B.3 Case Studies Revisited - Outline

Rat Pup Study

30 female rats received randomly assigned drug dose (control, low, or high). Birth weights of their pups are then compared.

2 level model: Level 1 is ratpup, level 2 is litter.

Dependent variable: rat pup birth weight

Fixed effect covariates: sex, treatment, litter size, treatment * sex interaction

Random effects: random intercept for each litter

Analysis Plan

We will use a 'top-down' strategy.

1. Start with 'loaded mean structure' that includes as many fixed efffect covariates as possible. Here, treatment, sex, litter size, and treatment * sex interaction.
2. Choose structure for random effects. Here that means having random intercepts. Test whether or not we need to include random intercepts
3. Select covariance structure for residuals. Test same for all treatment groups vs. different for all 3 vs. one variance for control and a 2nd for treatment groups
4. Reduce the model by testing whether or not certain fixed-effects are needed. Here, test if we can drop sex term, and test to see if we can drop treatment term.

Loaded Mean Structure Model (Model 3.1)

Include all possible fixed effects, plus random intercepts.
Model 3.1:

$$\begin{aligned}\text{Weight}_{ij} = \beta_0 &+ \beta_1 \times \text{High}_j + \beta_2 \times \text{Low}_j + \beta_3 \times \text{Female}_{ij} \\ &+ \beta_4 \times \text{Litsize}_j + \beta_5 \times \text{High}_j \times \text{Female}_{ij} \\ &+ \beta_6 \times \text{Low}_j \times \text{Female}_{ij} + u_j + \varepsilon_{ij} \\ u_j &\sim N(0, \sigma^2_{\text{litter}}) \\ \varepsilon_{ij} &\sim N(0, \sigma^2)\end{aligned}$$

This model has 9 parameters. β_0 through β_6 are the first 7, $\sigma^2, \sigma^2_{\text{litter}}$ are the other 2.
Want to test if we need to include random intercepts. Compare with model with only fixed effect terms. Or equivalently, $\sigma^2_{\text{litter}} = 0$

Model 3.1 Output

Our loaded means model 3.1 is represented in R as

```
> model3.1.fit <- lme(weight ~ treatment + sex1 + litsize +
    treatment*sex1, random = ~1 | litter, ratpup, method = "REML")
> summary(model3.1.fit)
Linear mixed-effects model fit by REML
 Data: ratpup
       AIC      BIC    logLik
  419.1043 452.8775 -200.5522
Random effects:
 Formula: ~1 | litter
        (Intercept) Residual
StdDev:   0.3106722 0.404337
Fixed effects: weight ~ treatment + sex1 + litsize + treatment * sex1
                     Value  Std.Error  DF   t-value p-value
(Intercept)       8.323340 0.27333009 292 30.451605  0.0000
treatmentHigh    -0.906057 0.19154238  23 -4.730320  0.0001
treatmentLow     -0.467040 0.15818328  23 -2.952521  0.0071
```

$\text{AIC} = -2 \times \text{logLik} + 2p$ where $p = \#$ of parameters
$p = (\text{AIC} + 2 \times \text{logLik})/2 = 9$

Model 3.1 Output (cont).

We will first work on adjusting our random factors. But peeking at the fixed effects,

```
> anova(model3.1.fit)
                 numDF denDF  F-value p-value
(Intercept)          1   292 9093.772  <.0001
treatment            2    23    5.082  0.0149
sex1                 1   292   52.602  <.0001
litsize              1    23   47.374  <.0001
treatment:sex1       2   292    0.466  0.6282
```

It looks like the treatment * sex terms will end up being discarded in the end.

Note that these fixed effects have 7 'numerator degrees of freedom' for β_0 through β_6.

Testing Hypothesis 3.1

Do we want random intercepts? Compare with a model with same fixed effects, no random intercept. Equivalently, $\sigma^2_{\text{litter}} = 0$

```
> model3.1a.fit <- gls(weight ~ treatment + sex1 + litsize
    + treatment*sex1, data = ratpup)
> summary(model3.1a.fit)
Generalized least squares fit by REML
  Model: weight ~ treatment + sex1 + litsize + treatment * sex1
  Data: ratpup
       AIC      BIC    logLik
  506.5099 536.5305 -245.255
> anova(model3.1.fit, model3.1a.fit)
              Model df      AIC      logLik   Test  L.Ratio p-value
model3.1.fit      1  9 419.1043 -200.5522
model3.1a.fit     2  8 506.5099 -245.2550 1 vs 2 89.40562  <.0001
```

By hand, use likelihood ratio test, with REML in both models because we are testing random effects.
LRT has 1 d.o.f. because different in models only is in σ^2_{litter}
Or: H_0 ($\sigma^2_{\text{litter}} = 0$) has 8 parameters, H_1 has 9 and $9 - 8 = 1$.

Adjustment for testing Variance = 0

Because σ^2_{litter} cannot be negative, the test statistic doesn't have a χ^2_1 distribution, but rather is a 50-50 mixture of χ^2_1 and 0.

Details are tricky, tl;dr: we need to divide p-value by 2.

Here, $< .0001/2$ is still tiny, we still reject H_0 and keep the random intercepts.

Suppose the likelihoods were interesting. E.g., -200.55 in alternative, -202.47 in null, for a test stat of $2[-200.55 - (-202.47)] = 3.84$.

That is the 95th percentile of a χ^2_1 variable, normally making $p = 0.05$. Dividing by 2 gives $p = 0.025$.

This only happens 'on the boundary of our parameter space' i.e., for testing variance = 0.

Choosing Residual Variance: Model 3.2A

Next, want to choose variance structure for residuals. Want independent residuals, but let's test allowing residual variance to vary based on treatment

```
> model3.2a.fit <- lme(weight ~ treatment + sex1 + litsize
    + treatment*sex1, random = ~1 | litter, ratpup, method = "REML",
    weights = varIdent(form = ~1 | treatment))
> summary(model3.2a.fit)
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | litter
        (Intercept)  Residual
StdDev:   0.3134846 0.5147948

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | treatment
 Parameter estimates:
  Control       Low      High
1.0000000 0.5649830 0.6394383
```

Interpreting Residual Variance Output

What does that mean?

```
        (Intercept)  Residual
StdDev:   0.3134846 0.5147948
  Control       Low      High
1.0000000 0.5649830 0.6394383
```

The intercept StdDev is the random intercept standard deviation. The factors under control, low, high are multipliers to scale residual standard deviation.

$$\widehat{\sigma}^2_{\text{litter}} = 0.3135^2$$
$$\widehat{\sigma}^2_{\text{Control}} = (0.5148 \cdot 1)^2$$
$$\widehat{\sigma}^2_{\text{Low}} = (0.5148 \cdot 0.5649)^2 = 0.2908^2$$
$$\widehat{\sigma}^2_{\text{High}} = (0.5148 \cdot 0.6394)^2 = 0.3292^2$$

For comparison, Model 3.1 had $\widehat{\sigma}^2_{\text{litter}} = 0.3107^2$ and $\widehat{\sigma}^2_{\text{Control}} = \widehat{\sigma}^2_{\text{High}} = \widehat{\sigma}^2_{\text{Low}} = \widehat{\sigma}^2 = 0.4043^2$

Testing Variance Structure

Should we use this more complicated variance structure?

Use LRT to compare our original loaded means model 3.1 with 3.2A in which the residual variance varied by treatment.

```
> anova(model3.1.fit, model3.2a.fit)
               Model df      AIC      logLik   Test  L.Ratio p-value
model3.1.fit       1  9 419.1043 -200.5522
model3.2a.fit      2 11 381.8847 -179.9423 1 vs 2 41.21964  <.0001
```

p-value is small, so we reject H_0 (model 3.1) in favor of the more complicated model. We don't divide the p-value by 2 because we aren't testing $\sigma^2 = 0$.

LRT has 2 dof: σ^2 is being replaced by 3 variance parameters, for an increase of 2 parameters.

Grouping Low and High Variances

In Model 3.2A, $\widehat{\sigma}^2_{\text{High}} \approx \widehat{\sigma}^2_{\text{Low}}$
Can we have just 2 residual variances instead of 3?
$\widehat{\sigma}^2_{\text{Control}}$ and $\widehat{\sigma}^2_{\text{High / Low}}$?

```
> ratpup$trtgrp[treatment =="Control"] <- 1
> ratpup$trtgrp[treatment =="Low" | treatment =="High"] <- 2
> ratpup
    pup_id weight    sex litter litsize treatment sex1 trtgrp
1        1   6.60   Male      1      12   Control    0      1
2        2   7.40   Male      1      12   Control    0      1
256    256   5.97 Female     20      16       Low    1      2
257    257   6.11 Female     20      16       Low    1      2
258    258   5.09   Male     21      14      High    0      2
259    259   5.57   Male     21      14      High    0      2
> model3.2b.fit <- lme(weight ~ treatment + sex1 + litsize
    + treatment*sex1, random = ~1 | litter, ratpup,
    method = "REML", weights = varIdent(form = ~1 | trtgrp))
```

Testing Variance Structures

Model 3.1 had 1 residual variance parameter
Model 3.2A had 3, one per treatment group
Model 3.2B has 2, control and high/low

```
> anova(model3.2a.fit, model3.2b.fit)
               Model df      AIC    logLik   Test  L.Ratio p-value
model3.2a.fit      1 11 381.8847 -179.9423
model3.2b.fit      2 10 381.0807 -180.5404 1 vs 2 1.196053  0.2741
```

The p-value is large, so we prefer the simpler model 3.2B (with 2 variances)

```
> anova(model3.1.fit, model3.2b.fit)
               Model df      AIC     ogLik   Test  L.Ratio p-value
model3.1.fit       1  9 419.1043 -200.5522
model3.2b.fit      2 10 381.0807 -180.5404 1 vs 2 40.02358  <.0001
```

The p-value is small, so we prefer the more complicated model 3.2B.
In neither case did we divide p-values by 2 because we aren't testing $\sigma^2 = 0$ but rather $\sigma_1^2 = \sigma_2^2$.

Testing Fixed Effects

We have chosen our random effect and variance structure.

Next, we choose what fixed effects to keep.
One way to do so is with F-tests from the R output.

```
> anova(model3.2b.fit)
                numDF denDF  F-value p-value
(Intercept)         1   292 9027.740  <.0001
treatment           2    23    4.241  0.0271
sex1                1   292   61.568  <.0001
litsize             1    23   49.577  <.0001
treatment:sex1      2   292    0.317  0.7288
```

The treatment * sex terms are not significant so we can omit them.

The treatment term is less clear. One approach would be to do a likelihood ratio test to see if we want to include it

LRT for fixed effect

A LRT for a fixed effect requires comparing two ML models.

```
> model3.3.ml.fit <- lme(weight ~ treatment + sex1 + litsize,
  random = ~ 1 | litter, ratpup, method = "ML",
  weights = varIdent(form = ~ 1 | trtgrp))
> model3.3a.ml.fit <- lme(weight ~ sex1 + litsize,
  random = ~ 1 | litter, ratpup, method = "ML",
  weights = varIdent(form = ~ 1 | trtgrp))
> # Test 3.3.ml vs 3.3a.ml: can we drop treatment term?
> anova(model3.3.ml.fit, model3.3a.ml.fit)
                Model df      AIC    logLik   Test  L.Ratio p-value
model3.3.ml.fit     1  8 353.7734 -168.8867
model3.3a.ml.fit    2  6 368.3706 -178.1853 1 vs 2 18.59723   1e-04
```

We have 2 degrees of freedom because there were two treatment indicator variables (for low and high dose). Equivalently, there were $(3-1)=2$ treatment classes. The p-value is small, so we keep the more complicated model with the treatment effects.
(We probably would anyways, since treatment effects are the point of the study)

Final Model

For our final model, we thus want to keep the treatment term. We want to fit with REML to get unbiased variance estimators.

```
> model3.3.reml.fit <- lme(weight ~ sex1 + litsize + treatment,
  random = ~ 1 | litter, ratpup, method = "REML",
  weights = varIdent(form = ~ 1 | trtgrp))
> summary(model3.3.reml.fit)
Fixed effects: weight ~ sex1 + litsize + treatment
                   Value  Std.Error  DF   t-value p-value
(Intercept)     8.327633 0.27406957 294 30.385106  0.0000
sex1           -0.343431 0.04204323 294 -8.168531  0.0000
litsize        -0.130681 0.01855194  23 -7.044036  0.0000
treatmentHigh  -0.862268 0.18293359  23 -4.713556  0.0001
treatmentLow   -0.433663 0.15226167  23 -2.848140  0.0091
> anova(model3.3.reml.fit)
            numDF denDF  F-value p-value
(Intercept)     1   294 9029.091  <.0001
sex1            1   294   63.596  <.0001
litsize         1    23   33.658  <.0001
treatment       2    23   11.387   4e-04
```

Recap

1. We started with Model 3.1 with all fixed effects, random intercepts, common residual variance.
2. Using LRT, rejected Model 3.1A: $\sigma^2_{\text{litter}} = 0$, kept random intercept. Divided p-value by 2 because testing a variance = 0.
3. Using LRTs, selected 3.2B with $\sigma^2_{\text{High}} = \sigma^2_{\text{Low}}$. Didn't have to divide p-value by 2 because testing 2 variances equaling each other, not 0.
4. Using F-test on 3.2B, chose to drop sex * treatment interaction term.
5. All of those tests so far used REML estimation
6. Tested significance of treatment term in Model 3.3. Used LRT with ML estimation because testing for fixed effect.
7. Re-ran Model 3.3 with REML estimation to get unbiased variance estimators

Exercise

You wish to decide between having a single variance for all residuals versus a variance structure that varies between 3 different classes. To do so, you wish to perform a likelihood ratio test. Given the following loglikelihoods under both REML and ML estimation, what is the outcome of the test at 5% and 2.5% significance levels?

Model	REML logLik	ML lokLik
1 residual variance	−200.5522	−200.4143
3 residual variances	−196.8517	−196.8158

Exercise

You wish to decide between having a single variance for all residuals versus a variance structure that varies between 3 different classes. To do so, you wish to perform a likelihood ratio test. Given the following loglikelihoods under both REML and ML estimation, what is the outcome of the test at 5% and 2.5% significance levels?

Model	REML logLik	ML lokLik
1 residual variance	-200.5522	-200.4143
3 residual variances	-196.8517	-196.8158

We are testing variances, not fixed effects, so want the REML numbers. We have $3 - 1 = 2$ degrees of freedom.
Don't have to divide p-value by 2 because aren't testing variance at 0.
The simpler, 1 variance model is the null.
$T = 2[(-196.8517) - (-200.5522)] = 7.401 > 7.38$
That exceeds the 5% and 2.5% significance level critical values, so we reject the null and use the more complicated 3 variance model.