

Смешанные линейные модели (вложенные случайные факторы)

Линейные модели...

Марина Варфоломеева, Вадим Хайтов

Кафедра Зоологии беспозвоночных, Биологический факультет, СПбГУ



Вы узнаете

- ▶ Что такое вложенные случайные факторы и в каких случаях они применяются

Вы сможете

- ▶ Объяснить, что такое вложенные случайные факторы
- ▶ Привести примеры иерархических случайных факторов
- ▶ Вычислить коэффициент внутриклассовой корреляции для случая с двумя вложенными случайными факторами
- ▶ Подобрать смешанную линейную модель со вложенными случайными факторами

Смешанные модели со вложенными случайными факторами

Вложенные факторы (Nested effects)



Вложенные факторы (Nested effects)

Факторы образуют иерархическую последовательность вложенности

- ▶ лес -> дерево в лесу -> ветка на дереве -> наблюдение (личинки насекомых)

Вложенные факторы (Nested effects)

Факторы образуют иерархическую последовательность вложенности

► лес → дерево в лесу → ветка на дереве → наблюдение (личинки насекомых)

Внутри каждого уровня главного фактора будут разные (нестрого сопоставимые) уровни вложенного фактора

Деревья, с которых собирали личинок, будут разные в разных лесах (разные экземпляры).

Вложенные факторы (Nested effects)

Факторы образуют иерархическую последовательность вложенности

► лес → дерево в лесу → ветка на дереве → наблюдение (личинки насекомых)

Внутри каждого уровня главного фактора будут разные (нестрого сопоставимые) уровни вложенного фактора

Деревья, с которых собирали личинок, будут разные в разных лесах (разные экземпляры).

Уровни вложенных факторов описывают иерархию взаимного сходства наблюдений

Личинки с разных деревьев из одного леса имеют право быть похожими друг на друга больше, чем на личинок из другого леса

Личинки на одном дереве имеют право быть похожими друг на друга больше, чем на личинок с другого дерева

И т.п.

Другие примеры вложенных факторов

- ▶ регион -> город -> больница -> наблюдение (пациент)
- ▶ самка -> выводок -> наблюдение (особь)
- ▶ лес -> дерево в лесу -> гнездо на дереве -> наблюдение (птенец)
- ▶ улитка -> спороциста в улитке -> наблюдение (редия)

Пример: Высота растений и выпас скота

Вообще-то, статья Gennet et al. 2017 о птицах, но чтобы про них что-то лучше понять, нужно разобраться с их местообитанием.

Как в разные годы высота растительного покрова зависит от выпаса скота, экспозиции склона и проективного покрытия местных растений?

Зависимая переменная:

- ▶ **height** - высота растительного покрова

Предикторы:

- ▶ **graze** - выпас коров (0, 1)
- ▶ **AspectCat** - экспозиция (S, N)
- ▶ **nativecov** - покрытие местной флоры %
- ▶ **slope** - наклон
- ▶ **year** - год наблюдений
- ▶ **Park** - парк
- ▶ **plotID** - уникальный идентификатор участка

Данные: Gennet, S., Spotswood, E., Hammond, M. and Bartolome, J.W., 2017. Livestock grazing supports native plants and songbirds in a California annual grassland. *PLoS one*, 12(6), p.e0176367.

Открываем данные

Исходные данные не опрятны (this data is not tidy)! Каждый участок в каждом году фигурирует два раза (т.к. информация о присутствии-отсутствии каких-то видов на участке записана в разных строках). Откроем и переформатируем данные так, чтобы не было дублирования и каждому участку соответствовала одна строка.

```
library(readxl)
library(tidyr)
gr <- read_excel("data/Grazing_native_plants_Gennet_et_al._2017_S1.xlsx")
graz <- gr %>% spread(Species, presence)
```

Знакомство с данными

Есть ли пропущенные значения?

```
sum(is.na(graz))
```

```
# [1] 0
```

Сколько участков было в каждом парке в каждый год?

```
with(graz, table(Park, year))
```

```
#      year
# Park 2004 2005 2006 2007 2008 2009 2010 2011
#  MT     6    10    10    10    10    10    10    10
#  PR     6     6     6     6     6     6     6     6
#  SU     0     9     9     9     9     9     9     9
#  VC    10    10    10    10    11    11    11    11
```

Как закодированы переменные?

```
str(graz)
```

```
# Classes 'tbl_df', 'tbl' and 'data.frame': 271 obs. of  18 variables:
# $ plotID      : chr  "MT1" "MT2" "MT3" "MT4" ...
# $ Plot       : num  1 2 3 4 5 6 4 5 6 7 ...
# $ Park       : chr  "MT" "MT" "MT" "MT" ...
# $ year       : num  2004 2004 2004 2004 2004 ...
# $ graze      : num  0 0 0 1 1 1 1 1 1 1 ...
# $ Aspect     : num  146 250 262 190 274 ...
# $ AspectCat  : chr  "S" "S" "S" "S" ...
# $ heatloadrel: num  0.03 1.35 2.06 0.31 1.83 0.95 0.03 2.06 0.7 1.22 ...
# $ slope      : num  37.8 41.1 35.4 28 58.9 ...
# $ nativecov  : num  0 0.36 1.43 0 1.07 ...
# $ litt       : num  28.21 31.43 11.07 8.93 18.57 ...
# $ bare       : num  0 0.357 9.286 4.643 7.857 ...
# $ height     : num  29.2 26.2 23.2 11.9 19.3 ...
# $ htstdev    : num  16.14 15.41 15.47 8.25 10.45 ...
# $ cov        : num  0.553 0.587 0.666 0.695 0.541 ...
# $ GRSP       : num  0 0 0 0 1 0 1 1 0 0 ...
# $ HOLA       : num  0 0 0 1 1 1 1 1 0 0 ...
# $ WEME       : num  0 1 0 1 1 0 0 1 0 1 ...
```

Наводим порядок

Сделаем факторами переменные, которые понадобятся для модели

```
graz$graze_f <- factor(graz$graze)
graz$AspectCat <- factor(graz$AspectCat)
graz$year_f <- factor(graz$year)
```

Извлечем корень из обилия местных видов

```
graz$nativecov_sq <- sqrt(graz$nativecov)
```

Модель

Вспомним главный вопрос исследования и подберем модель

Как в разные годы высота растительного покрова зависит от выпаса скота, экспозиции склона и проективного покрытия местных растений?

Модель

Вспомним главный вопрос исследования и подберем модель

Как в разные годы высота растительного покрова зависит от выпаса скота, экспозиции склона и проективного покрытия местных растений?

Нам нужно учесть, что в разные годы из-за кучи разных причин высота растений может различаться

Кроме того, нужно учесть, что в разных парках и на разных участках растения будут расти сходным образом в разные годы. У нас есть иерархические факторы парк и участок в парке

Подбираем модель методом максимального правдоподобия, т.к. она нам понадобится, чтобы проверить значимость фиксированных эффектов при помощи теста отношения правдоподобий.

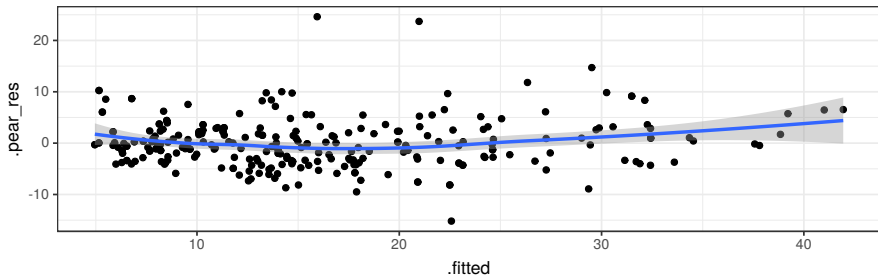
```
library(lme4)
ML1 <- lmer(height ~ graze_f*AspectCat + year_f +
             nativecov_sq + slope + (1|Park/plotID),
             data = graz, REML = FALSE)
```

Анализ остатков

```
# Данные для анализа остатков  
ML1_diag <- data.frame(  
  graz,  
  .pear_res = residuals(ML1, type = "pearson"),  
  .fitted = fitted(ML1, type = "response"))
```

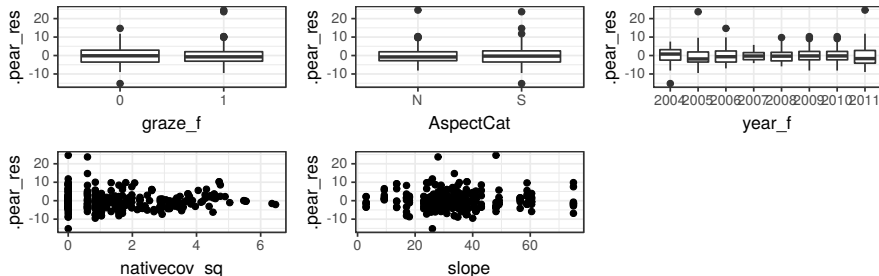

График остатков

```
library(ggplot2); library(cowplot); theme_set(theme_bw())
gg_res <- ggplot(data = ML1_diag, aes(y = .pear_res))
gg_res + geom_point(aes(x = .fitted)) +
  geom_smooth(aes(x = .fitted))
```



Графики остатков от переменных в модели

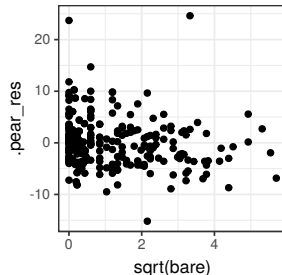
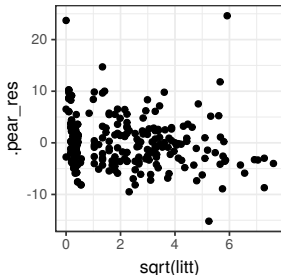
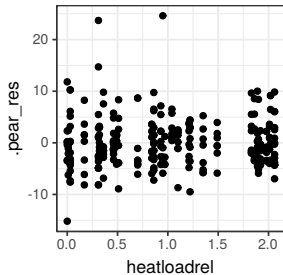
```
plot_grid(gg_res + geom_boxplot(aes(x = graze_f)),  
gg_res + geom_boxplot(aes(x = AspectCat)),  
gg_res + geom_boxplot(aes(x = year_f)),  
gg_res + geom_point(aes(x = nativecov_sq)),  
gg_res + geom_point(aes(x = slope)),  
ncol = 3)
```



► Паттерн на графике nativecov_sq. Возможно, здесь нужно использовать GAMM.

Графики остатков от переменных не в модели

```
plot_grid(  
  gg_res + geom_point(aes(x = heatloadrel)),  
  gg_res + geom_point(aes(x = sqrt(litt))),  
  gg_res + geom_point(aes(x = sqrt(bare))),  
  ncol = 3)
```



- Паттерн на графике `heatloadrel` - Возможно, есть тренд на графике `sqrt(litt)`

Тесты отношения правдоподобий для полной модели

Модель ML1 была подобрана при помощи ML, поэтому можно применять тесты отношения правдоподобий прямо к ней

```
drop1(ML1, test = 'Chi')
```

```
# Single term deletions
#
# Model:
# height ~ graze_f * AspectCat + year_f + nativecov_sq + slope +
# (1 | Park/plotID)
#
```

	Df	AIC	LRT	Pr(Chi)
# <none>		1729.4		
# year_f	7	1819.8	104.405	< 2.2e-16 ***
# nativecov_sq	1	1728.9	1.478	0.224030
# slope	1	1727.8	0.450	0.502531
# graze_f:AspectCat	1	1736.3	8.956	0.002765 **
# ---				

```
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Высота растительного покрова:

- ▶ на склонах разной экспозиции по-разному зависит от выпаса скота (достоверное взаимодействие)
- ▶ различается в разные годы
- ▶ не зависит от покрытия местных растений и крутизны склона

Задание 1

Рассчитайте внутриклассовую корреляцию

- ▶ Для наблюдений на одном и том же участке
- ▶ Для наблюдений в одном и том же парке

Внутриклассовая корреляция

Для расчета внутриклассовой корреляции нужна модель, подобранная при помощи REML

```
REML1 <- lmer(height ~ graze_f*AspectCat + year_f +  
              nativecov_sq + slope + (1|Park/plotID),  
              data = graz, REML = TRUE)
```

```
summary(REML1)
```

```
# Linear mixed model fit by REML ['lmerMod']  
# Formula: height ~ graze_f * AspectCat + year_f + nativecov_sq + slope +  
# (1 | Park/plotID)  
# Data: graz  
#  
# REML criterion at convergence: 1677.9  
#  
# Scaled residuals:  
#      Min       1Q   Median       3Q      Max   
# -2.9236 -0.6372 -0.0822  0.4329  4.7964   
#  
# Random effects:  
# Groups      Name      Variance Std.Dev.  
# plotID:Park (Intercept) 11.358   3.370  
# Park        (Intercept)  2.478   1.574  
# Residual                26.351   5.133  
# Number of obs: 271, groups: plotID:Park, 36; Park, 4  
#  
# Fixed effects:  
#              Estimate Std. Error t value  
# (Intercept)  18.44908    3.09233   5.966  
# graze_f1     -5.25433    2.71310  -1.937  
# graze_f2     -2.57165    2.75561  -0.933  
# graze_f3     -2.57165    2.75561  -0.933  
# graze_f4     -2.57165    2.75561  -0.933  
# graze_f5     -2.57165    2.75561  -0.933  
# graze_f6     -2.57165    2.75561  -0.933  
# graze_f7     -2.57165    2.75561  -0.933  
# graze_f8     -2.57165    2.75561  -0.933  
# graze_f9     -2.57165    2.75561  -0.933  
# graze_f10    -2.57165    2.75561  -0.933  
# AspectCat1    0.00000    0.00000   0.000  
# AspectCat2    0.00000    0.00000   0.000  
# AspectCat3    0.00000    0.00000   0.000  
# AspectCat4    0.00000    0.00000   0.000  
# AspectCat5    0.00000    0.00000   0.000  
# AspectCat6    0.00000    0.00000   0.000  
# AspectCat7    0.00000    0.00000   0.000  
# AspectCat8    0.00000    0.00000   0.000  
# AspectCat9    0.00000    0.00000   0.000  
# AspectCat10   0.00000    0.00000   0.000  
# year_f1       0.00000    0.00000   0.000  
# year_f2       0.00000    0.00000   0.000  
# year_f3       0.00000    0.00000   0.000  
# year_f4       0.00000    0.00000   0.000  
# year_f5       0.00000    0.00000   0.000  
# year_f6       0.00000    0.00000   0.000  
# year_f7       0.00000    0.00000   0.000  
# year_f8       0.00000    0.00000   0.000  
# year_f9       0.00000    0.00000   0.000  
# year_f10      0.00000    0.00000   0.000  
# nativecov_sq  0.00000    0.00000   0.000  
# slope         0.00000    0.00000   0.000
```

Внутриклассовая корреляция

Для наблюдений на одном и том же участке $\sigma_{plotID}^2 / (\sigma_{plotID}^2 + \sigma_{Park}^2 + \sigma^2)$

```
3.370^2 / (1.574^2 + 3.370^2 + 5.133^2)
```

```
# [1] 0.282636
```

Для наблюдений в одном и том же парке $\sigma_{Park}^2 / (\sigma_{plotID}^2 + \sigma_{Park}^2 + \sigma^2)$

```
1.574^2 / (1.574^2 + 3.370^2 + 5.133^2)
```

```
# [1] 0.06165626
```

В результатах summary(REML1)

Random effects:

Groups	Name	Variance	Std.Dev.
plotID:Park	(Intercept)	11.358	3.370
Park	(Intercept)	2.478	1.574
Residual		26.351	5.133

Number of obs: 271, groups: plotID:Park, 36; Park, 4

- ▶ Значения высоты травяного покрова похожи внутри участка. Сходство наблюдений внутри одного парка слабее.

Результаты полной модели

summary(REML1)

```
# Linear mixed model fit by REML ['lmerMod']
# Formula: height ~ graze_f * AspectCat + year_f + nativecov_sq + slope +
# (1 | Park/plotID)
# Data: graz
#
# REML criterion at convergence: 1677.9
#
# Scaled residuals:
#      Min       1Q   Median       3Q      Max
# -2.9236 -0.6372 -0.0822  0.4329  4.7964
#
# Random effects:
# Groups      Name                Variance Std.Dev.
# plotID:Park (Intercept) 11.358    3.370
# Park        (Intercept)  2.478    1.574
# Residual                26.351    5.133
# Number of obs: 271, groups: plotID:Park, 36; Park, 4
#
# Fixed effects:
#              Estimate Std. Error t value
# (Intercept) 18.44908    3.09233   5.966
# graze_f1    -5.25433    2.71310  -1.937
# AspectCatS   8.57165    2.75561   3.111
# year_f2005   6.80750    1.42386   4.781
# year_f2006   4.07972    1.41617   2.881
# year_f2007  -0.58734    1.42261  -0.413
# year_f2008  -2.70376    1.40999  -1.918
# year_f2009  -2.69192    1.40996  -1.909
# year_f2010  -2.69192    1.40996  -1.909
# year_f2011   2.72004    1.41800   1.924
```


Данные для графика предсказаний фиксированной части модели

Используем для визуализации модель, подобранную при помощи REML

Исходные данные

```
NewData_REML1 <- expand.grid(graze_f = levels(graze$graze_f),  
                             AspectCat = levels(graze$AspectCat),  
                             year_f = levels(graze$year_f))  
NewData_REML1$nativecov_sq <- mean(graze$nativecov_sq)  
NewData_REML1$slope <- mean(graze$slope)
```

Предсказанные значения при помощи матриц

```
X <- model.matrix(~ graze_f * AspectCat + year_f + nativecov_sq + slope,  
                  data = NewData_REML1)  
betas = fixef(REML1)  
NewData_REML1$fit <- X %*% betas
```

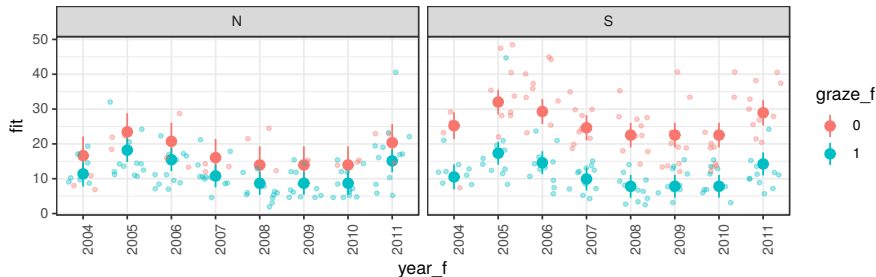
Стандартные ошибки и дов. интервалы

```
NewData_REML1$se <- sqrt( diag(X %*% vcov(REML1) %*% t(X)) )  
NewData_REML1$lwr <- NewData_REML1$fit - 2 * NewData_REML1$se  
NewData_REML1$upr <- NewData_REML1$fit + 2 * NewData_REML1$se
```

График предсказаний фиксированной части модели

На южных склонах высота травы выше там, где не пасут скот, а на северных нет.
(Строго говоря, нужен еще пост хок тест, чтобы это утверждать.)

```
ggplot(data = NewData_REML1, aes(x = year_f, y = fit, colour = graze_f)) +  
  geom_pointrange(aes(ymin = lwr, ymax = upr)) +  
  facet_wrap(~ AspectCat) +  
  geom_jitter(data = graz, aes(y = height), alpha = 0.35, size = 1) +  
  theme(axis.text.x = element_text(angle = 90))
```



Вариант решения с подбором оптимальной модели (самостоятельно)

Задание 2

Оптимизируйте модель с предыдущего шага

Сделайте анализ остатков

Опишите и визуализируйте финальную модель

Решение: Подбор оптимальной модели (1)

Для подбора оптимальной модели воспользуемся тестами отношения правдоподобий. Для него нужно использовать модели, подобранные при помощи ML

```
drop1(ML1, test = "Chi")
```

```
# Single term deletions
#
# Model:
# height ~ graze_f * AspectCat + year_f + nativecov_sq + slope +
# (1 | Park/plotID)
#
```

	Df	AIC	LRT	Pr(Chi)
# <none>		1729.4		
# year_f	7	1819.8	104.405	< 2.2e-16 ***
# nativecov_sq	1	1728.9	1.478	0.224030
# slope	1	1727.8	0.450	0.502531
# graze_f:AspectCat	1	1736.3	8.956	0.002765 **
# ---				

```
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Решение: Подбор оптимальной модели (2)

```
ML1.1 <- update(ML1, .~.-slope)
drop1(ML1.1, test = "Chi")
```

```
# Single term deletions
#
# Model:
# height ~ graze_f + AspectCat + year_f + nativecov_sq + (1 | Park/plotID) +
#   graze_f:AspectCat
#
```

	Df	AIC	LRT	Pr(Chi)
# <none>		1727.8		
# year_f	7	1818.1	104.277	< 2.2e-16 ***
# nativecov_sq	1	1727.5	1.681	0.194802
# graze_f:AspectCat	1	1734.4	8.535	0.003484 **

```
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Решение: Подбор оптимальной модели (3)

```
ML1.2 <- update(ML1.1, .~.-nativecov_sq)
drop1(ML1.2, test = "Chi")
```

```
# Single term deletions
#
# Model:
# height ~ graze_f + AspectCat + year_f + (1 | Park/plotID) + graze_f:AspectCat
#           Df      AIC      LRT   Pr(Chi)
# <none>             1727.5
# year_f             7 1818.7 105.152 < 2.2e-16 ***
# graze_f:AspectCat  1 1733.6   8.087  0.004458 **
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

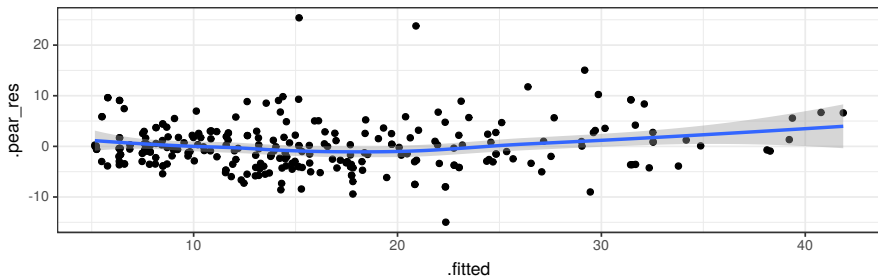
Решение: Анализ остатков

```
# Данные для анализа остатков
```

```
ML1.2_diag <- data.frame(  
  graz,  
  .pear_res = residuals(ML1.2, type = "pearson"),  
  .fitted = fitted(ML1.2, type = "response"))
```

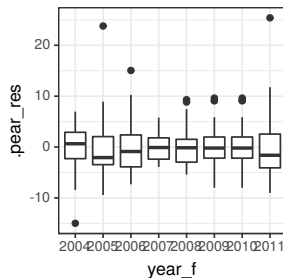
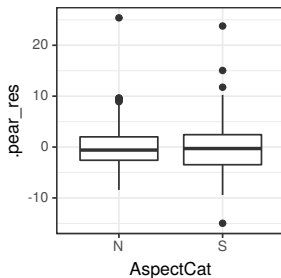
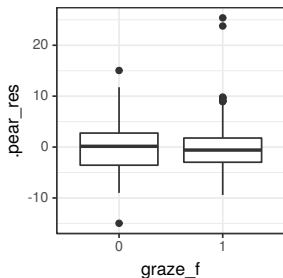

Решение: График остатков

```
gg_res <- ggplot(data = ML1.2_diag, aes(y = .pear_res))  
gg_res + geom_point(aes(x = .fitted)) +  
  geom_smooth(aes(x = .fitted))
```



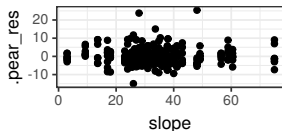
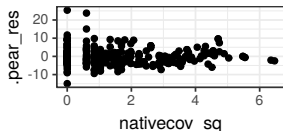
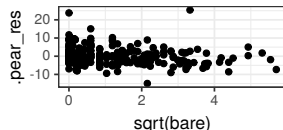
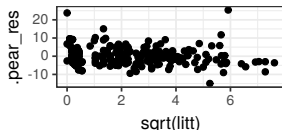
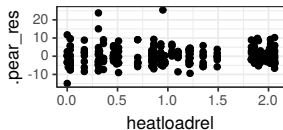
Решение: Графики остатков от переменных в модели

```
plot_grid(gg_res + geom_boxplot(aes(x = graze_f)),  
gg_res + geom_boxplot(aes(x = AspectCat)),  
gg_res + geom_boxplot(aes(x = year_f)),  
ncol = 3)
```



Решение: Графики остатков от переменных не в модели

```
plot_grid(  
  gg_res + geom_point(aes(x = heatloadrel)),  
  gg_res + geom_point(aes(x = sqrt(litt))),  
  gg_res + geom_point(aes(x = sqrt(bare))),  
  gg_res + geom_point(aes(x = nativecov_sq)),  
  gg_res + geom_point(aes(x = slope)),  
  ncol = 3)
```



- Паттерн на графике heatloadrel, nativecov_sq

Решение: Тестируем влияние факторов в финальной модели

Для тестов отношения правдоподобий используем финальную модель, подобранную при помощи ML

```
drop1(ML1.2, test = 'Chi')
```

```
# Single term deletions
#
# Model:
# height ~ graze_f + AspectCat + year_f + (1 | Park/plotID) + graze_f:AspectCat
#               Df      AIC      LRT   Pr(Chi)
# <none>                1727.5
# year_f                7 1818.7 105.152 < 2.2e-16 ***
# graze_f:AspectCat    1 1733.6   8.087  0.004458 **
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Высота растительного покрова:

- ▶ на склонах разной экспозиции по-разному зависит от выпаса скота (достоверное взаимодействие)
- ▶ различается в разные годы
- ▶ не зависит от покрытия местных растений и крутизны склона

Решение: Описываем результаты

Для описания результатов используем модель, подобранную при помощи REML, т.к. он дает более точные оценки случайных эффектов

```
REML1.2 <- update(ML1.2, REML = TRUE)
```

Решение: Результаты

```
summary(REML1.2)
```

```
# Linear mixed model fit by REML ['lmerMod']
# Formula:
# height ~ graze_f + AspectCat + year_f + (1 | Park/plotID) + graze_f:AspectCat
# Data: graz
#
# REML criterion at convergence: 1675.6
#
# Scaled residuals:
#      Min       1Q   Median       3Q      Max
# -2.9168 -0.6306 -0.0791  0.4264  4.9848
#
# Random effects:
# Groups      Name                Variance Std.Dev.
# plotID:Park (Intercept) 13.761    3.710
# Park        (Intercept)  1.384    1.177
# Residual                25.916    5.091
# Number of obs: 271, groups: plotID:Park, 36; Park, 4
#
# Fixed effects:
#              Estimate Std. Error t value
# (Intercept)   16.4076    2.7767   5.909
# graze_f1      -5.6780    2.7602  -2.057
# AspectCatS     9.4405    2.8750   3.284
# year_f2005     6.6263    1.4043   4.719
# year_f2006     4.1314    1.4043   2.942
# year_f2007    -0.3683    1.4043  -0.262
# year_f2008    -2.7040    1.3982  -1.934
# year_f2009    -2.6896    1.3982  -1.924
# year_f2010    -2.6896    1.3982  -1.924
# year_f2011     2.0873    1.3082   1.592
```

Решение: Внутрикласовая корреляция

Для расчета нужна модель, подобранная при помощи REML

Random effects:

Groups	Name	Variance	Std.Dev.
plotID:Park	(Intercept)	13.761	3.710
Park	(Intercept)	1.384	1.177
Residual		25.916	5.091

Number of obs: 271, groups: plotID:Park, 36; Park, 4

Для наблюдений на одном и том же участке $\sigma_{plotID}^2 / (\sigma_{plotID}^2 + \sigma_{Park}^2 + \sigma^2)$

```
3.710^2 / (1.177^2 + 3.710^2 + 5.091^2)
```

```
# [1] 0.3351563
```

Для наблюдений в одном и том же парке $\sigma_{Park}^2 / (\sigma_{plotID}^2 + \sigma_{Park}^2 + \sigma^2)$

```
1.177^2 / (1.177^2 + 3.710^2 + 5.091^2)
```

```
# [1] 0.03378768
```

- ▶ Значения высоты травяного покрова похожи внутри участка. Сходство наблюдений внутри одного парка слабее.

Решение: Данные для графика предсказаний фиксированной части модели

Используем для визуализации модель, подобранную при помощи REML

Исходные данные

```
NewData_REML1.2 <- expand.grid(graза_f = levels(graза$газа_f),  
                             AspectCat = levels(graза$AspectCat),  
                             year_f = levels(graза$year_f))  
NewData_REML1.2$nativecov_sq <- mean(graза$nativecov_sq)  
NewData_REML1.2$slope <- mean(graза$slope)
```

Предсказанные значения при помощи матриц

```
X <- model.matrix(~ граза_f * AspectCat + year_f, data = NewData_REML1.2)  
betas = fixef(REML1.2)  
NewData_REML1.2$fit <- X %*% betas
```

Стандартные ошибки и дов. интервалы

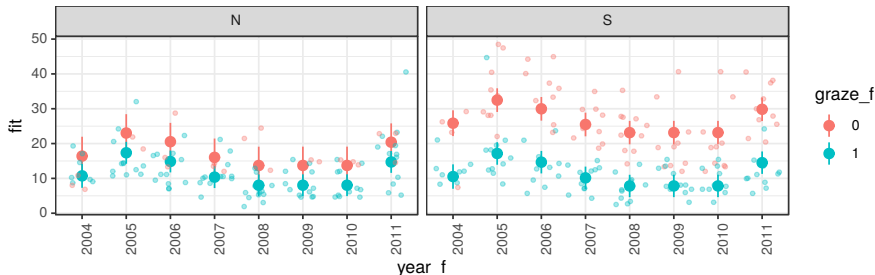
```
NewData_REML1.2$se <- sqrt( diag(X %*% vcov(REML1.2) %*% t(X)) )  
NewData_REML1.2$lwr <- NewData_REML1.2$fit - 2 * NewData_REML1.2$se  
NewData_REML1.2$upr <- NewData_REML1.2$fit + 2 * NewData_REML1.2$se
```


Решение: График предсказаний фиксированной части модели

На южных склонах высота травы выше там, где не пасут скот, а на северных нет.
(Строго говоря, нужен еще пост хок тест, чтобы это утверждать)

Не удивляйтесь тому, что график похож на предыдущий, т.к. те факторы, которые мы удалили из модели и так не влияли.

```
ggplot(data = NewData_REML1.2, aes(x = year_f, y = fit, colour = graze_f)) +  
  geom_pointrange(aes(ymin = lwr, ymax = upr)) +  
  facet_wrap(~ AspectCat) +  
  geom_jitter(data = graz, aes(y = height), alpha = 0.35, size = 1) +  
  theme(axis.text.x = element_text(angle = 90))
```



Take-home messages

- ▶ Случайные факторы в смешанных моделях могут быть вложены друг в друга
- ▶ Есть два способа подбора коэффициентов в смешанных моделях: ML и REML. Для разных этапов анализа важно, каким именно способом подобрана модель.

Дополнительные ресурсы

- ▶ Crawley, M.J. (2007). The R Book (Wiley).
- ▶ Faraway, J. J. (2017). Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models (Vol. 124). CRC press.
- ▶ Zuur, A. F., Hilbe, J., & Ieno, E. N. (2013). A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists. Highland Statistics.
- ▶ Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., and Smith, G.M. (2009). Mixed Effects Models and Extensions in Ecology With R (Springer)
- ▶ Pinheiro, J., Bates, D. (2000). Mixed-Effects Models in S and S-PLUS. Springer