

Outline

- 1 Description of the course
- 2 Linear Regression
- 3 Logistic Regression
- 4 Multilevel models: Longitudinal data
- 5 Multilevel models: Cluster data
- 6 Missing data

Models for cluster data

Ariel Alonso Abad

Catholic University of Leuven

Models for Clustered Data

The Rat Pup Example

The data come from a study in which 30 female rats were randomly assigned to receive one of three doses of an experimental compound (variable **treat** with levels: high, low or control). Although 10 female rats were initially assigned to receive each treatment dose, three of the female rats in the high-dose group died, so there are no data for their litters. In addition, litter sizes (variable **lts**) varied widely, ranging from 2 to 18 pups. The sex of the pups was also recorded (variable **sex**)

Objective of the study: To compare the birth weights (variable **w**) of pups from litters born to female rats that received the high- and low-dose treatments to the birth weights of pups from litters that received the control treatment.

Jose Pinheiro and Doug Bates, (2000) Mixed-Effects Models in S and S-PLUS.

The Rat Pup Example

- Two-level clustered data from a cluster randomized trial
- Each litter (cluster) was randomly assigned to a specific level of treatment
- Rat pups (units of analysis) nested within litters
- Birth weights of rat pups within the same litter are likely to be correlated because the pups shared the same maternal environment

Exploring the data in R

```
> ## Reading the data
> ratpup <- read.table("rat_pup.dat", h = T)
> ratpup$sex1[ratpup$sex == "Female"] <- 1
> ratpup$sex1[ratpup$sex == "Male"] <- 0
> attach(ratpup)
>
> ## Table describing the data
> g <- function(x)c(N=length(x),Mean=mean(x,na.rm=TRUE),
+ SD=sd(x,na.rm=TRUE), Min=min(x,na.rm=TRUE),Max=max(x,na.rm=TRUE))
> summarize(weight,by=llist(treatment,sex),g)
>
  treatment   sex weight      Mean       SD   Min   Max
1  Control Female    54 6.116111 0.6851179 3.68 7.57
2  Control  Male    77 6.471039 0.7537880 4.57 8.33
3    High Female    32 5.851562 0.6001887 4.48 7.68
4    High  Male    33 5.918485 0.6909058 5.01 7.70
5    Low Female    65 5.837538 0.4504964 4.75 7.73
6    Low  Male    61 6.025082 0.3803403 5.25 7.13
```

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

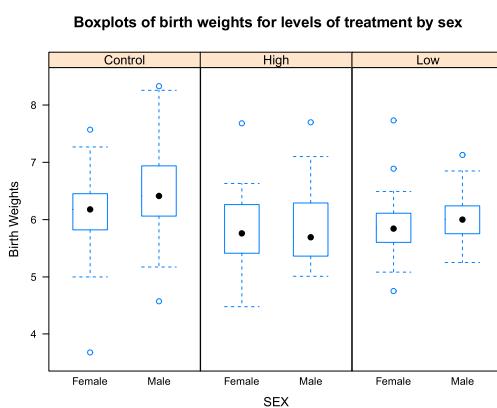
Treatment	Sex	N obs	Mean	SD	Minimum	Maximum
Control	Female	54.00	6.12	0.69	3.68	7.57
Control	Male	77.00	6.47	0.75	4.57	8.33
High	Female	32.00	5.85	0.60	4.48	7.68
High	Male	33.00	5.92	0.69	5.01	7.70
Low	Female	65.00	5.84	0.45	4.75	7.73
Low	Male	61.00	6.03	0.38	5.25	7.13

- The experimental treatments appear to have a negative effect on mean birth weight for males and females
- Sample mean birth weight of males are consistently higher than those of females within all levels of treatment

Exploring the data in R

```
> ## Comparing the distributions of birth weights
> ## for each treatment by sex combination
>
> library(lattice) # trellis graphics
> library(grid)
>
> bwplot(weight ~ sex|treatment, data=ratpup, aspect = 2,
+ ylab="Birth Weights", xlab="SEX",
+ main = "Boxplots of birth weights for levels of treatment by sex")
>
```

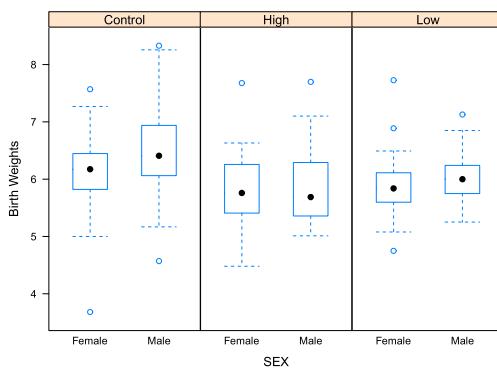
Birth weights for levels of treatment by sex



- Males appear to have a higher median birth weight than females in the low and control groups, but not in the high group
- The distribution of birth weight appears to be roughly symmetric at each level of treatment and sex

Birth weights for levels of treatment by sex

Boxplots of birth weights for levels of treatment by sex

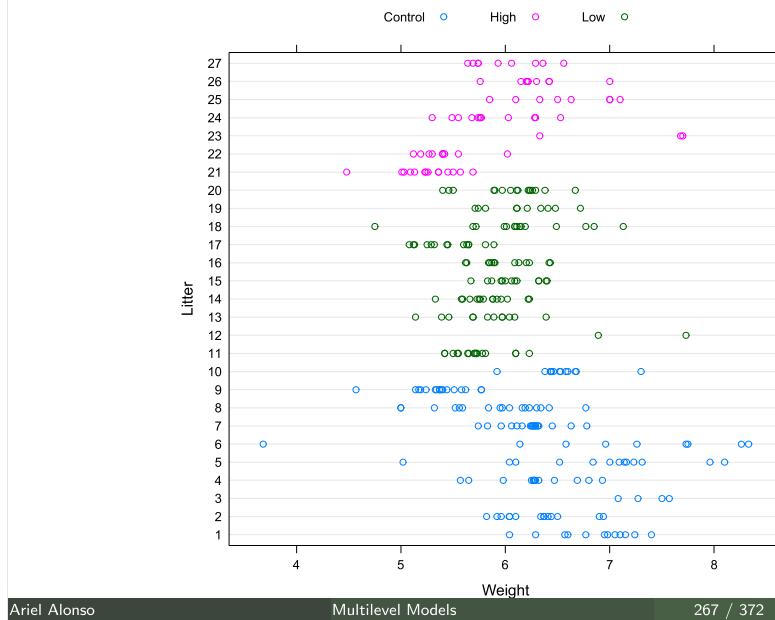


- Lower birth weight for the high- and low-dose treatments compared to the control group
- Variance of the birth weight is similar for males and females within each treatment but appears to differ across treatments

Exploring the data in R

```
> ## Comparing the distributions of birth weights for each treatment
>
> dotplot(litterid ~ weight, group=treatment, data = ratpup,
+ xlab="Weight", ylab="Litter",
+ auto.key=list(space="top", column=3, cex=.8, title="",
+ cex.title=1, lines=FALSE, points=TRUE) )
> with(ratpup, interaction.plot(treatment, sex, weight))
>
```

Exploring the data in R

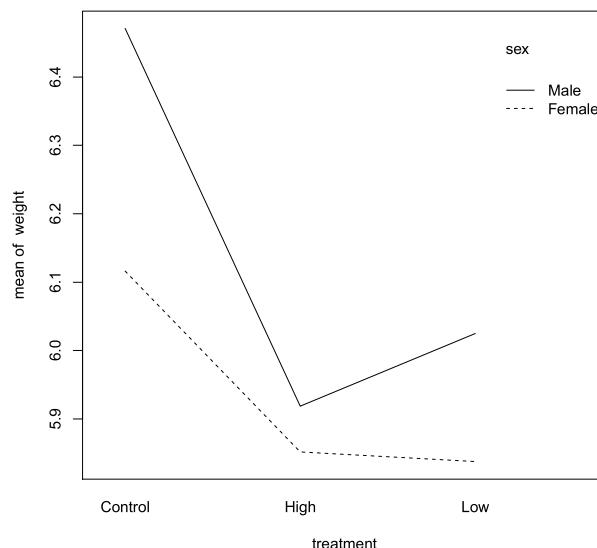


Ariel Alonso

Multilevel Models

267 / 372

Exploring the data in R



Ariel Alonso

Multilevel Models

268 / 372

Hierarchical model

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} \text{sex}_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

and sex_{ij} a level 1 indicator variable for female.

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} \text{treat}_{1i} + \gamma_{02} \text{treat}_{2i} + \gamma_{03} l_s + b_0; \\ \pi_{1i} = \gamma_{10} + \gamma_{20} \text{treat}_{1i} + \gamma_{30} \text{treat}_{2i} \end{cases}$$

where treat_{1i} and treat_{2i} are level 2 indicator variables for high and low treatment levels, l_s is the litter size and $b_0 \sim N(0, \sigma_b^2)$

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} \text{sex}_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01} \text{treat}_{1i} + \gamma_{02} \text{treat}_{2i} + \gamma_{03} l_s + b_0; \\ \pi_{1i} = \gamma_{10} + \gamma_{20} \text{treat}_{1i} + \gamma_{30} \text{treat}_{2i} \end{cases}$$

⇒ Birth weights of pups vary **within** litter due to differences in gender and in other unaccounted factors (ε_{ij})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i}sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01}treat_{1i} + \gamma_{02}treat_{2i} + \gamma_{03}ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20}treat_{1i} + \gamma_{30}treat_{2i} \end{cases}$$

⇒ Birth weights vary **between** litters due to differences in treatment, litter size and other litter-specific characteristics unaccounted for by the model (b_{0i})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i}sex_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Level 2:

$$\begin{cases} \pi_{0i} = \gamma_{00} + \gamma_{01}treat_{1i} + \gamma_{02}treat_{2i} + \gamma_{03}ls_i + b_{0i} \\ \pi_{1i} = \gamma_{10} + \gamma_{20}treat_{1i} + \gamma_{30}treat_{2i} \end{cases}$$

⇒ Birth weights vary **between** litters due to differences in treatment, litter size and other litter-specific characteristics unaccounted for by the model (b_{0i})

⇒ Notice that treatment may affect males and females pups differently

One single model

Model

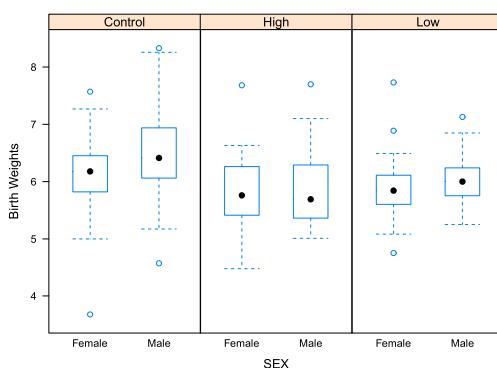
$$w_{ij} = \gamma_{00} + \gamma_{01} treat_{1i} + \gamma_{02} treat_{2i} + \gamma_{03} ls_i + \\ \gamma_{10} sex_{ij} + \gamma_{20} treat_{1i} sex_{ij} + \gamma_{30} treat_{2i} sex_{ij} + \\ b_{0i} + \varepsilon_{ij}$$

Distributional Assumptions

$$b_{0i} \sim N(0, \sigma_b^2) \text{ and } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Modeling the covariance structure

Boxplots of birth weights for levels of treatment by sex



- Previous model assumes that the within litter variability σ_ε^2 is constant across treatment
- The variances of the birth weights are similar for males and females within each treatment but appear to differ across treatments

Covariance structure: Testing homoscedasticity

Hence, one wants to test if the variance of the residuals (σ_{ε}^2) is the same (homogeneous) for the three treatment groups (high, low, and control)

$$H_0 : \sigma_{high}^2 = \sigma_{low}^2 = \sigma_{control}^2 = \sigma_{\varepsilon}^2$$

- REML-based likelihood ratio test to compare two models (**mean structure stays the same**):

Model 1: All three variances equal (meanfull.hom)

Model 2: All three variances different (meanfull.het)

- The asymptotic null distribution of this test statistic is a χ^2 with 2 degrees of freedom

Covariance structure: Testing homoscedasticity

- At this moment the `lmer()` function does not allow users to fit models with heterogeneous error variance structures
- Therefore, we will work with the function `lme()` from the package `nlme`
- `lme()` and `lmer()` are similar but there are some differences in syntax and output that will be explained in the following

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                         random = ~1 | litterid, ratpup, method = "REML")
>
```

- The factor() function is not necessary for treatment, because the original treatment variable has string values High, Low, and Control, and will therefore be considered as a factor automatically
- We also do not need to declare sex1 as a factor, because it is an indicator variable having only values of 0 and 1

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                         random = ~1 | litterid, ratpup, method = "REML")
>
```

- `lme()` treats the lowest level (alphabetically or numerically) of a factor as the reference category. This means that “Control” will be the reference category of treatment. The reference level can be changed using

```
treatment=relevel(treatment,ref="High")
```

Fitting the homocedastic model in R

Model 1

```
> ## Fitting a homocedastic model
>
> library(nlme)
>
> meanfull.hom <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                         random = ~1 | litterid, ratup, method = "REML")
>
```

- `random = 1 | litterid`, includes a random effect (intercept) for each level of litter in the model
- `method = "REML"`, specifies that the default REML estimation method is to be used

Fitting the homocedastic model in R

```
> summary(meanfull.hom)
>
Linear mixed-effects model fit by REML
Data: ratup
      AIC      BIC      logLik
419.1043 452.8775 -200.5522

Random effects:
Formula: ~1 | litterid
          (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
sex1 -0.411688 0.07315410 292 -5.627679 0.0000
litsize -0.128382 0.01875336 23 -6.845819 0.0000
treatmentHigh:sex1 0.107023 0.13176318 292 0.812239 0.4173
treatmentLow:sex1 0.083866 0.10568189 292 0.793568 0.4281
.....
Standardized Within-Group Residuals:
      Min        Q1        Med        Q3        Max
-7.47250744 -0.50014749  0.02911668  0.57348178  3.00962055

Number of Observations: 322
Number of Groups: 27
>
```

Fitting the homocedastic model in R

```
> anova(meanfull.hom)
>
  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
>
```

- The `anova()` function performs a series of Type I (or sequential) F-tests for the fixed effects in the model, each of which are conditional on the preceding terms in the model specification
- For example, the F-test for `sex1` is conditional on the treatment effects, but the F-test for `treatment` is not conditional on the `sex1` effect

Fitting the homocedastic model in R

```
> anova(meanfull.hom)
>
  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
>
```

Model fitted using REML

The model was fitted using REML and, therefore, different mean structures cannot be compared!

Homocedastic model: Predicting the random effects

```
> ## Display the random effects (EBLUPs) from the model.  
>  
> random.effects(meanfull.hom)  
   (Intercept)  
1  0.17480024  
2 -0.07362296  
3 -0.17490203  
4 -0.05376249  
5  0.34446954  
6 -0.05480208  
7  0.39153638  
8 -0.02616704  
9 -0.61772106  
10 0.09017150  
11 0.04136696  
12 0.02072931  
13 -0.35981737  
14 0.01847368  
15 0.03549783  
16 0.06416884  
17 -0.39636862  
18 0.42802095  
19 -0.18110865  
20 0.32903707  
21 -0.27813901  
22 -0.49096620  
23 0.26053476  
24 0.17537803  
25 0.23748827  
26 0.22966911  
27 -0.13396497
```

Fitting the heterocedastic model in R

Model 2

```
> ## Fitting a heterocedastic model  
>  
> meanfull.het <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,  
+                         random = ~1 | litterid, ratpup, method = "REML",  
+                         weights = varIdent(form = ~1 | treatment))  
>
```

- The arguments of the `lme()` function are the same as those used to fit Model 1, with the addition of the `weights` argument

- The argument

`weights = varIdent(form = ~ 1 |treatment)`

sets up a heterogeneous residual variance structure, with observations at different levels of treatment having different residual variance parameters

Fitting the heterocedastic model in R

```
> summary(meanfull.het)
>
Linear mixed-effects model fit by REML
Data: ratup
      AIC      BIC    logLik
 381.8847 423.163 -179.9423

Random effects:
 Formula: ~1 | litterid
          (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

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                Value Std.Error DF t-value p-value
(Intercept)     8.345294 0.27464753 292 30.385468 0.0000
treatmentHigh  -0.903277 0.19215903  23 -4.700672 0.0001
treatmentLow   -0.466292 0.15908908  23 -2.931013 0.0075
sex1           -0.408131 0.09303486 292 -4.386865 0.0000
litsize         -0.130007 0.01848708  23 -7.032332 0.0000
treatmentHigh:sex1 0.094666 0.12919527 292  0.732737 0.4643
treatmentLow:sex1  0.076013 0.10811858 292  0.703053 0.4826
.....
```

Fitting the heterocedastic model in R

```
Random effects:
 Formula: ~1 | litterid
          (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
```

- Random effects portion of the output: Estimated residual standard deviation equal to 0.5147948
- Parameter estimates: Values by which the residual standard deviation should be multiplied to obtain the estimated standard deviation of the residuals in each treatment group
- This multiplier is 1.0 for the control group (the reference). Multipliers for the low and high treatment groups are very similar

Heterocedastic versus homocedastic model

The variance of the residuals (σ_{ε}^2) is the same (homogeneous) for the three treatment groups

$$H_0 : \sigma_{high}^2 = \sigma_{low}^2 = \sigma_{control}^2 = \sigma_{\varepsilon}^2$$

```
> ## Heterocedastic versus homocedastic model
>
> anova(meanfull.hom,meanfull.het)
>
>          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
meanfull.hom     1 9 419.1043 452.8775 -200.5522
meanfull.het     2 11 381.8847 423.1630 -179.9423 1 vs 2 41.21964 <.0001
>
```

Heterocedastic model

```
Random effects:
Formula: ~1 | litterid
  (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
```

- $\sigma_{high} = 0.5147948 \cdot 0.6394383$, $\sigma_{low} = 0.5147948 \cdot 0.5649830$ and
 $\sigma_{control} = 0.5147948 \cdot 1$

Heterocedastic model

```
Random effects:  
  Formula: ~1 | litterid  
            (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
```

- $\sigma_{high} = 0.329179$, $\sigma_{low} = 0.290850$ and $\sigma_{control} = 0.5147948$
- Is $\sigma_{high}^2 = \sigma_{low}^2$?

High-low dose: Equal residual variance

Hence, one wants to test if the variance of the residuals in the high and low dose groups are the same

$$H_0 : \sigma_{high}^2 = \sigma_{low}^2$$

- REML-based likelihood ratio test to compare two models (**mean structure stays the same**):

Model 2: All three variances different (meanfull.het)

Model 3: $\sigma_{high}^2 = \sigma_{low}^2$ (meanfull.hilo)

- The asymptotic null distribution of this test statistic is a χ^2 with 1 degrees of freedom

High-low dose: Equal residual variance

```
> ## High-low dose: Equal residual variance
>
> ratpup$trtgrp[treatment=="Control"] <- 1
> ratpup$trtgrp[treatment == "Low" | treatment == "High"] <- 2
>
> meanfull.hilo <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,
+                         random = ~1 | litterid, ratpup, method = "REML",
+                         weights = varIdent(form = ~1 | trtgrp))
>
> summary(meanfull.hilo)
> anova(meanfull.hilo)
>
```

Fitting the heterocedastic model in R

```
> summary(meanfull.hilo)
>
Linear mixed-effects model fit by REML
Data: ratpup
      AIC      BIC      logLik
381.0807 418.6065 -180.5404

Random effects:
Formula: ~1 | litterid
          (Intercept) Residual
StdDev:   0.3145679 0.5147878

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | trtgrp
Parameter estimates:
           1         2
1.0000000 0.5905487

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                Value Std.Error DF t-value p-value
(Intercept)     8.350351 0.27567833 292 30.290196 0.0000
treatmentHigh  -0.901844 0.19140146  23 -4.711793 0.0001
treatmentLow   -0.466596 0.15999337  23 -2.916347 0.0078
sex1            -0.408198 0.09303540 292 -4.387529 0.0000
litsize          -0.130383 0.01856367  23 -7.023574 0.0000
treatmentHigh:sex1 0.092026 0.12461723 292  0.738473 0.4608
treatmentLow:sex1  0.076397 0.10939797 292  0.698337 0.4855
.....
>
```

High-low dose: Equal residual variance

Hence, one wants to test if the variance of the residuals in the high and low dose groups are the same

$$H_0 : \sigma_{high}^2 = \sigma_{low}^2$$

```
> ## High-low dose: Equal residual variance
>
> anova(meanfull.het,meanfull.hilo)
>
>             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
meanfull.het     1 11 381.8847 423.1630 -179.9423
meanfull.hilo     2 10 381.0807 418.6065 -180.5404 1 vs 2 1.196053  0.2741
```

Is there a litter effect?

- Can the random effects (b_{0i}) associated with the litter-specific intercepts be omitted from Model 3?
- One do not directly test the significance of the random litter-specific intercepts, but rather tests a hypothesis related to the variance of the random litter effects.
- The null and alternative hypotheses can be written as follows:

$$H_0 : \sigma_b^2 = 0 \text{ versus } H_1 : \sigma_b^2 > 0$$

Is there a litter effect?

- Although hypothesis tests are often phrased in terms of parameter restrictions, they basically compare the quality of the fit obtained from two nested models
- Likelihood ratio tests (LRTs) are a valuable tool to compare nested models
- An approximate reference distribution for a LRT is the χ^2_γ where γ , the degrees of freedom, is determined by the difference in the number of parameters for the models H_1 and H_0
- Hence, the LRT for testing $H_0 : \sigma_b^2 = 0$ versus $H_1 : \sigma_b^2 > 0$ has an approximate reference distribution χ^2_1

Is there a litter effect?

- However, the argument for using a χ^2_1 distribution **does not apply** when the parameter value being tested is on the boundary of the parametric space
- The asymptotic null distribution of the test statistic is a mixture of χ^2 distributions, with 0 and 1 degrees of freedom, and equal weights of 0.5
- As shown in Pinheiro and Bates (2000) Section 2.5, the p-value from the χ^2_1 distribution will be “conservative” in the sense that it is larger than a simulation-based p-value would be
- In the worst-case scenario the χ^2_1 -based p-value will be twice as large as it should be

Is there a litter effect?

```
> ## Is there a litter effect?  
>  
> meanfull.hilo.nolitter <- gls(weight ~ treatment + sex1 + litsize +  
+           treatment:sex1, data = ratpup, weights = varIdent(form = ~1 | trtgrp))  
>  
> summary(meanfull.hilo.nolitter)  
>
```

Is there a litter effect?

```
Generalized least squares fit by REML  
Model: weight ~ treatment + sex1 + litsize + treatment:sex1  
Data: ratpup  
      AIC      BIC    logLik  
489.6521 523.4252 -235.826  
  
Variance function:  
  Structure: Different standard deviations per stratum  
  Formula: ~1 | trtgrp  
Parameter estimates:  
      1          2  
1.0000000 0.7060188  
  
Coefficients:  
              Value Std.Error t-value p-value  
(Intercept) 8.201712 0.15902776 51.57409 0.0000  
treatmentHigh -0.976414 0.10624042 -9.19060 0.0000  
treatmentLow -0.456018 0.08700180 -5.24147 0.0000  
sex1          -0.339911 0.10616682 -3.20167 0.0015  
litsize        -0.121478 0.01008518 -12.04524 0.0000  
treatmentHigh:sex1 0.180960 0.14941228  1.21114 0.2267  
treatmentLow:sex1 0.076386 0.13035758   0.58597 0.5583  
  
.....  
Residual standard error: 0.5980885  
Degrees of freedom: 322 total; 315 residual
```

Is there a litter effect?

Is there a litter effect?

$$H_0 : \sigma_b^2 = 0 \text{ versus } H_1 : \sigma_b^2 > 0$$

```
> ## Is there a litter effect?  
>  
> anova(meanfull.hilo.nolitter,meanfull.hilo)  
>  
          Model df      AIC      BIC  logLik   Test L.Ratio p-value  
meanfull.hilo.nolitter    1  9 489.6521 523.4252 -235.8260  
meanfull.hilo            2 10 381.0807 418.6065 -180.5404 1 vs 2 110.5713 <.0001  
>
```

Modeling the mean structure

```
> ## Fitting the final model using ML  
>  
> meanfull.hilo.ml <- lme(weight ~ treatment + sex1 + litsize + treatment:sex1,  
+                           random = ~1 | litterid, ratpup, method = "ML",  
+                           weights = varIdent(form = ~1 | trtgrp))  
>  
> summary(meanfull.hilo.ml)
```

Modeling the mean structure

```
> summary(meanfull.hilo.ml)
>
Linear mixed-effects model fit by maximum likelihood
Data: ratpup
      AIC      BIC    logLik
357.1317 394.8773 -168.5659

Random effects:
Formula: ~1 | litterid
          (Intercept) Residual
StdDev:   0.2882595 0.5123784

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | trtgrp
Parameter estimates:
          1         2
1.0000000 0.5897706

Fixed effects: weight ~ treatment + sex1 + litsize + treatment:sex1
                Value Std.Error DF t-value p-value
(Intercept)     8.350608 0.26150064 292 31.93341 0.0000
treatmentHigh  -0.904757 0.18092616  23 -5.00070 0.0000
treatmentLow   -0.466869 0.15105108  23 -3.09080 0.0052
sex1           -0.406590 0.09357754 292 -4.34495 0.0000
litsize        -0.130402 0.01755814  23 -7.42689 0.0000
treatmentHigh:sex1 0.093026 0.12521954 292  0.74290 0.4581
treatmentLow:sex1  0.075602 0.10998665 292  0.68737 0.4924
.....
Number of Observations: 322
Number of Groups: 27
```

Modeling the mean structure

```
> anova(meanfull.hilo.ml)
>
      numDF denDF  F-value p-value
(Intercept)      1    292 10274.678 <.0001
treatment        2     23    4.810  0.0180
sex1            1    292    59.906 <.0001
litsize          1     23    55.438 <.0001
treatment:sex1   2    292     0.315  0.7303
>
```

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} \text{sex}_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{ Control} \\ N(0, 0.30^2), \text{ Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 \text{treat}_{1i} - 0.47 \text{treat}_{2i} - 0.13 \text{ls}_i + b_{0i}; \\ \pi_{1i} = -0.41 \\ b_{0i} \sim N(0, 0.29^2) \end{cases}$$

⇒ Birth weights of pups vary **within** litter due to differences in gender and in other unaccounted factors (ε_{ij})

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} \text{sex}_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{ Control} \\ N(0, 0.30^2), \text{ Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 \text{treat}_{1i} - 0.47 \text{treat}_{2i} - 0.13 \text{ls}_i + b_{0i}; \\ \pi_{1i} = -0.41 \\ b_{0i} \sim N(0, 0.29^2) \end{cases}$$

⇒ Litters in the high/low dose have pups with smaller average birth weights. In addition, litter size has a negative impact on the average birth weight and there is extra variability from other unknown factors

Hierarchical model interpretation

Level 1: ANOVA type model

$$w_{ij} = \pi_{0i} + \pi_{1i} \text{sex}_{ij} + \varepsilon_{ij}, \text{ with } \varepsilon_{ij} \sim \begin{cases} N(0, 0.51^2), \text{ Control} \\ N(0, 0.30^2), \text{ Low/High dose} \end{cases}$$

Level 2:

$$\begin{cases} \pi_{0i} = 8.35 - 0.90 \text{treat}_{1i} - 0.47 \text{treat}_{2i} - 0.13 \text{ls}_i + b_0; \\ \pi_{1i} = -0.41 \\ b_0 \sim N(0, 0.29^2) \end{cases}$$

⇒ Litters in the high/low dose have pups with smaller average birth weights. In addition, litter size has a negative impact on the average birth weight and there is extra variability from other unknown factors

⇒ Treatment affects males and females pups equally

Outline

- 1 Description of the course
- 2 Linear Regression
- 3 Logistic Regression
- 4 Multilevel models: Longitudinal data
- 5 Multilevel models: Cluster data
- 6 Missing data