# P&B MM Rail p4

*W Greg Alvord, & Nick Carchedi [Douglas M. Bates & Jose Pinheiro]*

*November 17, 2014*

## Opening Observations

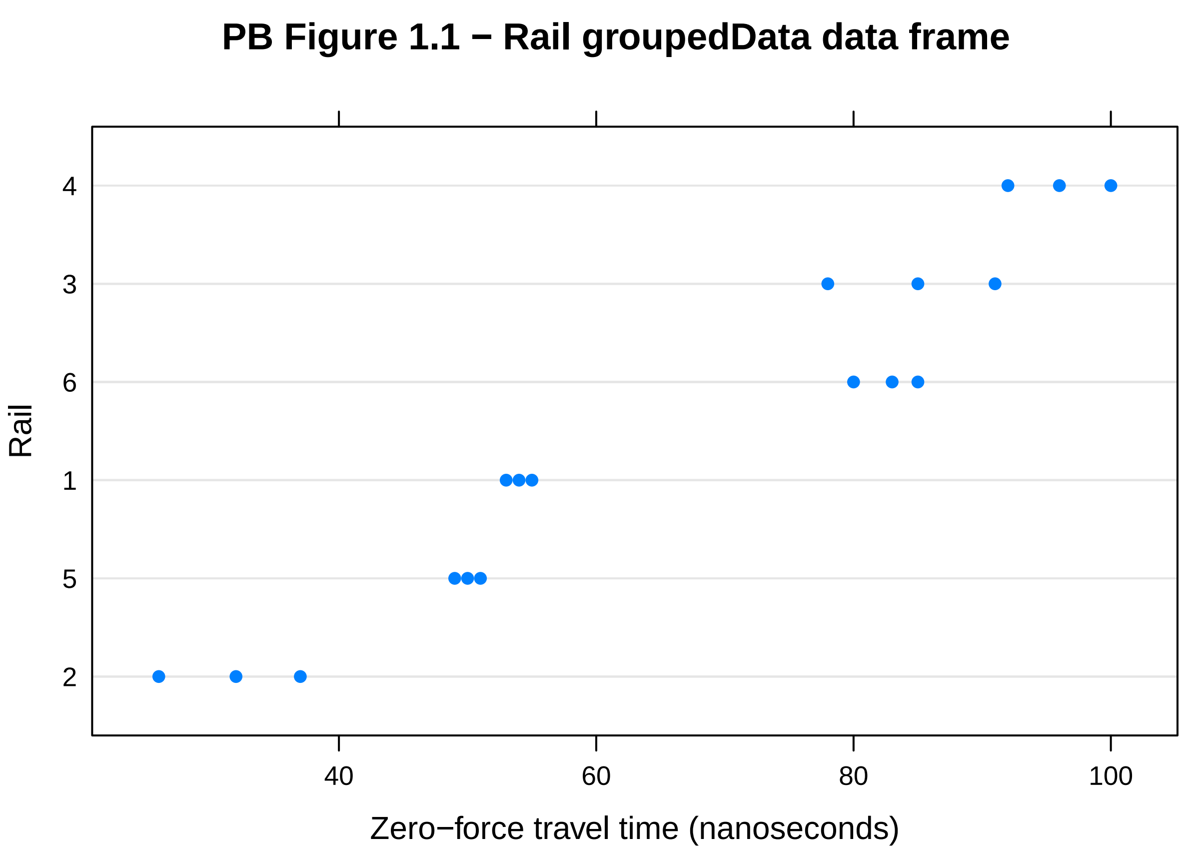
[Note: This document reflects initial attempts using RMarkdown for Pinheiro and Bates' Rail example. Much of the text is directly copied from Pinheiro and Bates (2000), designated as PB or P&B throughout. Commentary is provided in connection with certain RMarkdown commands, etc.]

## Mixed Models Rail Example, Pinheiro and Bates, page 4

### Introduction to the Rail Problem

We begin with the first example discussed by Pinheiro and Bates (2000). Six rails were chosen at random; each rail was tested three times by measuring the time it took for a certain type of ultrasonic wave to travel the length of the rail. The engineers were interested in: (1) the average travel time for a 'typical' rail (the *expected travel time*), (2) the *variation* in average travel time among the rails (*between-rail variability*), and (3) the *variation* in the observed travel times for a single rail (the *within-rail variability*). We see from Figure 1.1 that there is considerable variability in the mean travel times among the different rails.

```
plot(Rail, main = 'PB Figure 1.1 - Rail groupedData data frame')
```



Here is the `groupedData` data frame. . .

```
Rail
```

```
## Grouped Data: travel ~ 1 | Rail
##    Rail travel
## 1     1     55
## 2     1     53
## 3     1     54
## 4     2     26
## 5     2     37
## 6     2     32
## 7     3     78
## 8     3     91
## 9     3     85
## 10    4     92
## 11    4    100
## 12    4     96
## 13    5     49
## 14    5     51
## 15    5     50
## 16    6     80
## 17    6     85
## 18    6     83
```

The 'rail effects' indicated in Figure 1.2 may be incorporated into the model for the travel times by allowing the mean of each rail to be represented by a separate parameter. This *fixed-effects* model for the one-way classification is written

$$y_{ij} = \beta_i + \epsilon_{ij}, \quad i = 1, ..., M, \quad j = 1, ..., n_i \qquad \texttt{PB eq(1.2)}$$

where the $\beta_i$ represents the mean travel time of rail $i$ and, as in (PB eqn 1.1), the errors $\epsilon_{ij}$ are assumed to be independently distributed as $N(0, \sigma^2)$. [Note: I have used hspace{40 mm} in the front of this equation. I have also used texttt{} for the equation number at this early stage.]

Here is the model for P&B equation (1.2). [In an experiment with RMarkdown, only the coefficients have been printed with this output. The argument `output.lines = 9:16` was added in the RMarkdown code.]

```
fm2Rail.lm <- lm(travel ~ as.factor(Rail) - 1, data = Rail)
summary(fm2Rail.lm)
```

```
...
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## as.factor(Rail)2   31.667      2.321   13.64 1.15e-08
## as.factor(Rail)5   50.000      2.321   21.54 5.86e-11
## as.factor(Rail)1   54.000      2.321   23.26 2.37e-11
## as.factor(Rail)6   82.667      2.321   35.61 1.54e-13
## as.factor(Rail)3   84.667      2.321   36.47 1.16e-13
## as.factor(Rail)4   96.000      2.321   41.35 2.59e-14
...
```

The estimates here show the mean `travel` values for each of the rails. The means could also be obtained with command

```
with(Rail, tapply(travel, Rail, mean))
```

```
##        2        5        1        6        3        4
## 31.66667 50.00000 54.00000 82.66667 84.66667 96.00000
```

Even though the fixed-effects model (1.2) accounts for the rail effects, it does not provide a useful representation of the rails data. Its only models the *specific sample of rails used in this particular experiment*, while the main interest is in the *population* of rails from which the sample was drawn. In particular, `fm2Rail.lm` does *not* provide an estimate of the *between-rail* variability, which is one of the central quantities of interest in the rails experiment. Another drawback of this fixed-effects model is that the number of parameters in the model increases linearly with the number of rails.

A *random-effects* model circumvents these problems by treating the rail effects as random variations around a population mean. The following reparameterization of model `(1.2)` helps motivate the random-effects model for the rails data. We write

$$y_{ij} = \bar{\beta} + (\beta_i - \bar{\beta}) + \epsilon_{ij} \qquad \texttt{PB eq(1.3)} \tag{1}$$

Repeating this (i.e., experimenting with RMarkdown). . .

$$y_{ij} = \bar{\beta} + (\beta_i - \bar{\beta}) + \epsilon_{ij} \tag{2}$$

where $\bar{\beta} = \sum_{i=1}^{6} \beta i/6$ represents the average travel time for the rails in the experiment. The random-effects model replaces $\bar{\beta}$ with the mean travel time over the *population of rails* and replaces the deviations $\beta_i - \bar{\beta}$ by random variables whose distribution is to be estimated.

A random-effects model for the one-way classification used in the rails experiment is written. . . [Note also that when we used begin{equation} and end{equation}, commands from LaTeX, we got the equation number in parentheses (1).]

$$y_{ij} = \beta + b_i + \epsilon_{ij}, \qquad \mbox{PB (1.4)}$$

where $\beta$ is the mean travel time across the population of rails being sampled, $b_i$ is a random variable representing the deviation from the population mean of the mean travel time for the $i$th rail, and $\epsilon_{ij}$ is a random variable representing the deviation in travel time for observation $j$ on rail $i$ from the mean travel time for all rail $i$. [Note: In equation (1.4) I used mbox{}.]

To complete the statistical model, we must specify the distribution of the random variables $b_i, i = 1, ..., M$ and $\epsilon_{ij}, i = 1, ..., M; j = 1, ..., n_i$. We begin by modeling both of these as independent, constant variance, denoted by $\sigma_b^2$ for the $b_i$ or *between-rail* variability, and $\sigma^2$ for the $\epsilon_{ij}$ or *within-rail* variability. That is

$$b_i \sim N(0, \sigma_b^2), \quad \epsilon_{ij} \sim N(0, \sigma^2) \tag{3}$$

This model may be modified if it does not seem appropriate. As described in Chapter 4, PB encourage using graphical and numberical diagnostic tools to assess the validity of the model and to suggest ways in which it could be modified. To start, however, we use the simple model.

First, we fit the `lme` model for the Rail data.

```r
fm1Rail.lme <- lme(travel ~ 1, data = Rail, random = ~ 1 | Rail)
```

Here, we show the `print()` for the object `fm1Rail.lme`.

```r
fm1Rail.lme
```

```
## Linear mixed-effects model fit by REML
##   Data: Rail
##   Log-restricted-likelihood: -61.0885
##   Fixed: travel ~ 1
## (Intercept)
##        66.5
##
## Random effects:
##  Formula: ~1 | Rail
##         (Intercept) Residual
## StdDev:    24.80547 4.020779
##
## Number of Observations: 18
## Number of Groups: 6
```

Now,we show the `summary()` function for the same item. Here it is.

```r
summary(fm1Rail.lme)
```

```
## Linear mixed-effects model fit by REML
##  Data: Rail
##       AIC      BIC   logLik
##   128.177 130.6766 -61.0885
##
## Random effects:
##  Formula: ~1 | Rail
##         (Intercept) Residual
## StdDev:    24.80547 4.020779
##
## Fixed effects: travel ~ 1
##             Value Std.Error DF  t-value p-value
## (Intercept)  66.5  10.17104 12 6.538173       0
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.61882658 -0.28217671  0.03569328  0.21955784  1.61437744
##
## Number of Observations: 18
## Number of Groups: 6
```

This model, which has two sources of random variation, $b_i$ and $\epsilon_{ij}$, is sometimes called a *hierarchical* model (references, etc.) or a *multilevel* model. The $b_i$ are called *random* effects because they are associated with the particular experimental units - rails in this case - that are selected at random from the population of interest. They are *effects* because they represent a deviation from an overall mean. That is, the “effect” of choosing rail $i$ is to shift the mean travel time from $\beta$ to $\beta + b_i$. Because observations made on the same rail share the same random effect $b_i$, they are correlated. The covariance between observations on the same rail is $\sigma_b^2$ corresponding to a correlation of $\sigma_b^2/(\sigma_b^2 + \sigma^2)$.

The parameters of the statistical model created by combining (P&B 1.4) and (P&B 1.5) are $\beta$, $\sigma_b^2$, and $\sigma^2$. Note that the number of parameters for this particular problem will always be three (i.e., 3), irrespective of the number of rails in the experiment. Although the random effects, $b_i, i = 1, ..., M$ may behave like parameters, formally, they just represent another level of random variation in the model - so we do not 'estimate' them as such. We will, however, form predictions $\hat{\beta}$ of thge values of these random variables, given the data we observed.

The REML estimates for the parameters are calculated as

$$\hat{\beta} = 66.5, \quad \hat{\sigma_b} = 24.805, \quad \hat{\sigma} = 4.0208$$

Here is an equation presented, in matrix form, on page 14 of P&B. [We jump ahead just a little bit.]

$$y_i = X_i\beta + Z_i b_i + \epsilon_i \tag{4}$$

Now, we construct these quantities in matrix form and apply them to the data for the first rail.

```
X1 <- matrix(1, nrow = 3, ncol = 1)
X1
```

```
##      [,1]
## [1,]    1
## [2,]    1
## [3,]    1
```

```
beta.hat <- fixef(fm1Rail.lme)
beta.hat
```

```
## (Intercept)
##        66.5
```

```
b1 <- ranef(fm1Rail.lme)[3,1]
b1
```

```
## [1] -12.39148
```

```
Z1 <- X1
epsilon1 <- resid(fm1Rail.lme)[1:3]
epsilon1
```

```
##          1          1          1
##  0.8914756 -1.1085244 -0.1085244
```

Now we will display $y_1$ and $y_1$ as computed from the matrix equation.

```
## display y1
y1 <- Rail$travel[1:3]
y1
```

```
## [1] 55 53 54
```

```
##
## Compute y1 as predicted from lme model
y1.computed <- X1 %*% beta.hat + Z1 %*% b1 + epsilon1
y1.computed
```

```
##      [,1]
## [1,]   55
## [2,]   53
## [3,]   54
```

```
##
round(y1, 8) == round(y1.computed, 8)
```

```
##      [,1]
## [1,] TRUE
## [2,] TRUE
## [3,] TRUE
```

### Assessing the Fitted Model

This next plot reproduces Figure 1.4 in PB, (page 11).

## References

Pinheiro JC & Bates DM (2000) *Mixed Effects Models in S and S-Plus* Springer

## Redundant comments

This is equation 1.3, p. 7, in P&B.

$$y_{ij} = \bar{\beta} + \beta_i - \bar{\beta} + \epsilon_{ij}$$

This is equation 1.3, p. 7, in P&B. [Note: This equation was obtained with the LaTeX/RMarkdown command `boldsymbol`]

$$\boldsymbol{y_{ij} = \bar{\beta} + \beta_i - \bar{\beta} + \epsilon_{ij}}$$

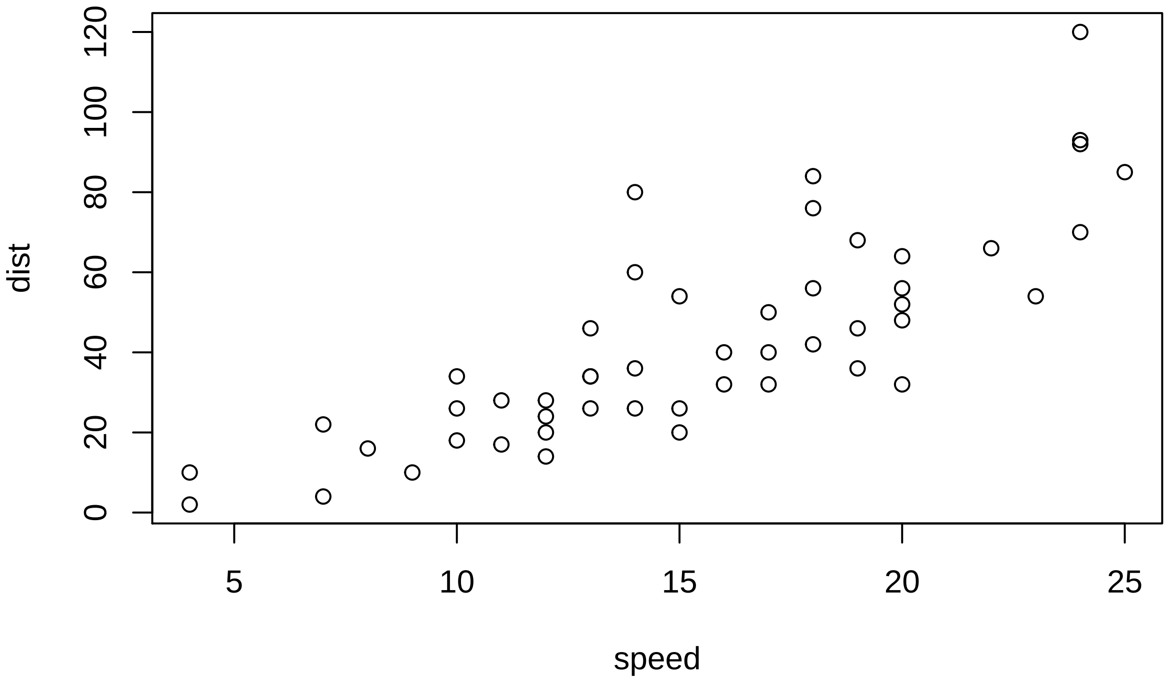
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

You can also embed plots, for example:




Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.