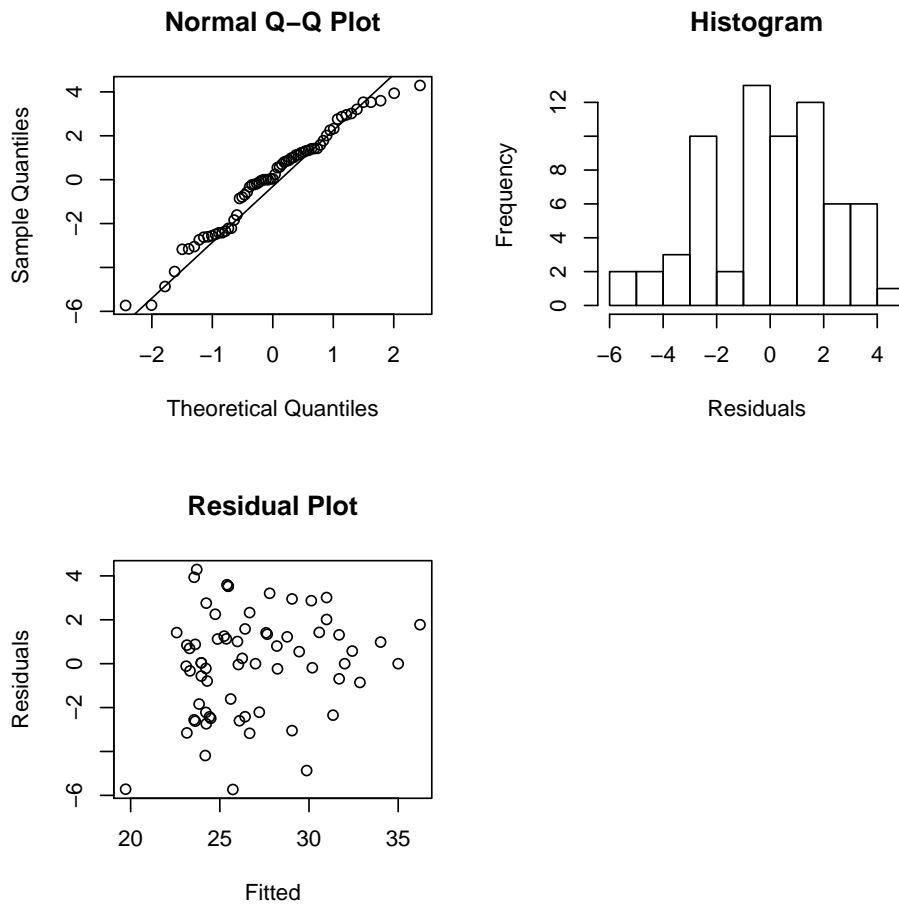
Note: The error bars represent the 95% confidence interval.

## 4.4 Tunnel length and Bored Internodes

### 4.4.1 Analysis

The model assumptions can be checked using the plots below:

Residual Plots



The Shapiro-Wilk test of normality is used to determine if the residuals are normally distributed. In this case we can conclude that the residuals are derived from a population that is normally distributed (as the Shapiro-Wilk  $p$ -value = 0.103). We can see in the residual plot that the residuals have a constant variance and we assume that the residuals are independent.

Note: The interaction of Variety and the number of Bored internodes was fitted and tested in the model and was not significant ( $p$ -value > 0.05).

The ANOVA table follows:

|                 | Df | denDF | F.inc | Pr    |
|-----------------|----|-------|-------|-------|
| Variety         | 5  | 18.1  | 6.00  | 0.002 |
| Boredinternodes | 11 | 41.9  | 1.03  | 0.434 |

From the ANOVA table we see there is not a significant effect due to Bored Internodes on Tunnel Length,  $p$ -value = 0.434. As shown previously varitey is also significant ( $p$ -value 0.002).

## References

- Butler, D. (2009). asreml: asreml() fits the linear mixed model [Computer software manual]. Retrieved from [www.vsni.co.uk](http://www.vsni.co.uk) (R package version 3.0)
- R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>