Random and Mixed Effects Models

Statistics 149

Spring 2006

Fixed Effects Versus Random Effects

Lets consider normal based one-away ANOVA model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

In most (all?) the cases we have dealt with before along this line, it has been assumed that the categorical factors used for prediction have been *fixed effects*. In these cases, the levels have been specifically chosen. For example, car type and vehicle age in the insurance example fit into this. In these cases, the parameters to be estimated are unknown constants.

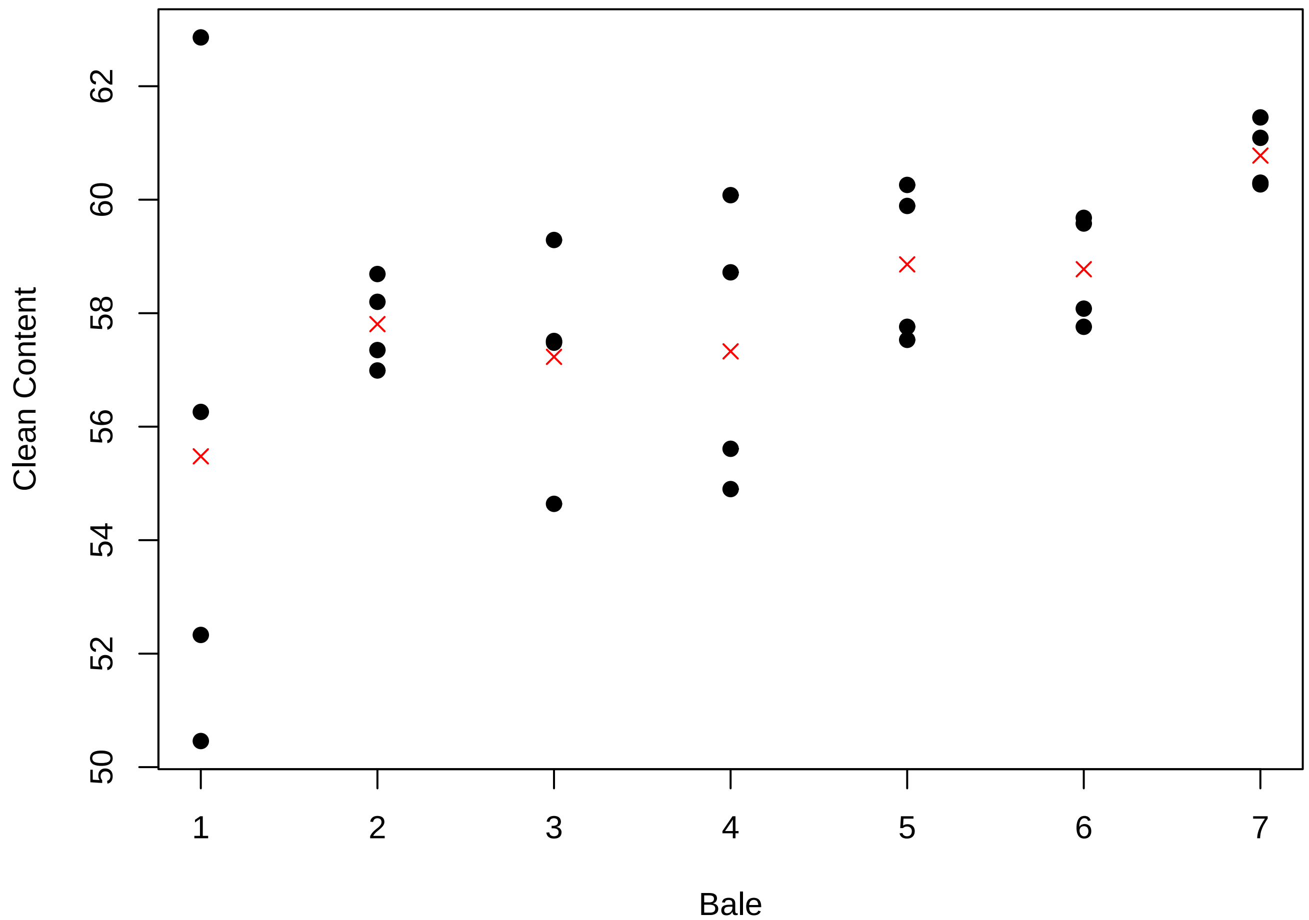
However, there are other cases where we are interested in a large population of possible levels of a treatment factor and the levels used in the study are a random sample from this population.

Example: Clean Wool Experiment (Dean & Vos, Example 17.2.2)

Raw wool contains varying amounts of grease, dirt, and foreign material which must be removed before manufacturing begins. The purchase price and customs levy of a shipment are based on the actual amount of wool present after cleaning (the clean content). The clean content (`clean`)is expressed as the percentage the weight of the clean wool is of the original weight of the raw wool.

The treatment factor was wool bale (`bale`) and its levels were the entire population of bales in a particular shipment. Seven bales were randomly sampled and 4 core samples from each bales had their clean content measures.

In this case bale is our treatment factor, though we aren't really interested in these particular 7 bales. The interest is in how much bales can differ.



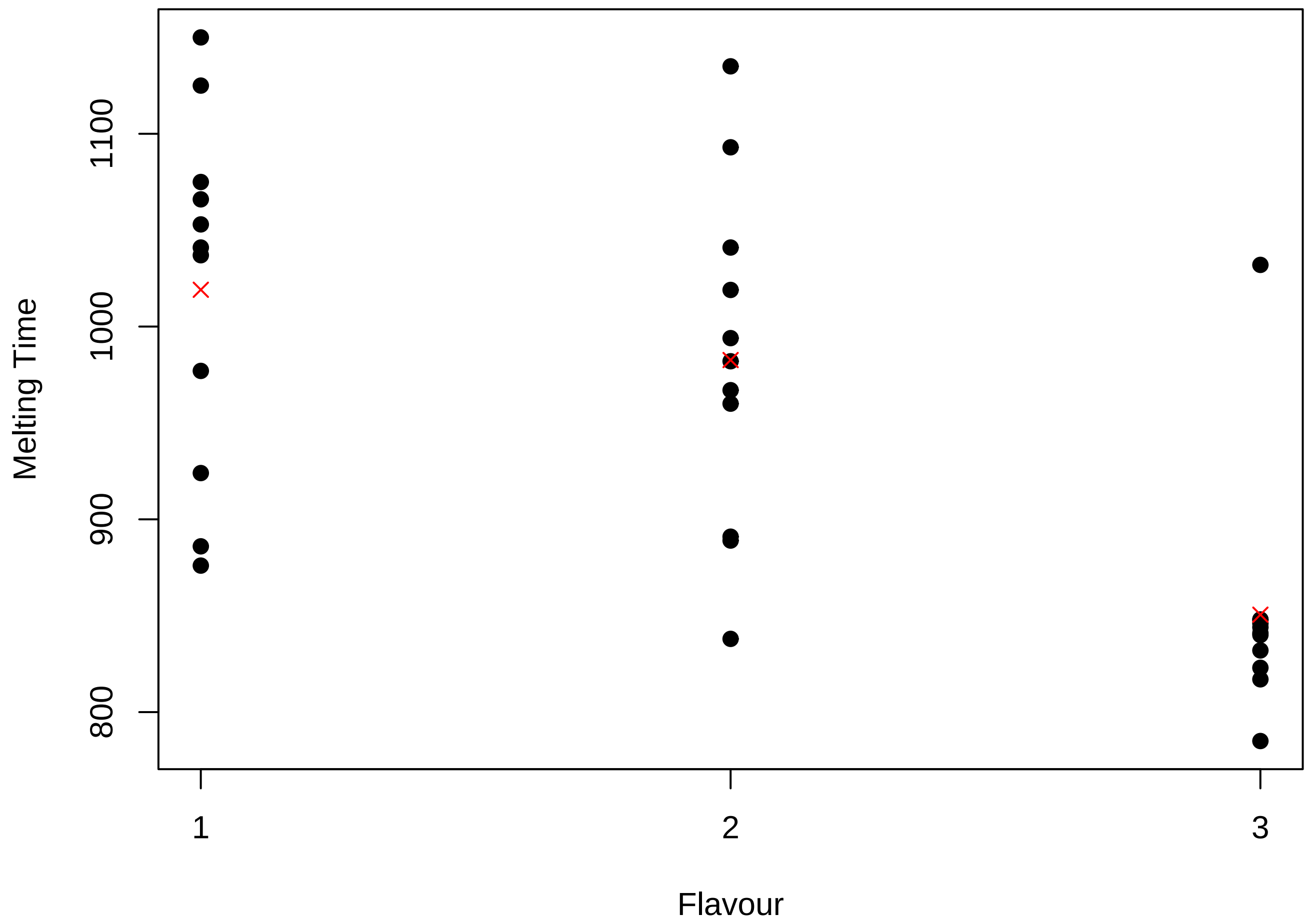
Clean Content
50
52
54
56
58
60
62
1
2
3
4
5
6
7
Bale

Example: Ice Cream Experiment (Dean & Vos, Example 17.3.1)

An experiment was run to examine whether or not different flavours of ice cream melt at different rates. A random sample of three flavours was selected from a large populations offered to the customer by a single manufacturer in May 1986. It is not obvious that the selected flavours are representative of all possible ice cream flavours, since some may include an ingredient that inhibits melting. The theoretical population is therefore the population of all flavours that could be made with ingredients similar to those available.

Three flavours of ice cream were stored in the same freezer in similar sized containers. For each observation, one teaspoonful of ice cream was taken from the freezer, transferred to a plate, and the melting time at room temperature was observed to the nearest second. Eleven observations were taken on each flavour and the order of observations was also recorded.

In this cases, we want to describe the variability in melting times between the different flavours.


Melting Time
1100
1000
900
800
1
2
3
Flavour

One-way Random Effects Model

We want to fit a model of the form

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where the α_is describe the variability between the different level of the factor of interest (e.g. bale or flavour) and the ϵ_{ij}s describe the variability of observations within a factor level (often measurement error).

As we often consider the factor levels a sample from a population, its reasonable to consider the α_is as draws from a population. The usual assumptions and model fit are

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$$\alpha_i \overset{iid}{\sim} N(0, \sigma^2_\alpha)$$

$$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

where the α_is and ϵ_{ij} are all mutually independent.

The assumption that $E[\alpha_i] = 0$ is needed for estimation and is similar to the constraints needed in the fixed effects case.

The distributional properties of the observations is a bit different than the fixed effects case. To start,

$$\begin{aligned} E[y_{ij}] &= E[\mu + \alpha_i + \epsilon_{ij}] \\ &= E[\mu] + E[\alpha_i] + E[\epsilon_{ij}] \\ &= \mu \end{aligned}$$

Next,

$$\begin{aligned} \mathrm{Var}(y_{ij}) &= \mathrm{Var}(\mu + \alpha_i + \epsilon_{ij}) \\ &= \mathrm{Var}(\alpha_i) + \mathrm{Var}(\epsilon_{ij}) \qquad \text{since } \alpha_i \text{ and } \epsilon_{ij} \text{ are independent} \\ &= \sigma_\alpha^2 + \sigma^2 \end{aligned}$$

Also, for two observations taken under the same factor level i,

$$\begin{aligned}\mathrm{Cov}(y_{ij}, y_{ik}) &= \mathrm{Cov}(\mu + \alpha_i + \epsilon_{ij}, \mu + \alpha_i + \epsilon_{ik}) \\ &= \mathrm{Cov}(\alpha_i, \alpha_i) + \mathrm{Cov}(\alpha_i, \epsilon_{ij}) + \mathrm{Cov}(\alpha_i, \epsilon_{ik}) + \mathrm{Cov}(\epsilon_{ij}, \epsilon_{ik}) \\ &= \mathrm{Var}(\alpha_i) = \sigma_\alpha^2\end{aligned}$$

So y_{ij} and y_{ik} are correlated with

$$\mathrm{Corr}(y_{ij}, y_{ik}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$$

In some settings, this is known as the interclass correlation.

If two observations come from different treatment factor levels

$$\mathrm{Cov}(y_{ij}, y_{lk}) = 0$$

So

$$y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2)$$

The quantities σ_α^2 and σ^2 are often referred to as variance components.

Note that one-way random effects model is a special case of the model

$$y_{ij} \sim N(\mu, \tau^2)$$

$$\text{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \rho\tau^2 & \text{if } i = i' \text{ and } j \neq j' \\ 0 & \text{if } i \neq i' \end{cases}$$

where $\rho \in (\rho_{\min}, 1)$ where $\rho_{\min} < 0$ and depends on the number of observation within each factor level.

The one-way random effects model has to have a non-negative correlation between observations within the same factor level.

Estimation in the One-way Random Effects Model

Assume that there are ν factor levels observed in the data and that for factor level i, there are m_i observations.

In this model, there are three parameters to estimate: μ, σ_α^2 and σ^2

- μ: Since

  $$E[\bar{y}_{++}] = \mu$$

  the usual estimate of μ is $\bar{y}_{++}$, the average of all the observations.

- σ^2: Let

$$SSE = \sum_{i=1}^{\nu}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{i+})^2$$

$$= \sum_{i=1}^{\nu}\sum_{j=1}^{m_i} y_{ij}^2 + \sum_{i=1}^{\nu} m_i \bar{y}_{i+}^2$$

be the usual SSE in a one-way fixed effects ANOVA.

It can be shown that

$$E[y_{ij}^2] = \mathrm{Var}(y_{ij}) + (E[y_ij])^2 = \sigma_\alpha^2 + \sigma^2 + \mu^2$$

and

$$E[\bar{y}_{i+}^2] = \mathrm{Var}(\bar{y}_{i+}) + (E[\bar{y}_{i+}])^2 = \sigma_\alpha^2 + \frac{\sigma^2}{m_i} + \mu^2$$

Then is can be shown that

$$\begin{aligned} E[SSE] &= \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} (\sigma_\alpha^2 + \sigma^2 + \mu^2) + \sum_{i=1}^{\nu} m_i \left(\sigma_\alpha^2 + \frac{\sigma^2}{m_i} + \mu^2 \right) \\ &= n\sigma^2 - \nu\sigma^2 \qquad \text{where } n = \sum m_i \\ &= (n - \nu)\sigma^2 \end{aligned}$$

So the MSE

$$MSE = \frac{SSE}{n - \nu}$$

is an unbiased estimator of σ^2

- σ_α^2: Again mimicing the analysis from the one-way fixed effects ANOVA, let

$$\begin{aligned} SST &= \sum_{i=1}^{\nu} m_i(\bar{y}_{i+} - \bar{y}_{++})^2 \\ &= \sum_{i=1}^{\nu} m_i \bar{y}_{i+}^2 + n\bar{y}_{++}^2 \end{aligned}$$

be the usual treatment sums of squares in a one-way fixed effects ANOVA.

It can be shown that

$$E[\bar{y}_{++}^2] = \mathrm{Var}(\bar{y}_{++}) + (E[\bar{y}_{++}])^2 = \frac{\sum m_i^2}{n^2}\sigma_\alpha^2 + \frac{\sigma^2}{n} + \mu^2$$

Then is can be shown that

$$E[SST] = \sum_{i=1}^{\nu} m_i \left(\sigma_\alpha^2 + \frac{\sigma^2}{m_i} + \mu^2\right) - n\left(\frac{\sum m_i^2}{n^2}\sigma_\alpha^2 + \frac{\sigma^2}{n} + \mu^2\right)$$
$$= \left(n - \frac{\sum m_i^2}{n}\right)\sigma_\alpha^2 - (\nu - 1)\sigma^2$$

Since $MST = \frac{SST}{\nu-1}$,

$$E[MST] = c\sigma_\alpha^2 + \sigma^2$$

where

$$c = \frac{n^2 - \sum m_i^2}{n(\nu - 1)}$$

If all the m_i are equal to m, then $n = \nu m$ and $c = m$

Thus an unbiased estimate of σ_α^2 is

$$\frac{MST - MSE}{c}$$

It is possible for this estimator to give a negative estimate even though σ_α^2 cannot be.

This is something that could happen when σ_α^2 is close to 0. If MSE is much smaller than MST, you probably want to question the adequacy of the model.

So if desired, we can get most of what we want from the standard one-way ANOVA analysis. Though you have to do some additional work to get c if the number of observations on each factor level varies.

Another approach in **R** is with the `lme4` package. Its a general package for fitted mixed models, which include random effects models. In fact it will handle generalized linear mixed models, so the normal assumptions can be relaxed

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice
```

```
> wool.re <- lmer(clean ~ 1 + (1 | bale) , data=wool)

> wool.re
Linear mixed-effects model fit by REML
Formula: clean ~ 1 + (1 | bale)
          Data: wool
      AIC     BIC   logLik MLdeviance REMLdeviance
 136.8586 139.523 -66.4293   133.7433     132.8586
Random effects:
 Groups   Name        Variance Std.Dev.
 bale     (Intercept) 1.1833   1.0878
 Residual             6.2606   2.5021
number of obs: 28, groups: bale, 7

Fixed effects:
            Estimate Std. Error t value
(Intercept) 58.03643    0.62661   92.62
```

```
> icecream.re <- lmer(time ~ 1 + (1 | flavour) , data=icecream)

> icecream.re
Linear mixed-effects model fit by REML
Formula: time ~ 1 + (1 | flavour)
         Data: icecream
      AIC      BIC    logLik MLdeviance REMLdeviance
 385.7047 388.6978 -190.8524   391.3986     381.7047
Random effects:
 Groups   Name        Variance Std.Dev.
 flavour  (Intercept) 7247.6   85.133
 Residual             6781.9   82.352
number of obs: 33, groups: flavour, 3

Fixed effects:
            Estimate Std. Error t value
(Intercept)  950.758     51.199  18.570
```

The general structure of the command is to described the fixed effects in the model. In these examples, there is only the intercept, indicated by `1`. Then the terms involving the random effects come. In this case we are looking at describing deviations around the intercept, which are described by `(1 | bale)` and `(1 | flavour)`.

As part of the output, it gives estimates and standard errors for the fixed effects. Note that these depend on both variance components, not just σ^2 as in the fixed effects case.

For example,

$$\begin{aligned}\mathrm{Var}(y_{++}) &= \mathrm{Var}\left(\sum m_i \alpha_i + \epsilon_{++}\right) \\ &= \sum m_i^2 \sigma_\alpha^2 + n\sigma^2\end{aligned}$$

If all the $m_i = m$ (as in the two examples)

$$\mathrm{Var}(\bar{y}_{++}) = \frac{m\sigma_\alpha^2 + \sigma^2}{n}$$

Testing in the One-way Random Effects Model

In the this model, to examine whether there is a treatment effect, we need to examine σ_α^2. In the testing framework, we need to examine the hypotheses

$$H_0 : \sigma_\alpha^2 = 0 \qquad \text{vs} \qquad H_A : \sigma_\alpha^2 > 0$$

It ends up that it's possible to show that

$$\frac{SST}{c\sigma_\alpha^2 + \sigma^2} \sim \chi^2_{\nu-1}$$

and

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-\nu}$$

and that SST and SSE are independent, so under the null hypothesis

$$\frac{MST}{MSE} \sim F_{\nu-1,n-\nu}$$

So the standard F test from the one-way ANOVA gives us the answer we want.

For the two examples

```
> anova(wool.fe, test="F")
Analysis of Variance Table

Response: clean
          Df  Sum Sq Mean Sq F value Pr(>F)
bale       6  65.963  10.994   1.756 0.1573
Residuals 21 131.472   6.261

> anova(icecream.fe, test="F")
Analysis of Variance Table

Response: time
          Df Sum Sq Mean Sq F value    Pr(>F)
flavour    2 173010   86505  12.755 9.799e-05 ***
Residuals 30 203456    6782
```

Confidence Intervals on the Variance Components

Often is useful to get confidence interval on various combinations of the variance components

- σ^2: This is the easiest situation, as the problem reduces to the fixed effects case. The interval is based on the pivotal quantity

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-\nu}$$

A one sided upper confidence bound is

$$\sigma^2 \leq \frac{SSE}{\chi^{2*}_{1-\alpha}}$$

where

$$P[\chi^2_{n-\nu} \geq \chi^{2*}_{1-\alpha}] = 1 - \alpha$$

Similarly a two-sided confidence interval is given by

$$\frac{SSE}{\chi^{2*}_{\alpha/2}} \leq \sigma^2 \leq \frac{SSE}{\chi^{2*}_{1-\alpha/2}}$$

- σ^2_α/σ^2: We can get a handle on this based on the result mentioned in testing

$$\frac{MST}{MSE(c\sigma^2_\alpha/\sigma^2+1)} \sim F_{\nu-1,n-\nu}$$

Thus

$$P\left[F^*_{1-\alpha/2} \leq \frac{MST}{MSE(c\sigma^2_\alpha/\sigma^2+1)} \leq F^*_{\alpha/2}\right] = 1-\alpha$$

Rearranging the left side gives

$$\frac{c\sigma^2_\alpha}{\sigma^2} \leq \frac{MST}{MSE\ F^*_{1-\alpha/2}} - 1$$

and similarly for the right side

$$\frac{c\sigma_\alpha^2}{\sigma^2} \geq \frac{MST}{MSE\ F^*_{\alpha/2}} - 1$$

Combining these give the interval

$$\frac{1}{c}\left[\frac{MST}{MSE\ F^*_{\alpha/2}} - 1\right] \leq \frac{\sigma_\alpha^2}{\sigma^2} \leq \frac{1}{c}\left[\frac{MST}{MSE\ F^*_{1-\alpha/2}} - 1\right]$$

Note that if MST isn't much larger than MSE, the left endpoint could be less than 0.

For the wool example, a 90% interval is

$$\left(\frac{1}{4}\left[\frac{1.756}{2.572} - 1\right], \frac{1}{4}\left[\frac{1.756}{0.259} - 1\right]\right) = (-0.079, 1.447)$$

So here is a case with a negative left endpoint. This shouldn't be two surprising, since we couldn't reject $\sigma_\alpha^2 = 0$ earlier.

For the ice cream example, a 90% interval is

$$\left(\frac{1}{4}\left[\frac{12.775}{3.32} - 1\right], \frac{1}{4}\left[\frac{12.775}{0.0513} - 1\right]\right) = (0.258, 22.513)$$

This interval is very wide, suggesting that $\sigma_a lpha^2$ could only be a quarter of σ^2 or it could be 22 times larger. This shouldn't be too surprising, as we don't have much information to estimate $\sigma_a lpha^2$ since only three flavours were chosen.

In trying to estimate a quantity like this, you need to find a balance between the number of levels examined and observations per level.

- σ^2_α: There are a number of procedures available for obtaining approximate confidence intervals for this variance. Unlike the other two situations, there are not nice exact distributional results to base a confidence interval on.

  One popular approach is the following. Let $\hat{\sigma}^2_\alpha$ be the estimate of σ^2_α discussed earlier

$$\hat{\sigma}^2_\alpha = \frac{MST - MSE}{c}$$

  The exact distribution of $\hat{\sigma}^2_\alpha$ is that of a linear combination of independent χ^2s. While this distribution is not standard, it can be shown that $\hat{\sigma}^2_\alpha / \sigma^2_\alpha$ can be well approximated by a χ^2_{df}/df where

$$df = \frac{(MST - MSE)^2}{\frac{MST^2}{\nu - 1} + \frac{MSE^2}{n - \nu}}$$

Another way of thinking of this, is that

$$\frac{df\,\hat{\sigma}^2_\alpha}{E[\hat{\sigma}^2_\alpha]} \overset{approx.}{\sim} \chi^2_{df}$$

Unraveling this gives an approximately confidence interval for σ^2_α of

$$\left(\frac{df\,\hat{\sigma}^2_\alpha}{\chi^{2*}_{\alpha/2}}, \frac{df\,\hat{\sigma}^2_\alpha}{\chi^{2*}_{1-\alpha/2}} \right)$$

So for the ice cream example

$$df = \frac{(86504.9 - 6781.9)^2}{\frac{86504.9^2}{2} + \frac{6781.9}{30}} = 1.7$$

This gives a 90% interval for σ_α^2 of

$$\left(\frac{1.7 \times 7247.5}{5.3}, \frac{1.7 \times 7247.5}{0.07}\right) = (2324.7, 176012.0)$$

If we take square roots and divide by 60 to convert to minutes, a 90% confidence interval for the standard deviation of melting times is approximately (0.8, 7) minutes.

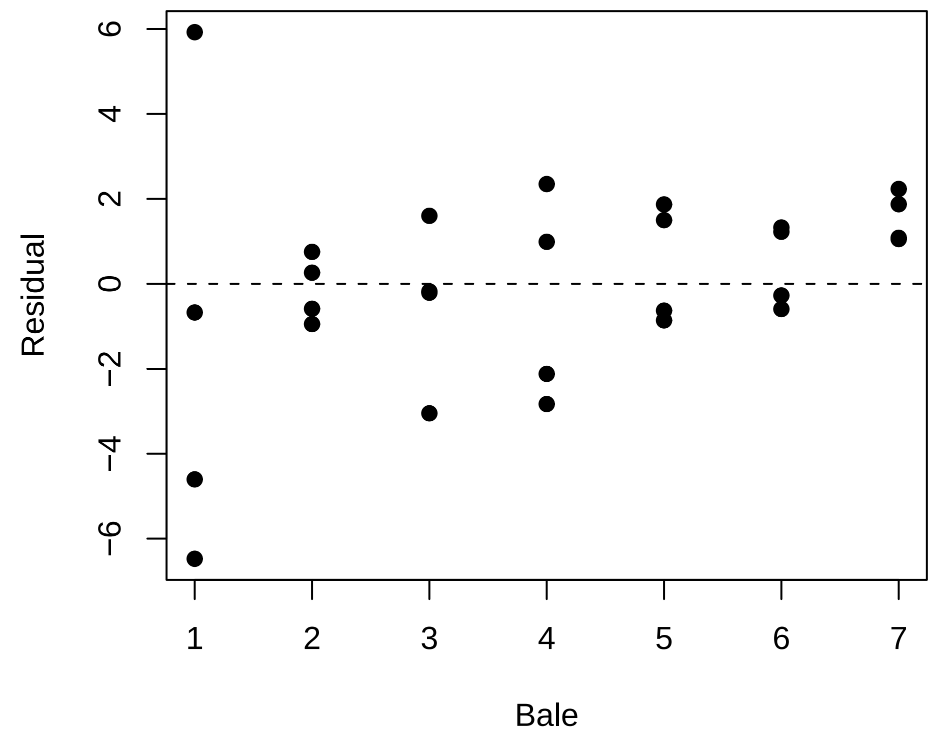
Checking Assumptions

As in the fixed effects case, we should check our modeling assumptions.

To check assumptions on the ϵ_{ij}s is easy. We can define residuals by

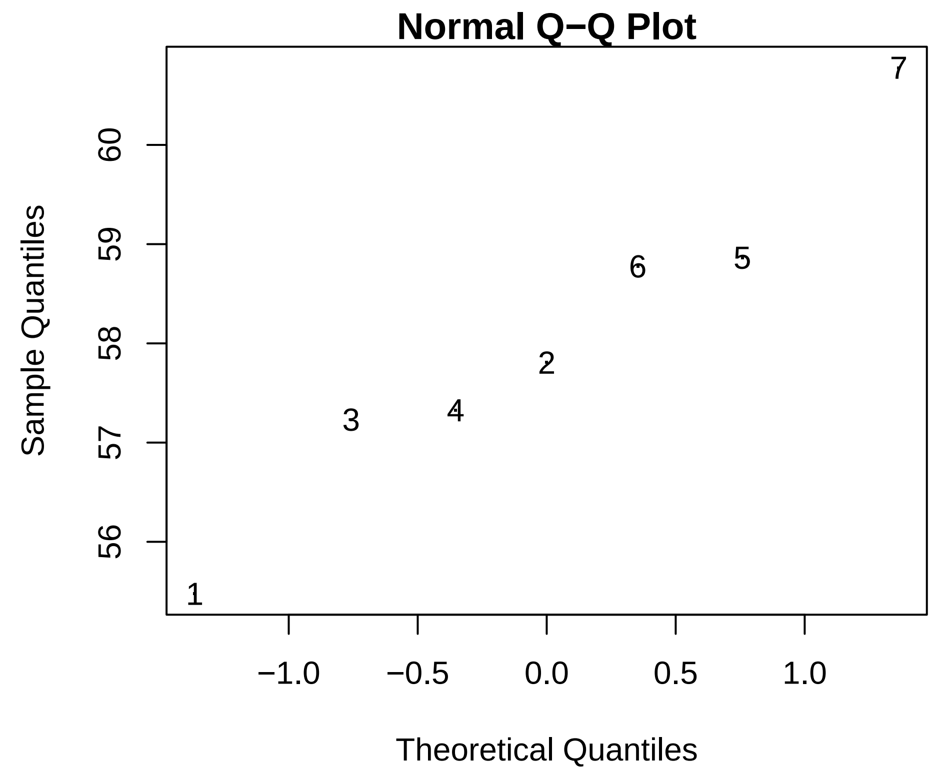
$$e_{ij} = y_{ij} - \bar{y}_{i+}$$

and check for outliers, constant variance, independence, and normality by standard techniques. (Note that this isn't what you get with `resid(lmer.object)`. I believe these are based on BLUEs of $\mu + \alpha_i$.



To check assumptions on the α_is, we can base this on

$$\bar{y}_{i+} \sim N(\mu, \sigma_\alpha^2 + \sigma^2/m_i)$$



So if all $m_i = m$, we can do a normal scores plot of the y_{i+}s. If the normality assumption is reasonable, these means should lie approximately on a straight line with x-intercept at about μ and slope about $\sqrt{\sigma_\alpha^2 + \sigma^2/m}$. The normality assumption is important as the procedures described are not robust to non-normality of the random effects. Unfortunately, this is often difficult to do as there often will not be many levels.

The $\bar{y}_{i+}$ can also be used to look for outliers as they can be easily standardized.

So in this example, there appears to be a variance problem in bale 1 and possibly extreme means for bales 1 and 7. Possibly there are multiple subpopulations here, which would be one possible explanation for the plots.