Blood Pressure Analysis

Soňa Molnárová

2024-06-16

Abstract

The aim of this experiment was to uncover the dependence of both systolic and diastolic blood pressure on various variables, namely coffee intake, thyroid hormone levels, physical activity, water intake and position and time of the day during which the pressure was measured.

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0
──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all con
flicts to become errors
## Loading required package: DoE.base
##
## Loading required package: grid
##
## Loading required package: conf.design
##
## Registered S3 method overwritten by 'DoE.base':
##   method           from
##   factorize.factor conf.design
##
##
## Attaching package: 'DoE.base'
##
##
## The following objects are masked from 'package:stats':
##
##     aov, lm
##
##
## The following object is masked from 'package:graphics':
##
##     plot.design
##
##
## The following object is masked from 'package:base':
##
##     lengths
```

```
## Warning: package 'olsrr' was built under R version 4.3.3
```

```
##
## Attaching package: 'olsrr'
##
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
## [[1]]
##  [1] "lubridate" "forcats"   "stringr"   "dplyr"     "purrr"     "readr"
##  [7] "tidyr"     "tibble"    "ggplot2"   "tidyverse" "stats"     "graphics"
## [13] "grDevices" "utils"     "datasets"  "methods"   "base"
##
## [[2]]
##  [1] "FrF2"        "DoE.base"    "conf.design" "grid"        "lubridate"
##  [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
## [11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
## [16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
## [21] "base"
##
## [[3]]
##  [1] "lattice"     "FrF2"        "DoE.base"    "conf.design" "grid"
##  [6] "lubridate"   "forcats"     "stringr"     "dplyr"       "purrr"
## [11] "readr"       "tidyr"       "tibble"      "ggplot2"     "tidyverse"
## [16] "stats"       "graphics"    "grDevices"   "utils"       "datasets"
## [21] "methods"     "base"
##
## [[4]]
##  [1] "olsrr"       "lattice"     "FrF2"        "DoE.base"    "conf.design"
##  [6] "grid"        "lubridate"   "forcats"     "stringr"     "dplyr"
## [11] "purrr"       "readr"       "tidyr"       "tibble"      "ggplot2"
## [16] "tidyverse"   "stats"       "graphics"    "grDevices"   "utils"
## [21] "datasets"    "methods"     "base"
```

Data Processing

Data are loaded from a public Github repository and subsequently converted to their corresponding numeric types. The meaning behind individual variables and their levels is described below.

coffee: High level of the variable coffee stands for drinking one small espresso and was measured in milliliters.

water.intake: Water intake depicts the amount of water drank before taking each measurement.

pill: Testing subject had an underactive thyroid glad - condition known as hypothyroidism, so the variable pill represents daily thyroid hormone dosage. One pill contains 25 μg of the hormone.

All observations with high (or positive) level of either of the three factors above were measured 30 minutes after ingestion. This time spacing was chosen based on Medicine Committee on Military Nutrition Research (2001), according to which peak plasma concentrations of caffeine occur between 15 minutes and 2 hours after intake.

squats: In integer variable numerically representing the amount of physical activity expended right before the measurement.

position: Standing and laying down were considered as the two position levels and measurements were always taken right after standing up.

time: Rather than choosing a specific hour to take the measurements, the low level of the time variable was taken as the time right after waking up. High level was then measured several hours later, usually between 3 and 5 pm. So instead of interpreting the variable in its literal sense, time should be understood as a level of physical activity expended throughout the day so far. High level therefore indicates that a certain degree of physical activity has already been expended during the day whereas low level means no physical activity at all.

```
data = read.table("https://raw.githubusercontent.com/molnason/NAEX/main/design.t
xt", header=TRUE, sep="\t")
head(data)
```

```
##   systolic diastolic coffee pill position squats time water.intake
## 1      101        59      0    0        0      0    0            0
## 2      124        58     50    0        0      0    1          400
## 3      102        54      0   25        0      0    1          400
## 4       97        66     50   25        0      0    0            0
## 5       92        60      0    0        1      0    1            0
## 6      113        70     50    0        1      0    0          400
```

```
data = data %>% transmute(systolic = systolic,
    diastolic = diastolic,
    coffee = as.numeric(coffee),
    pill = pill,
    position = as.factor(position),
    squats = as.integer(squats),
    time = as.factor(time),
    water.intake = as.numeric(water.intake))
```

Since the measurements were constrained by the time both caffeine and thyroid hormones needed to wear off, collection of data turned out to be quite time consuming. For this reason unreplicated fractional factorial design was run.

```
df = data[1:16,]

head(df)
```

```
##   systolic diastolic coffee pill position squats time water.intake
## 1      101        59      0    0        0      0    0            0
## 2      124        58     50    0        0      0    1          400
## 3      102        54      0   25        0      0    1          400
## 4       97        66     50   25        0      0    0            0
## 5       92        60      0    0        1      0    1            0
## 6      113        70     50    0        1      0    0          400
```

Two fractional factorial designs were constructed because blood pressure has a nature of a 2D response variable - systolic and diastolic pressure. Even though it would have been possible to add two responses to one design, the alternative with two designs seemed easier to work with and was therefore preferred.

```
bp_design = FrF2(2^(6-2), k = 6, replications = 1, factor.names = c("coffee", "p
ill", "position", "squats", "time", "water.intake"), randomize = F)

sys = df$systolic
dia = df$diastolic

sys_design = add.response(bp_design, sys)
dia_design = add.response(bp_design, dia)

summary(sys_design)
```

```
## Call:
## FrF2(2^(6 - 2), k = 6, replications = 1, factor.names = c("coffee",
##     "pill", "position", "squats", "time", "water.intake"), randomize = F)
##
## Experimental design of type  FrF2
## 16  runs
##
## Factor settings (scale ends):
##   coffee pill position squats time water.intake
## 1     -1   -1       -1     -1   -1           -1
## 2      1    1        1      1    1            1
##
## Responses:
## [1] sys
##
## Design generating information:
## $legend
## [1] A=coffee       B=pill         C=position     D=squats       E=time
## [6] F=water.intake
##
## $generators
## [1] E=ABC F=ABD
##
##
## Alias structure:
## $fi2
## [1] AB=CE=DF AC=BE    AD=BF    AE=BC    AF=BD    CD=EF    CF=DE
##
##
## The design itself:
##    coffee pill position squats time water.intake sys
## 1      -1   -1       -1     -1   -1           -1 101
## 2       1   -1       -1     -1    1            1 124
## 3      -1    1       -1     -1    1            1 102
## 4       1    1       -1     -1   -1           -1  97
## 5      -1   -1        1     -1    1           -1  92
## 6       1   -1        1     -1   -1            1 113
## 7      -1    1        1     -1   -1            1 114
## 8       1    1        1     -1    1           -1 113
## 9      -1   -1       -1      1   -1            1 130
## 10      1   -1       -1      1    1           -1 114
## 11     -1    1       -1      1    1           -1 125
## 12      1    1       -1      1   -1            1 128
## 13     -1   -1        1      1    1            1 133
## 14      1   -1        1      1   -1           -1 130
## 15     -1    1        1      1   -1           -1 119
```

```
## 16      1    1        1      1    1            1 131
## class=design, type= FrF2
```

```
summary(dia_design)
```

```
## Call:
## FrF2(2^(6 - 2), k = 6, replications = 1, factor.names = c("coffee",
##     "pill", "position", "squats", "time", "water.intake"), randomize = F)
##
## Experimental design of type  FrF2
## 16  runs
##
## Factor settings (scale ends):
##   coffee pill position squats time water.intake
## 1     -1   -1       -1     -1   -1           -1
## 2      1    1        1      1    1            1
##
## Responses:
## [1] dia
##
## Design generating information:
## $legend
## [1] A=coffee       B=pill         C=position     D=squats       E=time
## [6] F=water.intake
##
## $generators
## [1] E=ABC F=ABD
##
##
## Alias structure:
## $fi2
## [1] AB=CE=DF AC=BE    AD=BF    AE=BC    AF=BD    CD=EF    CF=DE
##
##
## The design itself:
##    coffee pill position squats time water.intake dia
## 1      -1   -1       -1     -1   -1           -1  59
## 2       1   -1       -1     -1    1            1  58
## 3      -1    1       -1     -1    1            1  54
## 4       1    1       -1     -1   -1           -1  66
## 5      -1   -1        1     -1    1           -1  60
## 6       1   -1        1     -1   -1            1  70
## 7      -1    1        1     -1   -1            1  70
## 8       1    1        1     -1    1           -1  77
## 9      -1   -1       -1      1   -1            1  66
## 10      1   -1       -1      1    1           -1  57
## 11     -1    1       -1      1    1           -1  59
## 12      1    1       -1      1   -1            1  65
## 13     -1   -1        1      1    1            1  63
## 14      1   -1        1      1   -1           -1  67
## 15     -1    1        1      1   -1           -1  74
```

```
## 16      1    1        1      1    1           1  68
## class=design, type= FrF2
```

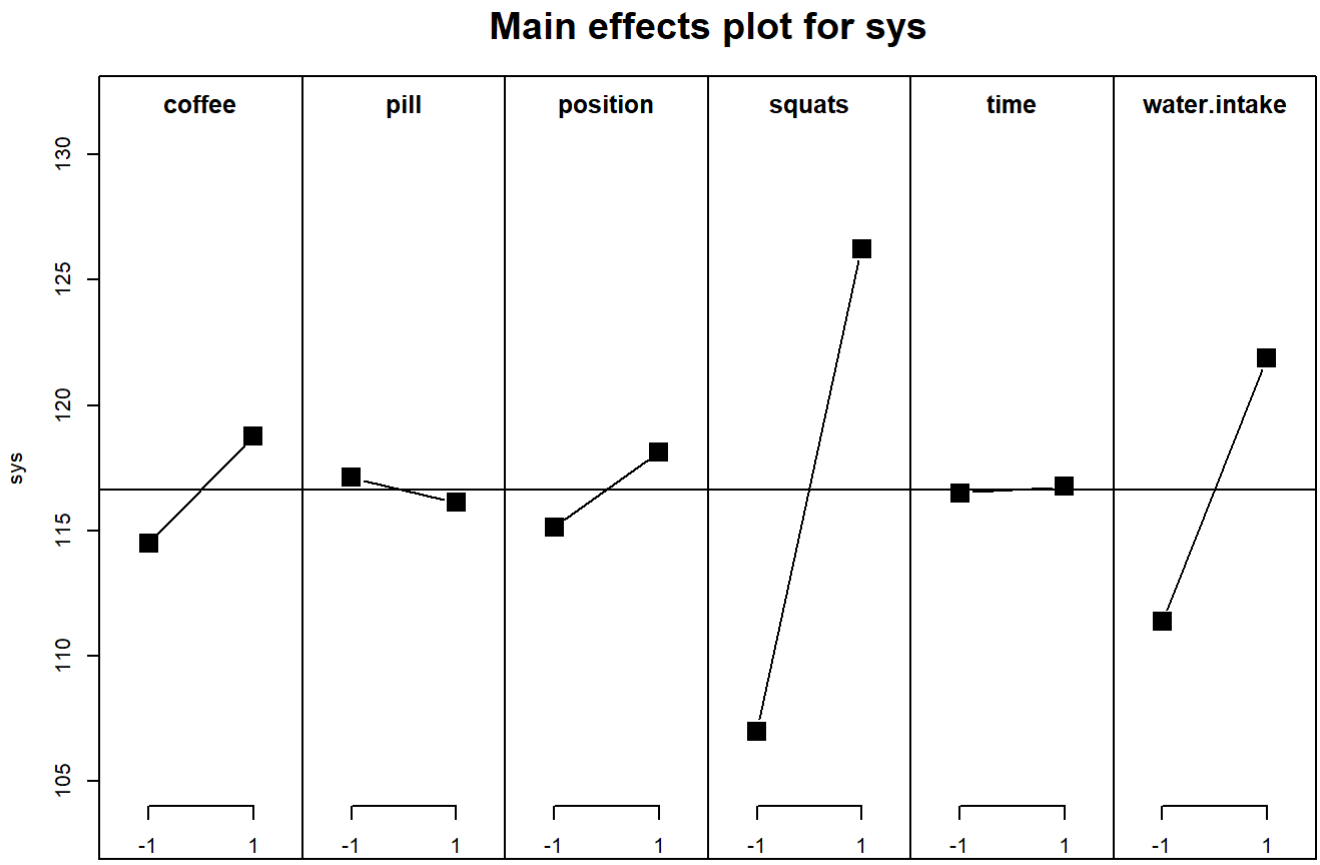
It can be seen from the designs' summary that they both have the same generators E = ABC and F = ABD. This means that out of all rows in the corresponding 2^6 full factorial design only the rows which satisfy the given conditions were kept in the design.

Another term mentioned above is alias. Alias of interactions AC and BE makes it impossible to differentiate between the two of them. Neither AC nor BE are therefore estimated alone. What is being estimated instead is the effect of AC + BE. In the above designs there are five pairs and one triplet of aliased first order interactions.

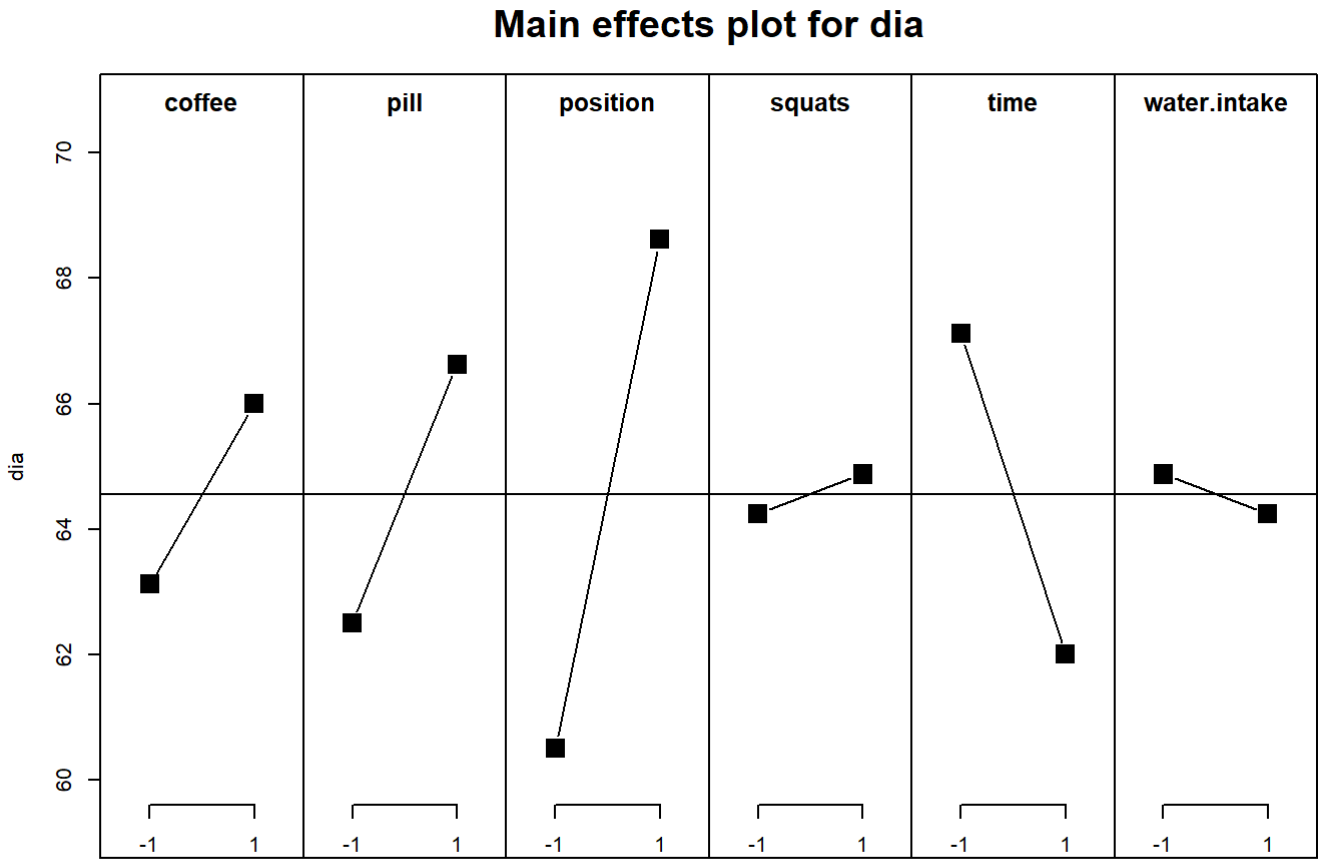
Visual Analysis

FrF2 designs can be used to generate main effects plots and interaction plots to reveal what effects and what interactions have the most significant effect on the response variables.

```
MEPlot(sys_design)
```

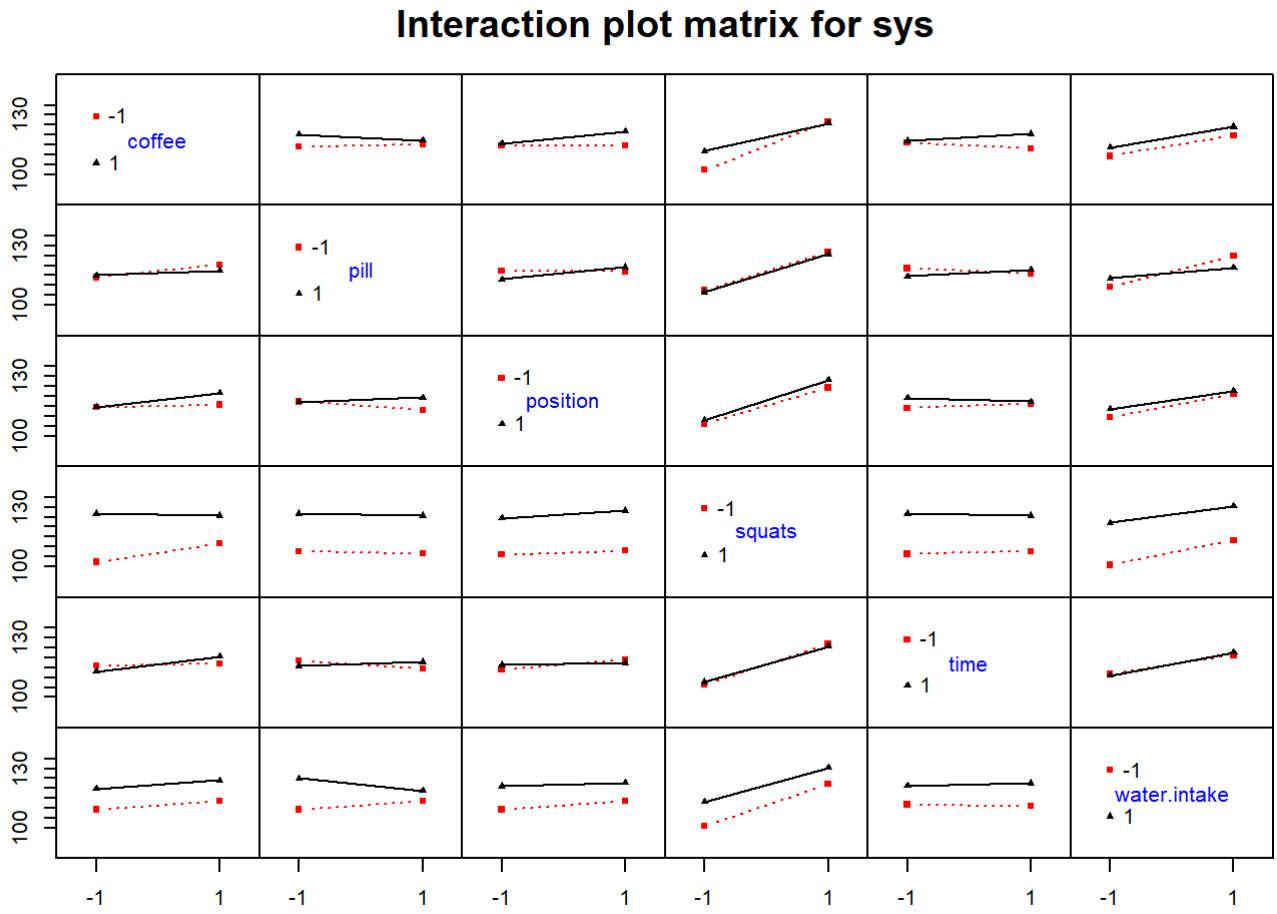



```
MEPlot(dia_design)
```

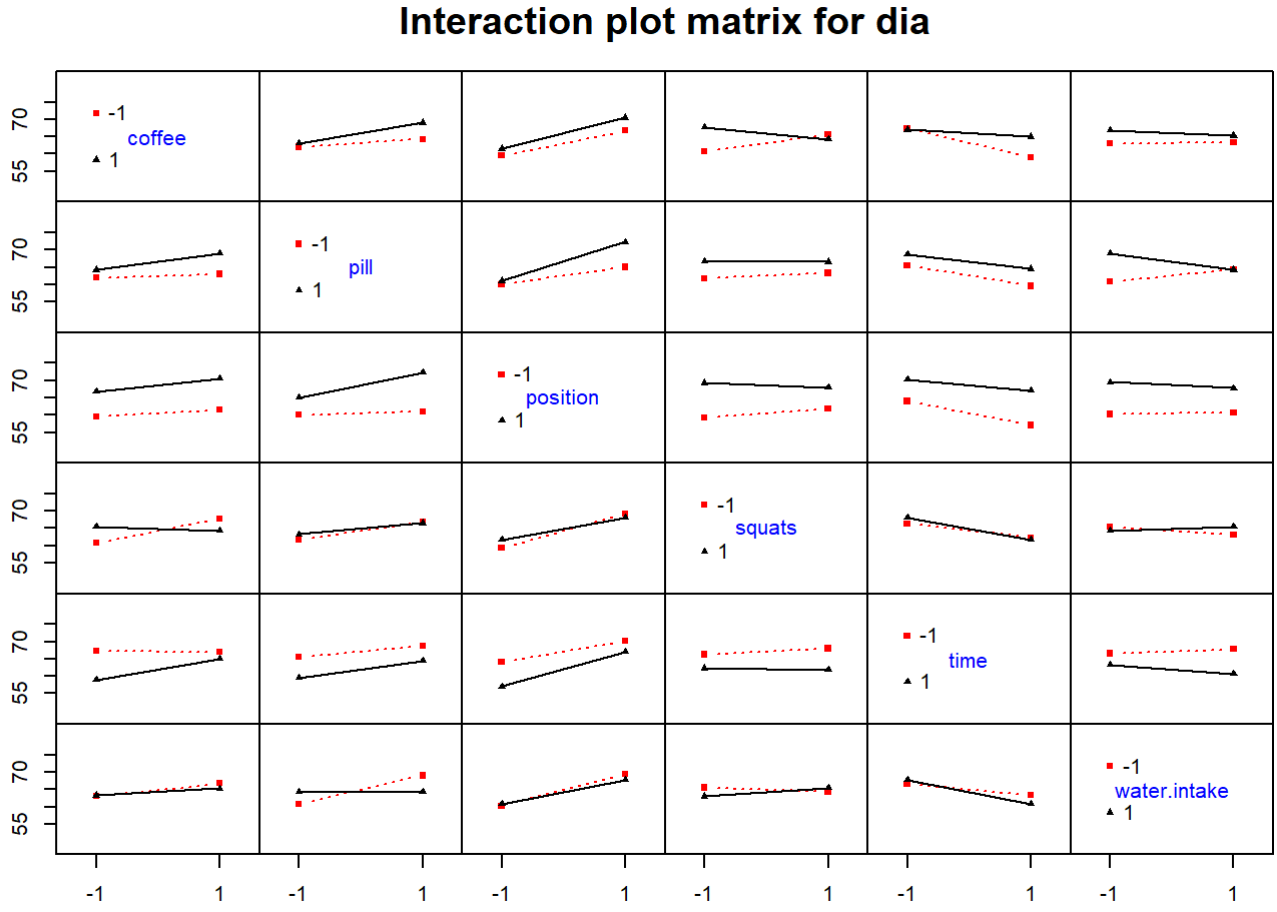


Physical activity manifested in form of squats together with water intake appeared to have an effect on systolic blood pressure. In case of diastolic blood pressure, factors time and position seemed to be significant based on their MEPlots.

```
IAPlot(sys_design)
```



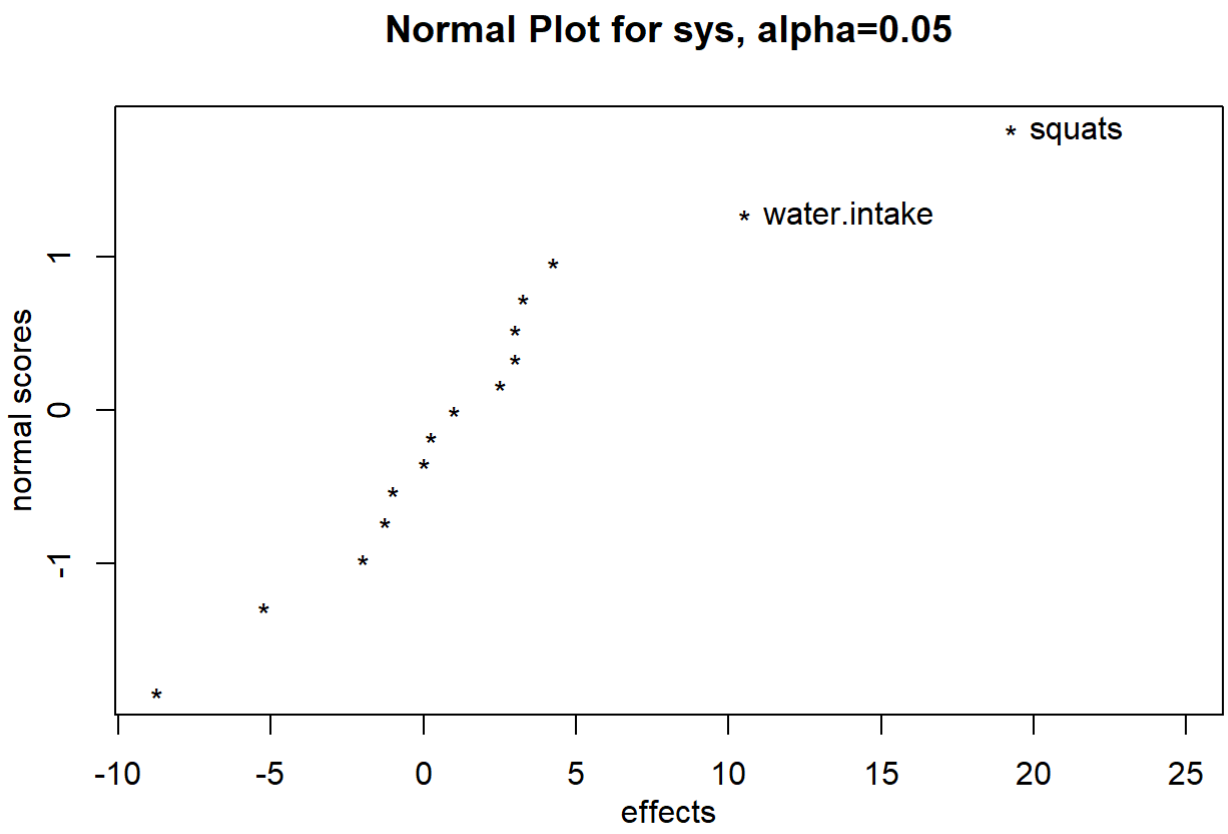

```
IAPlot(dia_design)
```



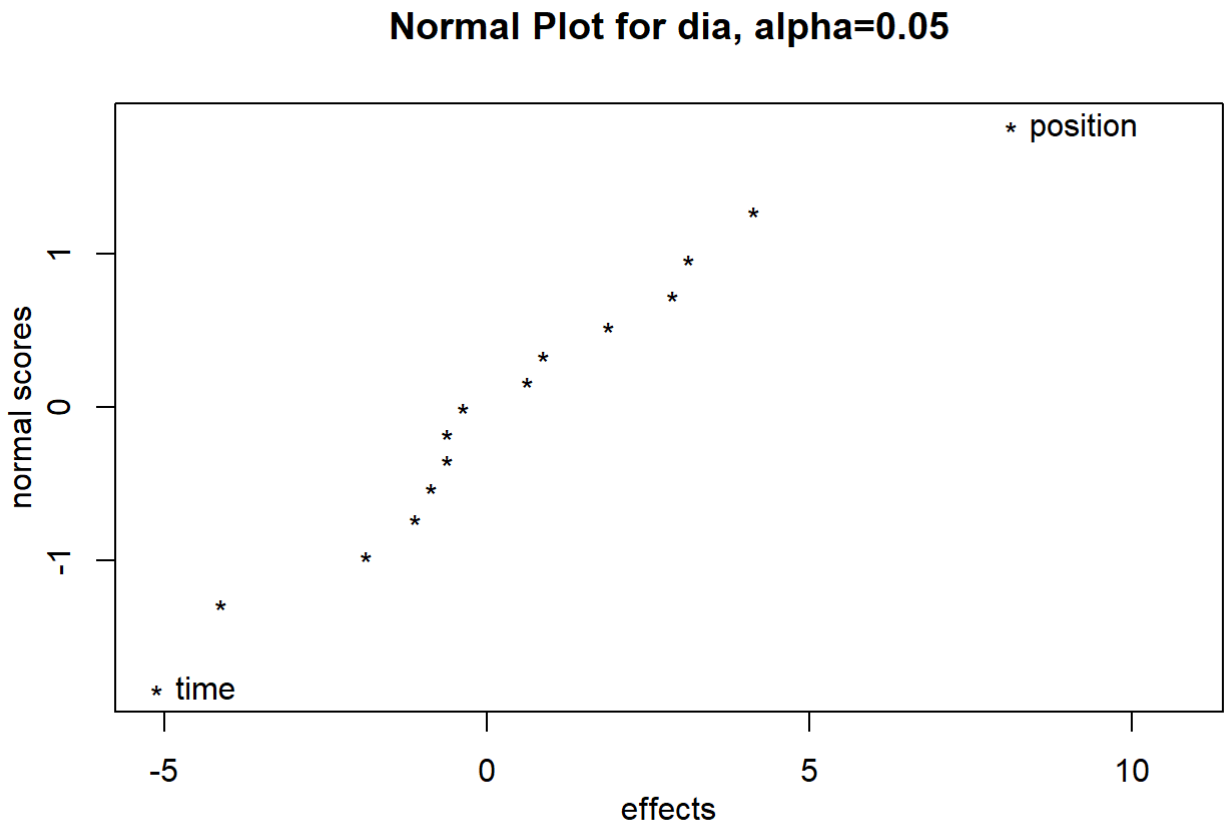
If two lines in interaction plots cross or are visibly non-parallel then their respective variables might form an interaction.

Another FrF2 visualization tool present Daniel plots. These are capable of detecting both significant main effects as well as significant interactions. The results obtained by this form of analysis correspond with the ones obtained using MEPlots.

```
FrF2::DanielPlot(sys_design)
```




```
FrF2::DanielPlot(dia_design)
```



Analysis Using ANOVA

Changes in both responses were rather small for all explanatory variables. For instance, the biggest change in diastolic pressure (in variable position, as seen in its main effects plot) is only by 9 mm Hg. The significance of the results obtained from visual analysis therefore needs to be verified by ANOVA.

```
summary(aov(sys ~ .^2 - coffee:water.intake - squats:pill - squats:position, sys
_design))
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## coffee                 1   72.2    72.2   0.654 0.4778
## pill                   1    4.0     4.0   0.036 0.8612
## position               1   36.0    36.0   0.326 0.6080
## squats                 1 1482.2  1482.2  13.424 0.0351 *
## time                   1    0.2     0.2   0.002 0.9650
## water.intake           1  441.0   441.0   3.994 0.1395
## coffee:pill            1   16.0    16.0   0.145 0.7288
## coffee:position        1   36.0    36.0   0.326 0.6080
## coffee:squats          1  110.2   110.2   0.998 0.3913
## coffee:time            1   42.3    42.3   0.383 0.5800
## position:water.intake  1    6.3     6.3   0.057 0.8273
## time:water.intake      1    4.0     4.0   0.036 0.8612
## Residuals              3  331.3   110.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(dia ~ .^2 - time:pill - squats:pill - water.intake:position, dia_des
ign))
```

```
##                     Df Sum Sq Mean Sq F value  Pr(>F)
## coffee               1  33.06   33.06  31.118 0.03067 *
## pill                 1  68.06   68.06  64.059 0.01525 *
## position             1 264.06  264.06 248.529 0.00400 **
## squats               1   1.56    1.56   1.471 0.34906
## time                 1 105.06  105.06  98.882 0.00996 **
## water.intake         1   1.56    1.56   1.471 0.34906
## coffee:pill          1  14.06   14.06  13.235 0.06795 .
## coffee:position      1   3.06    3.06   2.882 0.23165
## coffee:squats        1  68.06   68.06  64.059 0.01525 *
## coffee:time          1  39.06   39.06  36.765 0.02614 *
## coffee:water.intake  1   3.06    3.06   2.882 0.23165
## position:squats      1  14.06   14.06  13.235 0.06795 .
## squats:time          1   5.06    5.06   4.765 0.16075
## Residuals            2   2.13    1.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Only the squats variable appears significant in the systolic pressure model. On the other hand diastolic pressure model seems to be a lot more complex. Let us now fit the two (full) models and see their coefficients.

```
summary(lm(sys ~ .^2 - coffee:water.intake - squats:pill - squats:position, sys_
design))
```

```
##
## Call:
## lm.default(formula = sys ~ .^2 - coffee:water.intake - squats:pill -
##     squats:position, data = sys_design)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10     11
##  3.125  5.625 -5.625 -3.125 -3.125 -5.625  5.625  3.125 -3.125 -5.625  5.625
##     12     13     14     15     16
##  3.125  3.125  5.625 -5.625 -3.125
##
## Coefficients: (6 not defined because of singularities)
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              116.625      2.627  44.395 2.52e-05 ***
## coffee1                    2.125      2.627   0.809   0.4778
## pill1                     -0.500      2.627  -0.190   0.8612
## position1                  1.500      2.627   0.571   0.6080
## squats1                    9.625      2.627   3.664   0.0351 *
## time1                      0.125      2.627   0.048   0.9650
## water.intake1              5.250      2.627   1.998   0.1395
## coffee1:pill1             -1.000      2.627  -0.381   0.7288
## coffee1:position1          1.500      2.627   0.571   0.6080
## coffee1:squats1           -2.625      2.627  -0.999   0.3913
## coffee1:time1              1.625      2.627   0.619   0.5800
## pill1:position1               NA         NA      NA       NA
## pill1:time1                   NA         NA      NA       NA
## pill1:water.intake1           NA         NA      NA       NA
## position1:time1               NA         NA      NA       NA
## position1:water.intake1   -0.625      2.627  -0.238   0.8273
## squats1:time1                 NA         NA      NA       NA
## squats1:water.intake1         NA         NA      NA       NA
## time1:water.intake1        0.500      2.627   0.190   0.8612
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.51 on 3 degrees of freedom
## Multiple R-squared:  0.8717, Adjusted R-squared:  0.3585
## F-statistic: 1.698 on 12 and 3 DF,  p-value: 0.366
```

```
summary(lm(dia ~ .^2 - time:pill - squats:pill - water.intake:position, dia_desi
gn))
```

```
##
## Call:
## lm.default(formula = dia ~ .^2 - time:pill - squats:pill - water.intake:posit
ion,
##     data = dia_design)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10     11
##  0.500 -0.125  0.125 -0.500 -0.500  0.125 -0.125  0.500 -0.500  0.125 -0.125
##     12     13     14     15     16
##  0.500  0.500 -0.125  0.125 -0.500
##
## Coefficients: (5 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            64.5625     0.2577 250.539 1.59e-05 ***
## coffee1                 1.4375     0.2577   5.578  0.03067 *
## pill1                   2.0625     0.2577   8.004  0.01525 *
## position1               4.0625     0.2577  15.765  0.00400 **
## squats1                 0.3125     0.2577   1.213  0.34906
## time1                  -2.5625     0.2577  -9.944  0.00996 **
## water.intake1          -0.3125     0.2577  -1.213  0.34906
## coffee1:pill1           0.9375     0.2577   3.638  0.06795 .
## coffee1:position1       0.4375     0.2577   1.698  0.23165
## coffee1:squats1        -2.0625     0.2577  -8.004  0.01525 *
## coffee1:time1           1.5625     0.2577   6.063  0.02614 *
## coffee1:water.intake1  -0.4375     0.2577  -1.698  0.23165
## pill1:position1             NA         NA      NA       NA
## pill1:water.intake1         NA         NA      NA       NA
## position1:squats1      -0.9375     0.2577  -3.638  0.06795 .
## position1:time1             NA         NA      NA       NA
## squats1:time1          -0.5625     0.2577  -2.183  0.16075
## squats1:water.intake1       NA         NA      NA       NA
## time1:water.intake1         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.031 on 2 degrees of freedom
## Multiple R-squared:  0.9966, Adjusted R-squared:  0.9744
## F-statistic: 44.87 on 13 and 2 DF,  p-value: 0.022
```

There are six and five coefficients missing in the model for systolic and diastolic pressure respectively. As seen in the summary output, these are not defined because of singularities which means that the undefined variables are linear combinations of other variables and there is therefore no unique solution to the regression task.

Looking at the design summary it can be seen that the undefined coefficients belong to the aliased interaction terms. Instead of estimating the coefficient for coffee:squats what is really being estimated is the coefficient for coffee:squats + position:pill, which is obviously linearly dependent

on coffee:squats causing a singularity when trying to estimate position:pill coefficient.

Choosing generators as E = ABC and F = ABD caused the design to have every two-way interaction aliased with at least one other two-way interaction and it is hence impossible to estimate them independently.

```
m_sys_full = lm(sys ~ ., data = sys_design)
summary(m_sys_full)
```

```
##
## Call:
## lm.default(formula = sys ~ ., data = sys_design)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.750 -4.000  0.375  3.500 10.500
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    116.625      1.947  59.893 5.08e-13 ***
## coffee1          2.125      1.947   1.091 0.303493
## pill1           -0.500      1.947  -0.257 0.803125
## position1        1.500      1.947   0.770 0.460846
## squats1          9.625      1.947   4.943 0.000799 ***
## time1            0.125      1.947   0.064 0.950219
## water.intake1    5.250      1.947   2.696 0.024548 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.789 on 9 degrees of freedom
## Multiple R-squared:  0.7885, Adjusted R-squared:  0.6475
## F-statistic: 5.593 on 6 and 9 DF,  p-value: 0.01126
```

```
m_dia_full = lm(dia ~ ., data = dia_design)
summary(m_dia_full)
```

```
##
## Call:
## lm.default(formula = dia ~ ., data = dia_design)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1875 -1.5625 -0.5625  0.6562  7.4375
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    64.5625     1.0157  63.563 2.98e-13 ***
## coffee1         1.4375     1.0157   1.415  0.19065
## pill1           2.0625     1.0157   2.031  0.07287 .
## position1       4.0625     1.0157   4.000  0.00311 **
## squats1         0.3125     1.0157   0.308  0.76534
## time1          -2.5625     1.0157  -2.523  0.03262 *
## water.intake1  -0.3125     1.0157  -0.308  0.76534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.063 on 9 degrees of freedom
## Multiple R-squared:  0.7611, Adjusted R-squared:  0.6019
## F-statistic:  4.78 on 6 and 9 DF,  p-value: 0.01845
```

Removing the insignificant terms from the full model is good for reducing the length of confidence intervals. That is why all variables but squats and water.intake, whose p-values are lower than 0.05, will be omitted in the regression model for systolic pressure. Besides the statistically significant variables diastolic pressure model below also includes the pill variable.

```
m_sys = lm(sys ~ squats + water.intake, data = sys_design)
summary(m_sys)
```

```
##
## Call:
## lm.default(formula = sys ~ squats + water.intake, data = sys_design)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.250  -3.812  -0.625   2.312  11.750
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    116.625      1.779  65.546  < 2e-16 ***
## squats1          9.625      1.779   5.409 0.000119 ***
## water.intake1    5.250      1.779   2.951 0.011258 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.117 on 13 degrees of freedom
## Multiple R-squared:  0.7449, Adjusted R-squared:  0.7057
## F-statistic: 18.98 on 2 and 13 DF,  p-value: 0.000139
```

```
m_dia = lm(dia ~ pill + position + time, data = dia_design)
summary(m_dia)
```

```
##
## Call:
## lm.default(formula = dia ~ pill + position + time, data = dia_design)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0000 -2.0312 -0.1250  0.9375  8.8750
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  64.5625     0.9809  65.817  < 2e-16 ***
## pill1         2.0625     0.9809   2.103  0.05728 .
## position1     4.0625     0.9809   4.141  0.00137 **
## time1        -2.5625     0.9809  -2.612  0.02271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.924 on 12 degrees of freedom
## Multiple R-squared:  0.7029, Adjusted R-squared:  0.6287
## F-statistic: 9.465 on 3 and 12 DF,  p-value: 0.001739
```

As pill's p-value is slightly higher than 0.05, it is useful to check whether its addition actually improves the model or not. This can be done using ANOVA.

```
m_dia_check = lm(dia ~ position + time, data = dia_design)
summary(m_dia_check)
```

```
##
## Call:
## lm.default(formula = dia ~ position + time, data = dia_design)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0625 -3.2812 -0.4375  2.1563 10.9375
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   64.562      1.102  58.562  < 2e-16 ***
## position1      4.062      1.102   3.685  0.00275 **
## time1         -2.563      1.102  -2.324  0.03695 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.41 on 13 degrees of freedom
## Multiple R-squared:  0.5935, Adjusted R-squared:  0.531
## F-statistic:  9.49 on 2 and 13 DF,  p-value: 0.002876
```

```
anova(m_dia_check, m_dia)
```

```
## Analysis of Variance Table
##
## Model 1: dia ~ position + time
## Model 2: dia ~ pill + position + time
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     13 252.81
## 2     12 184.75  1    68.062 4.4208 0.05728 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtained ANOVA p-value is equal to the p-value of the pill coefficient in the more complex model and since its value is over 0.05, ANOVA suggests that the more complex model does not fit the data significantly better than the simpler model, which only includes time and position variables.

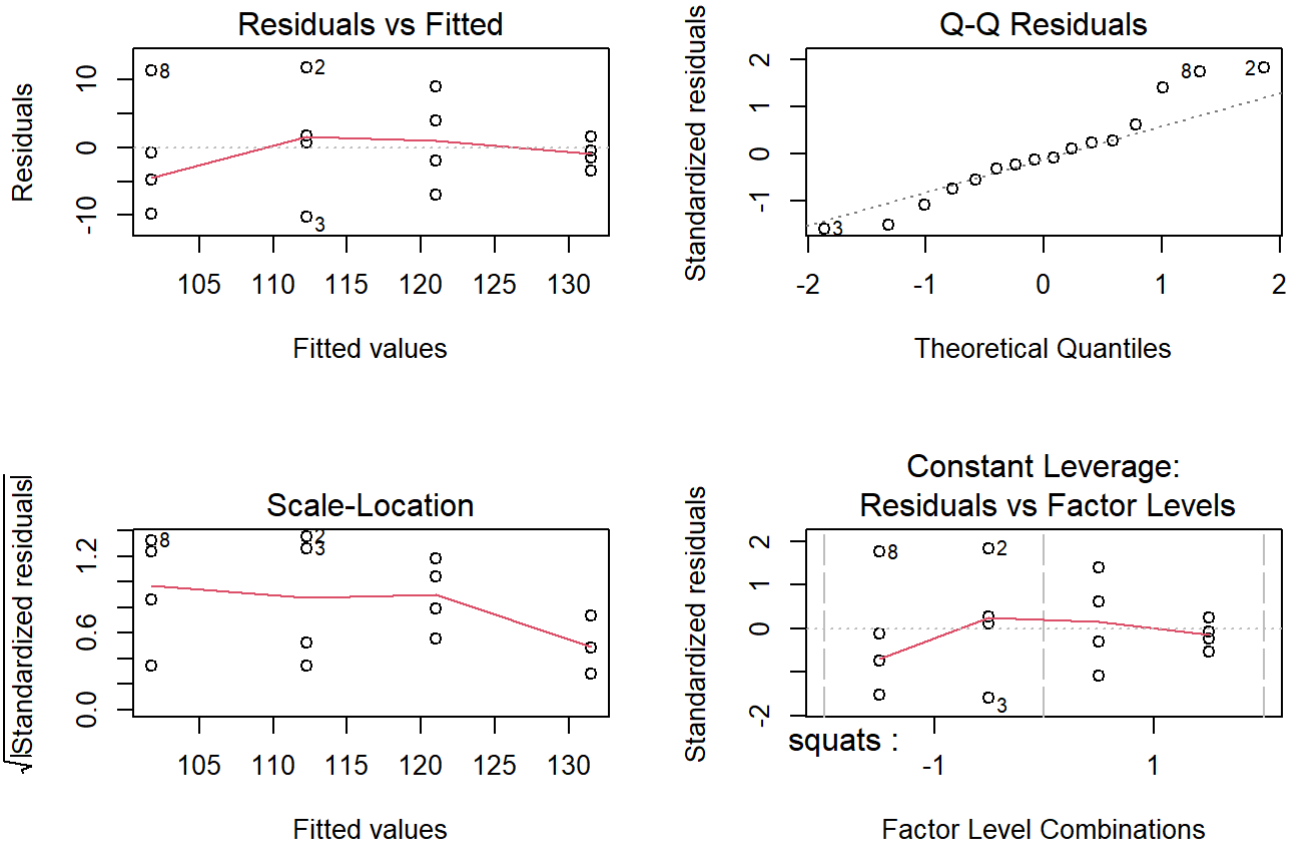
However, it can be observed from the model summary that the model's R-squared decreased rather significantly from 0.7 to 0.6 after omitting the pill variable. This is a larger drop than was observed after omitting all three other main effects from the full model when R-squared dropped from 0.76 to 0.7.

Furthermore, thyroid hormones (represented by the pill variable) have been documented to have an effect on diastolic blood pressure in people with hypothyroidism, see Danzi and Klein (2003). After combining the improved R-squared together with expert knowledge it would seem reasonable to keep the pill variable in the model despite the contradictory ANOVA results.

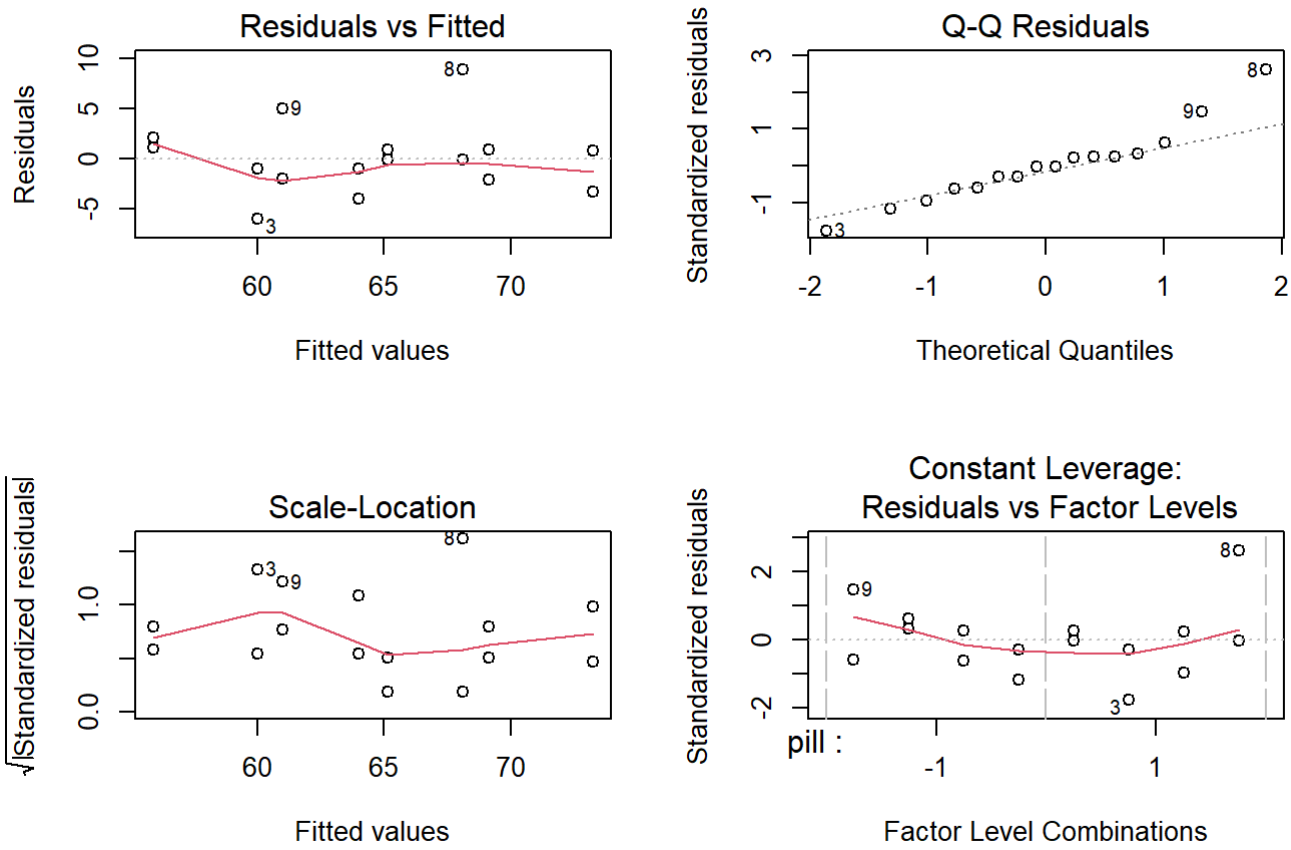
Checking Model Assumptions

After establishing the final form of both models it is necessary to check whether they follow regression model assumptions, specifically homoscedasticity and normality of residuals. This can be done by analyzing the residual plots. Linearity assumption will be focused on later.

```
par(mfrow = c(2,2))
plot(m_sys)
```



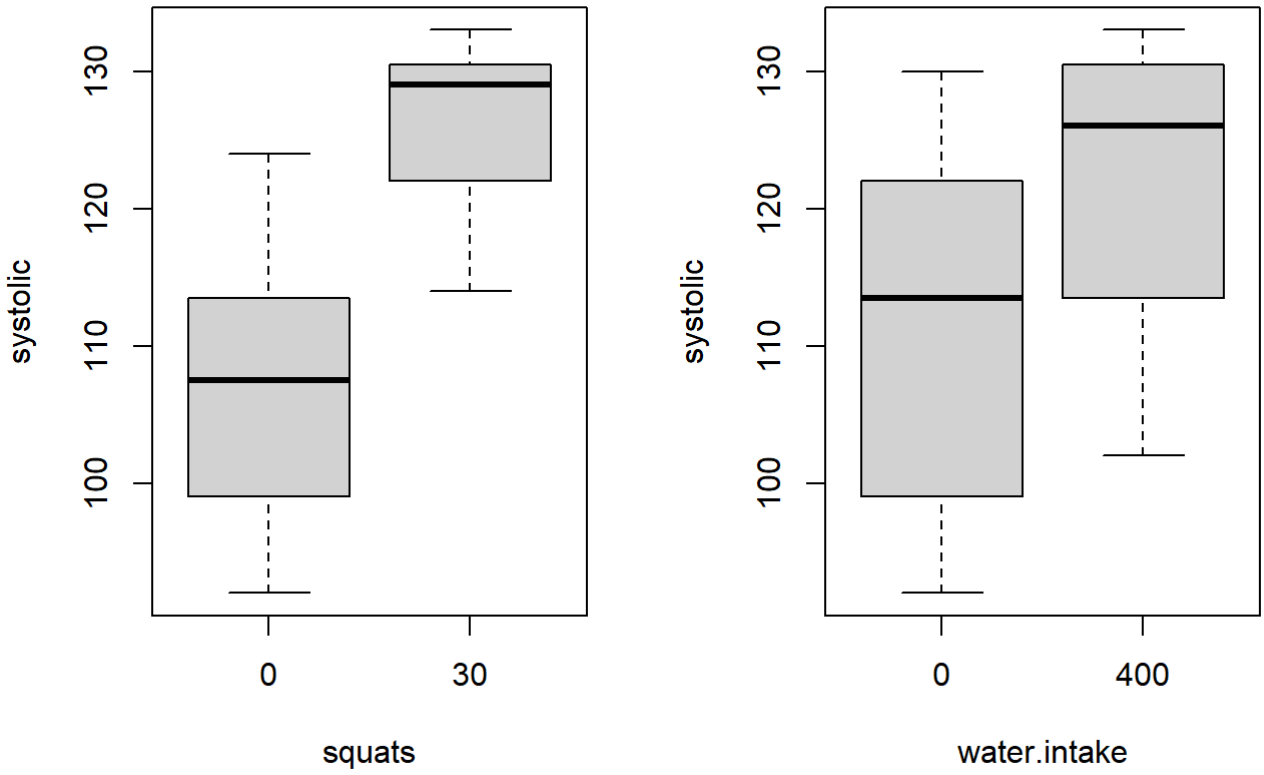

```
par(mfrow = c(2,2))
plot(m_dia)
```



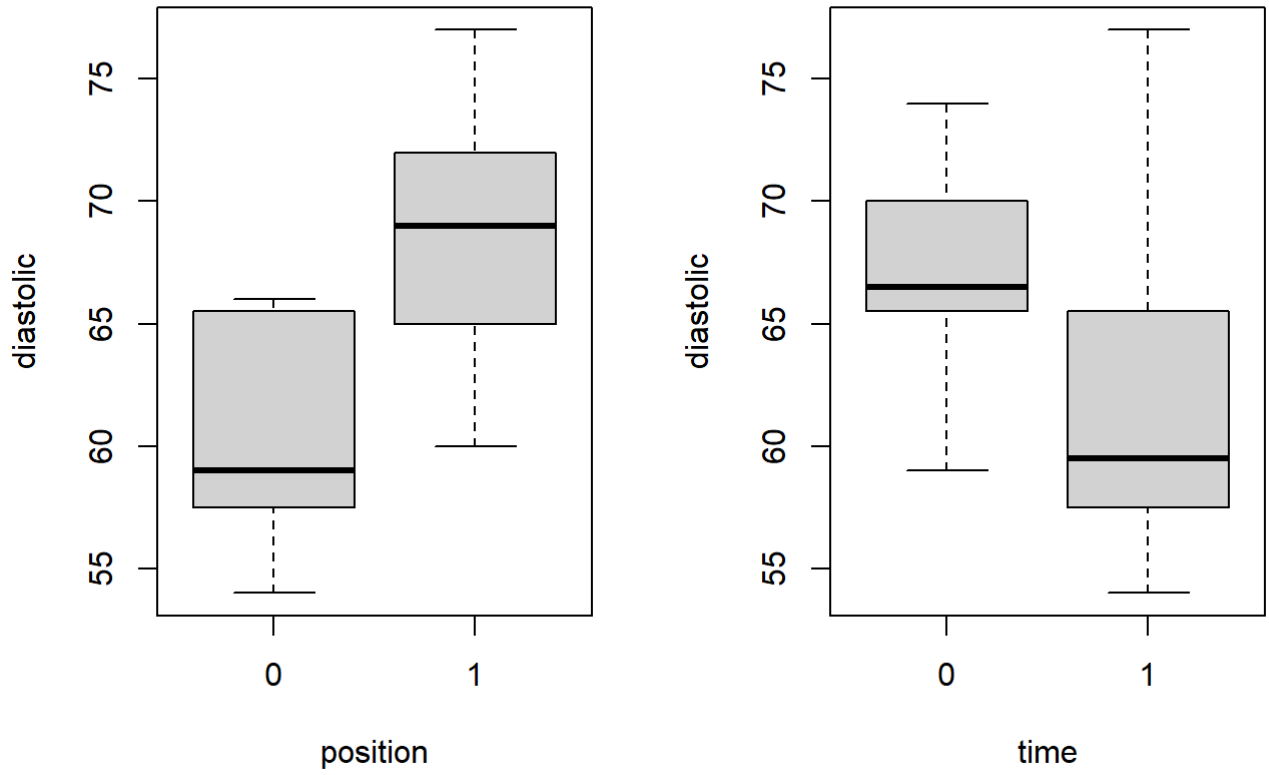
Normality assumption does not appear to be broken based on the QQ plots of both models. Only observation 8 falls rather far from the QQ line in both plots, which will be inspected later on.

Homoscedasticity seems to be followed in diastolic model, however, there is a noticeable decrease in variance with growing systolic pressure values in residuals vs fitted plot. This disproportion is also visible in box plots of both variables significant for the systolic model shown below. Diastolic pressure does not showcase this property.

```
par(mfrow = c(1,2))
boxplot(systolic ~ squats, data = df)
boxplot(systolic ~ water.intake, data = df)
```




```
par(mfrow = c(1,2))
boxplot(diastolic ~ position, data = df)
boxplot(diastolic ~ time, data = df)
```



Statistical tests can be performed to see whether this effect is significant enough or not. Breusch-Pagan test can be used for this purpose.

```
ols_test_breusch_pagan(m_sys)
```

```
##
##  Breusch Pagan Test for Heteroskedasticity
##  -----------------------------------------
##  Ho: the variance is constant
##  Ha: the variance is not constant
##
##                Data
##  -------------------------------
##  Response : sys
##  Variables: fitted values of sys
##
##          Test Summary
##  ----------------------------
##  DF            =    1
##  Chi2          =    2.22923
##  Prob > Chi2   =    0.1354214
```

Since the p-value is greater than the significance level there is insufficient evidence to conclude that there is heteroscedasticity in the data.

Inspecting Outliers

Observations 2, 3, 8 and 9 were detected to have the highest residuals according to the residuals vs fitted plots. It is therefore appropriate to inspect their validity.

```
influence.measures(m_sys)
```

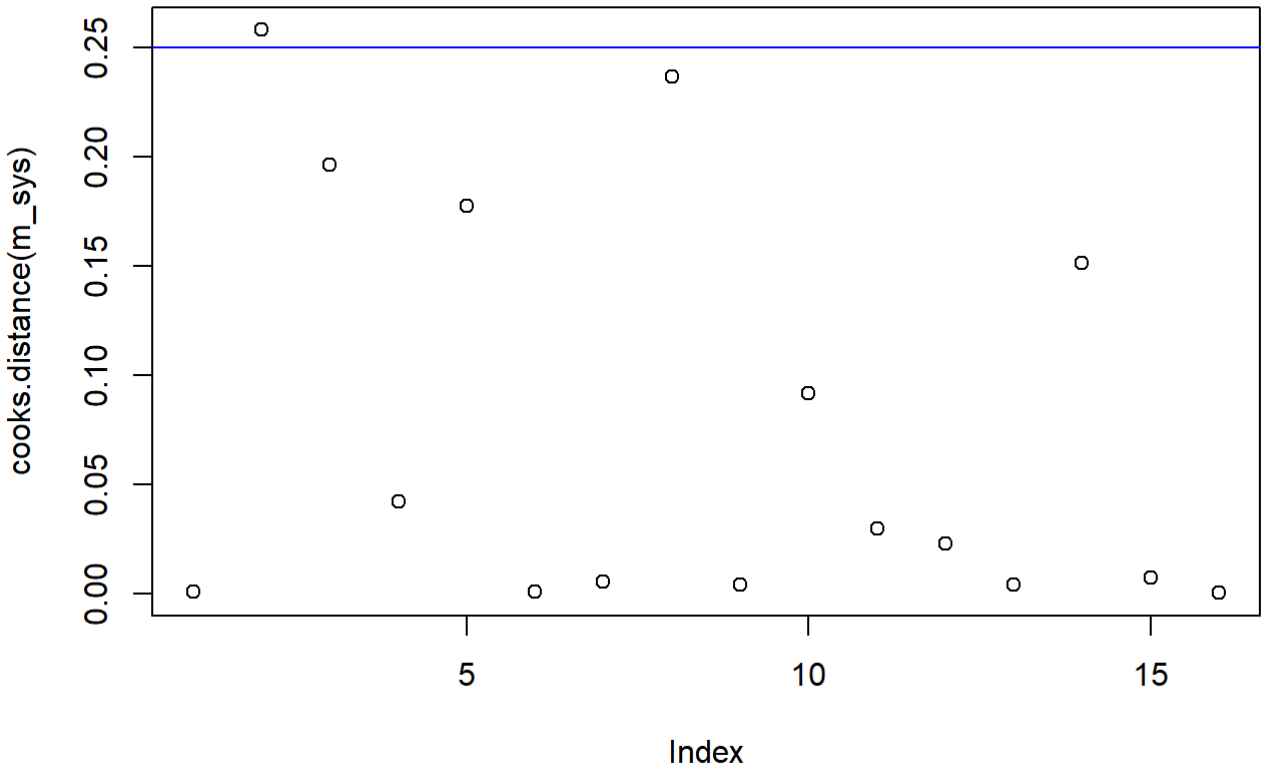
```
## Influence measures of
##   lm.default(formula = sys ~ squats + water.intake, data = sys_design) :
##
##     dfb.1_ dfb.sqt1 dfb.wt.1  dffit cov.r   cook.d   hat inf
## 1  -0.0312   0.0312   0.0312 -0.054 1.560 0.001051 0.187
## 2   0.5666  -0.5666   0.5666  0.981 0.639 0.258046 0.188
## 3  -0.4749   0.4749  -0.4749 -0.823 0.812 0.196367 0.188
## 4  -0.2016   0.2016   0.2016 -0.349 1.375 0.042170 0.188
## 5  -0.4466   0.4466   0.4466 -0.774 0.870 0.177677 0.188
## 6   0.0312  -0.0312   0.0312  0.054 1.560 0.001051 0.188
## 7   0.0729  -0.0729   0.0729  0.126 1.538 0.005724 0.188
## 8   0.5348  -0.5348  -0.5348  0.926 0.696 0.236552 0.188
## 9  -0.0624  -0.0624  -0.0624 -0.108 1.545 0.004205 0.188
## 10 -0.3051  -0.3051   0.3051 -0.528 1.173 0.091583 0.188
## 11  0.1687   0.1687  -0.1687  0.292 1.429 0.029905 0.188
## 12 -0.1471  -0.1471  -0.1471 -0.255 1.460 0.022896 0.188
## 13  0.0624   0.0624   0.0624  0.108 1.545 0.004205 0.188
## 14  0.4058   0.4058  -0.4058  0.703 0.956 0.151393 0.188
## 15 -0.0834  -0.0834   0.0834 -0.144 1.530 0.007476 0.188
## 16 -0.0208  -0.0208  -0.0208 -0.036 1.563 0.000467 0.188
```

```
influence.measures(m_dia)
```

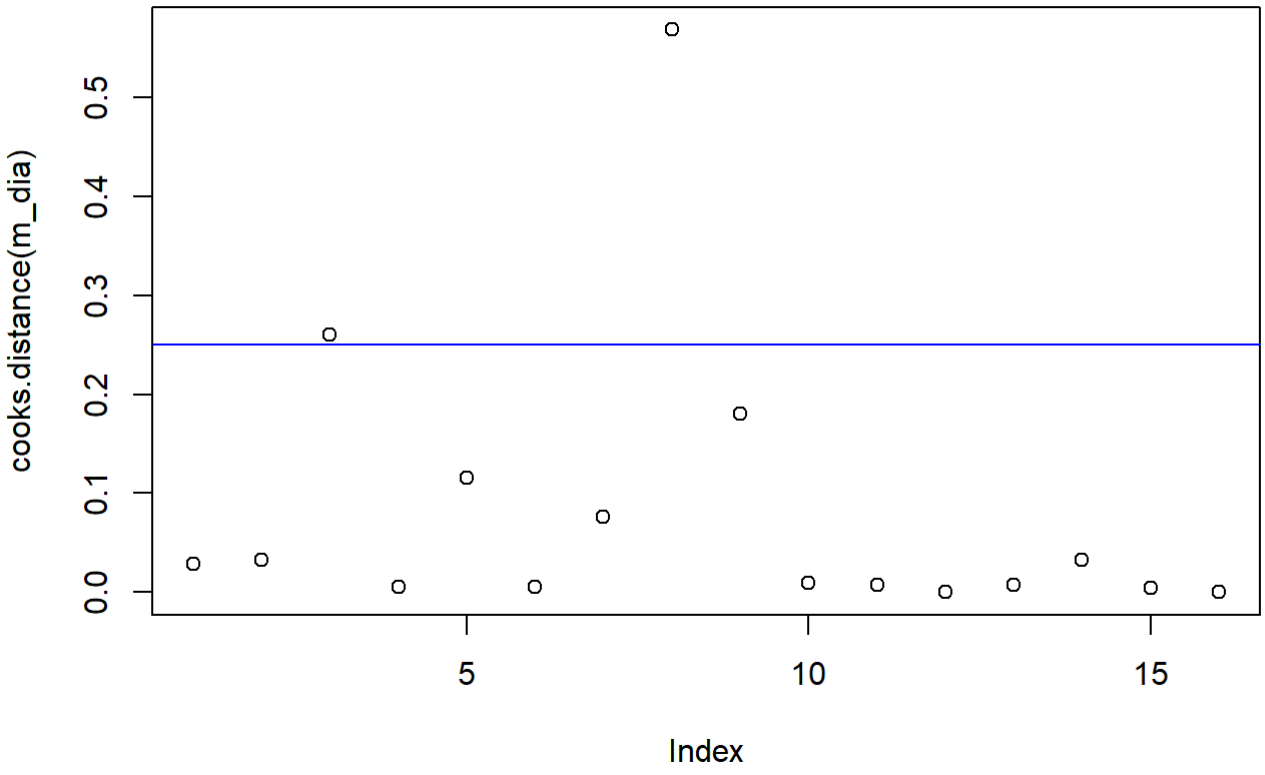
```
## Influence measures of
##   lm.default(formula = dia ~ pill + position + time, data = dia_design) :
##
##     dfb.1_ dfb.pll1 dfb.pst1 dfb.tim1   dffit  cov.r   cook.d  hat inf
## 1  -0.1651   0.1651   0.1651   0.1651 -0.3301 1.6796 0.028868 0.25
## 2   0.1757  -0.1757  -0.1757   0.1757  0.3515 1.6540 0.032589 0.25
## 3  -0.5672  -0.5672   0.5672  -0.5672 -1.1345 0.5668 0.259811 0.25
## 4   0.0714   0.0714  -0.0714  -0.0714  0.1427 1.8470 0.005525 0.25
## 5  -0.3459   0.3459  -0.3459  -0.3459 -0.6919 1.1560 0.115471 0.25
## 6   0.0714  -0.0714   0.0714  -0.0714  0.1427 1.8470 0.005525 0.25
## 7  -0.2750  -0.2750  -0.2750   0.2750 -0.5501 1.3751 0.076229 0.25
## 8   1.0988   1.0988   1.0988   1.0988  2.1977 0.0655 0.568448 0.25   *
## 9   0.4492  -0.4492  -0.4492  -0.4492  0.8984 0.8520 0.180424 0.25
## 10  0.0919  -0.0919  -0.0919   0.0919  0.1838 1.8203 0.009134 0.25
## 11 -0.0816  -0.0816   0.0816  -0.0816 -0.1633 1.8345 0.007217 0.25
## 12 -0.0102  -0.0102   0.0102   0.0102 -0.0203 1.8875 0.000113 0.25
## 13 -0.0816   0.0816  -0.0816  -0.0816 -0.1633 1.8345 0.007217 0.25
## 14 -0.1757   0.1757  -0.1757   0.1757 -0.3515 1.6540 0.032589 0.25
## 15  0.0611   0.0611   0.0611  -0.0611  0.1223 1.8579 0.004060 0.25
## 16 -0.0102  -0.0102  -0.0102  -0.0102 -0.0203 1.8875 0.000113 0.25
```

Only observation 8 was detected as influential for the diastolic model while no influential observations seem to occur in the systolic model. Cook's distance with $D_i > \frac{4}{n}$ criterion then detected observation 2 as influential for systolic pressure and observation 8 and 3 for diastolic pressure. Their plots can be seen below.

```
plot(cooks.distance(m_sys))
abline(h = 1/4, col="blue")
```




```
plot(cooks.distance(m_dia))
abline(h = 1/4, col="blue")
```



By comparing the design with the data as seen below, it can be confirmed that there was no error during data processing since all factor levels correspond with each other.

```
df[c(2,8),]
```

```
##   systolic diastolic coffee pill position squats time water.intake
## 2      124        58     50    0        0      0    1          400
## 8      113        77     50   25        1      0    1            0
```

```
bp_design[c(2,8),]
```

```
##   coffee pill position squats time water.intake
## 2      1   -1       -1     -1    1            1
## 8      1    1        1     -1    1           -1
```

Looking at the data frame summary it can be seen that the diastolic values in both detected observations are close to the recorded maximums. In case of the observation 2 the diastolic pressure is rather low whereas the systolic component remains at a healthy level. This is a state called diastolic hypotension.

```
summary(df)
```

```
##     systolic       diastolic         coffee        pill      position
##  Min.   : 92.0   Min.   :54.00   Min.   : 0   Min.   : 0.0   0:8
##  1st Qu.:110.2   1st Qu.:59.00   1st Qu.: 0   1st Qu.: 0.0   1:8
##  Median :116.5   Median :65.50   Median :25   Median :12.5
##  Mean   :116.6   Mean   :64.56   Mean   :25   Mean   :12.5
##  3rd Qu.:128.5   3rd Qu.:68.50   3rd Qu.:50   3rd Qu.:25.0
##  Max.   :133.0   Max.   :77.00   Max.   :50   Max.   :25.0
##      squats   time   water.intake
##  Min.   : 0   0:8   Min.   :  0
##  1st Qu.: 0   1:8   1st Qu.:  0
##  Median :15         Median :200
##  Mean   :15         Mean   :200
##  3rd Qu.:30         3rd Qu.:400
##  Max.   :30         Max.   :400
```

Observation 2 was also found to be influential for the systolic model based on Cook's distance. Its residual is positive meaning the model prediction was smaller than its real value. One possible reason for this could be insufficient amount of time spent in resting position before taking the measurement when taking multiple measurements during the same time of the day.

Observation 8 influential for the diastolic model belongs to the maximum response value and its model prediction is smaller than the actual value by approximately 9 mm Hg.

However, since there is no proof of some sort of error occurring during data processing or measurement, it is best to keep the observations to avoid artificially lowering the true variability of the data.

Testing for Quadratic Curvature

Linearity is another important assumption for all linear regression models. It can often be checked solely based on visual inspection of the residual plots in designs with multiple levels. This is not the case for two-factor designs as generally any two points can be fitted with a line. It is therefore necessary to add observations measured at different levels than the two given in the factorial design to see if the relationship is indeed linear.

```
cp = data[17:20,]
cp
```

```
##    systolic diastolic coffee pill position squats time water.intake
## 17      126        57     25 12.5        0     15    0          200
## 18      125        66     25 12.5        1     15    0          200
## 19      120        52     25 12.5        0     15    1          200
## 20      128        71     25 12.5        1     15    1          200
```

Observations most frequently used for this purpose are measured at the center, meaning exactly in between the given factor levels, and are hence called center points. Blood pressure data frame includes four center points which are measured for all combinations of the two position and time levels because their character is by definition non-numeric. This data frame was split into two - the first one being the data measured at the standard two levels and the second one being the four center points.

Presence of quadratic curvature can be tested by comparing the means of the response variable calculated over the two-level data and the center points and then comparing them to see if there is a significant difference.

```
sys_cp = mean(cp$systolic)
dia_cp = mean(cp$diastolic)

sys_nocp = mean(df$systolic)
dia_nocp = mean(df$diastolic)

sys_cp - sys_nocp
```

```
## [1] 8.125
```

```
dia_cp - dia_nocp
```

```
## [1] -3.0625
```

Difference in systolic pressure of the two data frames seems to be somewhat high which could potentially indicate curvature. t-test can be used to statistically check the significance of the difference.

```
t.test(df$systolic, cp$systolic)
```

```
##
##  Welch Two Sample t-test
##
## data:  df$systolic and cp$systolic
## t = -2.1989, df = 17.736, p-value = 0.0414
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.8962137  -0.3537863
## sample estimates:
## mean of x mean of y
##   116.625   124.750
```

```
t.test(df$diastolic, cp$diastolic)
```

```
##
##  Welch Two Sample t-test
##
## data:  df$diastolic and cp$diastolic
## t = 0.66816, df = 3.8883, p-value = 0.5416
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.808749 15.933749
## sample estimates:
## mean of x mean of y
##   64.5625   61.5000
```

p-value is greater than the chosen significance level of 0.05 in case of the diastolic response which means that the two group means are not very different. On the other hand p-value of the systolic means comparison was slightly under the given significance level so including a quadratic term into the systolic model might prove to be beneficial.

Summary

systolic model

Variables which were observed to have the highest influence on systolic blood pressure were discovered to be physical activity measured in squats and water intake. There was also a suspicion of heteroscedasticity, which was, however, refuted by Breusch-Pagan test. The model did not seem to break the linearity assumption. Deviation from normality did not seem to be overly significant either.

diastolic model

Diastolic blood pressure was revealed to be the most dependent on variables time and position. Variable pill was also included in the model despite ANOVA revealing it to be insignificant because of its positive influence on the model fit. Inclusion of an only numeric variable seemed to significantly improve the model's R-squared.

Another argument for keeping the pill variable in the model was the knowledge of influence that thyroid hormones have on diastolic blood pressure in hypothyroid patients, see S. Stabouli and Kotsis (2010).

According to the article, insufficient levels of thyroid hormone seem to have an elevating effect on diastolic blood pressure. However, looking at main effect plots in the Visual Analysis section it can be seen that the slope of the pill variable is positive. Based on the information above an opposite slope would be expected.

K. B. Ain and Banks (1993) say that thyroid hormone levels are significantly elevated from baseline levels at 2-9 hours from the time of dose ingestion. This would imply that timing the measurement 20 minutes after taking a pill was insufficient. On the other hand afternoon measurements could probably still register some residue effect of the released hormones on the blood pressure which might explain the reversed slope.

References

Danzi, S., and I. Klein. 2003. "Thyroid Hormone and Blood Pressure Regulation." *Current Hypertension Reports* 5 (6): 513–20.

K. B. Ain, T. M. Shiver, F. Pucino, and S. M. Banks. 1993. "Thyroid Hormone Levels Affected by Time of Blood Sampling in Thyroxine-Treated Patients." *Thyroid* 3 (2): 81–85.

Medicine Committee on Military Nutrition Research, Institute of. 2001. *Caffeine for the Sustainment of Mental Task Performance: Formulations for Military Operations*. National Academies Press.

S. Stabouli, S. Papakatsika, and V. Kotsis. 2010. "Hypothyroidism and Hypertension." *Thyroid* 8 (11): 1559–65.