

Chapter 8

```
library(tidyverse)

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():      dplyr, stats

library(AzureML)
ws <- workspace()

mlbook_red <- tbl_df(download.datasets(
  dataset = ws,
  name     = "mlbook2_r.csv"))
# what are the available variables
names(mlbook_red)

## [1] "schoolnr"      "pupilNR_new"  "langPOST"     "ses"          "IQ_verb"
## [6] "sex"           "Minority"     "denomina"     "sch_ses"      "sch_iqv"
## [11] "sch_min"

# Attach library
library(nlme)

##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
## collapse

Table 8.1 Model 1

mod81 <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
             random=~IQ_verb|schoolnr, data = mlbook_red, method="ML")
summary(mod81)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24510.82 24585.6 -12243.41
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev    Corr
## (Intercept) 2.8846515 (Intr)
## IQ_verb      0.3821388 -0.815
## Residual     5.9995464
##
```

```

## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
##               Value Std.Error   DF   t-value p-value
## (Intercept)   40.42614 0.26534574 3543 152.35272  0.0000
## IQ_verb        2.24874 0.06166254 3543  36.46845  0.0000
## ses            0.17142 0.01145165 3543  14.96874  0.0000
## sex            2.40671 0.20164298 3543  11.93548  0.0000
## sch_iqv        0.76889 0.29375384  207   2.61746  0.0095
## sch_ses       -0.09280 0.04201643  207  -2.20861  0.0283
## IQ_verb:ses   -0.01957 0.00480006 3543  -4.07781  0.0000
## sch_iqv:sch_ses -0.10523 0.03299836  207  -3.18909  0.0016
## Correlation:
##               (Intr) IQ_vrb ses    sex    sch_qv sch_ss IQ_vr:
## IQ_verb        -0.292
## ses             0.017 -0.254
## sex            -0.374  0.026 -0.022
## sch_iqv        -0.087 -0.166  0.061  0.004
## sch_ses         0.051  0.053 -0.265 -0.009 -0.497
## IQ_verb:ses    -0.076  0.076 -0.122 -0.038 -0.015 -0.141
## sch_iqv:sch_ses -0.361 -0.009  0.023  0.036  0.183 -0.017 -0.128
##
## Standardized Within-Group Residuals:
##               Min          Q1          Med          Q3          Max
## -4.14643392 -0.63402245  0.07982323  0.68121238  3.04626819
##
## Number of Observations: 3758
## Number of Groups: 211

Model 2

mod82 <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
             random=~IQ_verb|schoolnr, weights=varIdent(form=~1|sex),
             data = mlbook_red, method="ML")

summary(mod82)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24508.22 24589.23 -12241.11
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev    Corr
## (Intercept) 2.8746428 (Intr)
## IQ_verb      0.3826536 -0.824
## Residual     6.1523224
##
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | sex
## Parameter estimates:
##              0          1
## 1.0000000 0.9488492
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
##               Value Std.Error   DF   t-value p-value

```

```
## (Intercept)      40.43539 0.26654057 3543 151.70445 0.0000
## IQ_verb          2.24460 0.06166362 3543  36.40077 0.0000
## ses              0.17067 0.01143510 3543  14.92501 0.0000
## sex              2.40359 0.20140685 3543  11.93401 0.0000
## sch_iqv          0.74946 0.29262341  207   2.56117 0.0111
## sch_ses          -0.09067 0.04185758  207  -2.16623 0.0314
## IQ_verb:ses      -0.01948 0.00479239 3543  -4.06451 0.0000
## sch_iqv:sch_ses -0.10717 0.03295990  207  -3.25166 0.0013
## Correlation:
##              (Intr) IQ_vrb ses      sex      sch_qv sch_ss IQ_vr:
## IQ_verb      -0.293
## ses           0.017 -0.255
## sex          -0.391  0.026 -0.022
## sch_iqv      -0.085 -0.166  0.061  0.004
## sch_ses       0.051  0.053 -0.266 -0.009 -0.496
## IQ_verb:ses  -0.076  0.076 -0.122 -0.038 -0.015 -0.143
## sch_iqv:sch_ses -0.358 -0.009  0.024  0.037  0.179 -0.017 -0.128
##
## Standardized Within-Group Residuals:
##              Min          Q1          Med          Q3          Max
## -4.04381796 -0.62822093  0.07918544  0.68185125  2.97744655
##
## Number of Observations: 3758
## Number of Groups: 211
```

nlme offers a number of variance functions that do not include the linear variance function used in Model 3. These are treated in Pinheiro & Bates (2004), Section 5.2. It turns out that the variance functions offered by nlme, when used in the default way, represent this data set slightly less well than a linear variance function.

Therefore we work towards the results in Table 8.2 by also using the result obtained by another program (MLwiN), namely, that the linear variance function for level 1 was estimated as

```
IQ_verb <- mlbook_red$IQ_verb
#36.382 - 2*1.689*IQ_verb
```

We define the transformed variable

```
IQ_trans <- 36.382 - 2*1.689*mlbook_red$IQ_verb
```

We can use this variance function (giving it a freely estimated multiplicative parameter, which indeed will be estimated very close to 1) by using varFixed:

```
mod83 <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
             random=~ IQ_verb|schoolnr, weights=varFixed(~IQ_trans),
             data = mlbook_red, method="ML")
summary(mod83)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24454.22 24529 -12215.11
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev      Corr
## (Intercept) 2.8649245 (Intr)
## IQ_verb      0.3288457 -0.778
```

```

## Residual      1.0000001
##
## Variance function:
## Structure: fixed weights
## Formula: ~IQ_trans
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
##
##          Value Std.Error   DF   t-value p-value
## (Intercept)  40.51010 0.26314161 3543 153.94791 0.0000
## IQ_verb      2.20041 0.05820335 3543  37.80563 0.0000
## ses          0.17464 0.01145189 3543  15.24987 0.0000
## sex          2.31120 0.19832116 3543  11.65380 0.0000
## sch_iqv      0.68521 0.28960748  207   2.36601 0.0189
## sch_ses     -0.08687 0.04129195  207  -2.10385 0.0366
## IQ_verb:ses  -0.02232 0.00457096 3543  -4.88217 0.0000
## sch_iqv:sch_ses -0.10693 0.03258171  207  -3.28180 0.0012
## Correlation:
##          (Intr) IQ_vrb ses      sex      sch_qv sch_ss IQ_vr:
## IQ_verb      -0.308
## ses           0.025 -0.245
## sex          -0.371  0.037 -0.024
## sch_iqv      -0.078 -0.185  0.062  0.002
## sch_ses       0.052  0.062 -0.247 -0.011 -0.501
## IQ_verb:ses  -0.063  0.004 -0.242 -0.036 -0.014 -0.123
## sch_iqv:sch_ses -0.355 -0.013  0.044  0.034  0.172 -0.025 -0.152
##
## Standardized Within-Group Residuals:
##          Min          Q1          Med          Q3          Max
## -4.42011752 -0.63248539  0.09102577  0.69345000  2.86020972
##
## Number of Observations: 3758
## Number of Groups: 211

Another way is by using varPower

mod83a <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
              random=~ IQ_verb|schoolnr, weights=varPower(form=~IQ_trans),
              data = mlbook_red, method="ML")

summary(mod83a)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##          AIC          BIC      logLik
## 24456.2 24537.21 -12215.1
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev      Corr
## (Intercept) 2.8654184 (Intr)
## IQ_verb      0.3287286 -0.781
## Residual     1.0340154
##
## Variance function:
## Structure: Power of variance covariate
## Formula: ~IQ_trans

```

```

## Parameter estimates:
## power
## 0.4906359
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
##
## Value Std.Error DF t-value p-value
## (Intercept) 40.50824 0.26317705 3543 153.92011 0.0000
## IQ_verb 2.20136 0.05822917 3543 37.80505 0.0000
## ses 0.17456 0.01145027 3543 15.24527 0.0000
## sex 2.31319 0.19838936 3543 11.65987 0.0000
## sch_iqv 0.68610 0.28964222 207 2.36879 0.0188
## sch_ses -0.08693 0.04130328 207 -2.10478 0.0365
## IQ_verb:ses -0.02227 0.00457406 3543 -4.86948 0.0000
## sch_iqv:sch_ses -0.10689 0.03258255 207 -3.28074 0.0012
## Correlation:
## (Intr) IQ_vrb ses sex sch_qv sch_ss IQ_vr:
## IQ_verb -0.308
## ses 0.024 -0.245
## sex -0.371 0.036 -0.024
## sch_iqv -0.078 -0.185 0.062 0.002
## sch_ses 0.052 0.062 -0.247 -0.011 -0.501
## IQ_verb:ses -0.063 0.006 -0.240 -0.036 -0.014 -0.123
## sch_iqv:sch_ses -0.355 -0.013 0.044 0.034 0.172 -0.025 -0.152
##
## Standardized Within-Group Residuals:
## Min Q1 Med Q3 Max
## -4.4049968 -0.6328550 0.0914132 0.6934014 2.8579850
##
## Number of Observations: 3758
## Number of Groups: 211

and a further additional parameter is allowed by varConstPower

mod83b <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
              random=~ IQ_verb|schoolnr,
              weights=varConstPower(form=~IQ_trans),
              data = mlbook_red, method="ML")

summary(mod83b)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
## AIC BIC logLik
## 24458.2 24545.44 -12215.1
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
## StdDev Corr
## (Intercept) 2.8654106 (Intr)
## IQ_verb 0.3287414 -0.781
## Residual 1.0332895
##
## Variance function:
## Structure: Constant plus power of variance covariate
## Formula: ~IQ_trans
## Parameter estimates:

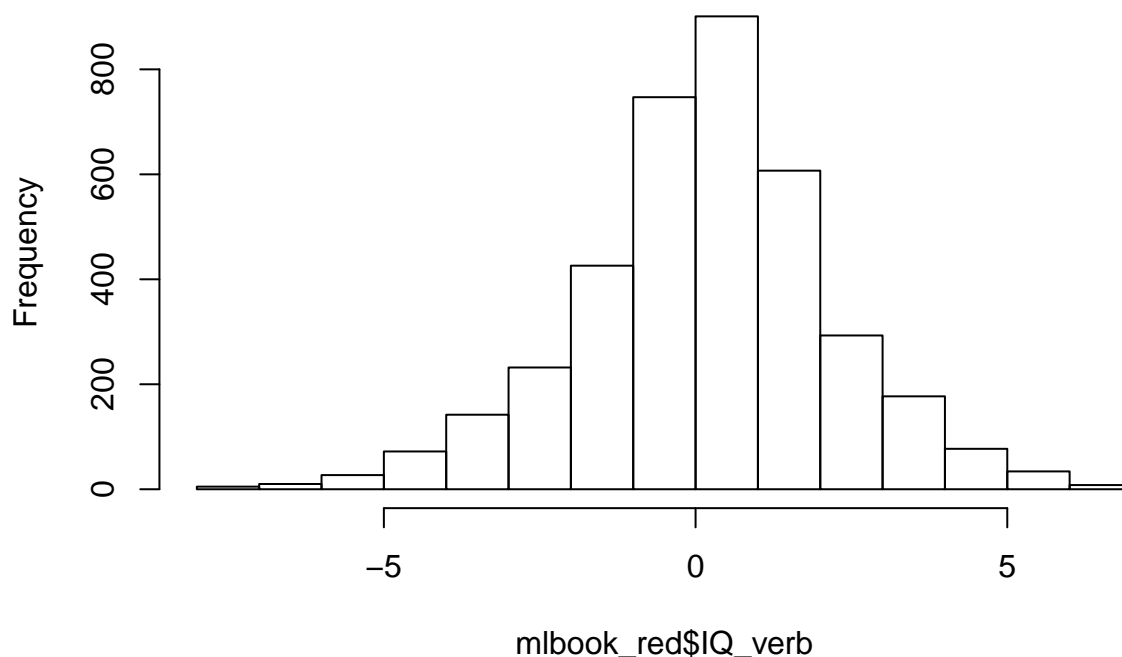
```

```
##          const          power
## 0.001440342 0.490762235
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
##              Value Std.Error   DF   t-value p-value
## (Intercept)   40.50824 0.26317664 3543 153.92035  0.0000
## IQ_verb        2.20136 0.05822974 3543  37.80475  0.0000
## ses            0.17456 0.01145027 3543  15.24526  0.0000
## sex            2.31319 0.19838940 3543  11.65986  0.0000
## sch_iqv        0.68611 0.28964212  207   2.36882  0.0188
## sch_ses       -0.08693 0.04130322  207  -2.10479  0.0365
## IQ_verb:ses   -0.02227 0.00457409 3543  -4.86942  0.0000
## sch_iqv:sch_ses -0.10689 0.03258257  207  -3.28073  0.0012
## Correlation:
##              (Intr) IQ_vrb ses      sex      sch_qv sch_ss IQ_vr:
## IQ_verb       -0.308
## ses           0.024 -0.245
## sex          -0.371  0.036 -0.024
## sch_iqv       -0.078 -0.185  0.062  0.002
## sch_ses        0.052  0.062 -0.247 -0.011 -0.501
## IQ_verb:ses   -0.063  0.006 -0.240 -0.036 -0.014 -0.123
## sch_iqv:sch_ses -0.355 -0.013  0.044  0.034  0.172 -0.025 -0.152
##
## Standardized Within-Group Residuals:
##          Min          Q1          Med          Q3          Max
## -4.4049404 -0.6328478  0.0914108  0.6934011  2.8579901
##
## Number of Observations: 3758
## Number of Groups: 211
```

A more satisfactory approach that does not assume previous knowledge, is to use a categorized version of IQ_verb: First, look at the distribution to find reasonable category bounds

```
hist(mlbook_red$IQ_verb)
```

Histogram of mlbook_red\$IQ_verb



```
IQ_cat <- cut(mlbook_red$IQ_verb, c(-10,-2,-1,0,1,2,10))
table(IQ_cat)
```

```
## IQ_cat
## (-10,-2] (-2,-1] (-1,0] (0,1] (1,2] (2,10]
##      488      426      747      901      607      589
```

Now we fit a model where the variance depends on categorized IQ

```
mod83c <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses,
              random=~ IQ_verb|schoolnr,
              weights=varIdent(form=~1|IQ_cat),
              data = mlbook_red, method="ML")
summary(mod83c)
```

```
## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24460.44 24566.38 -12213.22
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev    Corr
## (Intercept) 2.8564633 (Intr)
## IQ_verb      0.3691566 -0.718
## Residual     5.0157347
##
```

```
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | IQ_cat
## Parameter estimates:
## (2,10] (-10,-2] (-1,0] (1,2] (-2,-1] (0,1]
## 1.000000 1.275106 1.298808 1.094673 1.326459 1.185924
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses
## Value Std.Error DF t-value p-value
## (Intercept) 40.50986 0.26298405 3543 154.03921 0.0000
## IQ_verb 2.21973 0.05941152 3543 37.36191 0.0000
## ses 0.17266 0.01145585 3543 15.07173 0.0000
## sex 2.34700 0.19833644 3543 11.83344 0.0000
## sch_iqv 0.73007 0.29035690 207 2.51438 0.0127
## sch_ses -0.08754 0.04130055 207 -2.11966 0.0352
## IQ_verb:s ses -0.02088 0.00460322 3543 -4.53503 0.0000
## sch_iqv:sch_ses -0.10202 0.03269125 207 -3.12085 0.0021
## Correlation:
## (Intr) IQ_vrb ses sex sch_qv sch_ss IQ_vr:
## IQ_verb -0.309
## ses 0.026 -0.256
## sex -0.372 0.032 -0.021
## sch_iqv -0.082 -0.173 0.063 0.003
## sch_ses 0.052 0.060 -0.251 -0.011 -0.500
## IQ_verb:s ses -0.075 0.058 -0.238 -0.038 -0.009 -0.118
## sch_iqv:sch_ses -0.359 -0.006 0.040 0.037 0.178 -0.022 -0.134
##
## Standardized Within-Group Residuals:
## Min Q1 Med Q3 Max
## -4.21533798 -0.64363434 0.08025082 0.69012048 2.96378961
##
## Number of Observations: 3758
## Number of Groups: 211
```

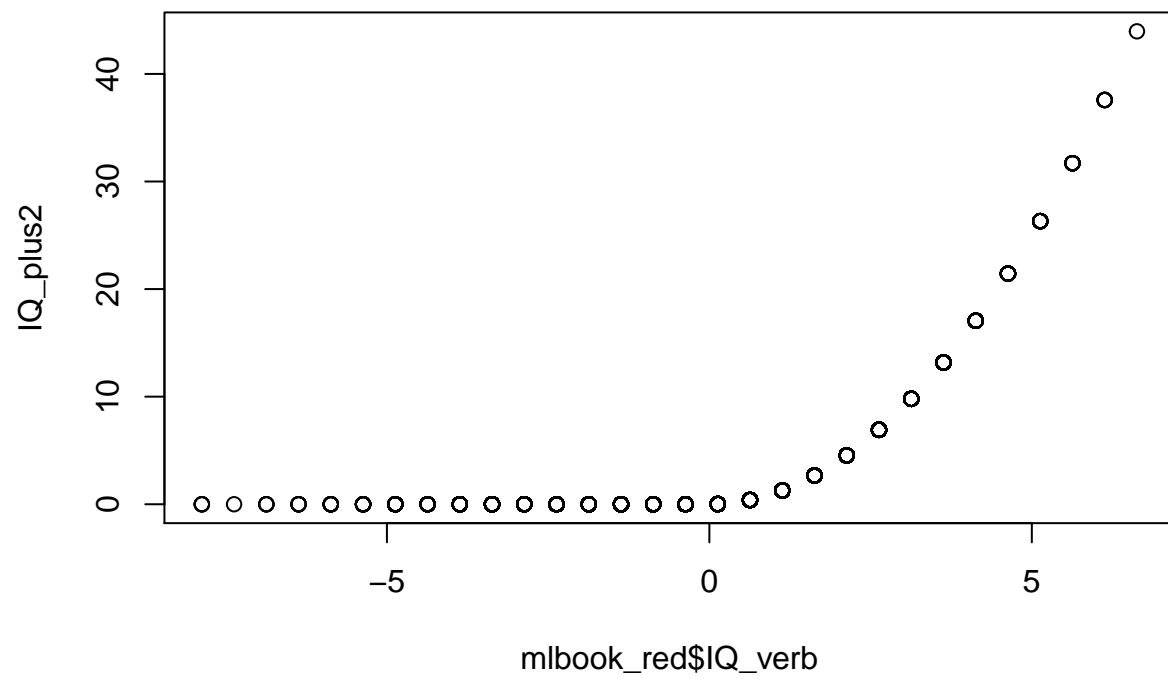
Parameter const is estimated close to 0 and power close to 0.5, with the log-likelihood -12215.1; this means that effectively the same model results as Model 3 in Table 8.2.

For Model 4 in Table 8.2, we first compute the transformed variables.

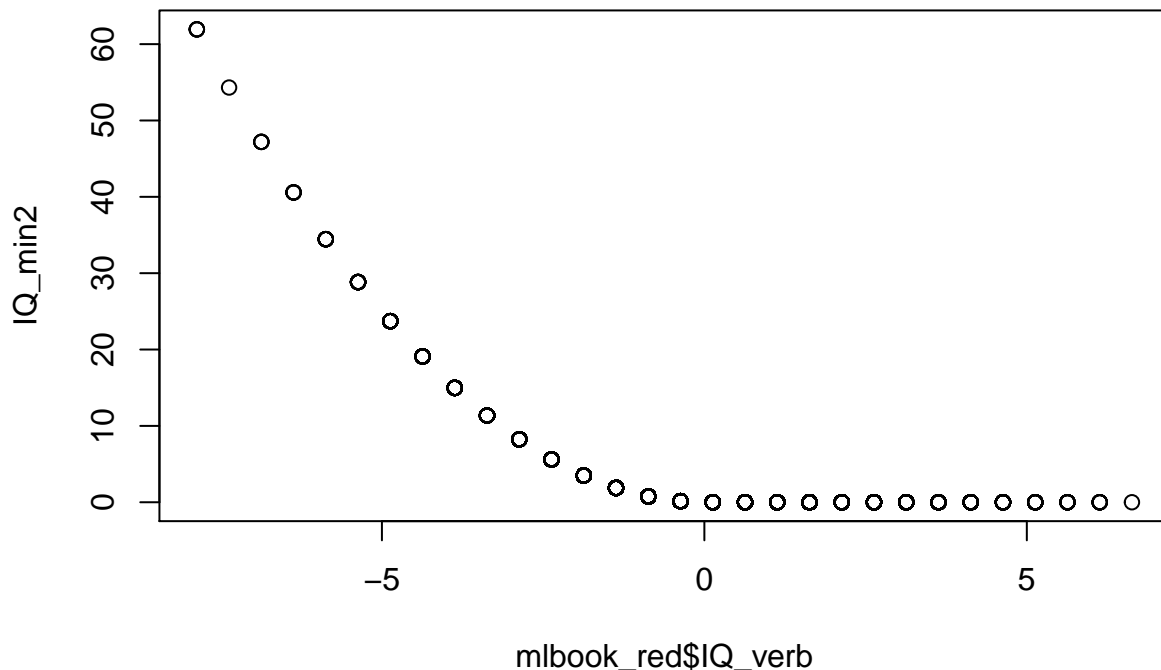
```
IQ_plus2 <- with(mlbook_red, ifelse(IQ_verb > 0, IQ_verb^2, 0))
IQ_min2 <- with(mlbook_red, ifelse(IQ_verb < 0, IQ_verb^2, 0))
```

Have a look at whether this does what we wanted:

```
plot(mlbook_red$IQ_verb, IQ_plus2)
```

```
plot(mlbook_red$IQ_verb, IQ_min2)
```



Again working toward the known answer, we transform `IQ_verb` to

```
IQ_trans2 <- 36.139 - 2*1.769*mlbook_red$IQ_verb
```

Estimating with `varFixed` turns out to require some higher control settings for the algorithm:

```
mod84 <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses +
              IQ_plus2 + IQ_min2,
              random=~ IQ_verb|schoolnr, weights=varFixed(~IQ_trans2),
              control=lmeControl(maxIter=400,msMaxIter=400,msMaxEval=400,
                                  niterEM=400,nlmStepMax=400,msTol=1E-9,msVerbose=TRUE),
              data = mlbook_red, method="ML")
```

```
## 0: 22490.732: 1.61187 2.88586 39.0130
## 1: 22341.734: 0.827962 2.26520 39.0295
## 2: 22324.900: 0.524463 1.97372 39.0368
## 3: 22320.274: 0.882021 1.75234 39.0205
## 4: 22319.469: 1.17143 1.45069 39.0088
## 5: 22319.450: 1.24673 1.38131 39.0040
## 6: 22319.441: 1.21373 1.41592 39.0058
## 7: 22319.441: 1.21644 1.41233 39.0060
## 8: 22319.441: 1.21663 1.41225 39.0063
```

```
summary(mod84)
```

```
## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24397.11 24484.35 -12184.56
```

```

##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.8276561 (Intr)
## IQ_verb      0.2433749 -0.995
## Residual     0.9991031
##
## Variance function:
## Structure: fixed weights
## Formula: ~IQ_trans2
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses + IQ_plus2 +      IQ_min2
##           Value Std.Error   DF   t-value p-value
## (Intercept)  40.51311 0.26548153 3541 152.60239 0.0000
## IQ_verb       3.04701 0.12578961 3541  24.22305 0.0000
## ses           0.16770 0.01149100 3541  14.59444 0.0000
## sex           2.25127 0.19670762 3541  11.44473 0.0000
## sch_iqv       0.80070 0.28526656  207   2.80685 0.0055
## sch_ses      -0.08379 0.04071565  207  -2.05788 0.0409
## IQ_plus2     -0.25992 0.03307122 3541  -7.85939 0.0000
## IQ_min2       0.19310 0.03786231 3541   5.10014 0.0000
## IQ_verb:ses  -0.01571 0.00501028 3541  -3.13469 0.0017
## sch_iqv:sch_ses -0.08925 0.03196439  207  -2.79202 0.0057
## Correlation:
##           (Intr) IQ_vrb ses    sex    sch_qv sch_ss IQ_pl2 IQ_mn2
## IQ_verb      -0.185
## ses          -0.001 -0.152
## sex          -0.368 -0.009 -0.018
## sch_iqv      -0.058 -0.049  0.047 -0.003
## sch_ses       0.043  0.045 -0.236 -0.010 -0.502
## IQ_plus2     -0.025 -0.829  0.098  0.039 -0.074  0.002
## IQ_min2      -0.156  0.744  0.041 -0.007 -0.024  0.043 -0.546
## IQ_verb:ses   0.036  0.050 -0.285 -0.050  0.032 -0.138 -0.232 -0.213
## sch_iqv:sch_ses -0.336  0.047  0.037  0.031  0.173 -0.030 -0.078  0.017
##           IQ_vr:
## IQ_verb
## ses
## sex
## sch_iqv
## sch_ses
## IQ_plus2
## IQ_min2
## IQ_verb:ses
## sch_iqv:sch_ses -0.117
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -4.37792295 -0.64085607  0.08891489  0.70081229  2.78148497
##
## Number of Observations: 3758
## Number of Groups: 211

```

Or, again with the categorized version of IQ:

```

mod84c <- lme(langPOST ~ IQ_verb*ses + sex + sch_iqv*sch_ses +
              IQ_plus2 + IQ_min2,
              random=~ IQ_verb|schoolnr, weights=varIdent(form=~1|IQ_cat),
              data = mlbook_red, method="ML")
summary(mod84c)

## Linear mixed-effects model fit by maximum likelihood
## Data: mlbook_red
##      AIC      BIC    logLik
## 24406.66 24525.06 -12184.33
##
## Random effects:
## Formula: ~IQ_verb | schoolnr
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.8218589 (Intr)
## IQ_verb      0.2876167 -0.887
## Residual     4.9719445
##
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | IQ_cat
## Parameter estimates:
##      (2,10] (-10,-2] (-1,0] (1,2] (-2,-1] (0,1]
## 1.000000 1.279763 1.308092 1.094504 1.326499 1.195896
## Fixed effects: langPOST ~ IQ_verb * ses + sex + sch_iqv * sch_ses + IQ_plus2 + IQ_min2
##           Value Std.Error DF t-value p-value
## (Intercept) 40.48811 0.26581745 3541 152.31545 0.0000
## IQ_verb      3.06454 0.12627484 3541 24.26878 0.0000
## ses          0.16749 0.01147842 3541 14.59195 0.0000
## sex          2.31031 0.19694474 3541 11.73073 0.0000
## sch_iqv      0.80631 0.28625199 207 2.81680 0.0053
## sch_ses     -0.08446 0.04077742 207 -2.07137 0.0396
## IQ_plus2    -0.26700 0.03564002 3541 -7.49171 0.0000
## IQ_min2      0.19815 0.03545232 3541 5.58930 0.0000
## IQ_verb:ses  -0.01658 0.00507468 3541 -3.26707 0.0011
## sch_iqv:sch_ses -0.08695 0.03207971 207 -2.71039 0.0073
## Correlation:
##           (Intr) IQ_vrb ses sex sch_qv sch_ss IQ_pl2 IQ_mn2
## IQ_verb      -0.182
## ses           0.000 -0.150
## sex          -0.369 -0.005 -0.017
## sch_iqv      -0.061 -0.055 0.050 0.001
## sch_ses       0.041 0.041 -0.242 -0.011 -0.502
## IQ_plus2     -0.039 -0.815 0.087 0.031 -0.067 0.008
## IQ_min2     -0.150 0.756 0.041 -0.005 -0.033 0.041 -0.547
## IQ_verb:ses   0.029 0.051 -0.271 -0.048 0.038 -0.138 -0.221 -0.226
## sch_iqv:sch_ses -0.339 0.042 0.033 0.035 0.181 -0.027 -0.070 0.014
## IQ_vr:
## IQ_verb
## ses
## sex
## sch_iqv
## sch_ses

```

```

## IQ_plus2
## IQ_min2
## IQ_verb:ses
## sch_iqv:sch_ses -0.100
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -4.29110406 -0.63793404  0.08608576  0.69427596  2.89597266
##
## Number of Observations: 3758
## Number of Groups: 211

```