Report (for colleagues)

David Meyer

27. Mai 2016

Starting point

In a software development company, three variables were collected for 21 bugs (no missing values):

- time needed to fix the bug (*duration*, metric)
- programmer (*programmer*, categorical)
- Bug type (*bugtype*)

Is there an influence of the respective factors given? Is there an interaction present?

Data management

The data were read using `read.table()`:

```
bugfixes2 = read.table("Y:/SS 2016/FH Technikum/BWI-2 DL DAS/Module_Books/Data/bugfixes2.csv",
                       header = TRUE)
head(bugfixes2)
```

```
##   duration programmer   bugtype
## 1      120 Eckkrammer       GUI
## 2      174      Meyer        DB
## 3      188      Meyer       GUI
## 4      161      Mandl       GUI
## 5      157 Eckkrammer Reporting
## 6      178      Meyer Reporting
```

Visualization

The distributions of the times needed to fix a bug, grouped by programmers and bug types, are visualized using parallel box plots (see figure 1). The five-number summaries for programmers and bug types are given by:

```
aggregate(duration ~ programmer + bugtype, data = bugfixes2, summary)
```

```
##   programmer   bugtype duration.Min. duration.1st Qu. duration.Median
## 1 Eckkrammer        DB         157.0            157.5           158.0
## 2      Mandl        DB         177.0            177.5           178.0
## 3      Meyer        DB         174.0            174.5           175.0
## 4 Eckkrammer       GUI         120.0            125.8           131.5
## 5      Mandl       GUI         161.0            164.2           167.5
## 6      Meyer       GUI         183.0            185.5           188.0
```

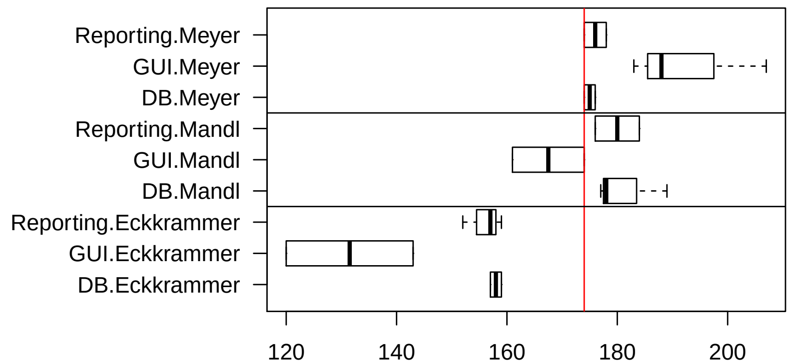


Figure 1: Boxplots of *duration*, given *programmer* and *bugtype*

```
## 7 Eckkrammer Reporting         152.0            154.5            157.0
## 8      Mandl Reporting         176.0            178.0            180.0
## 9      Meyer Reporting         174.0            175.0            176.0
##   duration.Mean duration.3rd Qu. duration.Max.
## 1         158.0            158.5         159.0
## 2         181.3            183.5         189.0
## 3         175.0            175.5         176.0
## 4         131.5            137.2         143.0
## 5         167.5            170.8         174.0
## 6         192.7            197.5         207.0
## 7         156.0            158.0         159.0
## 8         180.0            182.0         184.0
## 9         176.0            177.0         178.0
```

Programmer Eckkrammer seems to be considerably faster than his two colleagues, but bug type alone doesn't seem to have an influence. But we notice that Eckkrammer fixes GUI bugs particularly fast, whereas Meyer fixes these considerably slower. This could indicate the presence of an interaction.

Analysis of variance

To test the hypotheses "All programmers are equally fast" and "Bugs of different types are fixed equally fast" for the population, we use two-way analysis of variance. To begin with, we fit a regression model (with *duration* as dependent and *programmer* and *bugtype* as independent variables):

```
model = lm(duration ~ programmer + bugtype, data = bugfixes2)
```

The ANOVA table is given by:

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: duration
##            Df Sum Sq Mean Sq F value   Pr(>F)
## programmer  2 4420.7 2210.33 14.9635 0.000217 ***
## bugtype     2  145.1   72.56  0.4912 0.620804
## Residuals  16 2363.4  147.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Factor *programmer* is significant, but *bugtype* is not.

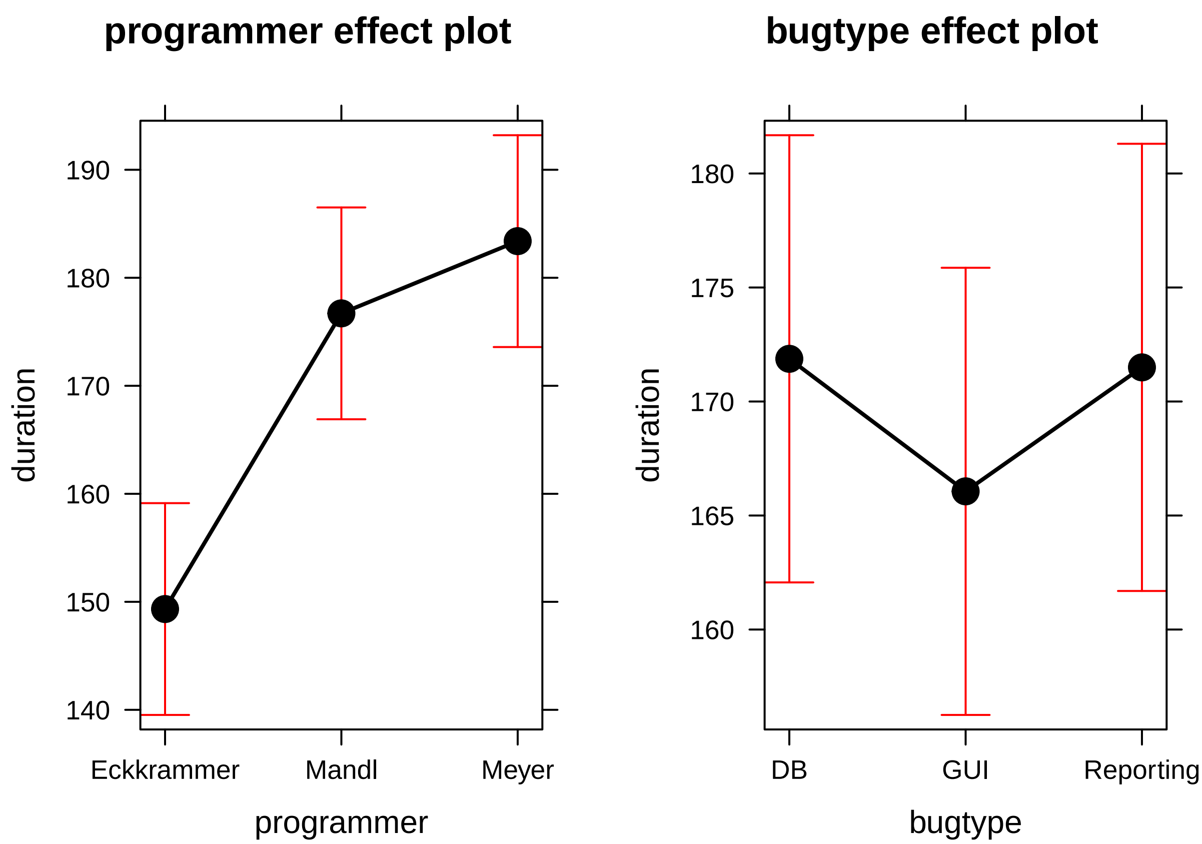
The summary of the regression model is given by:

```r
summary(model)
```

```
##
## Call:
## lm(formula = duration ~ programmer + bugtype, data = bugfixes2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.5804  -7.0804   0.9821   5.9821  27.3571
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       151.393      6.149  24.621 3.80e-14 ***
## programmerMandl    27.375      6.564   4.171 0.000722 ***
## programmerMeyer    34.062      6.564   5.189 8.95e-05 ***
## bugtypeGUI         -5.812      6.564  -0.886 0.388982
## bugtypeReporting   -0.375      6.564  -0.057 0.955148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.15 on 16 degrees of freedom
## Multiple R-squared:  0.6589, Adjusted R-squared:  0.5736
## F-statistic: 7.727 on 4 and 16 DF,  p-value: 0.001149
```

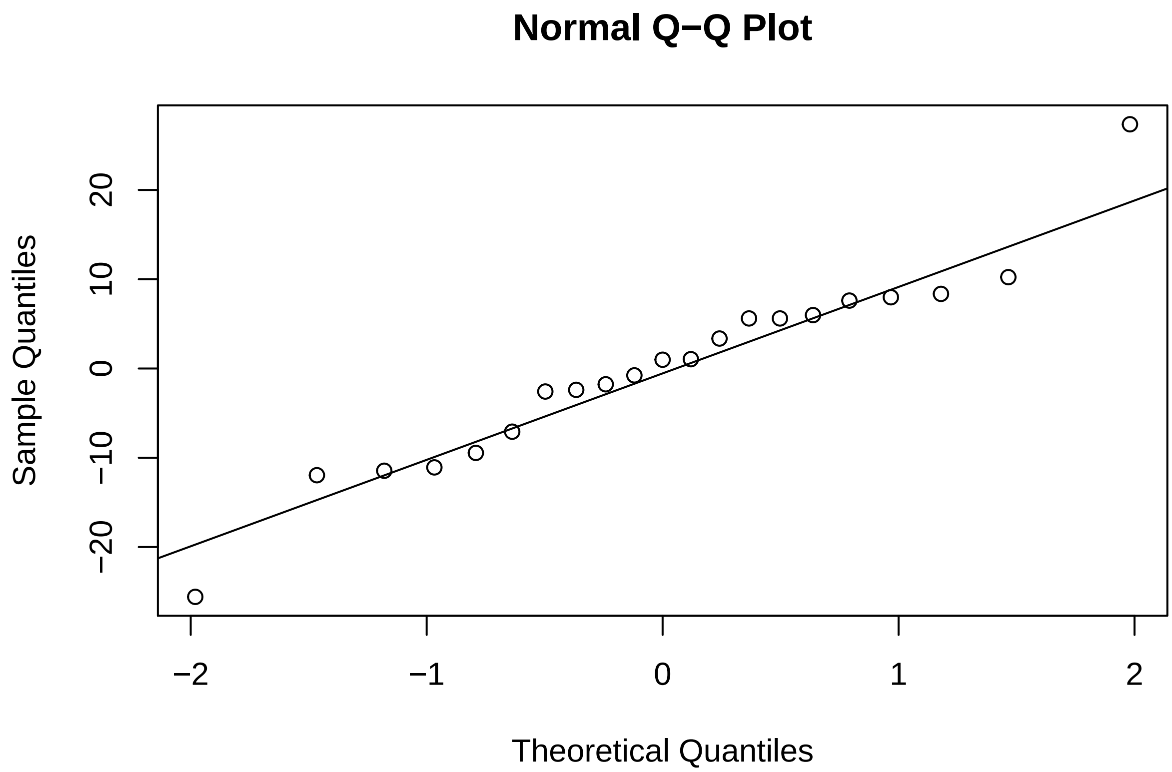
The explanatory power of the model is 65.9%, which is significant at the 0.05 significance level. The intercept represents the time needed to fix a DB bug for programmer "Eckkrammer" and is significant, just as the coefficients of the other two programmers – both of them are significantly slower than programmer Eckkrammer. The effect plot is used to summarize this graphically:

```r
library(effects)
plot(allEffects(model))
```



Finally, we check the normal distribution assumption of the residuals using a QQ Plot:

```
qqnorm(residuals(model))
qqline(residuals(model))
```



It doesn't show any peculiarities.

Now, we test for a possible interaction between *programmer* and *bugtype*. To do so, we fit a regression model with an additional interaction term:

```
model = lm(duration ~ programmer * bugtype, data = bugfixes2)
```

The ANOVA table is given by:

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: duration
##                    Df Sum Sq Mean Sq F value    Pr(>F)
## programmer          2 4420.7 2210.33 32.0209 1.544e-05 ***
## bugtype             2  145.1   72.56  1.0512  0.379600
## programmer:bugtype  4 1535.1  383.78  5.5598  0.009076 **
## Residuals          12  828.3   69.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The significances of the main effects do not change, the interaction term is significant. Thus, there is a joint influence of *programmer* and *bugtype* on *duration*.

The summary of the regression model is given by:

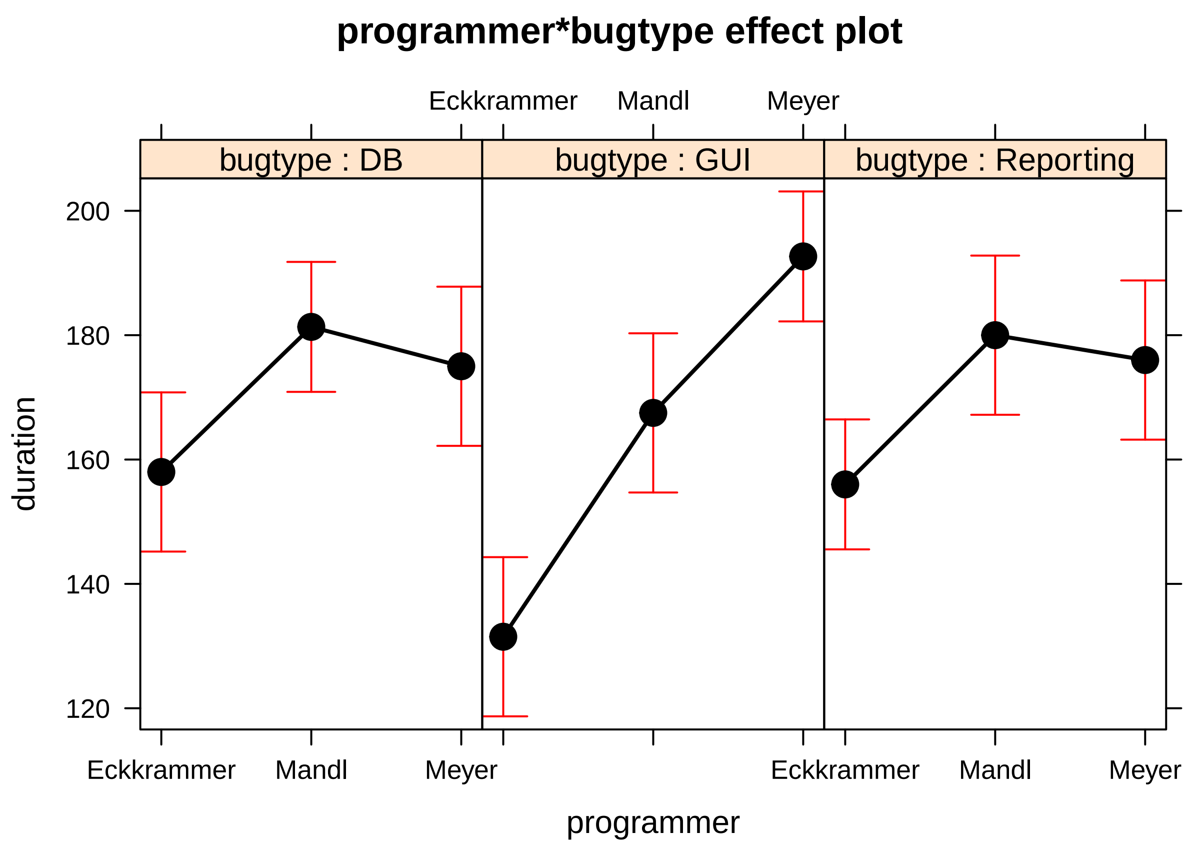
```r
summary(model)
```

```
##
## Call:
## lm(formula = duration ~ programmer * bugtype, data = bugfixes2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11.50  -4.00  -1.00   3.00  14.33
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      158.0000     5.8749  26.894 4.29e-12 ***
## programmerMandl                   23.3333     7.5844   3.076  0.00960 **
## programmerMeyer                   17.0000     8.3083   2.046  0.06330 .
## bugtypeGUI                       -26.5000     8.3083  -3.190  0.00778 **
## bugtypeReporting                  -2.0000     7.5844  -0.264  0.79649
## programmerMandl:bugtypeGUI        12.6667    11.2495   1.126  0.28219
## programmerMeyer:bugtypeGUI        44.1667    11.2495   3.926  0.00201 **
## programmerMandl:bugtypeReporting   0.6667    10.7260   0.062  0.95146
## programmerMeyer:bugtypeReporting   3.0000    11.2495   0.267  0.79424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.308 on 12 degrees of freedom
## Multiple R-squared:  0.8805, Adjusted R-squared:  0.8008
## F-statistic: 11.05 on 8 and 12 DF,  p-value: 0.0001774
```

The fitted model now has a (considerably higher) explanatory power of 88%, which is significant at the 0.05 significance level. The intercept represents the time needed to fix a DB bug for programmer "Eckkrammer" including the interaction and is significant, just as the coefficients of the other two programmers (main effects; Meyer is only significant at the 0.1 significance level) – both of them are significantly slower than programmer Eckkrammer. Moreover, the main effect of bug type GUI is significant, as well as the combination Meyer/GUI bugs.

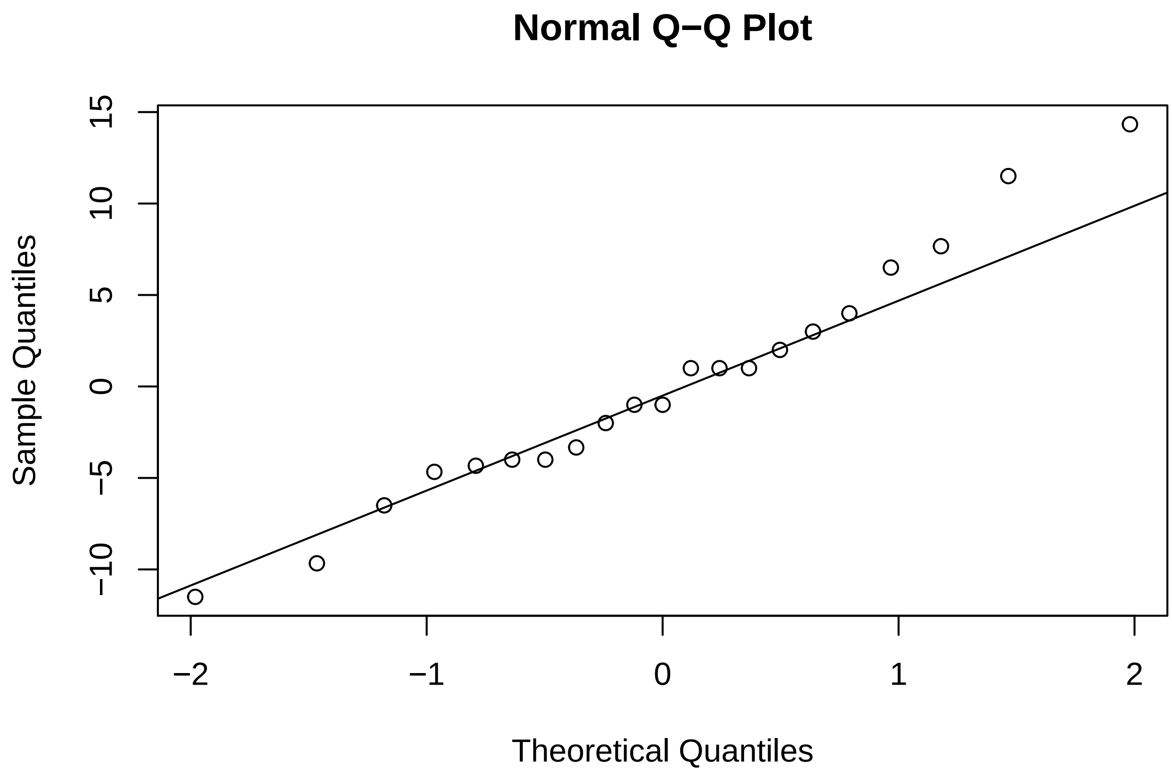
The effect plot is used to summarize this graphically:

```r
library(effects)
plot(allEffects(model))
```



Finally, we check the normal distribution assumption of the residuals using a QQ Plot:

```
qqnorm(residuals(model))
qqline(residuals(model))
```



It doesn't show any peculiarities.