STA 610L: Module 3.2

Linear mixed effects models (influence measures)

Dr. Olanrewaju Michael Akande

STA 610L

Residuals

Residual analysis and diagnostic methods are well developed for linear regression models (c.f., Cook and Weisberg, 1982), but they are somewhat less developed for mixed effects models.

This set of notes is based on Nieuwenhuis et al.

STA 610L

Example: Orthodontics Data

We'll consider the dental data with model

$$Y_{ij} = \beta_0 + \beta_1 I(\text{male})_i + \beta_2 t_j + \beta_3 I(\text{male})_i t_j + b_{0i} + b_{1i} t_j + \varepsilon_{ij},$$

where

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \overset{iid}{\sim} N\left(0, \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}\right) \perp \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2),$$

for illustration.

STA 610L

Residuals

Generally, the residuals $y_{ij} - \hat{y}_{ij}$ can be helpful in flagging outliers and in assessing the adequacy of most fitted models.

However, the definition of residuals is not always consistent in the case of mixed effects or hierarchical models:

- Many texts define residuals for subject/group i as $Y_i - X_i\widehat{\beta}$.
- Many software implementations define residuals as $Y_i - X_i\widehat{\beta} - Z_i\hat{b}_i$ (nice because these can then be analyzed using standard methods)

These are easy to compute and we already did the later in a previous module.

STA 610L

Review: residual analysis

That said, in any case, residual analysis is not always a great tool for detecting influential cases:

- Cases with high residuals or high standardized residuals are called outliers
- Outliers may or may not be influential in the model fit
- An influential case may dominate the regression model so that the line is drawn more closely towards the case (making it an inlier)

Review: influence

We hope that all data points have some amount of influence on our parameter estimates.

However, we may be concerned if a single case has disproportionate influence on model results.

If so, one observation or group of observations may pull the estimated regression line towards the group.

In such a case, excluding a single group might have a substantial effect on estimates.

This idea is behind the development of many popular influence diagnostics, often termed deletion diagnostics.

Review: leverage

The degree to which an observation has the *potential* to be influential is closely related to the leverage of the case, which is a measure of how extreme the case is in the X space.

Leverage is not simply defined as an outlying value in X space of a single variable but also in a multivariate sense.

For example, in a study of pregnancy outcomes, it may be relatively common to have mothers who are 40, or fathers who are 20, but babies who have a 40 year old mother and a 20 year old father may be fairly uncommon.

As you should already know, the leverage score for an observation i is the ith diagonal element of the projection or hat matrix.

Influence

It is not necessarily the case that outliers or cases with high leverage are influential.

So, how do we detect influential cases?

One popular approach is to use the principle that when a single case is removed from the data entirely, we would like for models based on the data not to give vastly different conclusions.

If parameter estimates change a lot after a single individual is excluded, then the individual may be considered influential.

Multi-level influence

Mixed effects and multilevel models estimate effects of lower-level and higher-level variables.

It is thus possible that in some cases a higher-level group is influential (more likely when you don't have very many groups), while in others, a single observation within a group is influential.

We will examine influence at both levels.

STA 610L

Influence of higher-Level observations

DFBETAS (standardized difference of the beta) measures the level of influence observations have on single parameter estimates.

It is calculated as the difference in magnitude of the parameter estimate between the model including and the model excluding the group (or individual in a longitudinal study), standardized by dividing by the standard error of the estimate that excludes the group (to prevent variance inflation from masking the level of influence).

For group i and parameter k,

$$\text{DFBETAS}_{ik} = \frac{\widehat{\gamma}_k - \widehat{\gamma}_{k(-i)}}{se(\widehat{\gamma}_{k(-i)})},$$

where $\widehat{\gamma}_k$ is the original estimate of the kth parameter, and $\widehat{\gamma}_{k(-i)}$ is the estimate of the same parameter after group i has been excluded from the data.

DFBETAS

Belsley (1980) recommends a cutoff of $\frac{2}{\sqrt{n}}$ for identifying overly influential observations.

Here n is defined as the number of groups at the level of removal $(-i)$ for the calculation.

(For the dental data we have 27 kids and 4 observations per kid, so at the group level $n = 27$.)

```
library(influence.ME)
m1.inf <- influence(m1,"Subject")
#use obs argument for observation-level deletion
print(2/sqrt(length(unique(Orthodont$Subject))))
dfbetas(m1.inf)
#note that there be issues with singularity when we start removing groups
```

STA 610L

DFBETAS

```r
m1.inf <- influence(m1,"Subject"); print(2/sqrt(length(unique(Orthodont$Subject))))
```

```
## [1] 0.3849002
```

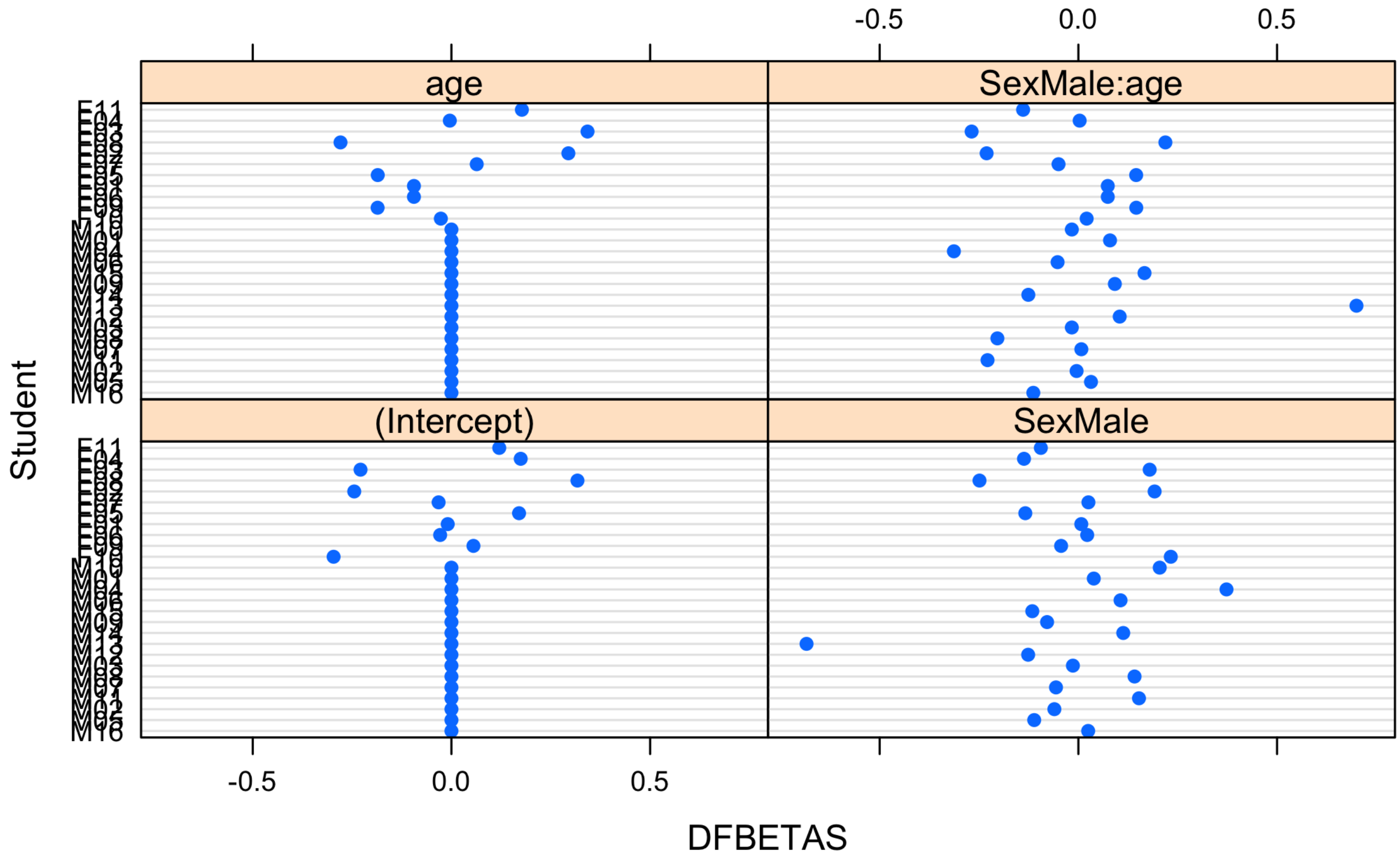
```r
round(dfbetas(m1.inf),4)
```

```
##     (Intercept) SexMale     age SexMale:age
## M16      0.0000  0.0247  0.0000     -0.1134
## M05      0.0000 -0.1111  0.0000      0.0317
## M02      0.0000 -0.0604  0.0000     -0.0045
## M11      0.0000  0.1525  0.0000     -0.2286
## M07      0.0000 -0.0563  0.0000      0.0075
## M08      0.0000  0.1414  0.0000     -0.2038
## M03      0.0000 -0.0138  0.0000     -0.0164
## M12      0.0000 -0.1264  0.0000      0.1040
## M13      0.0000 -0.6841  0.0000      0.6999
## M14      0.0000  0.1129  0.0000     -0.1259
## M09      0.0000 -0.0788  0.0000      0.0918
## M15      0.0000 -0.1159  0.0000      0.1664
## M06      0.0000  0.1061  0.0000     -0.0523
## M04      0.0000  0.3728  0.0000     -0.3132
## M01      0.0000  0.0388  0.0000      0.0796
## M10      0.0000  0.2047  0.0000     -0.0164
## F10     -0.2965  0.2326 -0.0266      0.0209
## F09      0.0554 -0.0434 -0.1858      0.1457
## F06     -0.0284  0.0222 -0.0944      0.0740
## F01     -0.0093  0.0073 -0.0943      0.0740
## F05      0.1702 -0.1335 -0.1854      0.1455
## F07     -0.0322  0.0253  0.0636     -0.0499
## F02     -0.2446  0.1919  0.2943     -0.2309
## F08      0.3171 -0.2488 -0.2792      0.2190
## F03     -0.2287  0.1794  0.3425     -0.2687
## F04      0.1744 -0.1368 -0.0041      0.0032
## F11      0.1204 -0.0944  0.1773     -0.1391
```

Here we see that M04 and M13 are influential on some of our estimates. What did these kids look like?

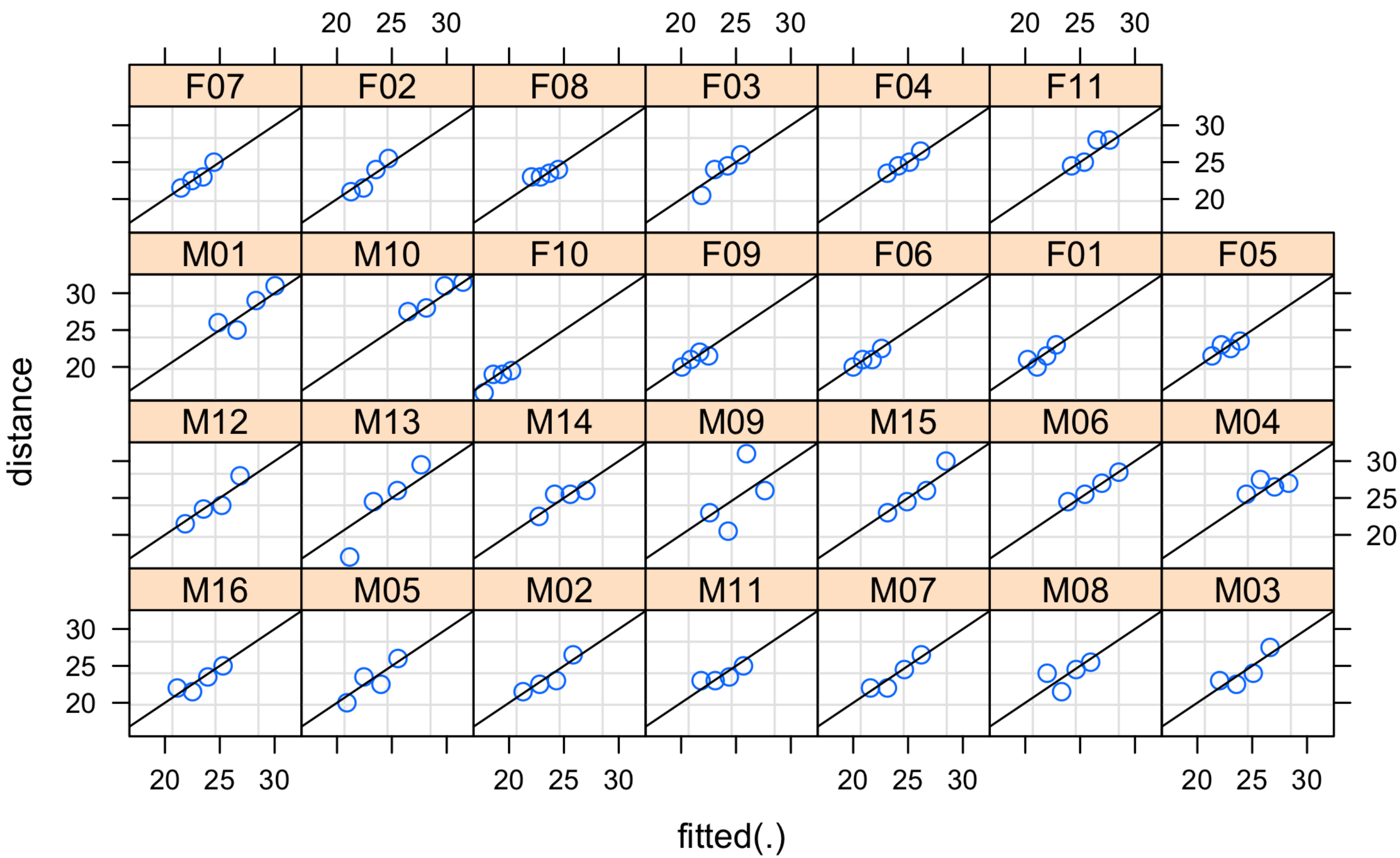
DFBETAS

```
plot(m1.inf,which="dfbetas",xlab="DFBETAS",ylab="Student")
```



DFBETAS

```r
plot(m1, distance ~ fitted(.) | Subject, abline = c(0,1))
```



STA 610L

DFBETAS

```
Orthodont$distance[Orthodont$Subject=="M04"]
```

```
## [1] 25.5 27.5 26.5 27.0
```

```
Orthodont$distance[Orthodont$Subject=="M13"]
```

```
## [1] 17.0 24.5 26.0 29.5
```

M04 had large measurements without a lot of growth over time -- pulling him out of the model reduced the intercept for boys and also decreased their slope.

M13 had a small measure at age 8 and then grew substantially. Leaving him out of the model changed the estimates significantly, greatly increasing the intercept for boys and also reducing the slope among boys.

Cook's distance

When the number of observations or predictors is large, it may take a while to wade through all the DFBETAS.

Cook's distance gives us a summary measure for influence on all parameter estimates.

It is defined as

$$C_i = \frac{1}{p}(\widehat{\gamma} - \widehat{\gamma}_{(-i)})'\widehat{\Sigma}^{-1}_{(-i)}(\widehat{\gamma} - \widehat{\gamma}_{(-i)})$$

where p is the length of β, and $\widehat{\Sigma}_{(-i)}$ is the covariance matrix of the parameter estimates excluding group i.

Van der Meer et al (2010) recommends a cutoff of $\frac{4}{n}$ where again n is the number of groups in the grouping factor being evaluated.

If there is just one parameter in the model, then Cook's distance is the DFBETAS squared for that parameter.

STA 610L

Cook's distance

```
print(4/length(unique(Orthodont$Subject)))
```

```
## [1] 0.1481481
```

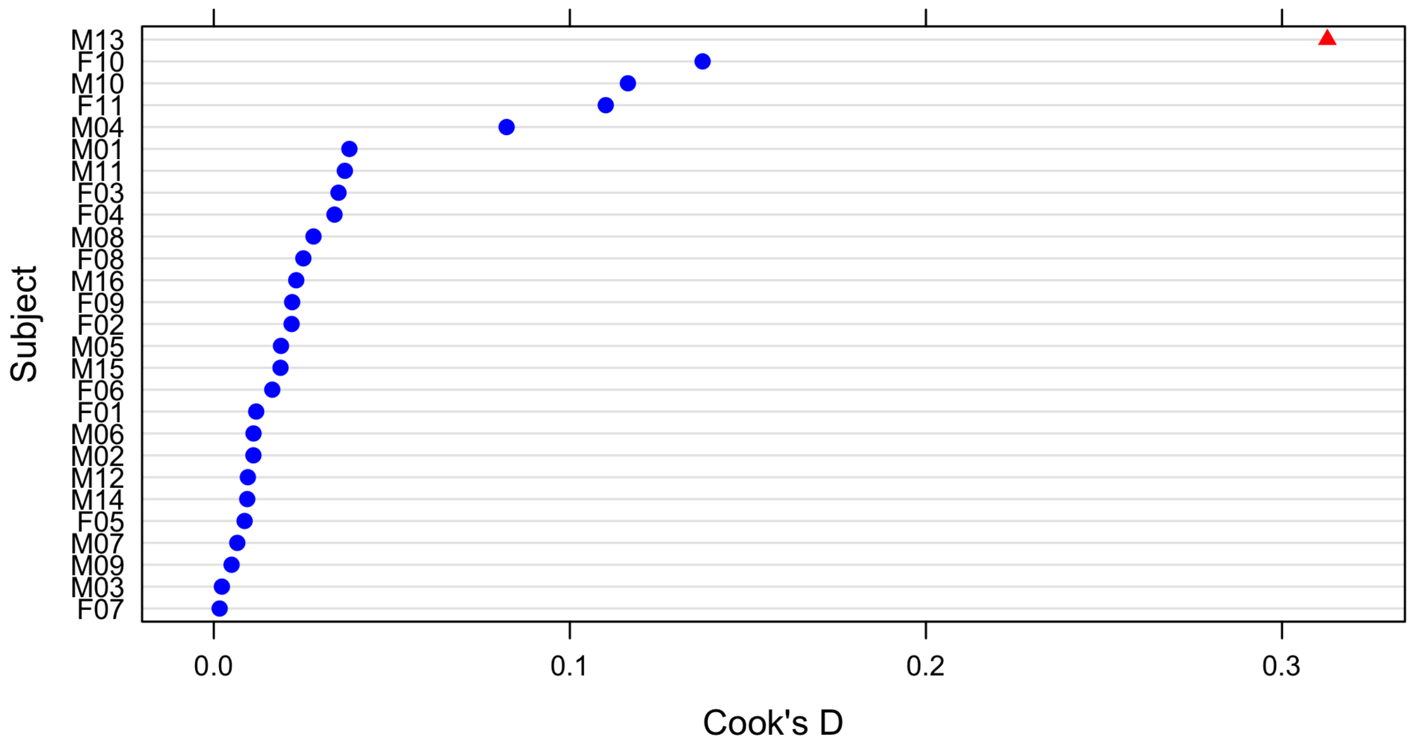
```
cooks.distance(m1.inf,sort=TRUE)
```

```
##            [,1]
## F07 0.001636431
## M03 0.002263374
## M09 0.004987931
## M07 0.006595594
## F05 0.008652440
## M14 0.009388496
## M12 0.009564971
## M02 0.011126080
## M06 0.011154742
## F01 0.011906428
## F06 0.016424195
## M15 0.018727158
## M05 0.018869728
## F02 0.021849514
## F09 0.022005758
## M16 0.023158496
## F08 0.025147749
## M08 0.027996778
## F04 0.033898438
## F03 0.035015311
## M11 0.036805752
## M01 0.038081771
## M04 0.082233065
## F11 0.110084164
## M10 0.116300386
## F10 0.137275747
## M13 0.312749412
```

STA 610L

Cook's distance

```
plot(m1.inf,which="cook",cutoff=4/length(unique(Orthodont$Subject)),
     sort=TRUE,xlab="Cook's D",ylab="Subject")
```



It's M13 again.

Other metrics

There are many other metrics we could use.

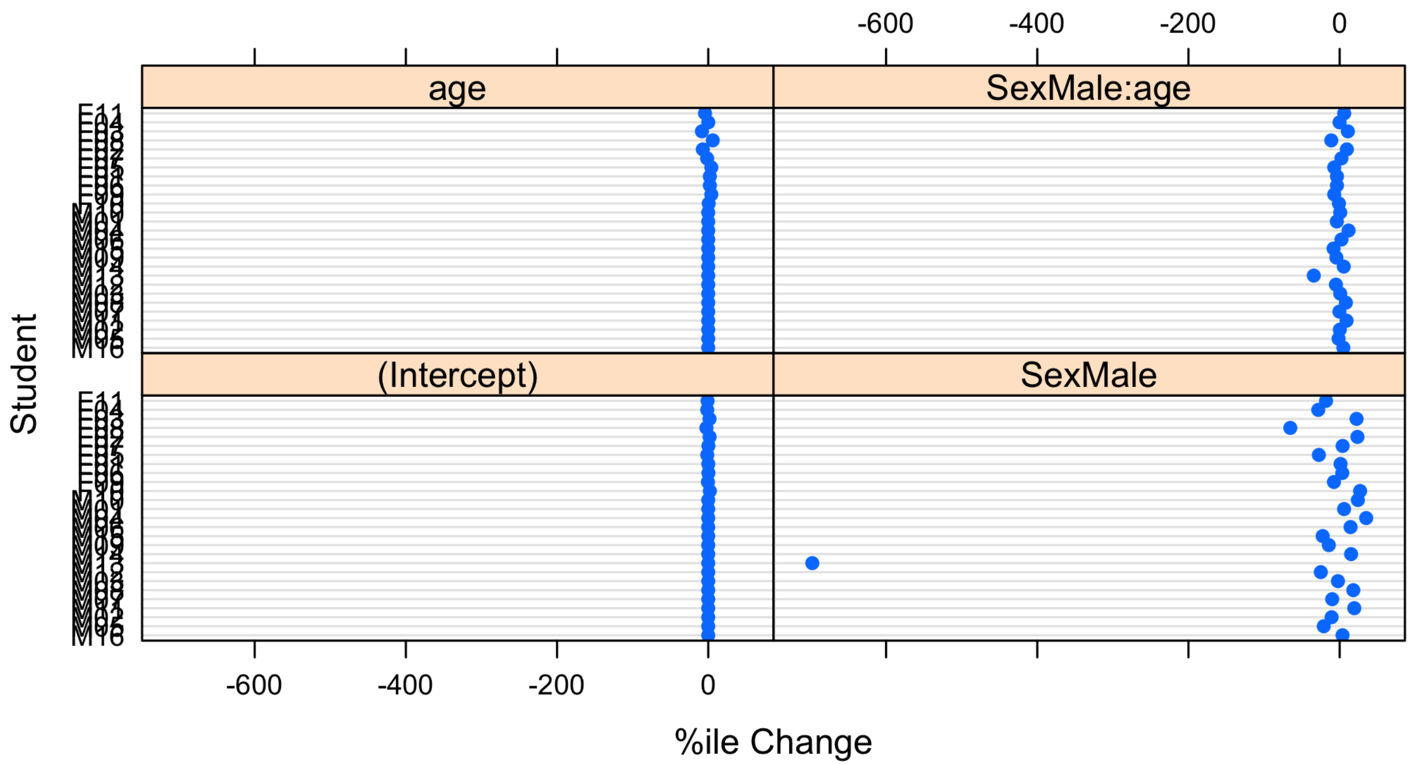
One option is the percentile change, which is defined as

$$\frac{\widehat{\gamma} - \widehat{\gamma}_{(-i)}}{\widehat{\gamma}} \times 100$$

STA 610L

Other metrics

```
plot(m1.inf,which="pchange",xlab="%ile Change",ylab="Student")
```



No surprise here!

Other metrics

Another metric is the changes in significance.

Basically, we evaluate whether excluding a group changes the statistical significance of any of the estimates in the model.

The user sets the critical value, and estimates that did not exceed it but do so when the group is removed, or *vice versa*, are flagged.

See the sigtest function.

STA 610L

Influence of lower-Level observations

We can also look at the influence of single lower-level observations.

They could be impactful in longitudinal data for example, when we have relatively few observations per individual.

Note however that the computational complexity of these deletion diagnostics will be increased in this case.

Here we look at Cook's Distance for the dental data on the individual observation level:

```
m1.inf.indiv <- influence(m1,obs=TRUE)
m1.cook <- cooks.distance(m1.inf.indiv)
infindiv <- m1.cook>4/length(Orthodont$distance)
```

Influence of lower-Level observations

```
data.frame(Orthodont$Subject,m1.cook,infindiv)[1:35,]
```

```
##    Orthodont.Subject      m1.cook infindiv
## 1                M01 1.169367e-02    FALSE
## 2                M01 2.774379e-03    FALSE
## 3                M01 4.697088e-04    FALSE
## 4                M01 7.815680e-03    FALSE
## 5                M02 4.198348e-04    FALSE
## 6                M02 9.929760e-05    FALSE
## 7                M02 1.636939e-03    FALSE
## 8                M02 3.592284e-03    FALSE
## 9                M03 7.489140e-03    FALSE
## 10               M03 1.259384e-03    FALSE
## 11               M03 1.122001e-03    FALSE
## 12               M03 6.194450e-03    FALSE
## 13               M04 1.190275e-02    FALSE
## 14               M04 3.784757e-03    FALSE
## 15               M04 2.816516e-04    FALSE
## 16               M04 1.718032e-02    FALSE
## 17               M05 6.509131e-03    FALSE
## 18               M05 1.211973e-03    FALSE
## 19               M05 2.141689e-03    FALSE
## 20               M05 1.591287e-03    FALSE
## 21               M06 2.957091e-03    FALSE
## 22               M06 6.019963e-06    FALSE
## 23               M06 4.049744e-07    FALSE
## 24               M06 6.023652e-06    FALSE
## 25               M07 1.482976e-03    FALSE
## 26               M07 1.400801e-03    FALSE
## 27               M07 2.536361e-05    FALSE
## 28               M07 6.789981e-04    FALSE
## 29               M08 3.310020e-02    FALSE
## 30               M08 3.653701e-03    FALSE
## 31               M08 1.862281e-05    FALSE
## 32               M08 1.793122e-03    FALSE
## 33               M09 1.297874e-03    FALSE
## 34               M09 1.601423e-02    FALSE
## 35               M09 2.421178e-02    FALSE
```

Influence of lower-Level observations

```
data.frame(Orthodont$Subject,m1.cook,infindiv)[36:72,]
```

```
##    Orthodont.Subject      m1.cook infindiv
## 36               M09 2.566986e-02    FALSE
## 37               M10 1.065729e-02    FALSE
## 38               M10 2.170243e-05    FALSE
## 39               M10 1.415284e-03    FALSE
## 40               M10 1.932585e-06    FALSE
## 41               M11 1.177022e-02    FALSE
## 42               M11 1.173807e-05    FALSE
## 43               M11 7.834679e-04    FALSE
## 44               M11 4.160895e-03    FALSE
## 45               M12 9.450493e-04    FALSE
## 46               M12 5.036144e-07    FALSE
## 47               M12 1.293549e-03    FALSE
## 48               M12 1.054221e-02    FALSE
## 49               M13 2.259216e-01     TRUE
## 50               M13 1.799954e-03    FALSE
## 51               M13 2.852317e-04    FALSE
## 52               M13 5.468146e-02     TRUE
## 53               M14 4.642457e-04    FALSE
## 54               M14 2.006172e-03    FALSE
## 55               M14 6.714294e-06    FALSE
## 56               M14 7.969978e-03    FALSE
## 57               M15 1.286213e-04    FALSE
## 58               M15 1.853975e-04    FALSE
## 59               M15 4.313165e-04    FALSE
## 60               M15 1.916918e-02    FALSE
## 61               M16 6.287876e-03    FALSE
## 62               M16 1.146909e-03    FALSE
## 63               M16 1.611651e-04    FALSE
## 64               M16 7.207025e-04    FALSE
## 65               F01 7.368422e-03    FALSE
## 66               F01 1.881906e-03    FALSE
## 67               F01 2.798211e-04    FALSE
## 68               F01 4.101701e-04    FALSE
## 69               F02 8.531024e-04    FALSE
## 70               F02 1.330744e-03    FALSE
## 71               F02 2.818954e-04    FALSE
## 72               F02 7.739192e-03    FALSE
```

Influence of lower-Level observations

```
data.frame(Orthodont$Subject,m1.cook,infindiv)[73:108,]
```

```
##     Orthodont.Subject      m1.cook infindiv
## 73                F03 2.241371e-02    FALSE
## 74                F03 1.442777e-03    FALSE
## 75                F03 9.758752e-05    FALSE
## 76                F03 4.093299e-03    FALSE
## 77                F04 1.943475e-03    FALSE
## 78                F04 2.522817e-04    FALSE
## 79                F04 2.100052e-05    FALSE
## 80                F04 1.540649e-03    FALSE
## 81                F05 2.795751e-04    FALSE
## 82                F05 1.065318e-03    FALSE
## 83                F05 3.979447e-04    FALSE
## 84                F05 1.723717e-03    FALSE
## 85                F06 3.054052e-08    FALSE
## 86                F06 3.035559e-05    FALSE
## 87                F06 7.255276e-04    FALSE
## 88                F06 8.462638e-05    FALSE
## 89                F07 2.439715e-05    FALSE
## 90                F07 2.393374e-06    FALSE
## 91                F07 2.942438e-04    FALSE
## 92                F07 3.162853e-03    FALSE
## 93                F08 1.036493e-02    FALSE
## 94                F08 3.252463e-05    FALSE
## 95                F08 4.482505e-05    FALSE
## 96                F08 2.786362e-03    FALSE
## 97                F09 6.698547e-05    FALSE
## 98                F09 2.252798e-05    FALSE
## 99                F09 1.276697e-04    FALSE
## 100               F09 1.178778e-02    FALSE
## 101               F10 2.092103e-02    FALSE
## 102               F10 3.245675e-04    FALSE
## 103               F10 2.257775e-04    FALSE
## 104               F10 7.369325e-03    FALSE
## 105               F11 1.083676e-03    FALSE
## 106               F11 2.344728e-04    FALSE
## 107               F11 3.109466e-03    FALSE
## 108               F11 1.202342e-03    FALSE
```

M13 once again!

Dealing with influential data

What to do with influential data is a much harder problem.

Reasonable strategies may include the following.

- Verify data recorded correctly
- Consider robust models
- Determine whether any lurking predictors should be added to the model
- Report results with and without overly influential results

What's next?

Move on to the readings for the next module!

STA 610L