MLM Nested Handout 4 (Spring 2022)

BLUPs, Residuals, Information Criteria, Level Notation (redux.), Pseudo-R^2

Predicting Random Effects: BLUPs

1. Predicting the *random effects* after fitting a model: these predictions are known as BLUPs.
   a. Why do this?
      i. We need to examine residuals that net out both fixed and random effects.
      ii. We should look at estimates of the random effects themselves to check our model assumptions (they should be reasonably normally distributed).
   b. Consider a much simpler model:

   $$MATHGAIN_{ijk} = b_0 + \zeta_k + \varepsilon_{ijk}, \text{ with } \zeta_k \sim N(0, \sigma_\zeta^2), \ \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \text{indep.}$$

      i. In this two-level model, there are two types of errors: subject-level and group level, with these sometimes being referred to as level 1 and 2, ε_{ijk} and ζ_k, respectively.
      ii. Residuals are *estimates* of these errors.
   c. After fitting a model, we have an overall sense of the impact of the fixed and random effects, the latter of which are described by estimated variance components, such as $\hat{\sigma}_\zeta^2$.
   d. We can use the parameter estimates from the model fits, along with the observed *outcomes* and *predictors* to make a best guess – a prediction[1] – as to where the ζ_k lie for each school k. Of course, this can be generalized in the presence of more fixed and random effects.

2. There is an explicit expression for 'best guesses' for ζ_k. These are

   $$\hat{\zeta}_k = E(\zeta_k | Y_{\cdot k}, X_{\cdot k}, \hat{\beta}, \hat{\sigma}_\zeta^2, \hat{\sigma}_\varepsilon^2),$$

   where $Y_{\cdot k}$ and $X_{\cdot k}$ denote all outcomes and predictors, respectively, associated with school k (including teachers and students in it).
   a. In words, this is the conditional expectation (mean) of the random effect, ζ_k, given the parameter estimates, the model, and the data relevant to school k.
      i. *Unconditionally,* meaning in the absence of any additional information, the mean is zero. I.e., $E(\zeta_k) = 0$; *conditional* on the observations and parameters, it is typically non-zero.
   b. It is known as the Best Linear Unbiased Predictor, or BLUP, for the random effects, because it is the "best" guess under certain criteria.
   c. We do not give the full formula for the BLUP (it requires matrix algebra), but we examine several of the simpler cases. The BLUP is a weighted average of what we might call 'fixed effects residuals' – the residuals we get after removing the population mean prediction.
      i. In our simple model, above, that would be $MATHGAIN_{ijk} - \hat{b}_0$. In a model such as $MATHGAIN_{ijk} = b_0 + b_1 SES_{ijk} + \zeta_k + \varepsilon_{ijk}$, the corresponding fixed effects residual would be: $MATHGAIN_{ijk} - (\hat{b}_0 + \hat{b}_1 SES_{ijk})$

      ii. The weighted average combines only the relevant observations: those that are in the same school, in our example.

---

[1]We call these predictions because they are made *after* estimating the parameters of the model. They are not estimates in the traditional sense.

iii. The formula for a single school effect is:

$$\hat{\zeta}_k | \hat{e}_{\cdot k}, \hat{\sigma}_\zeta^2, \hat{\sigma}_\varepsilon^2 = \frac{\hat{\sigma}_\zeta^2}{\hat{\sigma}_\varepsilon^2 + n_k \hat{\sigma}_\zeta^2} \hat{e}_{\cdot k},$$

where $\hat{e}_{\cdot k} = \sum_{i,j}(Y_{ijk} - X_{ijk}\hat{\beta})$ is the sum of the residuals associated with school k, n_k is the number of subjects in that school and $X_{ijk}\hat{\beta}$ is a shorthand for the prediction based on the fixed effects (you might be familiar with the term 'y-hat' used for such predictions. Y-hat is $\hat{b}_0$ in our model with a constant and $\hat{Y}_{ijk} = \hat{b}_0 + \hat{b}_1 SES_{ijk}$ in our model with SES).

iv. The math symbol '|' means "given," so the left-hand side of the formula reads, "zeta sub k given e-hat sub dot k, sigma-hat-squared sub zeta, sigma-hat-squared sub epsilon"

v. When $\hat{\sigma}_\zeta^2$ is large compared to $\hat{\sigma}_\varepsilon^2$, the formula reduces to approximately $\frac{\hat{\sigma}_\zeta^2}{n_k \hat{\sigma}_\zeta^2} \hat{e}_{\cdot k} = \frac{1}{n_k} \hat{e}_{\cdot k}$, which is simply the mean residual – exactly what we calculated in a prior handout.

vi. When $n_k = 1$, we have no *obvious* way to disentangle the single residual into structured and unstructured components (but we have to). The formula simplifies, though, and assists us in this procedure:

$$\hat{\zeta}_k | \hat{e}_{\cdot k}, \hat{\sigma}_\zeta^2, \hat{\sigma}_\varepsilon^2 = \frac{\hat{\sigma}_\zeta^2}{\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_\zeta^2} \hat{e}_{\cdot k} = ICC_\zeta \hat{e}_{\cdot k}$$

- The ICC (associated with component ζ) is a proportion between 0 and 1; the BLUP for the random effect is assigned as the ICC fraction of the residual for that case (the single residual is its own mean).
- A large ICC implies that the residual should be more strongly apportioned to the random effect and vice versa.
  - This makes sense: if we don't know how to partition a single observation into random effect and residual, we base the proportion on information from the full dataset (contained in the ICC).
- When variance components $\hat{\sigma}_\zeta^2$ and $\hat{\sigma}_\varepsilon^2$ are equal, the ICC=0.5, so exactly half of the residual is taken as structured, and the other half is the remainder, or residual. *Doesn't this make intuitive sense?*

vii. When n_k is much greater than 1, $\hat{\sigma}_\varepsilon^2$ will play a diminished role, as follows:
- Imagine (relatively speaking) large error variance $\hat{\sigma}_\varepsilon^2 \approx 10\hat{\sigma}_\zeta^2$, which is what we've seen in the classroom data. ICC<0.10. Then if we have a lot of observations in each school, say n_k=100, we can predict school effects quite precisely, and $\hat{\sigma}_\varepsilon^2$'s effect on the BLUP will be inconsequential.
- With a large within-group sample, the BLUP will tend to approach the value of the 'fixed effects (average) residual' for the school, since the fraction is approaching $1/n_k$.

d. If the model is a bit more complicated, such as:

$$MATHGAIN_{ijk} = b_0 + \eta_{jk} + \zeta_k + \varepsilon_{ijk},$$

with $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \eta_{jk} \sim N(0, \sigma_\eta^2), \zeta_k \sim N(0, \sigma_\zeta^2),$ indep.,

we can write an explicit expression for the BLUPs for a school k *with one classroom j in it*:

$$\hat{\eta}_{jk} | \hat{e}_{\cdot k}, \hat{\sigma}_\eta^2, \hat{\sigma}_\zeta^2, \hat{\sigma}_\varepsilon^2 = \frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\varepsilon^2 + n_k(\hat{\sigma}_\eta^2 + \hat{\sigma}_\zeta^2)} \hat{e}_{\cdot k}$$

$$\hat{\zeta}_k | \hat{e}_{\cdot k}, \hat{\sigma}_\eta^2, \hat{\sigma}_\zeta^2, \hat{\sigma}_\varepsilon^2 = \frac{\hat{\sigma}_\zeta^2}{\hat{\sigma}_\varepsilon^2 + n_k(\hat{\sigma}_\eta^2 + \hat{\sigma}_\zeta^2)} \hat{e}_{\cdot k},$$

2

where $\hat{e}_{\cdot k} = \sum_i (Y_{ijk} - X_{ijk}\hat{\beta})$ is the sum of the residuals associated with school k (and thus classroom j in that school as well), n_k and $X_{ijk}\hat{\beta}$ are as before.

    i. Most of the comments from the prior example hold here as well.

    ii. When $n_k = 1$, we see that the three fractions:

$$\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\varepsilon^2 + (\hat{\sigma}_\eta^2 + \hat{\sigma}_\zeta^2)}, \ \frac{\hat{\sigma}_\zeta^2}{\hat{\sigma}_\varepsilon^2 + (\hat{\sigma}_\eta^2 + \hat{\sigma}_\zeta^2)} \ and \ \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + (\hat{\sigma}_\eta^2 + \hat{\sigma}_\zeta^2)}$$

    divide the 'fixed effects residual' (sum) into three parts, proportional to the variance of that component.

       1. Since we are taking a residual and assigning portions of it to different types of error, it makes sense that if a specific unstructured error has the largest variance (relatively speaking), based on the model fit, then the largest portion of the residual should go to that component.

e. Matrix form (optional):

$$Y_{ijk} = b_0 + \zeta_k + \eta_{jk} + \varepsilon_{ijk}, \text{ with } \zeta_k \sim N(0, \sigma_\zeta^2), \ \eta_{jk} \sim N(0, \sigma_\eta^2) \ \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \text{indep.}$$

This translates, in matrix form (assuming complete balanced design for simplicity, and n_I students (each) in n_J classrooms in n_K schools) to:

$$Y = X\beta + Z\delta + \varepsilon \text{ with } \delta = \begin{vmatrix} \zeta & 0 \\ 0 & \eta \end{vmatrix}, \ \zeta \sim N(0_{n_K}, \sigma_\zeta^2 I_{n_K}), \ \eta \sim N(0_{n_J n_K}, \sigma_\eta^2 I_{n_J n_K}), \ \varepsilon \sim N(0_{n_I n_J n_K}, \sigma_\varepsilon^2 I_{n_I n_J n_K}),$$

independent of ea. other. Note that I_n is an $n \times n$ identity matrix and 0_n is a zero vector of length n. The design matrix X is simply a column of ones in our example, as it's just the intercept. The design matrix Z is more complex, but it identifies which component of the ζ or η corresponding to the school or classroom to 'pick.'

$$\text{Let } G = \begin{vmatrix} \sigma_\zeta^2 I_{n_K} & 0 \\ 0 & \sigma_\eta^2 I_{n_J n_K} \end{vmatrix}$$

$$\text{And let } \Sigma = ZGZ' + \sigma_\varepsilon^2 I_{n_I n_J n_K}, \text{ and } u = \begin{vmatrix} \zeta \\ \eta \end{vmatrix}$$

$$\text{Then } \hat{u} = \hat{G}\hat{Z}'\hat{\Sigma}^{-1}(Y - X\hat{\beta})$$

is the expression for the BLUPs of this model, where the 'hat' signifies estimation from a sample. You should recognize $(Y - X\hat{\beta})$ as the set of 'fixed effects residuals.'

3. Worked example:

$$MATHGAIN_{ijk} = b_0 + \zeta_k + \eta_{jk} + \varepsilon_{ijk}, \text{ with } \zeta_k \sim N(0, \sigma_\zeta^2), \ \eta_{jk} \sim N(0, \sigma_\eta^2) \ \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \text{indep.}$$

a. R code for model fit and prediction of BLUPs ($\hat{\zeta}_k, \hat{\eta}_{jk}$):

```
fit2 <- lmer(mathgain ~ 1 + (1 | schoolid/classid), data = dat)
print(summary(fit2))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathgain ~ 1 + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 11768.8
##
## Scaled residuals:
##    Min     1Q  Median     3Q    Max
```

```
## -4.6441 -0.5984 -0.0336  0.5334  5.6335
##
## Random effects:
##  Groups             Name          Variance Std.Dev.
##  classid:schoolid (Intercept)    99.22    9.961
##  schoolid         (Intercept)    77.50    8.804
##  Residual                      1028.23   32.066
## Number of obs: 1190, groups:  classid:schoolid, 312; schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)    57.427      1.443 93.851   39.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# generate BLUPs for random effs
ranefs <- ranef(fit2)
zetaM3 <- ranefs$schoolid[, 1]
etaM3 <- ranefs$classid[, 1]


# Alternative, using matrix alg.: reset classid to be 'in order'
Z <- cbind(model.matrix(~-1 + factor(schoolid), data = dat), model.matrix(~-1 + factor(classid),
    data = dat))
ehat <- dat$mathgain - predict(fit2, re.form = ~0)
sig2.eta <- as.numeric(VarCorr(fit2))[1]   #retrieves squared vals.
sig2.zeta <- as.numeric(VarCorr(fit2))[2]  #retrieves squared vals.
sig2.eps <- sigma(fit2)^2
# construct ZGZ'+I*sig.eps
nK <- length(unique(dat$schoolid))
nJ <- length(unique(dat$classid))
nI <- dim(dat)[1]
G <- diag(c(rep(sig2.zeta, nK), rep(sig2.eta, nJ)))
Sig <- Z %*% G %*% t(Z) + diag(rep(sig2.eps, nI))
Sig.inverse <- solve(Sig)
reffs <- G %*% t(Z) %*% Sig.inverse %*% ehat
zetaMatM3 <- reffs[1:nK]
etaMatM3 <- reffs[(nK + 1):(nK + nJ)]
cor(zetaM3, zetaMatM3)
```
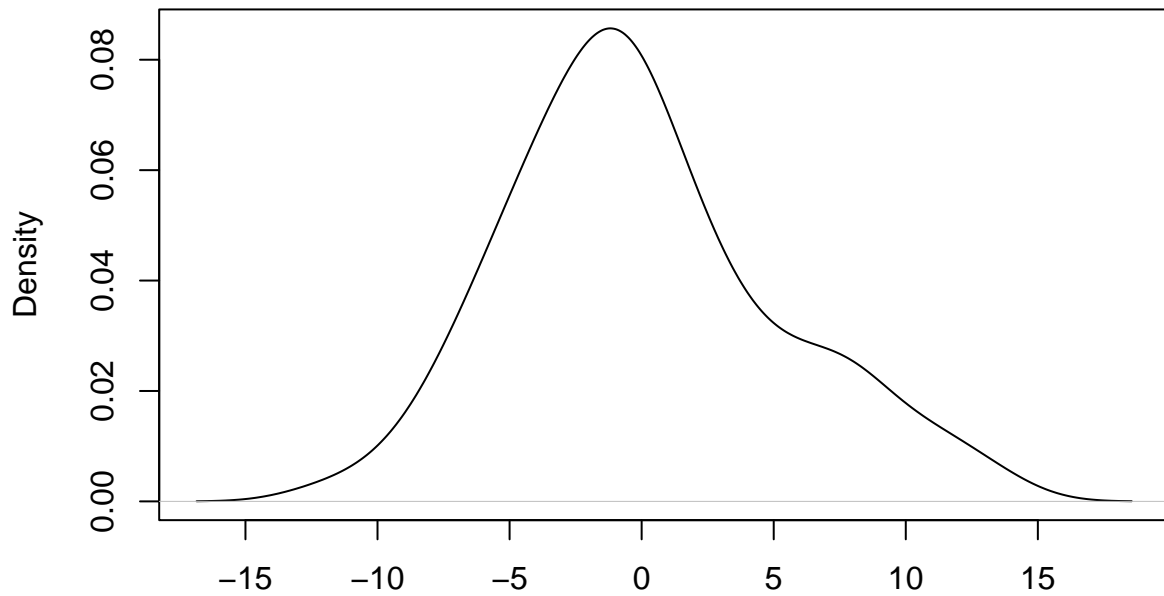
```
## [1] 1
```

```r
cor(etaM3, etaMatM3)
```

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

b. If we wish to examine the BLUPs themselves, we need only ONE per school or classroom depending on the BLUP (in R, this was how the raw values were produced).

```r
# NOTE: R stores the data in vectors that don't have the replication problem
plot(density(zetaM3))
```
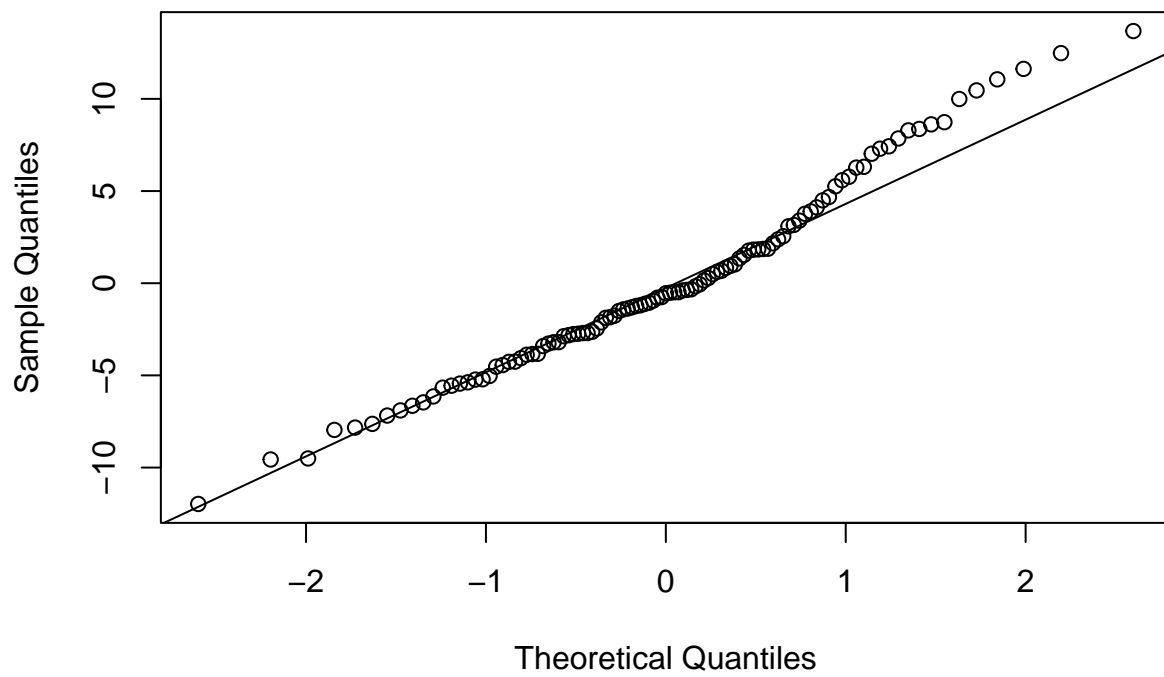
density.default(x = zetaM3)
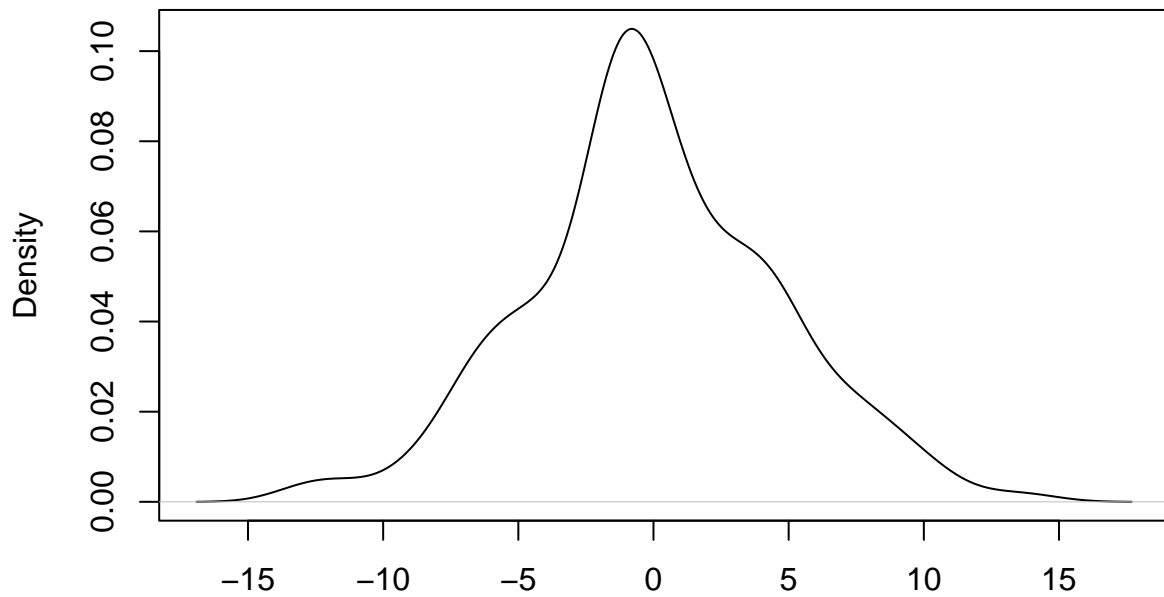

N = 107   Bandwidth = 1.625

```
qqnorm(zetaM3)
qqline(zetaM3)
```

Normal Q−Q Plot

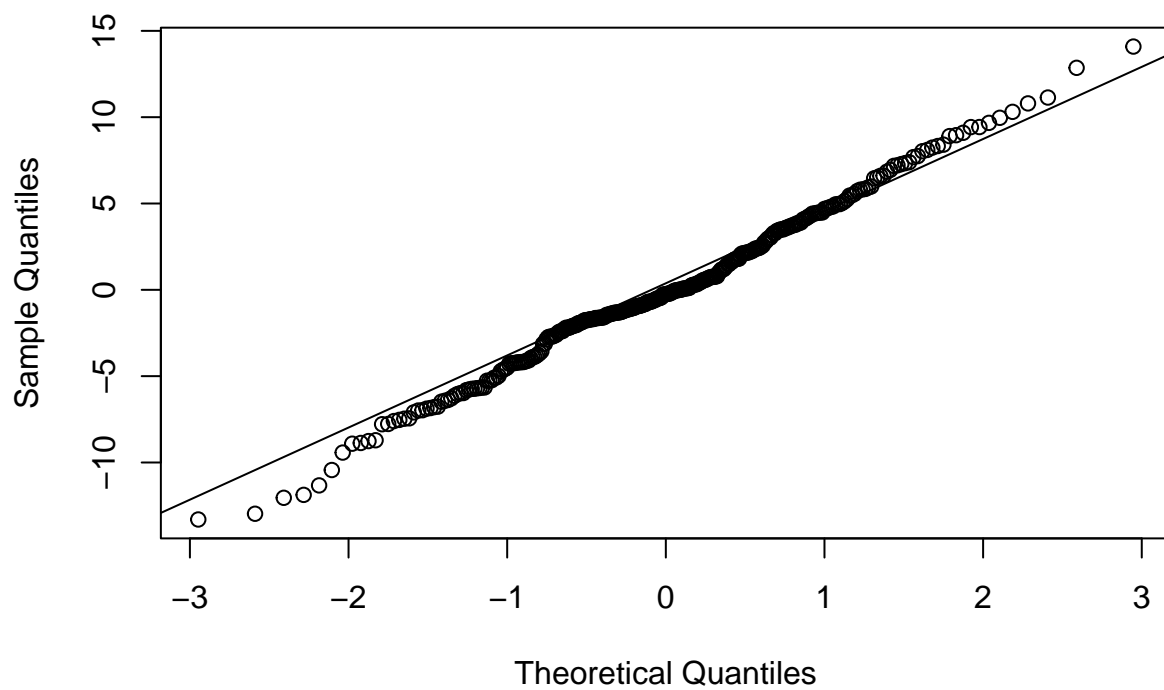Theoretical Quantiles

```
plot(density(etaM3))
```

## density.default(x = etaM3)

N = 312   Bandwidth = 1.2

```
qqnorm(etaM3)
qqline(etaM3)
```

## Normal Q–Q Plot

b. (cont.)
    i. There is some evidence of non-normality in the school effects, but it is probably tolerable (and sample is only 100 or so).
    ii. It seems the BLUPs for classroom effects are a bit more normal, at least looking at the q-q plot, though the q-q plot shows some concern with the left tail.

c. We often want to examine *residuals* based on the BLUPs (in other words, having netted out the random effects, or differences between schools and classrooms within them), and the next bit of code computes these directly from the model fit above.
    i. In the example, the *estimate* of the residual is: $\hat{\varepsilon}_{ijk} = MATHGAIN_{ijk} - (b_0 + \hat{\zeta}_k + \hat{\eta}_{jk})$
    ii. If we examine the density & q-q plots of those residuals, it's symmetric, but with heavier tails, especially the right tail.

```
# *now residuals predict resM3, residual
resM3 <- residuals(fit2)   #this has both school & classrm effects taken out
# kdensity resM3
plot(density(resM3))
```

density.default(x = resM3)

N = 1190   Bandwidth = 5.913
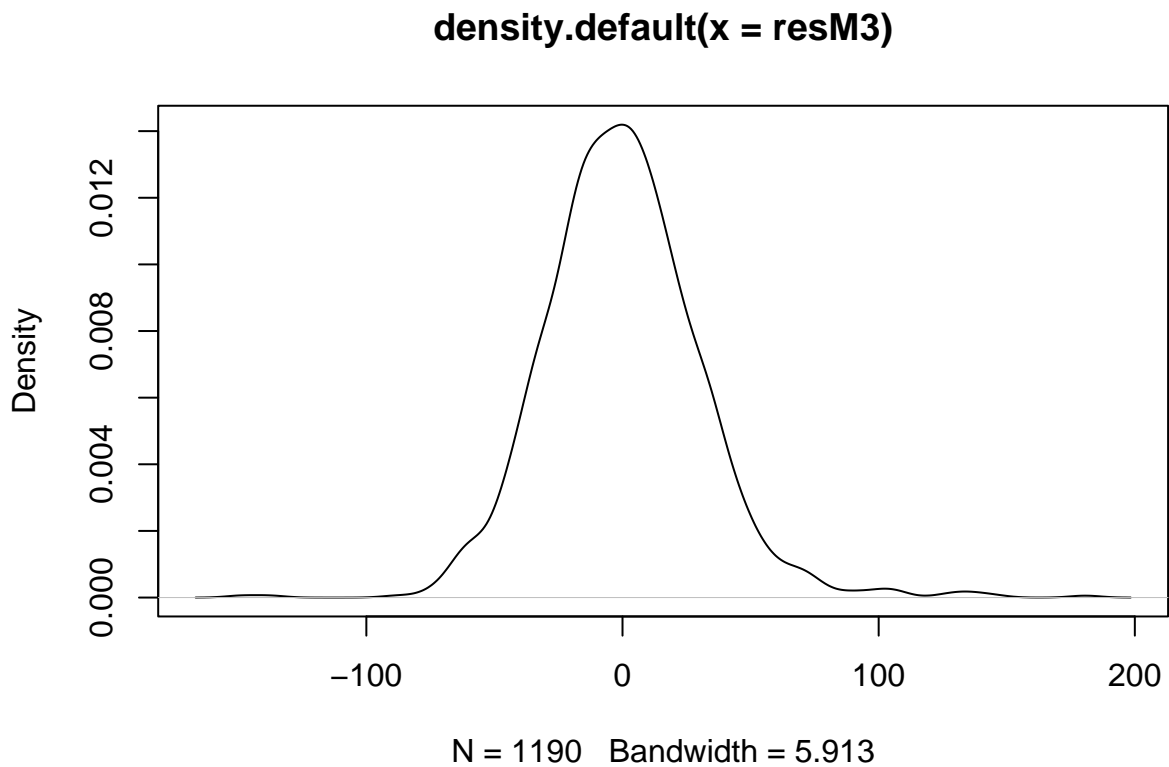
```
# qnorm resM3
qqnorm(resM3)
qqline(resM3)
```

Normal Q–Q Plot
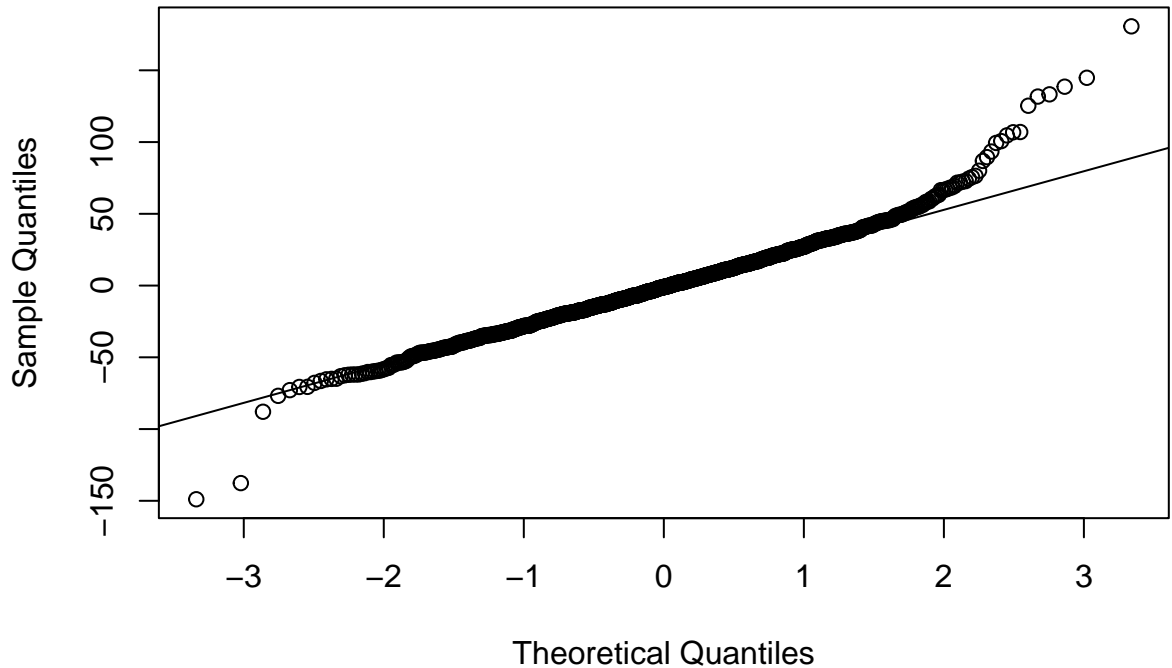
Examining residuals to uncover non-linearity

1. Just as in regression, residuals, particularly those associated with "level 1" or the subject level, are useful for uncovering non-linearity in the response to a covariate.
    a. For this discussion, our baseline model will consist of one layer of nesting and a single covariate: ses.

    $$Y_{ijk} = b_0 + b_1 SES_{ijk} + \zeta_k + \varepsilon_{ijk}, \text{ with } \zeta_k \sim N(0, \sigma_\zeta^2), \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \text{ indep.}$$

```
fit.ses <- lmer(mathkind ~ ses + (1 | schoolid), data = dat)
summary(fit.ses)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathkind ~ ses + (1 | schoolid)
##    Data: dat
##
## REML criterion at convergence: 12038.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7423 -0.5716  0.0188  0.6343  3.7763
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  309     17.58
##  Residual             1308     36.16
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
```

```
## (Intercept)   465.746      2.058   98.752 226.337  < 2e-16 ***
## ses            10.722      1.589 1187.995   6.747 2.35e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## ses 0.034
```

a. (cont.)
  i. Note that SES is significant – that doesn't mean that you are capturing the functional form appropriately.
  ii. It is still advisable to examine residuals:
  1. If a non-linear effect is present, the interpretation of that predictor changes.
  2. If the mean is misspecified, so is the covariance, and this can lead to improper inference.
b. Residuals in an MLM can be constructed to reflect different 'levels.'
  i. In a two-level model such as this, the BLUPs for $\hat{\zeta}_k$ can be understood as group-level (or level 2) residuals.
  ii. The residuals most familiar to us, and the default residual produced by R, are those that net out the estimated fixed *and* random effects: subject level (or level 1) defined as follows:
  1. Given $\hat{b}_0$, $\hat{b}_1$ and BLUP $\hat{\zeta}_k$, define a subject-level residual to be: $\hat{\varepsilon}_{ijk} = Y_{ijk} - (\hat{b}_0 + \hat{b}_1 SES_{ijk} + \hat{\zeta}_k)$.
  iii. The R code to estimate these (after fitting a model) is given next.
  1. Here, 'res.ses' is the name of the variable to which the residual is assigned.

```
res.ses <- residuals(fit.ses)
```

b. (cont.)
  i. In some situations, it is useful to standardize the residuals in some manner.
  1. It can be especially important when searching for outliers by examining BLUPs for the random effects.
  ii. For this discussion, we do not standardize.
  iii. Examining residuals grouped by categorical predictors, or in a scatterplot against a continuous predictor can reveal omitted structure in the functional form, such as non-linearities.
  iv. Use of a scatterplot smoother, such as 'loess', will help us assess the functional form.
  1. 'Local' smoothing functions can be understood as a running (weighted) mean of the outcome across different levels of a predictor.

```
ggplot(data = dat, aes(x = ses, y = residuals(fit.ses))) + geom_point() + stat_smooth(method = "loe
```

```
## `geom_smooth()` using formula 'y ~ x'
```
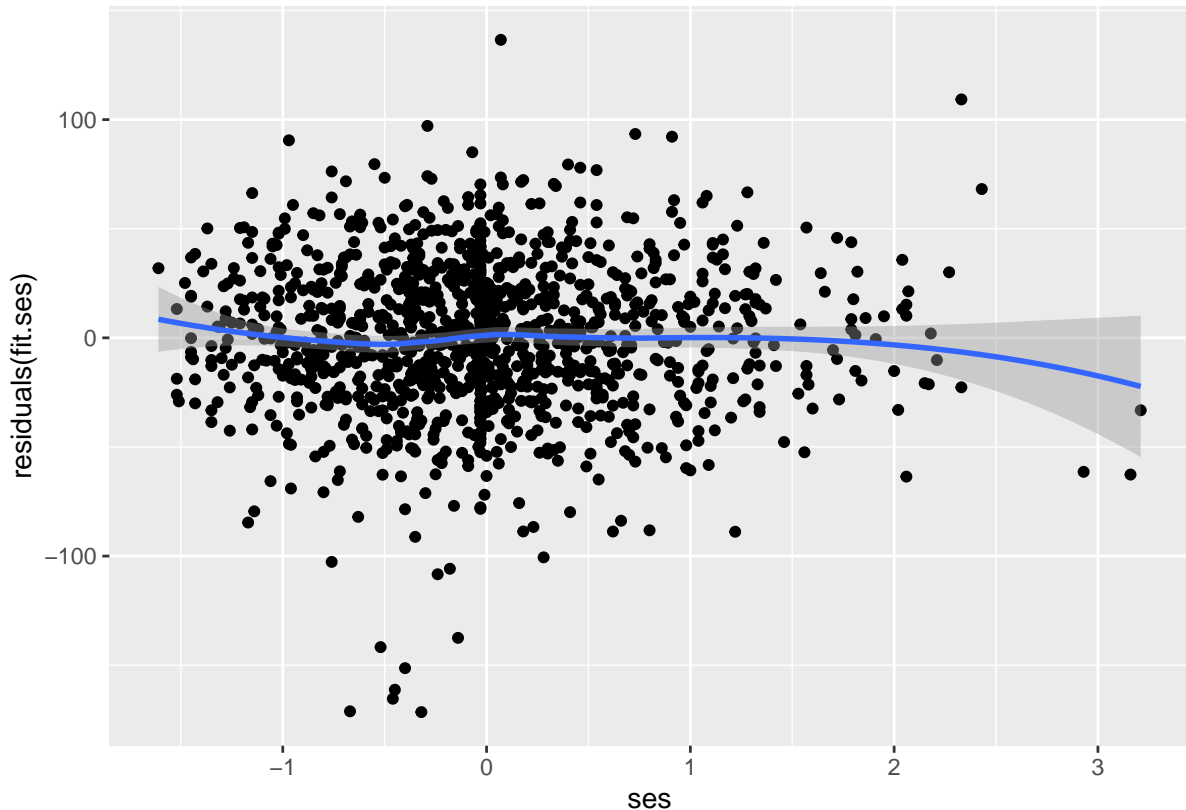
2. Notes on the residual plot:
   a. The loess (non-parametric) smooth suggests a cubic polynomial functional form.
   b. The confidence bounds suggest that the line $y = 0$ is fairly consistent with the residuals – implying weak to no evidence of non-linearity
   c. The area near $SES = 0$ could be deviating significantly from $y = 0$, so perhaps there is some non-linearity.
3. It is worth considering a polynomial functional form. These are relatively easy to specify and test.
   a. Given the residuals from the unconditional means model, we should at least consider a model with quadratic and then cubic terms for SES:

$$Y_{ijk} = b_0 + b_1 SES_{jk} + b_2 SES_{jk}^2 + b_3 SES_{jk}^3 + \zeta_k + \varepsilon_{ijk},$$

$$\text{with } \zeta_k \sim N(0, \sigma_\zeta^2), \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \text{ indep.}$$

   b. The code and results, including ANOVA to form the LR tests:

```
# try quadratic, then cubic NOTE: R can do this without adding new vars to the
# dataset using I()
fit.ses.2 <- lmer(mathkind ~ ses + I(ses^2) + (1 | schoolid), data = dat)
print(summary(fit.ses.2))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathkind ~ ses + I(ses^2) + (1 | schoolid)
##    Data: dat
##
## REML criterion at convergence: 12036.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7405 -0.5704  0.0182  0.6340  3.7741
```

10

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  309.2   17.58
##  Residual             1308.6   36.18
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error         df t value Pr(>|t|)
## (Intercept) 465.76629    2.18759  124.05297 212.913  < 2e-16 ***
## ses          10.73719    1.68649 1185.67733   6.367 2.76e-10 ***
## I(ses^2)     -0.03544    1.31759 1175.00854  -0.027    0.979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses
## ses       0.144
## I(ses^2) -0.339 -0.334
```

```
fit.ses.3 <- lmer(mathkind ~ ses + I(ses^2) + I(ses^3) + (1 | schoolid), data = dat)
print(summary(fit.ses.3))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathkind ~ ses + I(ses^2) + I(ses^3) + (1 | schoolid)
##    Data: dat
##
## REML criterion at convergence: 12030.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7133 -0.5662  0.0116  0.6388  3.8081
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  317.8   17.83
##  Residual             1302.5   36.09
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  464.930      2.241  130.641 207.445  < 2e-16 ***
## ses           13.822      2.259 1167.738   6.118 1.29e-09 ***
## I(ses^2)       2.479      1.795 1148.951   1.381   0.1674
## I(ses^3)      -2.068      1.002 1159.586  -2.064   0.0392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    I(s^2)
## ses      -0.015
## I(ses^2) -0.364  0.271
## I(ses^3)  0.179 -0.667 -0.680
```

```
anova(fit.ses, fit.ses.2, fit.ses.3)
```

```
## refitting model(s) with ML (instead of REML)
## Data: dat
## Models:
## fit.ses: mathkind ~ ses + (1 | schoolid)
## fit.ses.2: mathkind ~ ses + I(ses^2) + (1 | schoolid)
## fit.ses.3: mathkind ~ ses + I(ses^2) + I(ses^3) + (1 | schoolid)
##           npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.ses      4 12053 12073 -6022.4    12045
## fit.ses.2    5 12055 12080 -6022.4    12045 0.0006  1    0.98006
## fit.ses.3    6 12052 12083 -6020.2    12040 4.2248  1    0.03984 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
  b. (cont.) NOTES:
    i. The quadratic term is not significant, but the cubic term is.
    ii. The anova constructs sequential LR tests. These don't tell us much beyond the t-tests in the model summary.
    iii. The models were refit, since REML fits (the default) cannot be compared using LR tests.
    iv. Arguably, an important test is the *joint test* $H_0 : b_2 = b_3 = 0$. We can request this by including only the cubic model as a comparison in the anova call:

```
anova(fit.ses, fit.ses.3)
```

```
## refitting model(s) with ML (instead of REML)
## Data: dat
## Models:
## fit.ses: mathkind ~ ses + (1 | schoolid)
## fit.ses.3: mathkind ~ ses + I(ses^2) + I(ses^3) + (1 | schoolid)
##           npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.ses      4 12053 12073 -6022.4    12045
## fit.ses.3    6 12052 12083 -6020.2    12040 4.2254  2     0.1209
```

b. (cont.) NOTES:
    v. We have a non-significant joint test. Is this a contradiction? Not really, as it costs 2 degrees of freedom to include a cubic (and quadratic) term in the model, and the single term tests are 'hiding' that cost.

3. Examine residuals once more for the two different models. Was there any gain?

```
ggplot(data = dat, aes(x = ses, y = residuals(fit.ses.2))) + geom_point() + stat_smooth(method = "l
    ggtitle("Residuals from the Quadratic Model")
```
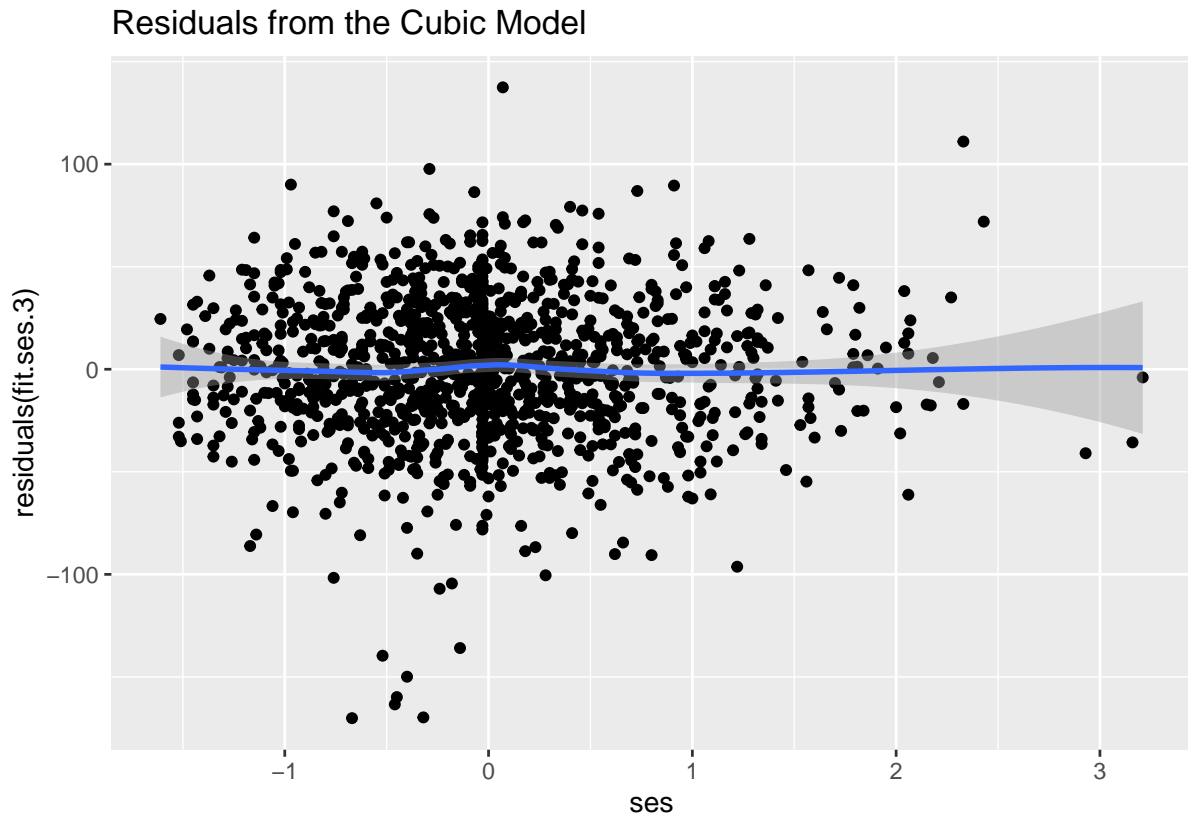
```
## `geom_smooth()` using formula 'y ~ x'
```

Residuals from the Quadratic Model
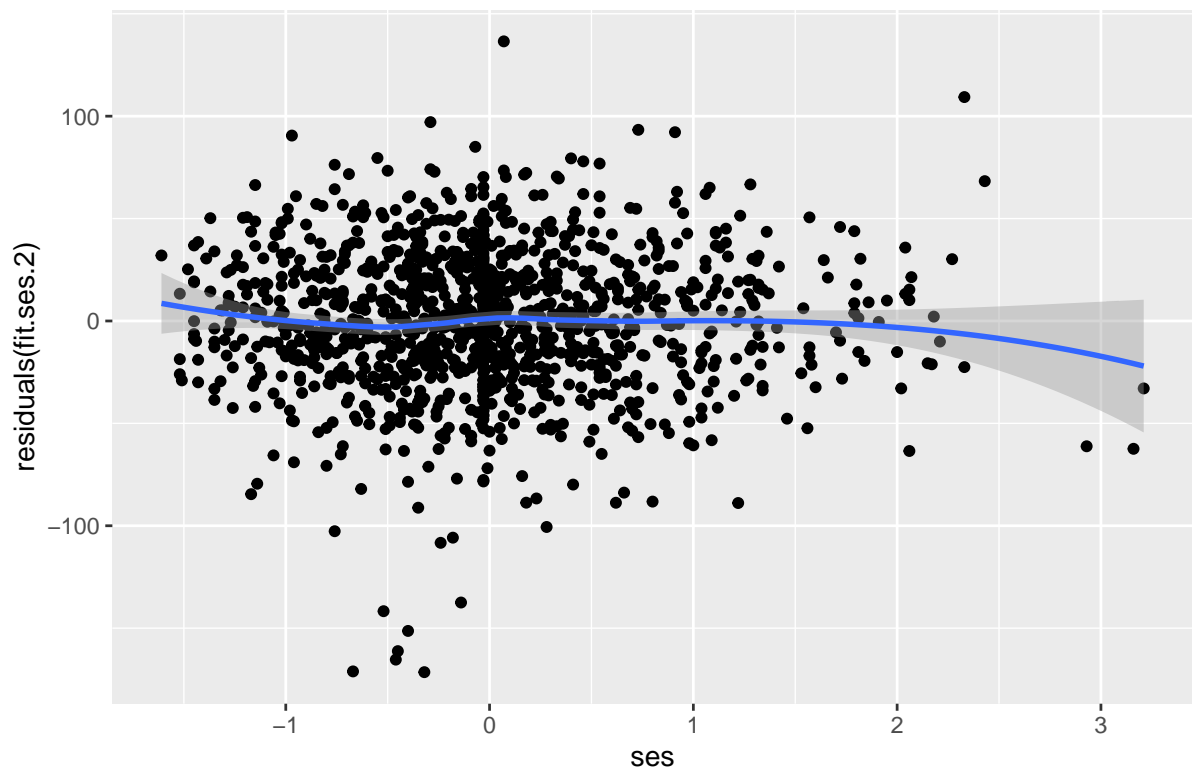
```
ggplot(data = dat, aes(x = ses, y = residuals(fit.ses.3))) + geom_point() + stat_smooth(method = "l
    ggtitle("Residuals from the Cubic Model")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Residuals from the Cubic Model

4. (cont.)

    i. Arguably, the cubic term helped, while the quadratic clearly didn't.

    ii. It is possible to defer to an Information Criterion approach (next section)

    iii. Re-examining the anova result, we see that AIC and BIC were also listed. We now examine these in greater detail.

Information Criteria

1. The models we've explored here have all been nested. When this is not the case, we need some additional model selection tools to compare non-nested models. We will use of Information Criteria (IC) in model selection in this situation. Even with nested models, there is some advantage to using IC, as it is closer to a "hands off" approach to model selection, reducing "researcher degrees of freedom."

    a. An information criterion is a measurement, based on the log-likelihood, used to assess goodness of fit (the better fitting model is selected).

        i. As long as the estimation methods are the same (REML vs. ML, e.g.) and the observations included in the estimation are identical, *this approach allows us to compare non-nested models (non-nested with respect to parameters).*

    b. The idea is to compare log-likelihoods, but a simple comparison won't work, because the likelihood should always improve, or at least won't get worse with the addition of new predictors.

        i. The idea is to 'penalize' the log-likelihood for the number of parameters.

        ii. Given two models with the same log-likelihoods and different numbers of parameters, an IC will favor the model with fewer parameters, satisfying the goal of 'parsimony' (the simpler model is preferred).

    c. We will describe the AIC (Akaike's IC) and BIC (Bayesian, or Schwarz's IC), but these are only two possibilities – there are many different choices – but these are the most common in use right now.[2]

---

[2]Note: there are emerging alternatives to AIC and BIC that are more suitable in the MLM setting. For now, AIC & BIC are

    i. The formula used in the AIC calculation is $AIC = -2\ln(L) + 2p$, where $\ln(L)$ is the loglikelihood of the model evaluated at the MLEs for the parameters and p is the number of parameters in the model (including the constant, all fixed effects and all random effects variances/covariances).
        1. We choose the model with the smaller AIC.
   ii. The formula used in the BIC calculation is $BIC = -2\ln(L) + p\ln(N)$, where $\ln(L)$ is the loglikelihood of the model evaluated at the MLEs for the parameters, p is the number of parameters in the model and N is the number of observations.
        1. We choose the model with the smaller BIC.
        2. There is some debate over whether N should be the number of observations or the number of subjects, when there are repeated measures for subjects. R uses the number of observations, and this yields a larger penalty term.
        3. A difference of 2 BIC points is understood as weakly significantly different. Larger differences are moderate and stronger evidence of one model's superiority to another.
  iii. There is a concern that the expression for $\ln(L)$ may vary with software implementation under REML estimation.
        1. The number of parameters in the model is also less clear in this instance, because REML estimation is conditional on the fixed effects.
        2. Given the above controversy, when comparing models with different fixed effects, ML not REML estimation is slightly preferred.
        3. In R, there are several ways to reveal the ICs.
- On a model fit, one can call the function: `BIC`
- To compare several models, store the estimates from each, and then call `anova` on the fit objects, listing them in increasing order of complexity.

d. Below we add a new model to our set that examine SES. Based on the idea that two different "regimes" or types of response, are present, we use a "switching regression" form in which the response is linear in SES until SES=-0.5 and then its slope is allowed to change. We also refit the models to make IC comparison more robust.

```
# switching regression:
fit.ses.switch <- lmer(mathkind ~ I(ses * (ses < -0.5)) + I(ses * (ses >= -0.5)) +
    (1 | schoolid), data = dat)
print(summary(fit.ses.switch))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathkind ~ I(ses * (ses < -0.5)) + I(ses * (ses >= -0.5)) + (1 |
##      schoolid)
##    Data: dat
##
## REML criterion at convergence: 12033.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7382 -0.5699  0.0107  0.6400  3.7928
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  308.3   17.56
##  Residual             1308.2   36.17
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##                         Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)              465.128      2.210 129.755 210.450  < 2e-16 ***
```

pretty good tools, but be aware that this is an active area of research and is subject to change.

```
## I(ses * (ses < -0.5))    8.965      2.797 1170.178   3.205 0.00139 **
## I(ses * (ses >= -0.5))   11.837      2.156 1178.756   5.490 4.91e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) I(*(<-
## I(*(<-0.5))  0.320
## I(*(>=-0.5) -0.224 -0.137
```

```
anova(fit.ses, fit.ses.switch, fit.ses.3, refit = T)
```

```
## refitting model(s) with ML (instead of REML)
## Data: dat
## Models:
## fit.ses: mathkind ~ ses + (1 | schoolid)
## fit.ses.switch: mathkind ~ I(ses * (ses < -0.5)) + I(ses * (ses >= -0.5)) + (1 | schoolid)
## fit.ses.3: mathkind ~ ses + I(ses^2) + I(ses^3) + (1 | schoolid)
##                 npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit.ses            4 12053 12073 -6022.4    12045
## fit.ses.switch     5 12054 12080 -6022.1    12044 0.5860  1    0.44397
## fit.ses.3          6 12052 12083 -6020.2    12040 3.6394  1    0.05643 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

   e.   i. If one were only using AIC, one might conclude that the cubic SES model was best.
      ii. BIC penalizes more heavily for too many parameters, and so it actually prefers a *linear* model in ses.
     iii. Neither IC chooses the switching regression, but we were at least able to compare these non-nested models!
     iv. There is no way to resolve the AIC/BIC difference. Each IC serves a different function. Many researchers prefer the use of AIC (based on more 'classical' assumptions) to BIC (based on Bayesian approach).
        1. I tend to use BIC when exploring a limited class of models (not looking for a 'true' model).
        2. Some statisticians would argue that BIC is useful when your goal is description (simpler stories are better) and that AIC is useful when your goal is prediction (more complex models may do a better job).

Level Notation and targeting variation – one example

1. Example of "Bryk/Raudenbush Level Notation" using Rat pup data
   a. Data description: 30 female rats were *randomly assigned* to receive one of three doses (none *control*, low and high) of an experimental compound expected to change birthweight. Several things happened during the study:
      i. Three of the rats given high doses died
     ii. The number of pups per litter varied from 2 to 18
    iii. ...the data are thus unbalanced in two ways
   b. Predictors / identifiers:
      i. Pup-based (level 1):
        1. Weight (birth weight – the outcome)
        2. Sex (male, female)
     ii. Litter-based (level 2 – these differ at the group level):
        1. Treatment (control, high, low, in that order)
        2. Litsize: litter size
        3. Litter: litter ID

c. First fit an unconditional means model M0; here it is, in level notation, for pup i in litter j:

$$\text{Level 1: } WEIGHT_{ij} = \beta_{0j} + \varepsilon_{ij}$$
$$\text{Level 2: } \beta_{0j} = b_{00} + \zeta_{0j}$$
$$\zeta_{0j} \sim N(0, \sigma_{\zeta_0}^2), \ \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2), \text{ indep.}$$

   i. NOTE: the extra zeros are part of the notation. The 1st subscript links all others to that equation on levels 2 and above.

```
require(foreign)
require(lme4)
require(lmerTest)
ratdat <- read.table("../Datasets/rat_pup.dat", header = T)
fit.M0 <- lmer(weight ~ 1 + (1 | litter), data = ratdat)
summary(fit.M0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ 1 + (1 | litter)
##    Data: ratdat
##
## REML criterion at convergence: 468.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.2978 -0.4358 -0.0119  0.4837  3.1972
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.3004   0.5481
##  Residual             0.1963   0.4431
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   6.1953     0.1091 24.1543   56.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

   d. The variance components are: $\sigma_{\zeta_0}^2 = 0.30$; $\sigma_{\varepsilon}^2 = 0.20$. There is more between-litter variation than within.

2. There were two treatment levels associated with the dosage, and it is clearly of interest to see whether they had an effect. This will be our model M1. These are entered in level 2, since treatment was assigned at the litter-level:

$$\text{Level 1: } WEIGHT_{ij} = \beta_{0j} + \varepsilon_{ij}$$
$$\text{Level 2: } \beta_{0j} = b_{00} + b_{01}TREAT2_j + b_{02}TREAT3_j + \zeta_{0j}$$
$$\zeta_{0j} \sim N(0, \sigma_{\zeta_0}^2), \ \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2), \text{ indep.}$$

   a. $TREAT$n is an indicator for treatment code 2 (high) or 3 (low).
      i. Note: the reference is the control group, but treatment codes are based on the alphabetical ordering of 'High' and 'Low'.

```
fit.M1 <- lmer(weight ~ factor(treatment) + (1 | litter), data = ratdat)
summary(fit.M1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ factor(treatment) + (1 | litter)
##    Data: ratdat
##
## REML criterion at convergence: 467.1
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -7.3265 -0.4246 -0.0269  0.4876  3.1621
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.2770   0.5263
##  Residual             0.1966   0.4433
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##                       Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)             6.4533     0.1716 21.1614  37.598   <2e-16 ***
## factor(treatment)High  -0.3944     0.2696 21.8031  -1.463   0.1577
## factor(treatment)Low   -0.4287     0.2435 21.3452  -1.761   0.0926 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) fct()H
## fctr(trtm)H -0.637
## fctr(trtm)L -0.705  0.449
```

```
# test _Itreatment_2 _Itreatment_3
print(wald.test(b = fixef(fit.M1), Sigma = summary(fit.M1)$vcov, Terms = 2:3))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 3.7, df = 2, P(> X2) = 0.16
```

    b. The Wald test is not significant.

        i. A preliminary conclusion might be that the treatment was ineffective.

        ii. But there are several *potential confounders* (even though this is a designed experiment) that we are ignoring.

3. The litters vary in their sex-ratios. If birth weights tend to be higher for one sex, then litter effects and sex-ratios are confounded, and this could spillover to the estimated treatment effects.[3]

    a. The simplest way to control for sex-specific birth weight is to add sex as a control. Sex is a level-1 control, as each pup is a different sex (and we are targeting level 1 variance).

    b. The model M2 is adjusted as follows:

$$\text{Level 1: } WEIGHT_{ij} = \beta_{0j} + \beta_{1j}SEX_{ij} + \varepsilon_{ij}$$
$$\text{Level 2: } \beta_{0j} = b_{00} + b_{01}TREAT2_j + b_{02}TREAT3_j + \zeta_{0j}$$
$$\beta_{1j} = b_{10}$$
$$\zeta_{0j} \sim N(0, \sigma_{\zeta_0}^2), \ \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2), \text{ indep.}$$

---

[3]We are of course assuming that treatment did not change the sex ratios.

c. Here, we use R's factor() command to make the (sex) reference group clearer in the output.

```
fit.M2 <- lmer(weight ~ factor(treatment) + factor(sex) + (1 | litter), data = ratdat)
summary(fit.M2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ factor(treatment) + factor(sex) + (1 | litter)
##    Data: ratdat
##
## REML criterion at convergence: 419.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.4648 -0.4501  0.0175  0.5420  3.1383
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.3259   0.5709
##  Residual             0.1636   0.4045
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##                        Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)              6.2450     0.1867   22.5619  33.452  < 2e-16 ***
## factor(treatment)High   -0.3547     0.2893   22.0459  -1.226    0.233
## factor(treatment)Low    -0.3747     0.2617   21.7403  -1.432    0.166
## factor(sex)Male          0.3613     0.0478  295.0721   7.558 5.18e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) fct()H fct()L
## fctr(trtm)H -0.633
## fctr(trtm)L -0.701  0.450
## factr(sx)Ml -0.150  0.017  0.026
```

c. (cont.)
  i. Sex is significant (and positive for males).
  ii. The variance components are: $\sigma_{\zeta_0}^2 = 0.33$; $\sigma_{\varepsilon}^2 = 0.16$.
  iii. The residual variance has decreased, and this is what was targeted by the new model.
  iv. But notably, there was an increase in between-litter variation (re-shuffling is common – once we get the mean right)
d. The Wald test still fails:

```
# test _Itreatment_2 _Itreatment_3
print(wald.test(b = fixef(fit.M2), Sigma = summary(fit.M2)$vcov, Terms = 2:3))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 2.5, df = 2, P(> X2) = 0.29
```

e. Maybe there are additional confounders to consider?

4. Litter size is also a potential confounder (here, it's more complex, in that litter size itself could be an

outcome[4], but we ignore this).

    a. Litter size is a level-2 control, as each litter is a different size.

        i. We are targeting level 2 variation, since this size affects all rat pups in the litter (modeled as a shift in level β_{0j}).

    b. The model M3 is adjusted as follows:

$$\text{Level 1: } WEIGHT_{ij} = \beta_{0j} + \beta_{1j}SEX_{ij} + \varepsilon_{ij}$$
$$\text{Level 2: } \beta_{0j} = b_{00} + b_{01}TREAT2_j + b_{02}TREAT3_j + b_{03}LITSIZE_j + \zeta_{0j}$$
$$\beta_{1j} = b_{10}$$
$$\zeta_{0j} \sim N(0, \sigma^2_{\zeta_0}), \ \varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon}), \ \text{indep.}$$

    c. Here is the R code and fit, with an adjustment to the reference category for sex:

```
# reverse coding:
ratdat$sex <- factor(ratdat$sex, levels = c("Male", "Female"))
fit.M3 <- lmer(weight ~ factor(treatment) + factor(sex) + litsize + (1 | litter),
    data = ratdat)
summary(fit.M3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ factor(treatment) + factor(sex) + litsize + (1 | litter)
##    Data: ratdat
##
## REML criterion at convergence: 397
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.5743 -0.4699  0.0111  0.5712  3.0603
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.0974   0.3121
##  Residual             0.1628   0.4035
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##                        Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)             8.30987    0.27371   32.61126  30.360  < 2e-16 ***
## factor(treatment)High  -0.85870    0.18181   24.97854  -4.723 7.65e-05 ***
## factor(treatment)Low   -0.42850    0.15040   22.90425  -2.849   0.0091 **
## factor(sex)Female      -0.35908    0.04749  301.82484  -7.562 4.81e-13 ***
## litsize                -0.12900    0.01879   31.67409  -6.864 9.63e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) fct()H fct()L fct()F
## fctr(trtm)H -0.581
## fctr(trtm)L -0.300  0.423
## fctr(sx)Fml -0.111 -0.009 -0.040
## litsize     -0.920  0.390  0.035  0.043
```

    c. (cont.)

        i. Everything becomes significant. *Interpretation (... all else equal):*

---

[4]We are of course assuming that treatment did not change the sex ratios.

1. LITSIZE: the bigger the litter, the smaller the birthweight.
2. SEX: female rat pups are smaller than males.
3. Low dose TREATMENT reduces birthweight less than high dose.
   ii. The variance components are: $\sigma^2_{\zeta_0} = 0.10$; $\sigma^2_\varepsilon = 0.16$.
   1. The targeted, level-2 variance drops dramatically.
   iii. The Wald test is highly significant:

```r
print(wald.test(b = fixef(fit.M3), Sigma = summary(fit.M3)$vcov, Terms = 2:3))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 23.2, df = 2, P(> X2) = 9.2e-06
```

5. Exploring interactions
   i. Treatment effects are not always homogeneous. They might vary by sex.
   1. So far, our sex effect simply allows male & female pups to have a different birthweight, regardless of treatment group.
   2. Now we allow for *treatment* effects to vary by sex
   ii. The new model is:

$$\text{Level 1: } WEIGHT_{ij} = \beta_{0j} + \beta_{1j}SEX_{ij} + \varepsilon_{ij}$$
$$\text{Level 2: } \beta_{0j} = b_{00} + b_{01}TREAT2_j + b_{02}TREAT3_j + b_{03}LITSIZE_j + \zeta_{0j}$$
$$\beta_{1j} = b_{10} + b_{11}TREAT2_j + b_{12}TREAT3_j$$
$$\zeta_{0j} \sim N(0, \sigma^2_{\zeta_0}), \ \varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon), \text{ indep.}$$

iii. The interaction terms follow from multiplying out the terms multiplying β_{1j} in the level 1 equation.

```r
fit.M4 <- lmer(weight ~ factor(treatment) * factor(sex) + litsize + (1 | litter),
    data = ratdat)
summary(fit.M4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ factor(treatment) * factor(sex) + litsize + (1 | litter)
##    Data: ratdat
##
## REML criterion at convergence: 401.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.4725 -0.5001  0.0291  0.5735  3.0096
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.09652  0.3107
##  Residual             0.16349  0.4043
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##                                     Estimate Std. Error       df t value
## (Intercept)                          8.32334    0.27333 32.87352  30.452
## factor(treatment)High               -0.90606    0.19154 30.81114  -4.730
```

```
## factor(treatment)Low                         -0.46704    0.15818  28.15456  -2.953
## factor(sex)Female                            -0.41169    0.07315 295.30143  -5.628
## litsize                                      -0.12838    0.01875  31.79751  -6.846
## factor(treatment)High:factor(sex)Female       0.10702    0.13176 303.74417   0.812
## factor(treatment)Low:factor(sex)Female        0.08387    0.10568 298.58691   0.794
##                                             Pr(>|t|)
## (Intercept)                                 < 2e-16 ***
## factor(treatment)High                       4.71e-05 ***
## factor(treatment)Low                          0.0063 **
## factor(sex)Female                           4.24e-08 ***
## litsize                                     9.94e-08 ***
## factor(treatment)High:factor(sex)Female       0.4173
## factor(treatment)Low:factor(sex)Female        0.4281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) fct()H fct()L fct()F litsiz f()H:(
## fctr(trtm)H -0.562
## fctr(trtm)L -0.297  0.404
## fctr(sx)Fml -0.111  0.158  0.191
## litsize     -0.916  0.363  0.022  0.001
## fctr()H:()F  0.043 -0.323 -0.106 -0.555  0.020
## fctr()L:()F  0.044 -0.096 -0.320 -0.692  0.036  0.385
```

```r
print(wald.test(b = fixef(fit.M4), Sigma = summary(fit.M4)$vcov, Terms = 6:7))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 0.93, df = 2, P(> X2) = 0.63
```

a. The Wald test on the interaction proves to be non-significant.

7. We revert to the model without the treatment×sex interaction and take a quick look at residuals (male=red; female=blue).

```
ratdat$res1M3 <- residuals(fit.M3)
bool <- ratdat$sex == "Female"
ggplot(data = ratdat, aes(y = res1M3, x = litsize)) + geom_point(aes(colour = factor(sex))) +
    geom_smooth(aes(colour = factor(sex)))

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```
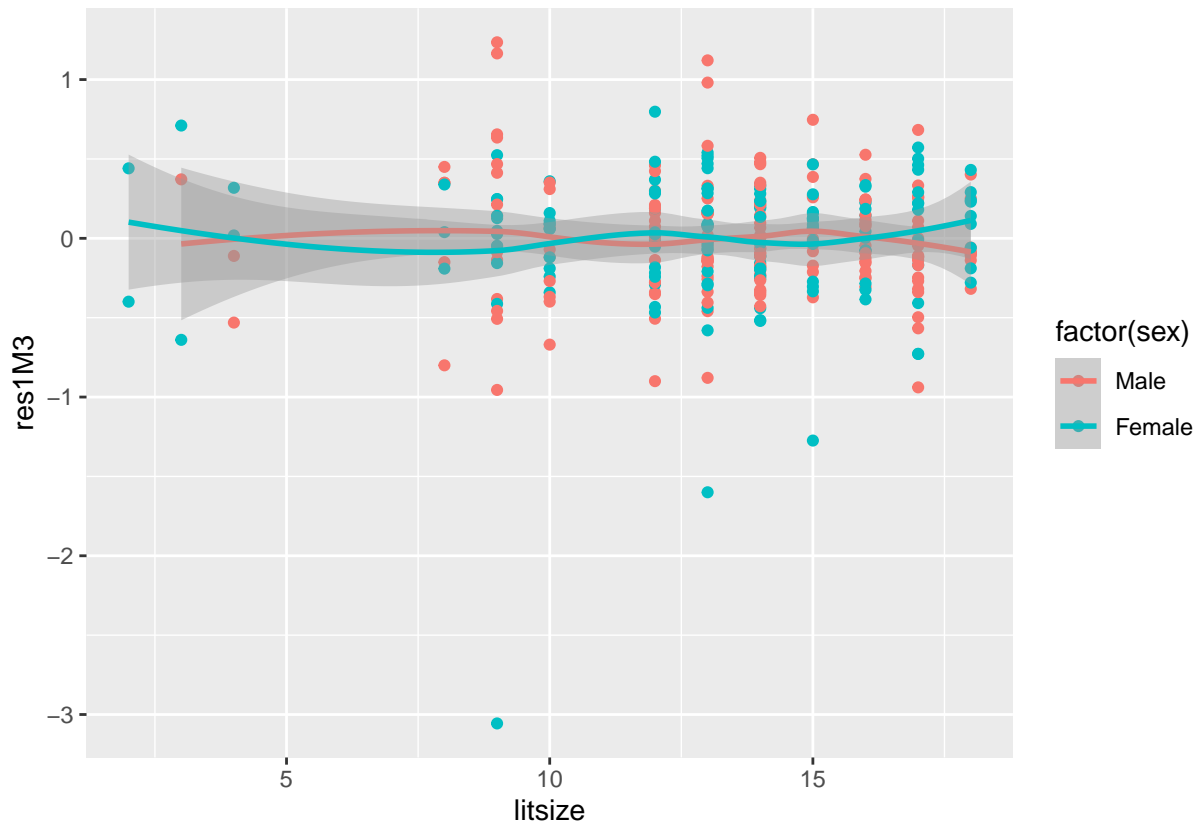
a. While some minor patterning is suggested, overall, this does not seem like a non-linear effect has been omitted, nor is there strong evidence of heteroscedasticity.

8. If we wanted to look at treatment X littersize interactions, these would be entered on level 2:

```
fit.M5 <- lmer(weight ~ treatment + litsize + factor(treatment):litsize + (1 | litter),
    data = ratdat)
summary(fit.M5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ treatment + litsize + factor(treatment):litsize + (1 |
##     litter)
##    Data: ratdat
##
## REML criterion at convergence: 450.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.4343 -0.4659 -0.0136  0.5037  3.0826
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  litter   (Intercept) 0.07426  0.2725
##  Residual             0.19549  0.4421
## Number of obs: 322, groups:  litter, 27
##
## Fixed effects:
##                                Estimate Std. Error       df t value Pr(>|t|)
```

23

```
## (Intercept)                          8.07191    0.35968 28.72248   22.442   < 2e-16 ***
## treatmentHigh                       -0.30939    0.54000 31.55602   -0.573    0.571
## treatmentLow                        -0.88554    0.54233 37.14040   -1.633    0.111
## litsize                             -0.12220    0.02579 26.76752   -4.738 6.28e-05 ***
## litsize:factor(treatment)High -0.05855    0.04745 29.42653   -1.234    0.227
## litsize:factor(treatment)Low   0.03131    0.03949 34.47790    0.793    0.433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) trtmnH trtmnL litsiz lt:()H
## treatmntHgh -0.666
## treatmentLw -0.663  0.442
## litsize     -0.964  0.642  0.639
## ltsz:fct()H  0.524 -0.947 -0.348 -0.544
## ltsz:fct()L  0.630 -0.419 -0.968 -0.653  0.355
```

```
print(wald.test(b = fixef(fit.M5), Sigma = summary(fit.M5)$vcov, Terms = 5:6))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 3.3, df = 2, P(> X2) = 0.2
```

Pseudo-R^2 calculations

1. One way to assess the fit of an MLM is through changes in the variance components.
   a. Singer & Willett suggest using these two measures:
      i. $R_W^2 = \frac{\hat{\sigma}_\varepsilon^2(M_0) - \hat{\sigma}_\varepsilon^2(M_1)}{\hat{\sigma}_\varepsilon^2(M_0)}$ – this is the proportion reduction in *residual* (within) variance as we move from model M_0 to M_1.
      ii. $R_B^2 = \frac{\hat{\sigma}_\zeta^2(M_0) - \hat{\sigma}_\zeta^2(M_1)}{\hat{\sigma}_\zeta^2(M_0)}$ – this is the proportion reduction in between-subject variance as we move from model M_0 to M_1.
      iii. Clearly this approach works well when the only systematic variation is in the intercept term.
         1. It can be used in models with random slopes in one of two ways: ignoring the random slope component, *or as a function of the predictors that they modify.*
   b. Referring to the Rat Pup Data, if we compare model M0 to M3:
      i. $R_W^2 = \frac{.20 - .16}{.20} = 0.20$
      ii. $R_B^2 = \frac{.30 - .10}{.30} = 0.67$
      iii. The within-litter variation was reduced by about 20% with the introduction of the full set of predictors in M3.
      iv. The between-litter variation was reduced by 67% with the introduction of the full set of predictors in M3 – mostly with the introduction of litter size.