How do you implement the stuff we talked about in class in R?
## Growth Data

- Data taken from Verbeke and Molenberghs page 241
- Data is a summary of brain size for boys and girls measured at ages 8, 10, 12, 14
- We create a “long format” data set like

```r
> dat[1 : 8,]
growth gender subject age
1  21.0     1    1    8
2  20.0     1    1   10
3  21.5     1    1   12
4  23.0     1    1   14
5  21.0     1    2    8
6  21.5     1    2   10
7  24.0     1    2   12
8  25.5     1    2   14
```
 Libraries to load

Load `nlme` and `gee`

```r
> library(nlme)
> library(gee)
```

Note if `gee` is not already installed

```r
> install.packages("gee")
```
Fitting a random intercept model

- random intercept only

```r
randomItc <- lme(growth ~ age + gender,
                 random = ~ 1 | subject,
                 data = dat,
                 method = "ML")
```

- The `random` statement tells R that there is a random intercept at the levels of the variable subject. *Make sure your data set is sorted by the subject variable.*

- The `method = "ML"` statement uses ML not REML estimation for the variance components.

- This fits the model

\[
Growth_{ij} = Itc + u_i + Age_{ij} + Gender_i + \epsilon_{ij}
\]

where \( u_i \sim N(0, \sigma_u^2) \) and \( \epsilon \sim N(0, \sigma^2) \).
Model fitting notes

- If a variety of baseline ages had been observed it is preferable to have separate effects for the (cross-sectional) baseline age and (longitudinal) years from baseline; this is not an issue in this data set where every child has the same baseline age.

- `lme` returns an object of class `lme`, for which there are many methods defined: plot, print, summary, coef, resid, fitted, coef, fixed.effects, AIC, intervals, anova.

- You can also just grab what you want directly:

  ```r
  > round(summary(randomItc)$tTable, 3)
   Value Std.Error DF  t-value p-value
  (Intercept) 17.707     0.832  80  21.294  0.000
   age        0.660     0.062  80  10.632  0.000
 gender      -2.321     0.743  25  -3.124  0.004
  ```
> intervals(randomItc)
Approximate 95% confidence intervals
  Fixed effects:
          lower     est.      upper
(Intercept)  16.0750295  17.7067130  19.3383964
       age       0.5383446   0.6601852   0.7820257
      gender     -3.8299924  -2.3210227   -0.8120531
attr(,"label")
[1] "Fixed effects:"

  Random Effects:
    Level: subject
          lower     est.      upper
sd((Intercept))  1.265289  1.730079   2.365603

  Within-group standard error:
          lower     est.      upper
1.219684  1.422728   1.659572
More lme functionality

- Look at the code to see different ways to use the `plot.lme` function. `lme` uses `trellis` graphics.
- `random.effects(randomItc)` will return the empirical BLUP estimates of the $u_i$.
- You can include `offset` terms which, in principle, can be used to calculate a profile likelihoods for slope parameters. I was unsuccessful when I tried this.
- You can use `method = "REML"` to get the REML estimates. Remember to use ML if you are going to use likelihood ratio tests for the slope parameters.
Example

Likelihood ratio test for an Age Gender interaction

```r
randomItcInt <- {
    update(randomItc,
    growth ~ age * gender)
}
> anova(randomItc, randomItcInt)
          Model df    Test  L.Ratio p-value
randomItc 1 5     1 vs 2 6.217427 0.0126
randomItcInt 2 6
```
(I edited the output a bit, it also gives the AIC, BIC and likelihoods)
Fitting a random slope

- Though the data do not really seem to display the need for a random slope term, here’s how I fit one anyway

```r
lme(growth ~ age + gender,
    random = ~ 1 + age | subject,
    data = dat,
    method = "ML"
)
```

Which fits the model

\[ \text{Growth}_{ij} = Itc + u_{1i} + \text{Age}_{ij} \times u_{2i} + \text{Gender}_i + \epsilon_{ij} \]

automatically allowing \( u_{1i} \) and \( u_{2i} \) to be correlated
More model fitting

- If you just want an unstructured correlation matrix, you can use `gls`
  
  ```r
  gls(growth ~ age * gender,
      correlation=corSymm(form=~1|subject),
      data = dat)
  ```

- If you want compound symmetric correlation matrix, you can change `corSymm` to `corCompSymm`

- You can have random effects and a general covariance matrix

- These models assume the assume a known variance structure, we can use `gee` to get robust variance estimates
Here is some sample `gee` code

```r
gee(growth ~ age + gender,
    id = subject,
    data = dat,
    corstr = "unstructured")
```

you can also use

```r
corstr = "independence"
```

and

```r
corstr = "exchangeable"
```

try `?gee` to see all of the available working correlation matrices
### Effect estimates as the models change

<table>
<thead>
<tr>
<th>Model</th>
<th>Itc</th>
<th>age</th>
<th>gender</th>
<th>age:gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>noCor</td>
<td>16.341</td>
<td>0.784</td>
<td>1.032</td>
<td>-0.305</td>
</tr>
<tr>
<td>randomItc</td>
<td>16.341</td>
<td>0.784</td>
<td>1.032</td>
<td>-0.305</td>
</tr>
<tr>
<td>exchLm</td>
<td>16.341</td>
<td>0.784</td>
<td>1.032</td>
<td>-0.305</td>
</tr>
<tr>
<td>unstrLm</td>
<td>15.933</td>
<td>0.824</td>
<td>1.474</td>
<td>-0.348</td>
</tr>
<tr>
<td>unstrGee</td>
<td>16.324</td>
<td>0.788</td>
<td>1.074</td>
<td>-0.310</td>
</tr>
<tr>
<td>idpendGee</td>
<td>16.341</td>
<td>0.784</td>
<td>1.032</td>
<td>-0.305</td>
</tr>
<tr>
<td>exchGee</td>
<td>16.341</td>
<td>0.784</td>
<td>1.032</td>
<td>-0.305</td>
</tr>
</tbody>
</table>

Apart from the linear model with an unstructured covariance matrix, the results are all nearly identical. (The odd result may be due to lack of convergence of estimates.)
### Standard error estimates by model

<table>
<thead>
<tr>
<th>Model</th>
<th>Itc</th>
<th>age</th>
<th>gender</th>
<th>age:gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>noCor</td>
<td>1.416</td>
<td>0.126</td>
<td>2.219</td>
<td>0.198</td>
</tr>
<tr>
<td>randomItc</td>
<td>0.981</td>
<td>0.078</td>
<td>1.538</td>
<td>0.122</td>
</tr>
<tr>
<td>exchLm</td>
<td>0.981</td>
<td>0.078</td>
<td>1.537</td>
<td>0.121</td>
</tr>
<tr>
<td>unstrLm</td>
<td>0.998</td>
<td>0.082</td>
<td>1.563</td>
<td>0.129</td>
</tr>
<tr>
<td>unstrGee</td>
<td>1.170</td>
<td>0.098</td>
<td>1.376</td>
<td>0.117</td>
</tr>
<tr>
<td>idpendGee</td>
<td>1.171</td>
<td>0.098</td>
<td>1.378</td>
<td>0.117</td>
</tr>
<tr>
<td>exchGee</td>
<td>1.171</td>
<td>0.098</td>
<td>1.378</td>
<td>0.117</td>
</tr>
</tbody>
</table>
No interaction model

To illustrate a point, consider the standard errors of the (time invariant) gender effect when there is no interaction with the (time varying) age.

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>age</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>noCor</td>
<td>1.112</td>
<td>0.098</td>
<td>0.445</td>
</tr>
<tr>
<td>randomItc</td>
<td>0.832</td>
<td>0.062</td>
<td>0.743</td>
</tr>
<tr>
<td>exchLm</td>
<td>0.834</td>
<td>0.062</td>
<td>0.761</td>
</tr>
<tr>
<td>unstrLm</td>
<td>0.897</td>
<td>0.070</td>
<td>0.757</td>
</tr>
<tr>
<td>unstrGee</td>
<td>0.895</td>
<td>0.070</td>
<td>0.730</td>
</tr>
<tr>
<td>idpendGee</td>
<td>0.889</td>
<td>0.070</td>
<td>0.750</td>
</tr>
<tr>
<td>exchGee</td>
<td>0.889</td>
<td>0.070</td>
<td>0.750</td>
</tr>
</tbody>
</table>