Aufgabe 1:

In dieser Aufgabe sollen Sie die kieferorthopädischen Wachstumsdaten aus dem in `nlme` enthaltenenen Datensatz `Orthodont` analysieren.

(a) Benutzen Sie die Funktion `lme`, um ein Random-Effects-Modell `m` der Form

$$y_{ij} = \beta_0 + b_{0i} + \beta_1(t_j - 11) + \beta_2 x(t_j - 11) + \varepsilon_{ij}$$

anzupassen. Die Variable x repräsentiert das Geschlecht, t das Alter. Visualisieren Sie die Modellanpassung von `m` mit den Befehlen

```
R>plot(augPred(m))
R>plot(m, resid(.) ~ age | Subject, abline = 0)
```

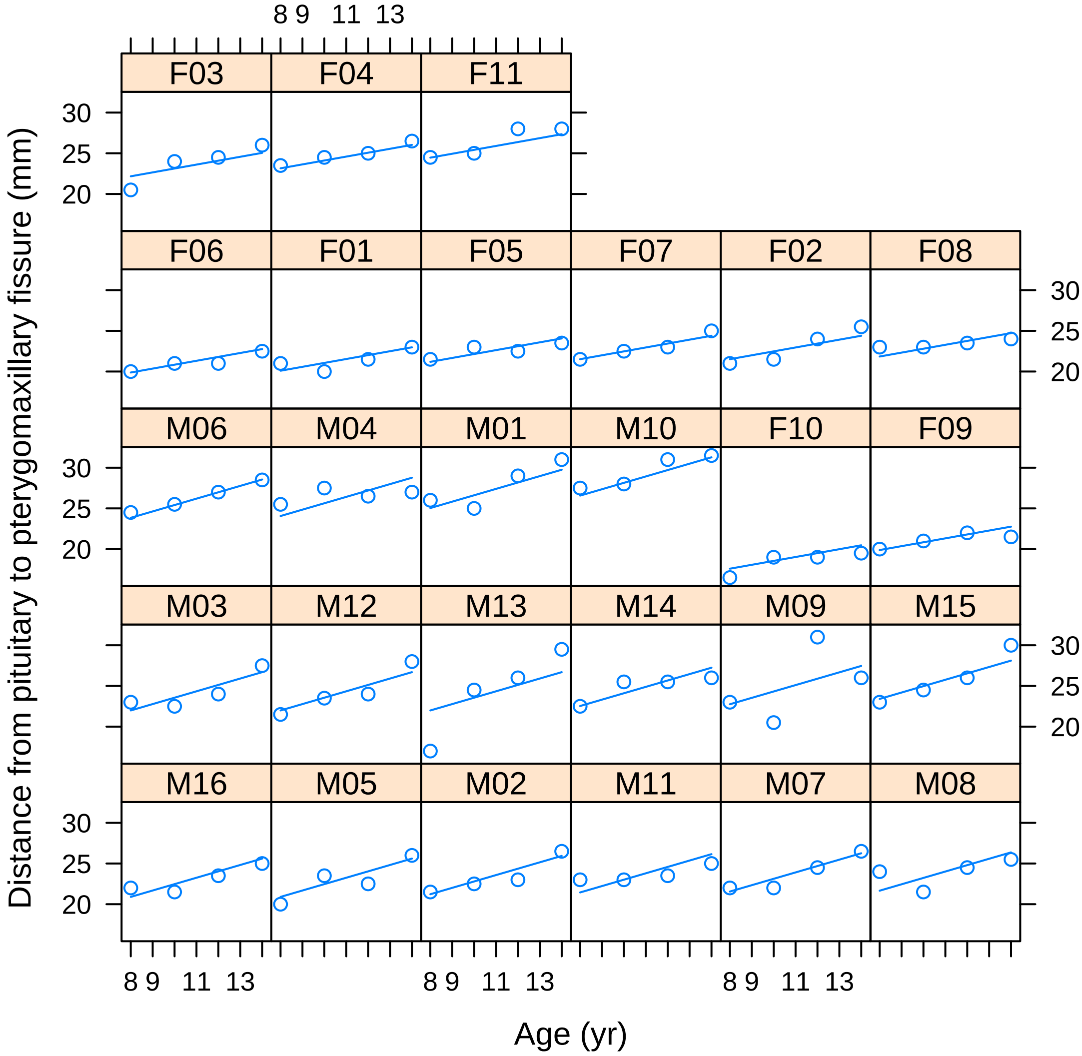
Ist das Modell ausreichend?

Lösung:

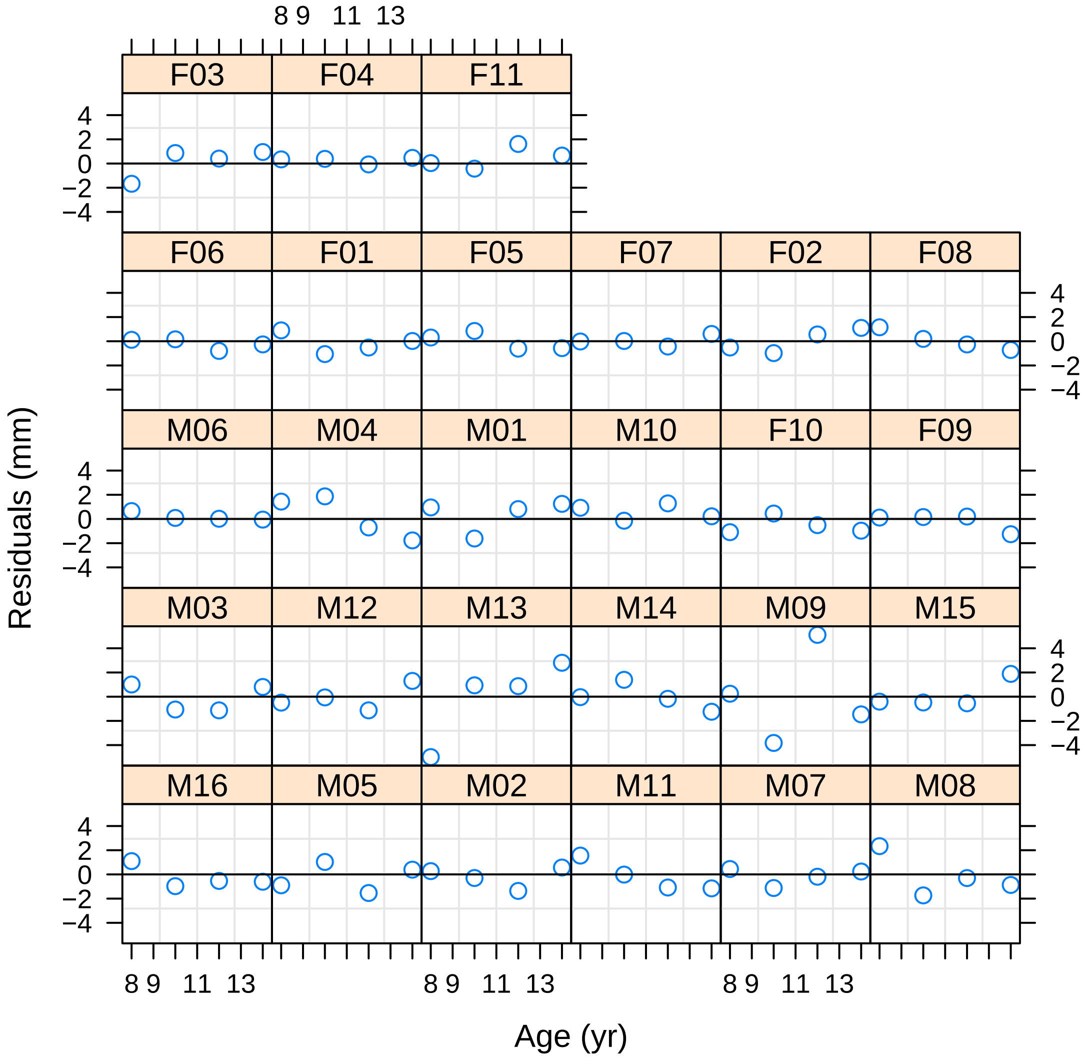
```
R>library(nlme)
R>data(Orthodont)
R>m <- lme(distance ~ I(age - 11) * Sex, random = ~1 | Subject,
+     data = Orthodont)
R>summary(m)

Linear mixed-effects model fit by REML
 Data: Orthodont
       AIC      BIC    logLik
  445.7572 461.6236 -216.8786

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.816214 1.386382

Fixed effects: distance ~ I(age - 11) * Sex
                          Value Std.Error DF  t-value p-value
(Intercept)           24.968750 0.4860008 79 51.37595  0.0000
I(age - 11)            0.784375 0.0775011 79 10.12082  0.0000
SexFemale             -2.321023 0.7614168 25 -3.04829  0.0054
I(age - 11):SexFemale -0.304830 0.1214209 79 -2.51052  0.0141
 Correlation:
                      (Intr) I(g-11) SexFml
I(age - 11)            0.000
SexFemale             -0.638  0.000
I(age - 11):SexFemale  0.000 -0.638   0.000

Standardized Within-Group Residuals:
```

```
       Min          Q1         Med          Q3         Max 
-3.59804400 -0.45461690  0.01578365  0.50244658  3.68620792 

Number of Observations: 108
Number of Groups: 27
```

R>print(plot(augPred(m)))



R>print(plot(m, resid(.) ~ age | Subject, abline = 0))

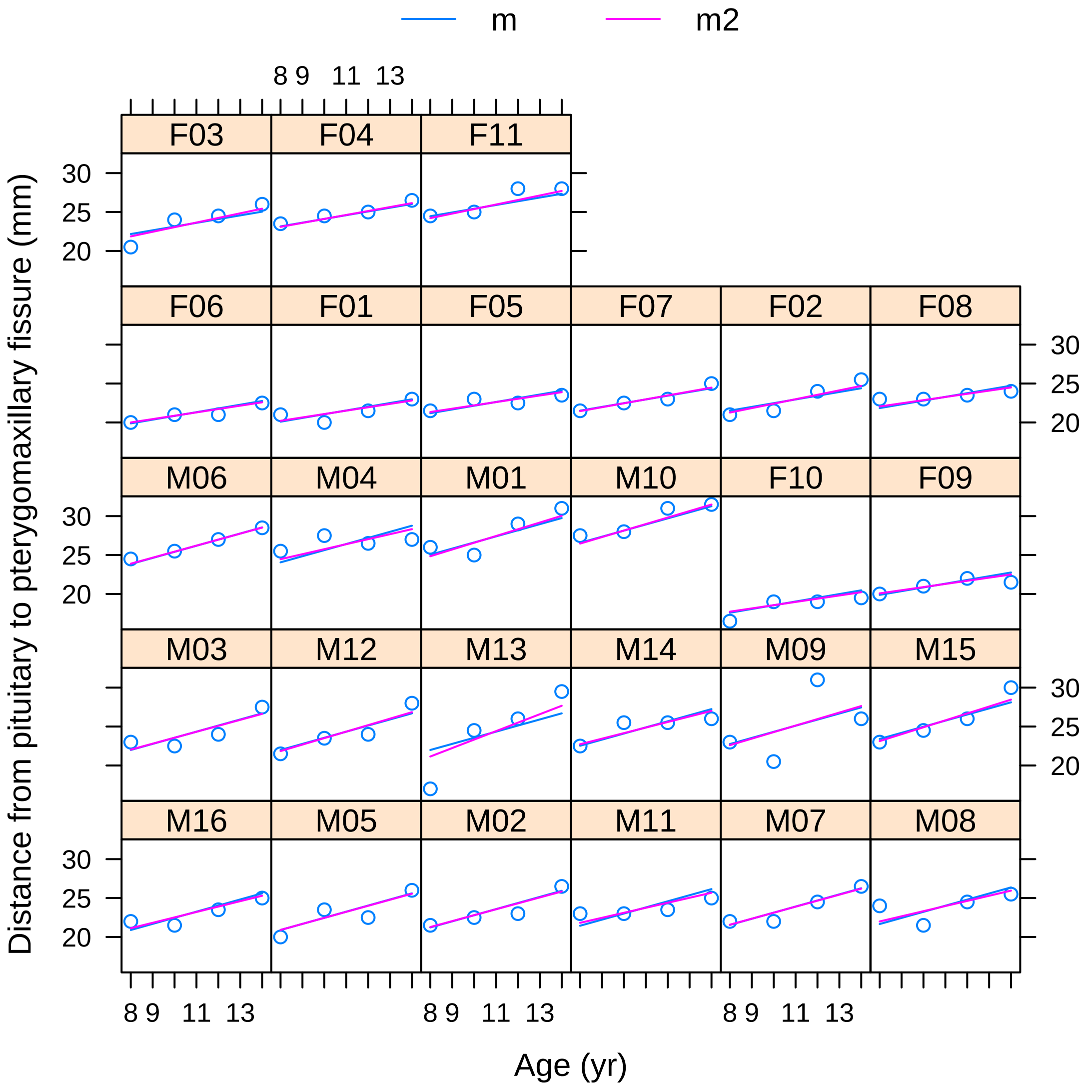


⇒ teilweise deutlicher Trend in subjektspezifischen Residuen, Random Slope evtl. sinnvoll.

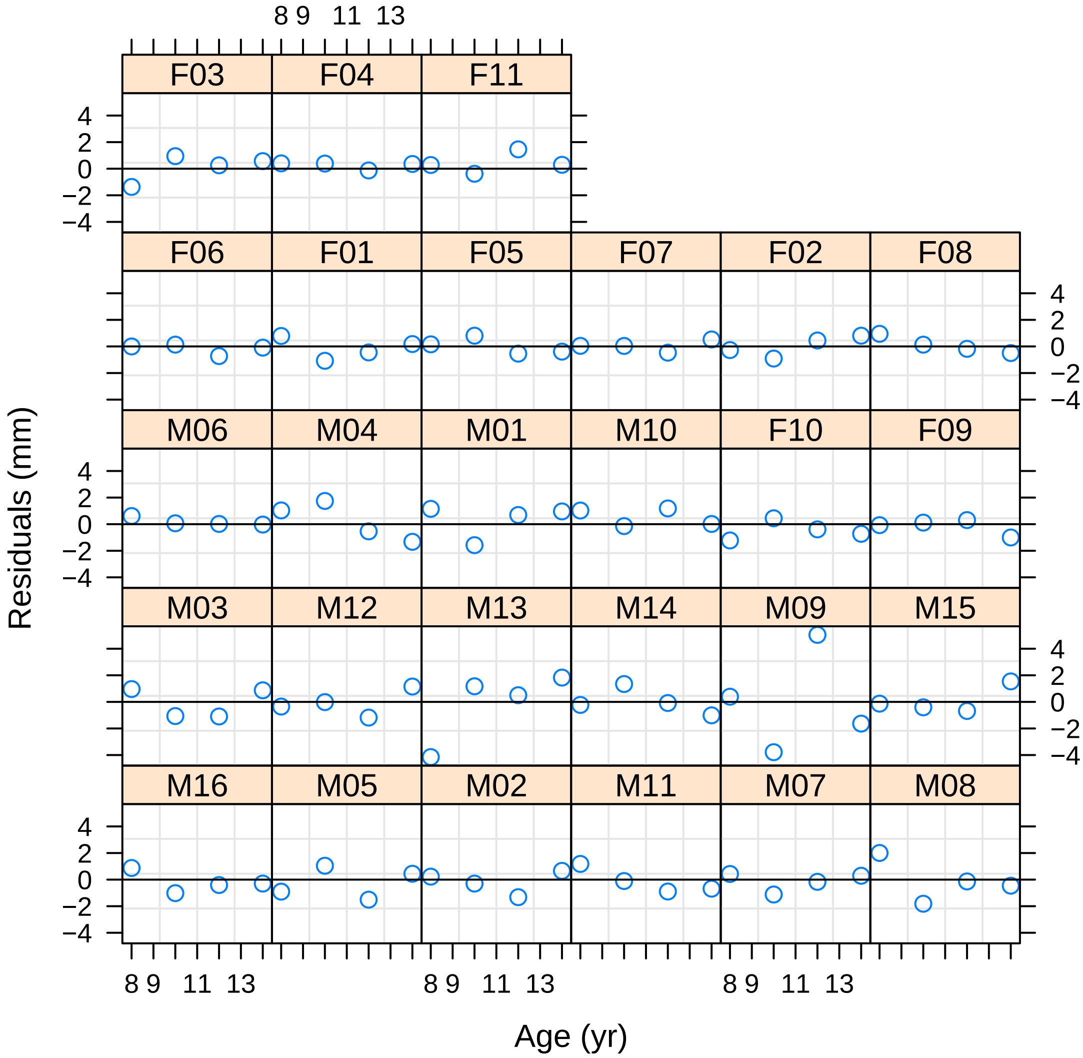
(b) Erweitern Sie das Modell um einen zufälligen subjektspezifischen Zeittrend. Vergleichen Sie (graphisch) die Schätzungen und Residuen aus den beiden Modellen.
Hinweis: `comparePred` erstellt ein plotbares Objekt.

Lösung:

```
R>m2 <- update(m, random = ~I(age - 11) | Subject)
R>print(plot(comparePred(m, m2)))
```



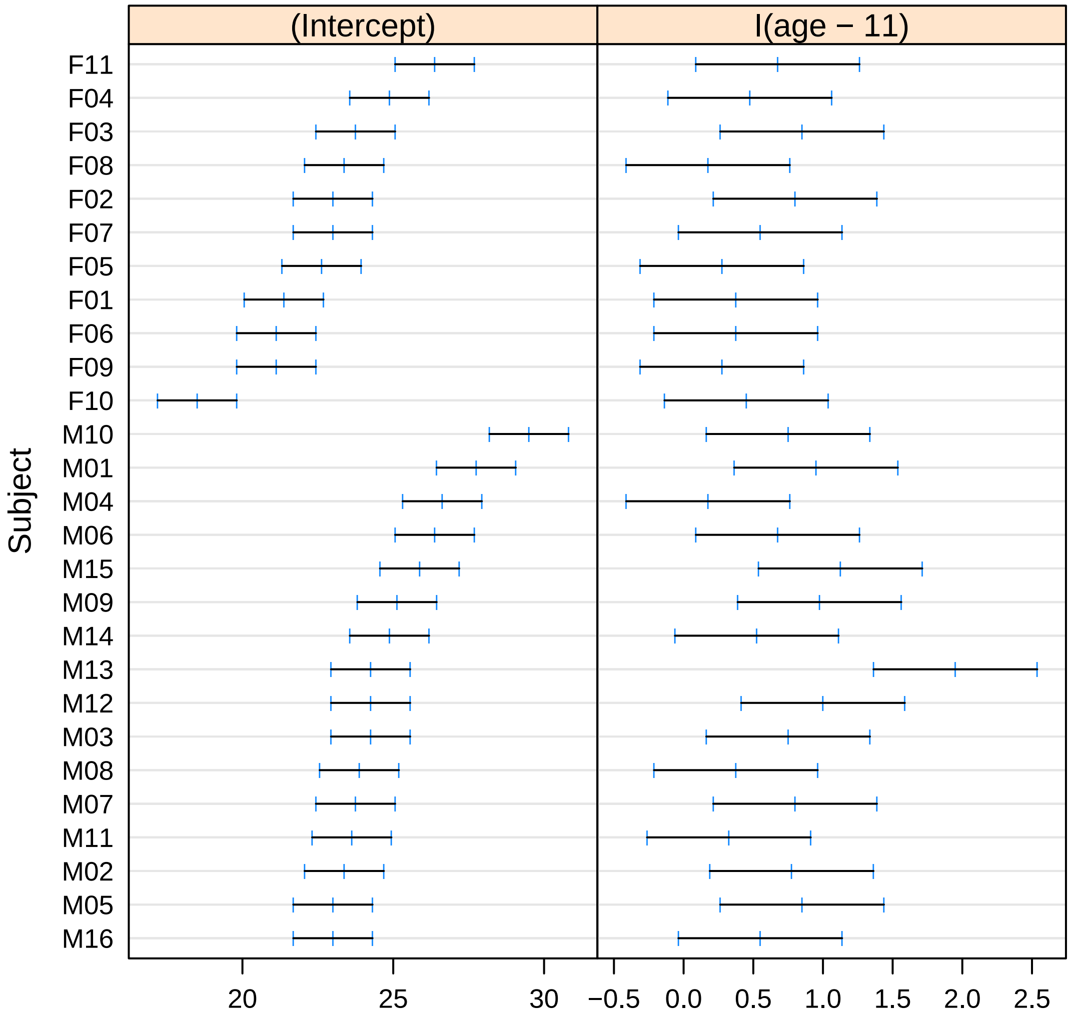

```
R>print(plot(m2, resid(.) ~ age | Subject, abline = 0))
```



⇒ wenig Veränderung - Warum?

Heuristisch: Subjektspezifischen Modelle zeigen dass Unterschiede in der Slope mit große Unsicherheit behaftet sind:

```
R>lmlist <- lmList(distance ~ I(age - 11), data = Orthodont)
R>print(plot(intervals(lmlist)))
```



Dementsprechend stark werden die subjektspezifischen Slope-Unterschiede durch die im Verhältnis sehr kleine geschätze Varianz der Random Slopes ($\widehat{\text{Var}}(\boldsymbol{b}_1) = 0.0325$, $\hat{\beta}_1 = 0.784$) geschrumpft ($\Rightarrow$ Shrinkage-Effekt).

(c) Schätzen Sie das gleiche Modell mit ML (d.h. `method='ML'`) und vergleichen Sie die resultierende Kovarianzmatrix der zufälligen Effekte mit der aus der REML-Schätzung. Vergleichen Sie auch für beide Methoden die subjektspezifischen Koeffizienten.
Hinweis: `getVarCov` produziert $\text{Cov}(\boldsymbol{b}_i)$. Die Koeffizienten bekommen Sie mit der Funktion `coef`, die Methode `compareFits` erstellt ein plotbares Objekt.

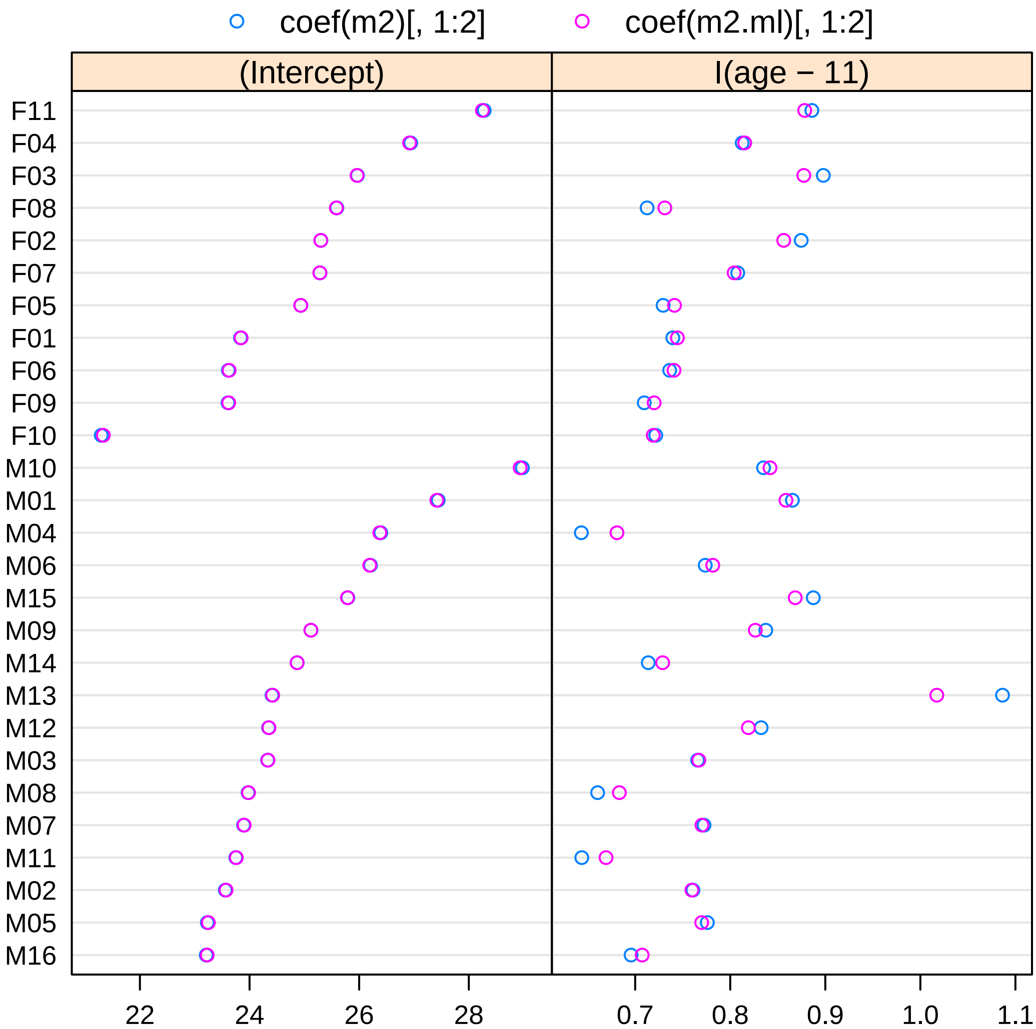
Lösung:

```
R>m2.ml <- update(m2, method = "ML")
R>getVarCov(m2)

Random effects variance covariance matrix
            (Intercept) I(age - 11)
(Intercept)    3.350100    0.068142
I(age - 11)    0.068142    0.032524
  Standard Deviations: 1.8303 0.18035
```

```
R>getVarCov(m2.ml)

Random effects variance covariance matrix
            (Intercept) I(age - 11)
(Intercept)    3.070200    0.063094
I(age - 11)    0.063094    0.023759
  Standard Deviations: 1.7522 0.15414
```

⇒ REML-Schätzungen größer als ML-Schätzungen (ML ist nach unten verzerrt, s. Vorlesung).

```
R>print(plot(compareFits(coef(m2)[, 1:2], coef(m2.ml)[, 1:2])))
```



⇒ unterschiedliche Schätzung der Varianzkomponenten wirkt sich bei Intercepts praktisch nicht aus, bei Slopes etwas deutlicher. REML-Schätzungen der Slopes streuen stärker da REML-Varianz deutlich größer als ML-Varianz.

Aufgabe 2:

Im folgenden sollen Sie die Funktion `gls` benutzen um marginale Modelle mit verschiedenen Korrelationstrukturen zu vergleichen.

(a) Schätzen Sie für die Orthodont-Daten ein Modell mit festen Effekten wie in in Modell `m` (s.o.), in dem die Beobachtungen verschiedener Subjekte voneinander unabhängig sein sollen und die Beobachtungen eines Subjektes eine unstrukturierte Korrelationstruktur (Korrelationsstruktur `corSymm`) besitzen. Vergleichen Sie die geschätzte Korrelationsmatrix mit der Korrelationsmatrix die sich aus der marginalen Betrachtungsweise von Modell `m` ergibt. Benutzen Sie die Funktion `intervals` um die Variabilität der geschätzen Korrelationsparameter zu bestimmen.

Lösung:

```
R>m.mar <- gls(distance ~ I(age - 11) * Sex, correlation = corSymm(form = ~1 |
+     Subject), data = Orthodont)
R>summary(m.mar)

Generalized least squares fit by REML
  Model: distance ~ I(age - 11) * Sex
  Data: Orthodont
       AIC      BIC    logLik
  448.1706 477.2589 -213.0853

Correlation Structure: General
 Formula: ~1 | Subject
 Parameter estimate(s):
 Correlation:
  1     2     3
2 0.575
3 0.638 0.574
4 0.515 0.749 0.721

Coefficients:
                         Value Std.Error  t-value p-value
(Intercept)          25.000655 0.4852207 51.52429  0.0000
I(age - 11)           0.824340 0.0824257 10.00101  0.0000
SexFemale            -2.355437 0.7601948 -3.09847  0.0025
I(age - 11):SexFemale -0.348101 0.1291362 -2.69561  0.0082

 Correlation:
                      (Intr) I(g-11) SexFml
I(age - 11)            0.070
SexFemale             -0.638 -0.045
I(age - 11):SexFemale -0.045 -0.638   0.070

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.41707511 -0.64439569 -0.07388949  0.58480652  2.26288068

Residual standard error: 2.286910
Degrees of freedom: 108 total; 104 residual

R>cov.m.mar <- getVarCov(m.mar)
R>varb.m <- getVarCov(m)[1, 1]
R>cov2cor(cov.m.mar)
```

```
Marginal variance covariance matrix
        [,1]    [,2]    [,3]    [,4]
[1,] 1.00000 0.57541 0.63757 0.51478
[2,] 0.57541 1.00000 0.57428 0.74888
[3,] 0.63757 0.57428 1.00000 0.72060
[4,] 0.51478 0.74888 0.72060 1.00000
  Standard Deviations: 1 1 1 1
```

```
R>varb.m/(varb.m + m$sigma^2)
```

```
[1] 0.6318388
```

Die letzte Zeile berechnet die (marginale) geschätzte Korrelation zweier Beobachtungen des selben Subjektes (s.a. Blatt 3: Varianz der zufälligen Intercepts geteilt durch Summe der Varianzen der zufälligen Intercepts und der Residuen) in Modell `m` - diese ist immer gleich da das Random Intercept Modell einem CompoundSymmetry Modell mit positiven Korrelationen entspricht. Man sieht hier dass die unrestringiert geschätze Korrelationstruktur aus `gls` der CompoundSymmetry recht ähnlich ist.

```
R>intervals(m.mar, which = "var-cov")
```

```
Approximate 95% confidence intervals

 Correlation structure:
             lower      est.     upper
cor(1,2) 0.2408361 0.5754124 0.7877569
cor(1,3) 0.3759471 0.6375722 0.8050547
cor(1,4) 0.1675088 0.5147787 0.7484040
cor(2,3) 0.2694383 0.5742825 0.7745177
cor(2,4) 0.5226129 0.7488760 0.8765910
cor(3,4) 0.5000277 0.7205955 0.8533694
attr(,"label")
[1] "Correlation structure:"

 Residual standard error:
   lower     est.    upper
1.863055 2.286910 2.807195
```

Die Intervalle der Korrelationen überlappen sich großteils- das spricht dafürt dass man die Korrelationsstruktur zu einer CompundSymmetry-Struktur vereinfachen kann.

(b) Schätzen Sie ein Modell mit vereinfachter Korrelationstruktur so, dass diese der marginalen Korrelationstruktur von Modell `m` entspricht.

Lösung:

```
R>m.mar2 <- gls(distance ~ I(age - 11) * Sex, correlation = corCompSymm(form = ~1 |
+     Subject), data = Orthodont)
R>summary(m.mar2)
```

```
Generalized least squares fit by REML
  Model: distance ~ I(age - 11) * Sex
  Data: Orthodont
```

```
      AIC      BIC    logLik
  445.7572 461.6236 -216.8786

Correlation Structure: Compound symmetry
 Formula: ~1 | Subject
 Parameter estimate(s):
      Rho
0.6318381

Coefficients:
                          Value Std.Error  t-value p-value
(Intercept)           24.968750 0.4860003 51.37600  0.0000
I(age - 11)            0.784375 0.0775011 10.12082  0.0000
SexFemale             -2.321023 0.7614161 -3.04830  0.0029
I(age - 11):SexFemale -0.304830 0.1214209 -2.51052  0.0136

 Correlation:
                      (Intr) I(g-11) SexFml
I(age - 11)            0.000
SexFemale             -0.638  0.000
I(age - 11):SexFemale  0.000 -0.638   0.000

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.45773173 -0.57853118 -0.07360637  0.58204364  2.29634478

Residual standard error: 2.284881
Degrees of freedom: 108 total; 104 residual
```

R>getVarCov(m.mar2)

```
Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]
[1,] 5.2207 3.2986 3.2986 3.2986
[2,] 3.2986 5.2207 3.2986 3.2986
[3,] 3.2986 3.2986 5.2207 3.2986
[4,] 3.2986 3.2986 3.2986 5.2207
  Standard Deviations: 2.2849 2.2849 2.2849 2.2849
```

R>cov2cor(getVarCov(m.mar2))

```
Marginal variance covariance matrix
        [,1]    [,2]    [,3]    [,4]
[1,] 1.00000 0.63184 0.63184 0.63184
[2,] 0.63184 1.00000 0.63184 0.63184
[3,] 0.63184 0.63184 1.00000 0.63184
[4,] 0.63184 0.63184 0.63184 1.00000
  Standard Deviations: 1 1 1 1
```