02441 Applied Statistics and Statistical Software

Exercise 5B - KFM2

The dataset kfm contains measurements of the newborn babies, their mother and milk consumption

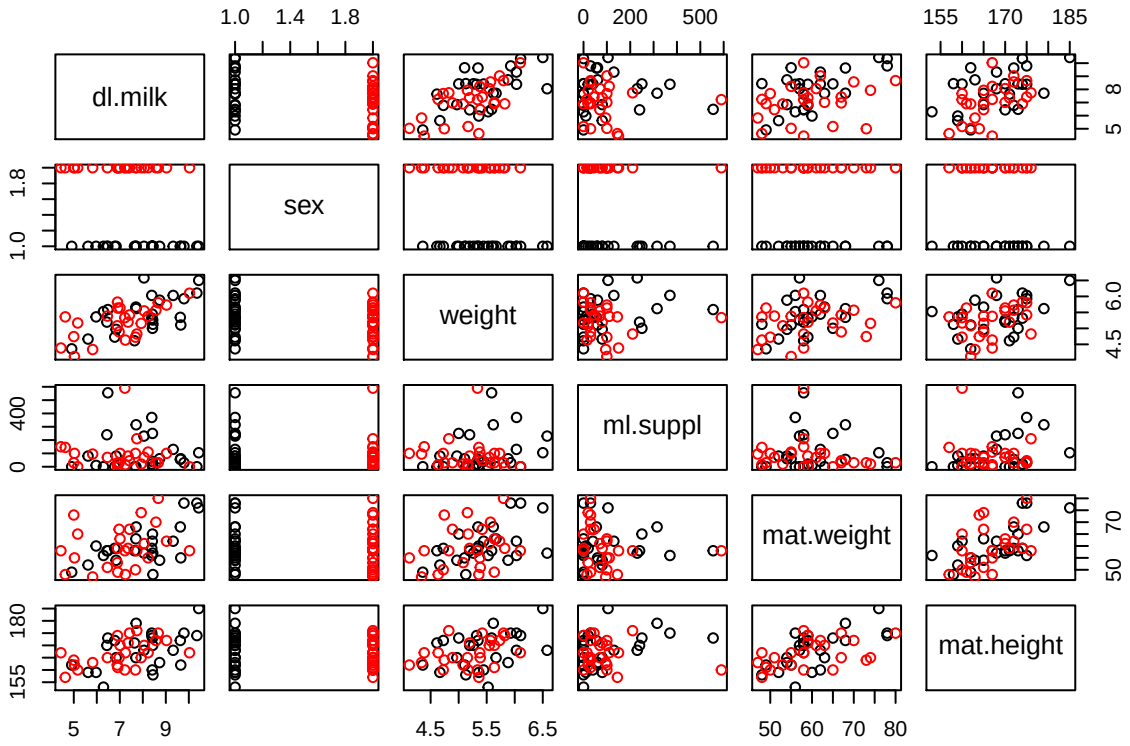
Variable name	Description
dl.milk	amount of breast milk (dl)
sex	gender of body
weight	baby weight (kg)
ml.suppl	amount of milk supplement (ml)
mat.weight	mothers weight (kg)
mat.height	mothers height (cm)

1. The purpose of this assignment is to build a model and to test predictors of the child's weight.

Of special interest is to investigate whether the effect of dl.milk, ml.suppl, mat.weight and mat.height are different for boys and girls. Remember that the stages in model building includes appropriate descriptive statistics, parameter estimation and testing, as well as model validation.

Load and visualize the data

```
kfm <- read.table("kfm.txt", header = TRUE)
# remove no and case
kfm <- kfm[,-c(1,2)]
pairs(kfm, col = kfm$sex)
```



There is multicolinearity in the data. mat.height, mat.weight and dl.milk appear to be correlated as it can be seen from the pairs plot. Proceed with maximal model and reduce stepwise.

Establish linear model

We only have 50 observations, so we have to reason our choice of the maximal model. For the continuous variables, let's start with all main effects and all possible interactions. On top of that we want to include all possible interaction combinations with sex.

```
lm1a <- lm(weight~sex*(dl.milk+ml.suppl+mat.weight+mat.height)^2, kfm)
library(car)
```

```
## Loading required package: carData
```

```
Anova(lm1a)
```

```
## Anova Table (Type II tests)
##
## Response: weight
##                          Sum Sq Df F value   Pr(>F)
## sex                      0.0000  1  0.0000 0.997655
## dl.milk                  2.2220  1 12.5652 0.001403 **
## ml.suppl                 0.6130  1  3.4665 0.073152 .
## mat.weight               0.0236  1  0.1333 0.717813
## mat.height               0.2379  1  1.3451 0.255929
## dl.milk:ml.suppl         0.0554  1  0.3130 0.580290
## dl.milk:mat.weight       0.3460  1  1.9564 0.172875
## dl.milk:mat.height       0.5262  1  2.9754 0.095564 .
## ml.suppl:mat.weight      0.2286  1  1.2928 0.265173
## ml.suppl:mat.height      0.3722  1  2.1045 0.157976
```

```
## mat.weight:mat.height     0.0806  1  0.4558 0.505139
## sex:dl.milk               0.0002  1  0.0009 0.975710
## sex:ml.suppl              1.1029  1  6.2364 0.018667 *
## sex:mat.weight            0.4361  1  2.4662 0.127553
## sex:mat.height            0.3115  1  1.7612 0.195197
## sex:dl.milk:ml.suppl      0.0537  1  0.3035 0.586048
## sex:dl.milk:mat.weight    0.0844  1  0.4774 0.495290
## sex:dl.milk:mat.height    0.0754  1  0.4265 0.519040
## sex:ml.suppl:mat.weight   0.2291  1  1.2957 0.264640
## sex:ml.suppl:mat.height   0.2053  1  1.1612 0.290421
## sex:mat.weight:mat.height 0.0518  1  0.2927 0.592779
## Residuals                 4.9516 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Let's reduce our model automatically using step
lm1b <- step(lm1a, k = 3.8)
```

```
## Start:  AIC=-32.02
## weight ~ sex * (dl.milk + ml.suppl + mat.weight + mat.height)^2
##
##                             Df Sum of Sq    RSS     AIC
## - sex:mat.weight:mat.height  1  0.051760 5.0033 -35.296
## - sex:dl.milk:ml.suppl       1  0.053676 5.0052 -35.277
## - sex:dl.milk:mat.height     1  0.075422 5.0270 -35.060
## - sex:dl.milk:mat.weight     1  0.084424 5.0360 -34.971
## - sex:ml.suppl:mat.height    1  0.205341 5.1569 -33.784
## - sex:ml.suppl:mat.weight    1  0.229140 5.1807 -33.554
## <none>                                   4.9516 -32.016
##
## Step:  AIC=-35.3
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:ml.suppl + dl.milk:mat.weight + dl.milk:mat.height +
##     ml.suppl:mat.weight + ml.suppl:mat.height + mat.weight:mat.height +
##     sex:dl.milk + sex:ml.suppl + sex:mat.weight + sex:mat.height +
##     sex:dl.milk:ml.suppl + sex:dl.milk:mat.weight + sex:dl.milk:mat.height +
##     sex:ml.suppl:mat.weight + sex:ml.suppl:mat.height
##
##                             Df Sum of Sq    RSS     AIC
## - sex:dl.milk:ml.suppl       1  0.036973 5.0403 -38.728
## - sex:dl.milk:mat.weight     1  0.039220 5.0425 -38.706
## - sex:dl.milk:mat.height     1  0.040866 5.0442 -38.689
## - mat.weight:mat.height      1  0.080601 5.0839 -38.297
## - sex:ml.suppl:mat.weight    1  0.198778 5.2021 -37.148
## - sex:ml.suppl:mat.height    1  0.211520 5.2148 -37.026
## <none>                                   5.0033 -35.296
##
## Step:  AIC=-38.73
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:ml.suppl + dl.milk:mat.weight + dl.milk:mat.height +
##     ml.suppl:mat.weight + ml.suppl:mat.height + mat.weight:mat.height +
##     sex:dl.milk + sex:ml.suppl + sex:mat.weight + sex:mat.height +
##     sex:dl.milk:mat.weight + sex:dl.milk:mat.height + sex:ml.suppl:mat.weight +
##     sex:ml.suppl:mat.height
##
```

```
##                            Df Sum of Sq    RSS     AIC
## - sex:dl.milk:mat.height    1  0.024586 5.0649 -42.285
## - dl.milk:ml.suppl          1  0.033001 5.0733 -42.202
## - mat.weight:mat.height     1  0.061089 5.1014 -41.926
## - sex:dl.milk:mat.weight    1  0.085079 5.1254 -41.691
## - sex:ml.suppl:mat.weight   1  0.181019 5.2213 -40.764
## - sex:ml.suppl:mat.height   1  0.241196 5.2815 -40.191
## <none>                                  5.0403 -38.728
##
## Step:  AIC=-42.28
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:ml.suppl + dl.milk:mat.weight + dl.milk:mat.height +
##     ml.suppl:mat.weight + ml.suppl:mat.height + mat.weight:mat.height +
##     sex:dl.milk + sex:ml.suppl + sex:mat.weight + sex:mat.height +
##     sex:dl.milk:mat.weight + sex:ml.suppl:mat.weight + sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - dl.milk:ml.suppl          1   0.02425 5.0891 -45.846
## - mat.weight:mat.height     1   0.05527 5.1201 -45.542
## - sex:ml.suppl:mat.weight   1   0.16050 5.2254 -44.525
## - sex:dl.milk:mat.weight    1   0.17877 5.2436 -44.350
## - sex:ml.suppl:mat.height   1   0.21939 5.2843 -43.964
## <none>                                  5.0649 -42.285
## - dl.milk:mat.height        1   0.68597 5.7508 -39.734
##
## Step:  AIC=-45.85
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.weight +
##     ml.suppl:mat.height + mat.weight:mat.height + sex:dl.milk +
##     sex:ml.suppl + sex:mat.weight + sex:mat.height + sex:dl.milk:mat.weight +
##     sex:ml.suppl:mat.weight + sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - mat.weight:mat.height     1   0.03979 5.1289 -49.257
## - sex:ml.suppl:mat.weight   1   0.15174 5.2409 -48.177
## - sex:dl.milk:mat.weight    1   0.16710 5.2562 -48.031
## - sex:ml.suppl:mat.height   1   0.23236 5.3215 -47.414
## <none>                                  5.0891 -45.846
## - dl.milk:mat.height        1   0.99777 6.0869 -40.694
##
## Step:  AIC=-49.26
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.weight +
##     ml.suppl:mat.height + sex:dl.milk + sex:ml.suppl + sex:mat.weight +
##     sex:mat.height + sex:dl.milk:mat.weight + sex:ml.suppl:mat.weight +
##     sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - sex:ml.suppl:mat.weight   1   0.12504 5.2539 -51.852
## - sex:dl.milk:mat.weight    1   0.15119 5.2801 -51.604
## - sex:ml.suppl:mat.height   1   0.30563 5.4345 -50.162
## <none>                                  5.1289 -49.257
## - dl.milk:mat.height        1   1.25802 6.3869 -42.089
##
```

```
## Step:  AIC=-51.85
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.weight +
##     ml.suppl:mat.height + sex:dl.milk + sex:ml.suppl + sex:mat.weight +
##     sex:mat.height + sex:dl.milk:mat.weight + sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - sex:dl.milk:mat.weight    1   0.06413 5.3181 -55.046
## - ml.suppl:mat.weight       1   0.18221 5.4362 -53.948
## - sex:ml.suppl:mat.height   1   0.28396 5.5379 -53.020
## <none>                                  5.2539 -51.852
## - dl.milk:mat.height        1   1.16463 6.4186 -45.641
##
## Step:  AIC=-55.05
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.weight +
##     ml.suppl:mat.height + sex:dl.milk + sex:ml.suppl + sex:mat.weight +
##     sex:mat.height + sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - sex:dl.milk               1   0.00006 5.3181 -58.845
## - ml.suppl:mat.weight       1   0.18799 5.5061 -57.109
## - sex:mat.weight            1   0.33593 5.6540 -55.783
## - sex:ml.suppl:mat.height   1   0.35511 5.6732 -55.614
## - dl.milk:mat.weight        1   0.39401 5.7121 -55.272
## <none>                                  5.3181 -55.046
## - dl.milk:mat.height        1   1.10500 6.4231 -49.406
##
## Step:  AIC=-58.84
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.weight +
##     ml.suppl:mat.height + sex:ml.suppl + sex:mat.weight + sex:mat.height +
##     sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - ml.suppl:mat.weight       1   0.19018 5.5083 -60.888
## - sex:ml.suppl:mat.height   1   0.36126 5.6794 -59.359
## - sex:mat.weight            1   0.36503 5.6832 -59.326
## - dl.milk:mat.weight        1   0.40465 5.7228 -58.978
## <none>                                  5.3181 -58.845
## - dl.milk:mat.height        1   1.10656 6.4247 -53.194
##
## Step:  AIC=-60.89
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.height +
##     sex:ml.suppl + sex:mat.weight + sex:mat.height + sex:ml.suppl:mat.height
##
##                            Df Sum of Sq    RSS     AIC
## - sex:ml.suppl:mat.height   1   0.21169 5.7200 -62.803
## - sex:mat.weight            1   0.29964 5.8080 -62.040
## - dl.milk:mat.weight        1   0.30728 5.8156 -61.974
## <none>                                  5.5083 -60.888
## - dl.milk:mat.height        1   1.11442 6.6227 -55.476
##
```

```
## Step:  AIC=-62.8
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.height +
##     sex:ml.suppl + sex:mat.weight + sex:mat.height
##
##                        Df Sum of Sq    RSS     AIC
## - sex:mat.weight        1   0.24703 5.9670 -64.489
## - dl.milk:mat.weight    1   0.28719 6.0072 -64.153
## <none>                              5.7200 -62.803
## - ml.suppl:mat.height   1   0.57127 6.2913 -61.843
## - sex:mat.height        1   0.76617 6.4862 -60.317
## - dl.milk:mat.height    1   1.17507 6.8951 -57.261
## - sex:ml.suppl          1   1.37671 7.0967 -55.820
##
## Step:  AIC=-64.49
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.weight + dl.milk:mat.height + ml.suppl:mat.height +
##     sex:ml.suppl + sex:mat.height
##
##                        Df Sum of Sq    RSS     AIC
## - dl.milk:mat.weight    1   0.11890 6.0859 -67.302
## - ml.suppl:mat.height   1   0.41541 6.3824 -64.924
## <none>                              5.9670 -64.489
## - sex:mat.height        1   0.51958 6.4866 -64.114
## - dl.milk:mat.height    1   1.06510 7.0321 -60.077
## - sex:ml.suppl          1   1.15849 7.1255 -59.417
##
## Step:  AIC=-67.3
## weight ~ sex + dl.milk + ml.suppl + mat.weight + mat.height +
##     dl.milk:mat.height + ml.suppl:mat.height + sex:ml.suppl +
##     sex:mat.height
##
##                        Df Sum of Sq    RSS     AIC
## - mat.weight            1   0.06930 6.1552 -70.536
## - ml.suppl:mat.height   1   0.33587 6.4218 -68.416
## <none>                              6.0859 -67.302
## - sex:mat.height        1   0.48967 6.5756 -67.233
## - sex:ml.suppl          1   1.09623 7.1822 -62.821
## - dl.milk:mat.height    1   1.12498 7.2109 -62.621
##
## Step:  AIC=-70.54
## weight ~ sex + dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     ml.suppl:mat.height + sex:ml.suppl + sex:mat.height
##
##                        Df Sum of Sq    RSS     AIC
## - ml.suppl:mat.height   1   0.39262 6.5479 -71.244
## <none>                              6.1552 -70.536
## - sex:mat.height        1   0.54223 6.6975 -70.115
## - sex:ml.suppl          1   1.18323 7.3385 -65.545
## - dl.milk:mat.height    1   1.44342 7.5987 -63.803
##
## Step:  AIC=-71.24
## weight ~ sex + dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     sex:ml.suppl + sex:mat.height
```

```
##
##                       Df Sum of Sq    RSS     AIC
## - sex:mat.height       1   0.34998 6.8978 -72.441
## <none>                              6.5479 -71.244
## - sex:ml.suppl         1   0.91485 7.4627 -68.505
## - dl.milk:mat.height   1   1.25279 7.8006 -66.291
##
## Step:  AIC=-72.44
## weight ~ sex + dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     sex:ml.suppl
##
##                       Df Sum of Sq    RSS     AIC
## <none>                              6.8978 -72.441
## - sex:ml.suppl         1   0.78593 7.6838 -70.846
## - dl.milk:mat.height   1   0.91353 7.8114 -70.022
```

```
drop1(lm1b, test = "F")
```

```
## Single term deletions
##
## Model:
## weight ~ sex + dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     sex:ml.suppl
##                     Df Sum of Sq    RSS     AIC F value  Pr(>F)
## <none>                           6.8978 -85.041
## dl.milk:mat.height   1   0.91353 7.8114 -80.822  5.6948 0.02149 *
## sex:ml.suppl         1   0.78593 7.6838 -81.646  4.8993 0.03222 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(lm1b)
```

```
## Anova Table (Type II tests)
##
## Response: weight
##                     Sum Sq Df F value    Pr(>F)
## sex                 0.0005  1  0.0029   0.95760
## dl.milk             4.1150  1 25.6521 8.185e-06 ***
## ml.suppl            0.4928  1  3.0719   0.08678 .
## mat.height          0.0157  1  0.0981   0.75559
## dl.milk:mat.height 0.9135  1  5.6948   0.02149 *
## sex:ml.suppl        0.7859  1  4.8993   0.03222 *
## Residuals           6.8978 43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

One could choose to stop here.

However, one could remove sex eventhough it is part of an interaction. In this case the interaction is with a continuous predictor (ml.suppl) so the term sex is for a difference in intercept while sex:ml.suppl is for differences in slopes for boys and girls. It does make sence to have lines with different slopes and same intercept.

As in the case above it may sometimes make sence to remove a factor that is part of a higher order interaction. But only when the higher order interaction doesn't include other factors.

```
drop1(lm1c <- update(lm1b, ~.-sex), test = "F")
```

```
## Single term deletions
##
## Model:
## weight ~ dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     sex:ml.suppl
##                    Df Sum of Sq    RSS     AIC F value  Pr(>F)
## <none>                          7.2295 -84.693
## dl.milk:mat.height  1   0.73422 7.9637 -81.856  4.4686 0.04023 *
## ml.suppl:sex        1   0.45472 7.6842 -83.643  2.7675 0.10330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm1d <- update(lm1c, ~.-ml.suppl:sex), test = "F")
```

```
## Single term deletions
##
## Model:
## weight ~ dl.milk + ml.suppl + mat.height + dl.milk:mat.height
##                    Df Sum of Sq    RSS     AIC F value  Pr(>F)
## <none>                          7.6842 -83.643
## ml.suppl            1   0.49665 8.1809 -82.511  2.9085 0.09501 .
## dl.milk:mat.height  1   0.64808 8.3323 -81.594  3.7953 0.05765 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm1e <- update(lm1d, ~.-ml.suppl), test = "F")
```

```
## Single term deletions
##
## Model:
## weight ~ dl.milk + mat.height + dl.milk:mat.height
##                    Df Sum of Sq    RSS     AIC F value  Pr(>F)
## <none>                          8.1809 -82.511
## dl.milk:mat.height  1   0.50585 8.6867 -81.511  2.8443 0.09847 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm1f <- update(lm1e, ~.-dl.milk:mat.height), test = "F")
```

```
## Single term deletions
##
## Model:
## weight ~ dl.milk + mat.height
##            Df Sum of Sq     RSS     AIC F value    Pr(>F)
## <none>                   8.6867 -81.511
## dl.milk     1    3.8409 12.5277 -65.204 20.7817 3.675e-05 ***
## mat.height  1    0.0849  8.7716 -83.025  0.4594    0.5012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm1g <- update(lm1f, ~.-mat.height), test = "F")
```

```
## Single term deletions
##
## Model:
```

```
## weight ~ dl.milk
##         Df Sum of Sq     RSS     AIC F value    Pr(>F)
## <none>                8.7716 -83.025
## dl.milk  1    5.9595 14.7312 -59.103  32.612 6.915e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(lm1g)
```

```
## Anova Table (Type II tests)
##
## Response: weight
##           Sum Sq Df F value    Pr(>F)
## dl.milk   5.9595  1  32.612 6.915e-07 ***
## Residuals 8.7716 48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm1b,lm1g)
```

```
## Analysis of Variance Table
##
## Model 1: weight ~ sex + dl.milk + ml.suppl + mat.height + dl.milk:mat.height +
##     sex:ml.suppl
## Model 2: weight ~ dl.milk
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     43 6.8978
## 2     48 8.7716 -5   -1.8738 2.3362 0.05806 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because `lm1g` and `lm1b` are nested models we can compare the fit using the `anova` function (from base R). The p-value is greater than 0.05. That means that the fit has not significantly improved for the complex model, hence, we keep the simpler model `lm1g`.

The result shows that sex does not influence the weight of new-born babies.

Side note: Once could have also tried to square root transform ml.suppl and repeated the model selction process.