

Коррекция гетерогенности дисперсий

Линейные модели...

Вадим Хайтов, Марина Варфоломеева

СПбГУ



Способы моделирования “поведения” дисперсии

Вы сможете

- ▶ Построить модели для данных, демонстрирующих высокую степень гетерогенности дисперсии
- ▶ Подобрать оптимальную модель, учитывающую ковариаты дисперсии
- ▶ Построить смешанную модель учитывающую не только группирующие (случайные) факторы, но и гетерогенность дисперсии.



Гетерогенность дисперсии



Пример: Способствуют ли взрослые мидии притоку молоди?

Данные взяты из работы Khaitov, 2013

```
myt <- read.table("data/myt.csv", sep=";", header = TRUE)  
head(myt, 12)
```

#	Year	Bank	Sample	Recruits	Small	Large
# 1	1997	vor2	1	0	12	25
# 2	1997	vor2	2	0	23	27
# 3	1997	vor2	3	0	17	36
# 4	1997	vor2	4	0	17	27
# 5	1997	vor2	5	0	30	25
# 6	1997	vor2	6	0	12	27
# 7	1999	vor2	1	12	12	41
# 8	1999	vor2	2	2	16	38
# 9	1999	vor2	3	1	13	49
# 10	1999	vor2	4	6	19	52
# 11	1999	vor2	5	1	7	31
# 12	1999	vor2	6	2	9	34



В качестве зависимой переменной будем анализировать

$$\sqrt{N_{recruits}}$$

```
myt$Sq_Recruits <- sqrt(myt$Recruits)  
myt$fYear <- factor(myt$Year)
```



Строим обычную регрессионную модель

```
mod_formula <- Sq_Recruits ~ Large + fYear + Bank +  
  Large:fYear + Large:Bank
```

```
M1_lm <- lm(mod_formula, data = myt)
```

```
library(car)
```

```
Anova(M1_lm)
```

```
# Anova Table (Type II tests)
```

```
#
```

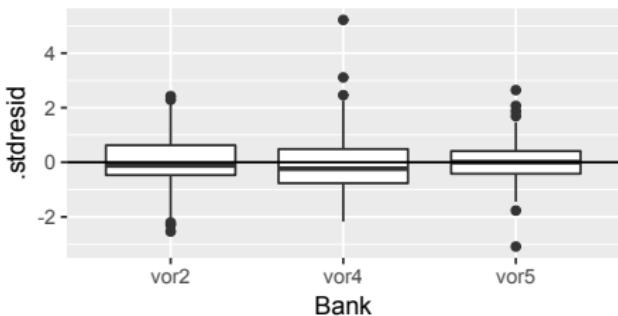
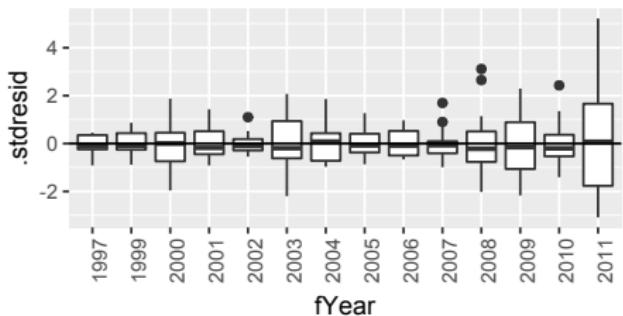
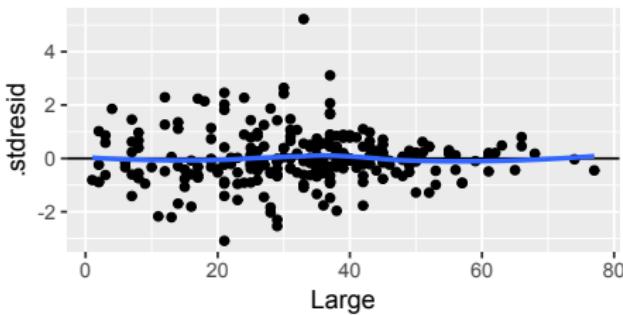
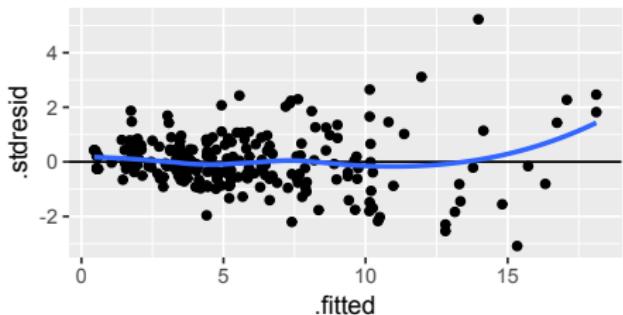
```
# Response: Sq_Recruits
```

	Sum Sq	Df	F value	Pr(>F)
# Large	32	1	5.85	0.01637 *
# fYear	2286	13	31.95	< 2e-16 ***
# Bank	263	2	23.85	0.00000000043 ***
# Large:fYear	220	13	3.08	0.00033 ***
# Large:Bank	2	2	0.14	0.86594
# Residuals	1194	217		
# ---				
# Signif. codes:	0	'***'	0.001	'**'
			0.01	'*'
			0.05	'. '
			0.1	' '
			1	

Можем ли мы доверять этим результатам?



Диагностика модели

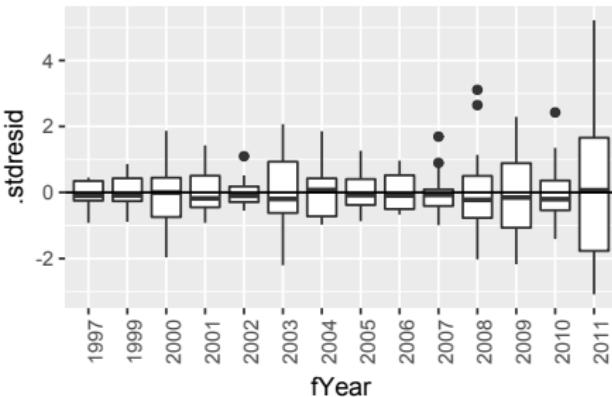
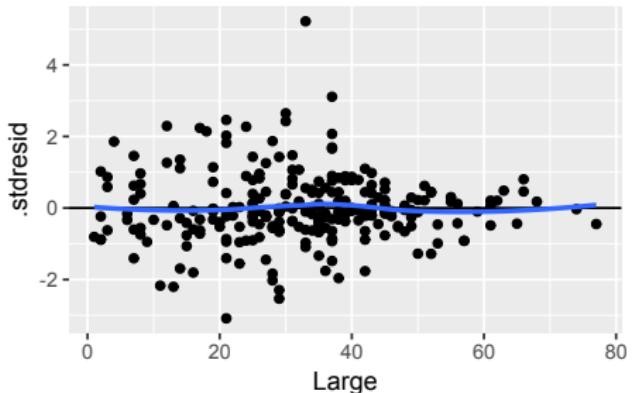


Мы не можем доверять величинам уровня значимости!

Нарушается условие применимости линейных моделей, основанных на нормальном распределении остатков с неизменной дисперсией.



Ковариата дисперсии (Variance covariate)



Видно, что дисперсия остатков убывает по мере увеличения значения переменной Large и меняется год от года

Переменные Large и fYear являются ковариатами дисперсии



Моделирование гетерогенности дисперсии



В начале курса мы записывали “обычную” регрессионную модель в таком виде

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

Фиксированная часть модели: $\mathbf{X}\boldsymbol{\beta}$

Случайная часть модели: ε

Обычно мы считали, что $\varepsilon \sim N(0, \sigma^2)$



В начале курса мы записывали “обычную” регрессионную модель в таком виде

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

Фиксированная часть модели: $\mathbf{X}\boldsymbol{\beta}$

Случайная часть модели: ε

Обычно мы считали, что $\varepsilon \sim N(0, \sigma^2)$

т.е. остатки независимы и одинаково распределены со средним 0 и дисперсией σ^2 , одинаковой для всех уровней y_i . То есть остатки - это шум, в котором нет каких-то паттернов.



Теория смешанных моделей расширяет наши представления о случайной части модели

$$Y = \underbrace{\text{fixed part}}_{\alpha + \beta_1 X_1 + \dots + \beta_q X_q} + \underbrace{\text{random part}}_{\alpha + f_1(X_1) + \dots + f_q(X_q)}$$

Heterogeneity
Nested data (random effects)
Temporal correlation
Spatial correlation
Random noise

Рис. из Zuur et al. 2009



Модели на языке матриц

Простая линейная модель

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$$

Смешанная линейная модель с группирующими факторами

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\varepsilon}_i \sim N(0, \Sigma_i)$$

$$\mathbf{b}_i \sim N(0, \mathbf{D})$$

Расширенная смешанная линейная модель

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 \Lambda_i)$$

$$\mathbf{b}_i \sim N(0, \mathbf{D})$$



Ковариата дисперсии (Variance covariate)

Расширенная модель может включать еще один компонент

$$\varepsilon \sim N(0, \sigma^2 \times f(VC))$$

VC - ковариата дисперсии

$f(VC)$ - функция, вводящая поправку, стабилизирующую дисперсию

В зависимости от формы функции $f(VC)$ мы получим разную структуру дисперсии в модели



Различные формы структуры дисперсии



Для дальнейших вычислений необходимо использовать функцию `gls` из пакета `nlme`

```
library(nlme)
M1_gls <- gls(mod_formula, data = myt)
```

Если ничего не менять, то эта функция дает результаты полностью идентичные результатам функции `lm()`.

Для оценки параметров по умолчанию используется Restricted Maximum Likelihood (REML). Этот метод дает более точные оценки случайных факторов, чем обычный ML

Осторожно! Модели, подобранные с помощью REML, можно сравнивать только если у них одинаковая фиксированная часть!

```
Anova(M1_gls)
```

```
# Analysis of Deviance Table (Type II tests)
#
# Response: Sq_Recruits
#          Df  Chisq Pr(>Chisq)
# Large      1   5.85   0.01555 *
# fYear     13  415.39   < 2e-16 ***
# Bank       2   47.71   4.4e-11 ***
# Large:fYear 13  40.05   0.00014 ***
# Large:Bank   2   0.29   0.86586
#
```



1. Фиксированная структура дисперсии

Дисперсия изменяется пропорционально значениям ковариаты дисперсии

$$\varepsilon_i \sim N(0, \sigma^2 \times VC_i)$$

Предположим, что дисперсия меняется пропорционально численности больших мидий (Large)

```
M2_gls <- gls(mod_formula, data = myt, weights = varFixed( ~ Large))
```



Можем сравнить две модели при помощи AIC

```
AIC(M1_gls, M2_gls)
```

```
#      df  AIC
# M1_gls 33 1228
# M2_gls 33 1290
```



Можем сравнить две модели при помощи AIC

```
AIC(M1_gls, M2_gls)
```

```
#      df  AIC
# M1_gls 33 1228
# M2_gls 33 1290
```

Стало хуже. Но дисперсия явно убывала по мере увеличения значений переменной Large...



2. Разные дисперсии для разных уровней категориальных предикторов

$$\varepsilon_{ij} \sim N(0, \sigma_j^2)$$

При построении моделей с такой структурой дисперсии подбирается $k - 1$ новых параметров, где k — количество уровней категориального предиктора.

```
M3_gls <- gls(mod_formula, data = myt, weights = varIdent(form = ~1|fYear))
```



Сравним модели при помощи LRT

Важно! Модели M1_gls и M3_gls вложенные, поэтому их можно сравнивать LRT

M1_gls: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$

M3_gls: $k_1\sigma_1^2 = k_2\sigma_2^2 = \dots = k_m\sigma_m^2$

```
anova(M1_gls, M3_gls)
```

```
#          Model df  AIC  BIC logLik   Test L.Ratio p-value
# M1_gls     1 33 1228 1340    -581
# M3_gls     2 46 1131 1287    -520 1 vs 2      123  <.0001
```



Сравним модели при помощи LRT

Важно! Модели M1_gls и M3_gls вложенные, поэтому их можно сравнивать LRT

M1_gls: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$

M3_gls: $k_1\sigma_1^2 = k_2\sigma_2^2 = \dots = k_m\sigma_m^2$

```
anova(M1_gls, M3_gls)
```

```
#      Model df  AIC  BIC logLik   Test L.Ratio p-value
# M1_gls     1 33 1228 1340    -581
# M3_gls     2 46 1131 1287    -520 1 vs 2       123  <.0001
```

Модель M3_gls лучше!



Для второго дискретного предиктора можно построить аналогичную модель

```
M3_gls2 <- gls(mod_formula, data = myt, weights = varIdent(form = ~1|Bank))  
anova(M1_gls, M3_gls2)
```

```
#          Model df  AIC  BIC logLik   Test L.Ratio p-value  
# M1_gls     1 33 1228 1340    -581  
# M3_gls2    2 35 1216 1334    -573 1 vs 2      16.5  0.0003
```



Для второго дискретного предиктора можно построить аналогичную модель

```
M3_gls2 <- gls(mod_formula, data = myt, weights = varIdent(form = ~1|Bank))

anova(M1_gls, M3_gls2)
```

```
#          Model df  AIC  BIC logLik   Test L.Ratio p-value
# M1_gls     1 33 1228 1340    -581
# M3_gls2    2 35 1216 1334    -573 1 vs 2     16.5  0.0003
```

Модель M3_gls2 лучше, но не настолько.

Дальше не будем пока сравнивать модели с разными структурами дисперсий по одной. Сначала построим все, а потом все сравним.



Что произошло в результате работы функции varIdent()?

```
summary(M3_gls)
```

Часть вывода summary(M3_gls)

Variance function:

Structure: Different standard deviations per stratum

Formula: ~1 | fYear

Parameter estimates:

1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1.00	2.62	4.39	3.47	2.84	5.85	4.93	3.21	2.95	3.87	7.98
2009	2010	2011								
9.26	5.97	13.59								

$$\varepsilon_{ij} \sim N(0, \sigma_j^2)$$

Т.е. в выводе summary() присутствуют оценки σ_j^2



3. Степенная зависимость дисперсии от ковариаты

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times |VC|^{2\delta})$$

Параметр δ неизвестен и требует оценки

Если $\delta = 0$, то структура дисперсии будет аналогична структуре дисперсии в "обычной" регрессионной модели, где $\varepsilon \sim N(0, \sigma^2)$

Важно! Если значения ковариаты дисперсии могут принимать значение равное нулю, то такая форма структуры дисперсии не определена и использоваться не может.

```
M4_gls <- gls(mod_formula, data = myt, weights = varPower(form = ~ Large))
```



Что произошло в результате работы функции varPower()?

```
summary(M4_gls)
```

Часть вывода summary(M4_gls)

Variance function:

Structure: Power of variance covariate

Formula: ~Large

Parameter estimates:

power

-0.214

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times |VC|^{2\delta})$$

Оценка параметра δ

```
M4_gls$modelStruct
```

```
# varStruct parameters:  
# power  
# -0.214
```



Задание

Степенная зависимость дисперсии от ковариаты может учитывать и взаимодействие ковариаты дисперсии с категориальными предикторами

Напишите код, с помощью которого в модели будет учтена степенная зависимость дисперсии от переменной Large, но разная для каждого уровня фактора fYear. Аналогичный код напишите для фактора Bank

Подсказка: Изучите справку по функции varPower()



Решение: разная степенная зависимость дисперсии от ковариаты для разных уровней дискретного предиктора

```
M5_gls <- gls(mod_formula, data = myt,  
                weights = varPower(form = ~ Large|fYear))  
M6_gls <- gls(mod_formula, data = myt,  
                weights = varPower(form = ~ Large|Bank))
```

```
M5_gls$modelStruct
```

```
# varStruct parameters:  
#   1997   1999   2000   2001   2002   2003   2004   2005  
# -0.4236 -0.3033 -0.1106 -0.1966 -0.2511 -0.0680 -0.1480 -0.2494  
#   2006   2007   2008   2009   2010   2011  
# -0.2805 -0.1650  0.0359  0.0860 -0.0536  0.1860
```

```
M6_gls$modelStruct
```

```
# varStruct parameters:  
#   vor2   vor4   vor5  
# -0.1980 -0.0488 -0.2299
```



4. Экспоненциальная зависимость дисперсии от ковариаты

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times e^{2\delta \times VC_i})$$

Эта форма структуры дисперсии может применяться для случаев, когда $VC = 0$

Если $\delta = 0$, то структура дисперсии будет аналогична структуре дисперсии в "обычной" регрессионной модели, то есть $\varepsilon_{ij} \sim N(0, \sigma^2)$

```
M7_gls <- gls(mod_formula, data = myt, weights = varExp(form = ~ Large))
M8_gls <- gls(mod_formula, data = myt, weights = varExp(form = ~ Large | fYear))
M9_gls <- gls(mod_formula, data = myt, weights = varExp(form = ~ Large | Bank))
```



Что произошло в результате работы функции varExp()?

Оцененные параметры

M7_gls\$modelStruct

```
# varStruct parameters:  
# expon  
# -0.02
```

M8_gls\$modelStruct

```
# varStruct parameters:  
# 1997 1999 2000 2001 2002 2003 2004  
# -0.03093 -0.02296 -0.00946 -0.01591 -0.01679 -0.00395 -0.01906  
# 2005 2006 2007 2008 2009 2010 2011  
# -0.02190 -0.02417 -0.01310 0.00743 0.00939 -0.00347 0.02475
```

M9_gls\$modelStruct

```
# varStruct parameters:  
# vor2 vor4 vor5  
# -0.02360 -0.00896 -0.02477
```



5. Усложненная степенная зависимость дисперсии от ковариаты

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times (\delta_1 + |VC|^{2\delta_2})^2)$$

Вопрос:

При каких значениях параметров функции $f(VC)$ структура дисперсии будет аналогична структуре дисперсии в "обычной" регрессионной модели?



5. Усложненная степенная зависимость дисперсии от ковариаты

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times (\delta_1 + |VC|^{2\delta_2})^2)$$

Вопрос:

При каких значениях параметров функции $f(VC)$ структура дисперсии будет аналогична структуре дисперсии в "обычной" регрессионной модели?

Ответ:

При $\delta_1 = 0$ и $\delta_2 = 0$ выражение $\varepsilon_{ij} \sim N(0, \sigma^2 \times (0 + |VC|^0)^2)$ будет эквивалентно $\varepsilon_{ij} \sim N(0, \sigma^2)$

```
M10_gls <- gls(mod_formula, data = myt,
                  weights = varConstPower(form = ~ Large))
# M11_gls <- gls(mod_formula, data = myt,
#                   weights = varConstPower(form = ~ Large | fYear))
M12_gls <- gls(mod_formula, data = myt,
                  weights = varConstPower(form = ~ Large | Bank))
```



Что произошло в результате работы функции varConstPower()?

$$\varepsilon_{ij} \sim N(0, \sigma^2 \times (\delta_1 + |VC|^{2\delta_2})^2)$$

```
M10_gls$modelStruct
```

```
# varStruct parameters:  
#   const power  
# -17.005 -0.214
```

```
M12_gls$modelStruct
```

```
# varStruct parameters:  
# const.vor2 const.vor4 const.vor5 power.vor2 power.vor4 power.vor5  
# -1.3195 -17.2798 -1.9165 -0.2939 0.0127 -0.2608
```



6. Комбинированная структура дисперсии

```
M13_gls <- gls(mod_formula, data = myt,  
                 weights = varComb(varIdent(form = ~ fYear),  
                                     varPower(form = ~ Large)))  
  
M14_gls <- gls(mod_formula, data = myt,  
                 weights = varComb(varIdent(form = ~ Bank),  
                                     varPower(form = ~ Large)))  
  
M15_gls <- gls(mod_formula, data = myt,  
                 weights = varComb(varIdent(form = ~ fYear),  
                                     varExp(form = ~ Large)))  
  
M16_gls <- gls(mod_formula, data = myt,  
                 weights = varComb(varIdent(form = ~ Bank),  
                                     varExp(form = ~ Large)))
```



Задание

Найдите модель с наилучшей структурой дисперсии

Решение

```
AICs <- AIC(M1_gls, M2_gls, M3_gls,  
             M4_gls, M5_gls, M6_gls,  
             M7_gls, M8_gls, M9_gls,  
             M10_gls, M12_gls,M13_gls,  
             M14_gls, M15_gls, M16_gls)
```



Решение

```
AICs[AICs$AIC == min(AICs$AIC), ]
```

```
#      df  AIC  
# M5_gls 47 1131
```

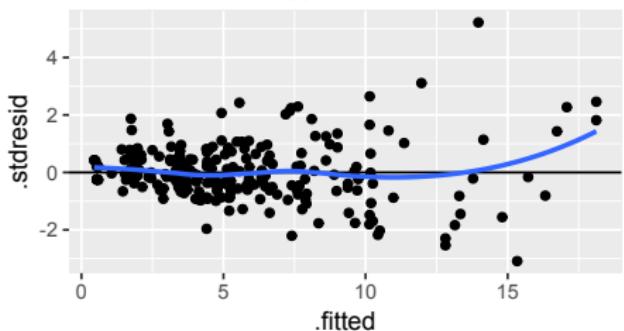
```
M5_gls$call
```

```
# gls(model = mod_formula, data = myt, weights = varPower(form = ~Large |  
#       fYear))
```

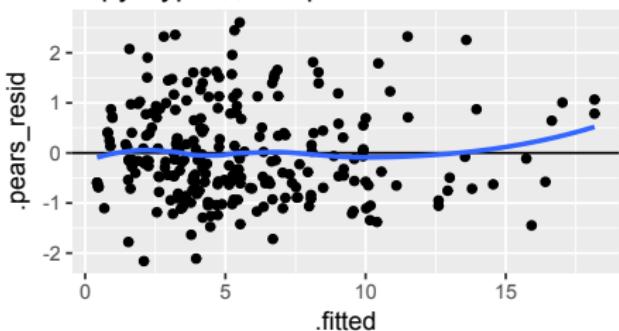


Диагностика модели с оптимальной структурой дисперсии

Было
в начальной модели

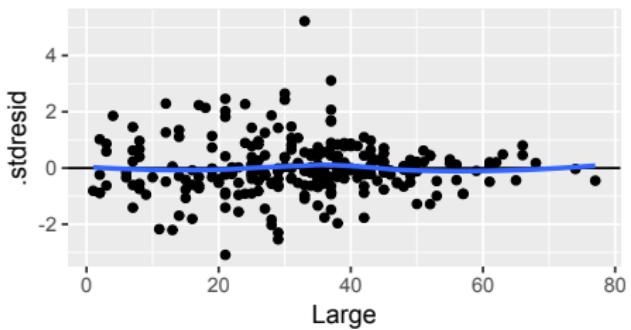


Стало после моделирования
структуры дисперсии

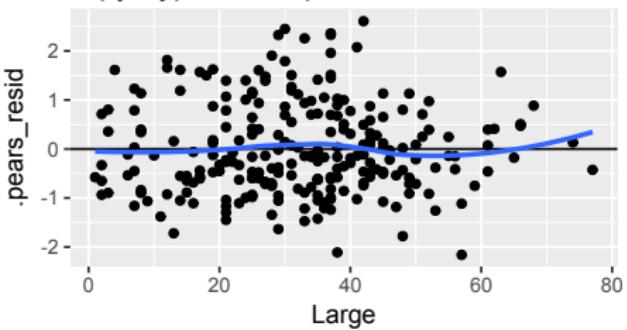


Диагностика модели с оптимальной структурой дисперсии

Было
в начальной модели

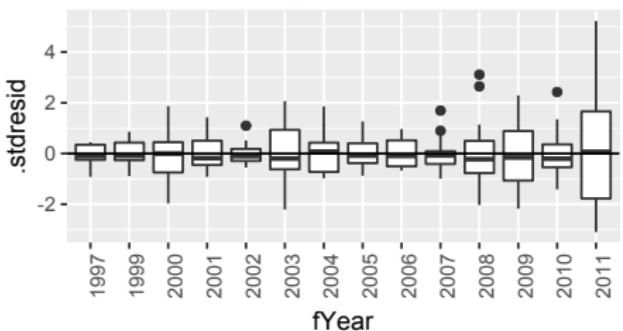


Стало после моделирования
структуры дисперсии

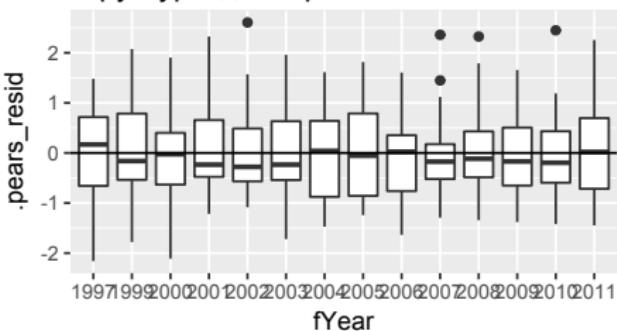


Диагностика модели с оптимальной структурой дисперсии

Было
в начальной модели

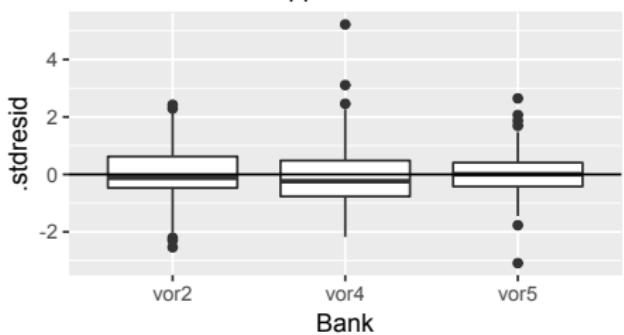


Стало после моделирования
структурой дисперсии

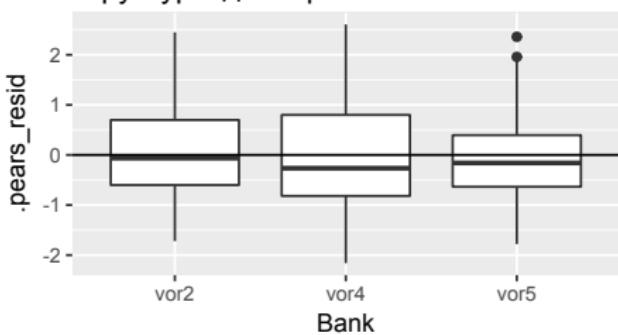


Диагностика модели с оптимальной структурой дисперсии

Было
в начальной модели



Стало после моделирования
структурой дисперсии



Можно ли упростить модель?

Обратите внимание, для оптимизации фиксированной части модели используем ML

```
M5_gls_ML <- update(M5_gls, method = "ML")
drop1(M5_gls_ML, test = "Chi")
```

```
# Single term deletions
#
# Model:
# Sq_Recruits ~ Large + fYear + Bank + Large:fYear + Large:Bank
#           Df  AIC   LRT Pr(>Chi)
# <none>      1044
# Large:fYear 13 1048  30.02   0.0047 ***
# Large:Bank   2 1050   9.68   0.0079 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

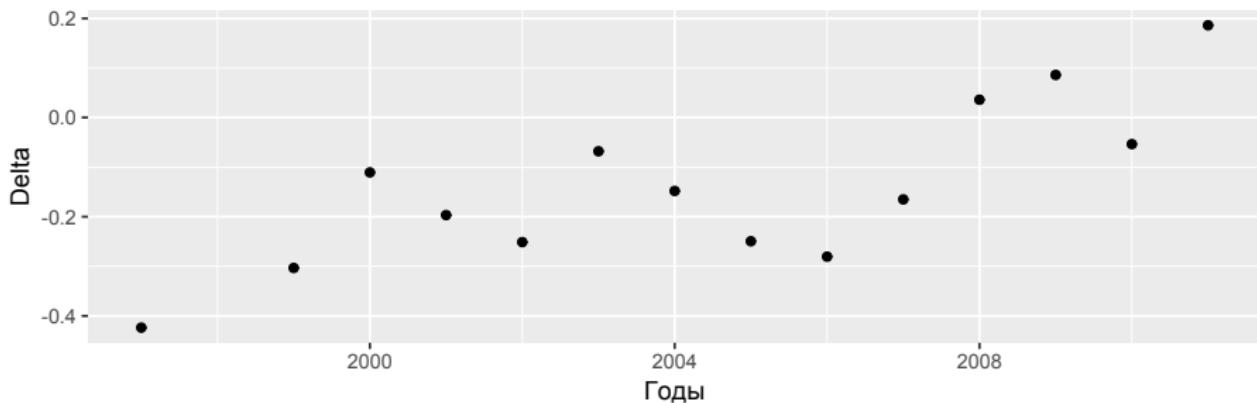
Эту модель упростить нельзя!

То есть, изменение структуры дисперсии заставляет формулировать иные биологические выводы.



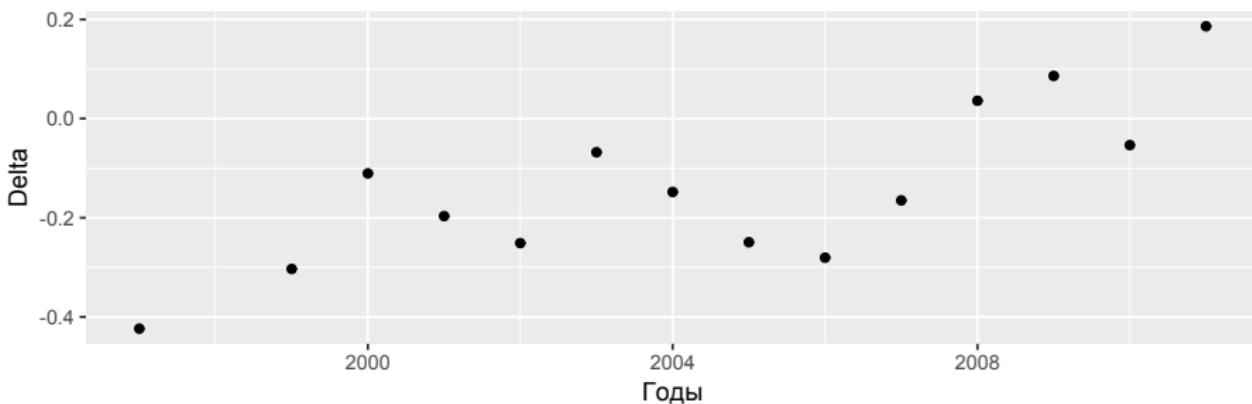
Структура дисперсии может иметь определенный биологический смысл

```
dat <- data.frame(x = c(1997, 1999:2011),  
                   y = as.vector(unlist(M5_gls$modelStruct)))  
ggplot(dat, aes(x = x, y = y)) + geom_point() + xlab("Годы") + ylab("Delta")
```



Структура дисперсии может иметь определенный биологический смысл

```
dat <- data.frame(x = c(1997, 1999:2011),  
                  y = as.vector(unlist(M5_gls$modelStruct)))  
ggplot(dat, aes(x = x, y = y)) + geom_point() + xlab("Годы") + ylab("Delta")
```



- ▶ В большинстве случаев параметр $\delta < 0$
- ▶ Чем больше обилие взрослых мидий, тем меньше варьирует обилие молоди
- ▶ Есть какая-то многолетняя динамика влияния обилия взрослых на "пятнистость" распределения молоди



Моделирование структуры дисперсии при наличии группирующих (случайных) факторов



Рост крыс при разной диете

Пример взят из книги Pinheiro & Bates, 2000 (Hand and Crowder (1996))

Три группы крыс, содержались при разных условиях кормления 64 дня.
Каждую крысу взвешивали с определенной периодичностью.

```
data("BodyWeight")
bw <- as.data.frame(BodyWeight)
head(bw, 14)
```

```
#   weight Time Rat Diet
# 1    240     1   1    1
# 2    250     8   1    1
# 3    255    15   1    1
# 4    260    22   1    1
# 5    262    29   1    1
# 6    258    36   1    1
# 7    266    43   1    1
# 8    266    44   1    1
# 9    265    50   1    1
# 10   272    57   1    1
# 11   278    64   1    1
# 12   225     1   2    1
# 13   230     8   2    1
# 14   230    15   2    1
```



Задание

Постройте модель, которая дала бы ответ на вопрос, изменяется ли характер роста крыс в зависимости от типа диеты?



Решение: Неправильная модель

```
M1 <- gls(weight ~ Time*Diet, data = bw)
```



Решение: Неправильная модель

```
M1 <- gls(weight ~ Time*Diet, data = bw)
```

Важно! Строить простую линейную модель в данном случае *некорректно*!

- ▶ Дизайн эксперимента изначально включает случайный фактор. Здесь мы имеем дело с повторными наблюдениями одного и того же объекта.
- ▶ Однако мы рассмотрим M1 для демонстрации того, что происходит, если не учитывать этой особенности экспериментального дизайна.

Anova(M1)

```
# Analysis of Deviance Table (Type II tests)
#
# Response: weight
#          Df  Chisq Pr(>Chisq)
# Time       1 19.55  0.0000098 ***
# Diet       2 2228.76   < 2e-16 ***
# Time:Diet  2   3.59     0.17
# ...
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Решение: Модель со случайными факторами

Важно! в этом эксперименте присутствует случайный (группирующий) фактор Rat, который необходимо учесть в модели.

```
M2 <- lme(weight ~ Time*Diet, data = bw, random = ~1|Rat)
M3 <- lme(weight ~ Time*Diet, data = bw, random = ~1 + Time|Rat)
```



Решение: Модель со случайными факторами

Важно! в этом эксперименте присутствует случайный (группирующий) фактор Rat, который необходимо учесть в модели.

```
M2 <- lme(weight ~ Time*Diet, data = bw, random = ~1|Rat)
M3 <- lme(weight ~ Time*Diet, data = bw, random = ~1 + Time|Rat)
```

Какую из моделей выбрать?

```
AIC(M2, M3)
```

```
#      df  AIC
# M2    8 1248
# M3   10 1172
```



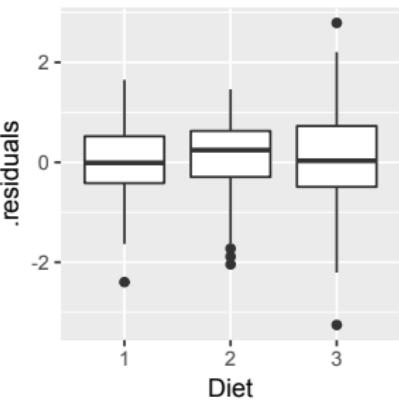
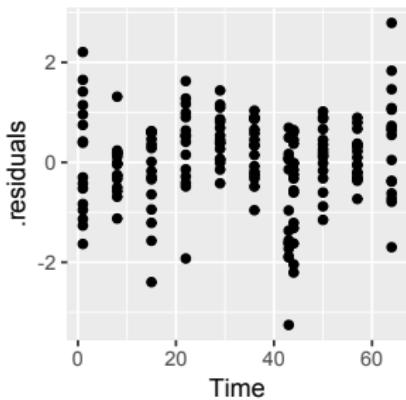
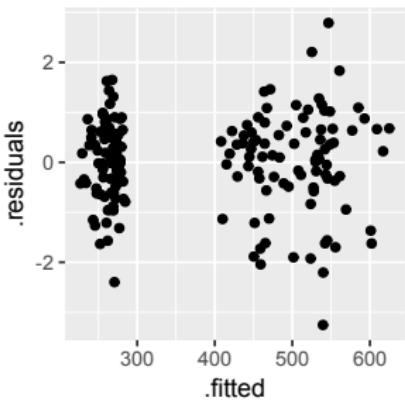
Решение: Пытаемся ответить на вопрос исследования

Anova(M3)

```
# Analysis of Deviance Table (Type II tests)
#
# Response: weight
#          Chisq Df Pr(>Chisq)
# Time      82.6  1    < 2e-16 ***
# Diet      170.7  2    < 2e-16 ***
# Time:Diet 15.2  2    0.00051 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Диагностика модели



Моделируем структуру дисперсии

```
M3_1 <- update(M3, weights = varIdent(form = ~ 1|Diet))
M3_2 <- update(M3, weights = varPower(form = ~Time))
M3_3 <- update(M3, weights = varPower(form = ~Time|Diet))
M3_4 <- update(M3, weights = varExp(form = ~Time))
M3_5 <- update(M3, weights = varExp(form = ~Time|Diet))
M3_6 <- update(M3, weights = varComb(varExp(form = ~Time),
                                         varIdent(form = ~1|Diet)))
```



Выбираем лучшую модель

```
AIC(M3, M3_1, M3_2, M3_3, M3_4, M3_5, M3_6)
```

```
#      df  AIC
# M3    10 1172
# M3_1  12 1164
# M3_2  11 1173
# M3_3  13 1158
# M3_4  11 1174
# M3_5  13 1155
# M3_6  13 1165
```



Отвечаю на вопрос

Anova(M3_5)

```
# Analysis of Deviance Table (Type II tests)
#
# Response: weight
#           Chisq Df Pr(>Chisq)
# Time      83.2  1    < 2e-16 ***
# Diet     169.3  2    < 2e-16 ***
# Time:Diet 17.3  2    0.00018 ***
# ---
# Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

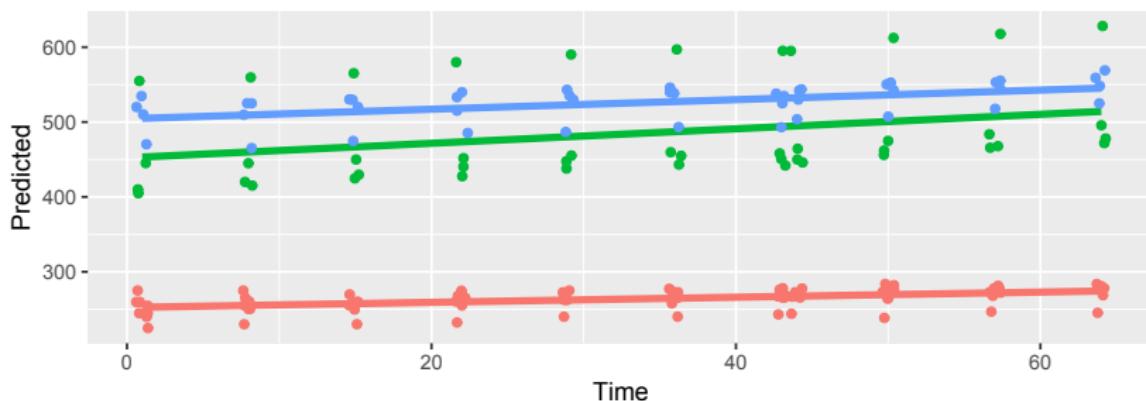


Смотрим на предсказания модели

```
MyData <- expand.grid(Time = unique(bw$Time), Diet = factor(1:3))

MyData$Predicted <- predict(M3_5, newdata = MyData, level = 0)

ggplot(MyData, aes(x = Time, y = Predicted, color = Diet)) +
  geom_line( size = 1.5) +
  geom_point(data = bw, aes(x = Time, y = weight),
             position = position_jitter())
```



Summary

Проблему гетерогенности дисперсии можно решить двумя способами:

1. Преобразование переменных
2. Введение в модель той или иной структуры дисперсии, учитывающей тот или иной набор ковариат дисперсии.



Что почитать

- ▶ Zuur, A.F. et al. 2009. Mixed effects models and extensions in ecology with R. - Statistics for biology and health. Springer, New York, NY.
- ▶ Pinheiro J, Bates D (2000) Mixed effects models in S and S-Plus. Springer-Verlag, New York, USA

