

# Линейные модели с дискретными предикторами

Линейные модели...

Марина Варфоломеева, Вадим Хайтов

Кафедра Зоологии беспозвоночных, Биологический факультет, СПбГУ



# Линейные модели с дискретными предикторами (дисперсионный анализ)

## Вы сможете

- ▶ Объяснить, в чем опасность множественных сравнений, и как с ними можно бороться
- ▶ Рассказать, как в дисперсионном анализе моделируются значения зависимой переменной
- ▶ Перечислить и проверить условия применимости дисперсионного анализа
- ▶ Интерпретировать и описать результаты, записанные в таблице дисперсионного анализа
- ▶ Провести множественные попарные сравнения при помощи post hoc теста Тьюки, представить и описать их результаты
- ▶ Построить график результатов дисперсионного анализа

## Множественные сравнения

## Пример: яйца кукушек

- ▶ species — вид птиц-хозяев (фактор)
- ▶ length — длина яиц кукушек в гнездах хозяев (зависимая переменная)

```
library(DAAG)
data("cuckoos")
# Положим данные в переменную с коротким названием, чтобы меньше печатать
cu <- cuckoos
head(cu, 3)
```

```
#   length breadth      species id
# 1   21.7    16.1 meadow.pipit 21
# 2   22.6    17.0 meadow.pipit 22
# 3   20.9    16.2 meadow.pipit 23
```



# Исследуем данные

```
# Пропущенных значений нет  
sum(is.na(cu))
```

```
# [1] 0
```

```
# Данные не сбалансированы (размеры групп разные)  
table(cu$species)
```

```
#  
# hedge.sparrow meadow.pipit pied.wagtail robin tree.pipit  
#           14           45           15           16           15  
#           wren  
#           15
```

## Изменим названия уровней фактора, чтобы было легче понять о каких птицах речь

```
levels(cu$species)
```

```
# [1] "hedge.sparrow" "meadow.pipit"  "pied.wagtail"  "robin"  
# [5] "tree.pipit"    "wren"
```

```
levels(cu$species) <- c("лес_зав", "луг_кон", "бел_тряс",  
                        "малин", "лес_кон", "крапив")
```



## Задание: Постройте график

Постройте график зависимости размера яиц кукушек от вида птиц-хозяев, в гнездах которых были обнаружены яйца (используйте `geom_boxplot`).

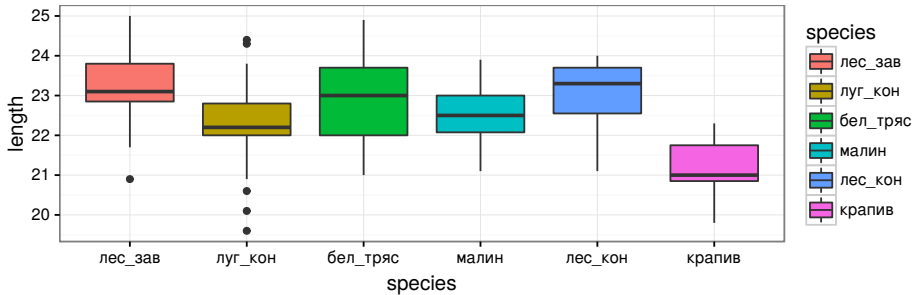
Раскрасьте график в зависимости от вида птиц-хозяев (используйте эстетики `fill` или `colour`)

### Дополнительное задание:

Попробуйте сменить палитру раскраски, используя `scale_colour_brewer` (варианты можно посмотреть в справке в подразделе примеров или в интернете [Colors \(ggplot2\): раздел RColorBrewer palette chart](#))

## Решение

```
library(ggplot2)
theme_set(theme_bw())
ggplot(data = cu, aes(x = species, y = length)) +
  geom_boxplot(aes(fill = species))
```



- Сейчас боксы расположены в беспорядке. Хорошо бы поменять порядок уровней



## Меняем порядок уровней

Упорядочим уровни по убыванию средней длины яиц.

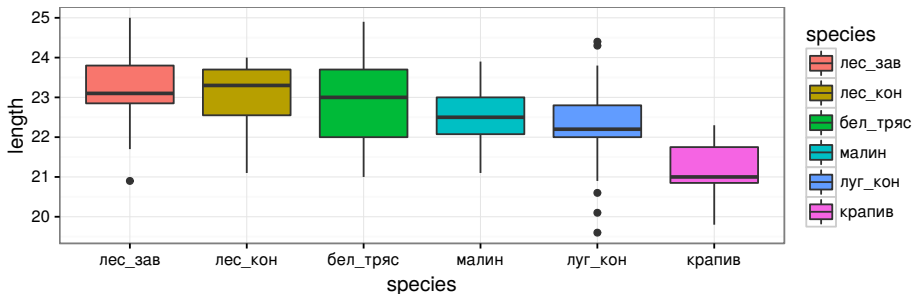
1. при помощи функции `reorder()` упорядочиваем по возрастанию средней длины яиц
2. меняем порядок уровней на противоположный при помощи функции `rev()`.

```
cu$species <- reorder(cu$species, cu$length, FUN = mean)
cu$species <- factor(cu$species, levels = rev(levels(cu$species)))
```

## График с новым порядком уровней

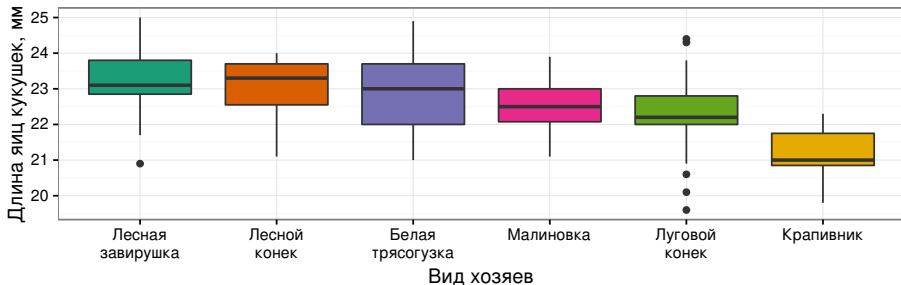
Поскольку изменив порядок уровней мы внесли изменения в исходные данные, придется полностью обновить график (т.к. `ggplot()` хранит данные внутри графика).

```
ggplot(data = cu, aes(x = species, y = length)) +  
  geom_boxplot(aes(fill = species))
```



Понравившийся график, если понадобится, можно в любой момент довести до ума, а остальные удалить

```
ggplot(data = cu, aes(x = species, y = length)) +  
  geom_boxplot(aes(fill = species)) +  
  labs(x = "Вид хозяев", y = "Длина яиц кукушек, мм") +  
  scale_fill_brewer(name = "Вид \nхозяев", palette = "Dark2") +  
  scale_x_discrete(labels = c("Лесная\nзавирушка", "Лесной\nконек",  
"Белая\nтрясогузка", "Малиновка", "Луговой\nконек", "Крапивник")) +  
  theme(legend.position = "none")
```



# Множественные сравнения

Мы могли бы сравнить длину яиц в гнездах резных хозяев при помощи t-критерия.

- ▶ 6 групп
- ▶ 15 сравнений

Если для каждого сравнения вероятность ошибки первого рода будет  $\alpha_{per\ comparison} = 0.05$ , то для группы — ?

# Множественные сравнения

Мы могли бы сравнить длину яиц в гнездах резных хозяев при помощи t-критерия.

- ▶ 6 групп
- ▶ 15 сравнений

Если для каждого сравнения вероятность ошибки первого рода будет  $\alpha_{per\ comparison} = 0.05$ , то для группы — ?

$$\alpha_{family\ wise} = 0.05 * 15 = 0.75$$

Мы рискуем найти различия там где их нет с 75% вероятностью!!!

# Поправка Бонферрони

Если нужно много сравнений, можно снизить  $\alpha_{per\ comparison}$

$$\alpha_{per\ comparison} = \frac{\alpha_{family\ wise}}{n}$$



# Поправка Бонферрони

Если нужно много сравнений, можно снизить  $\alpha_{per\ comparison}$

$$\alpha_{per\ comparison} = \frac{\alpha_{family\ wise}}{n}$$

Например, если хотим зафиксировать  $\alpha_{family\ wise} = 0.05$

С поправкой Бонферрони  $\alpha_{per\ comparison} = 0.05/15 = 0.003$

Очень жесткий критерий!

## Модель однофакторного дисперсионного анализа



## Модель фиктивных переменных, `contr.treatment`

Модель  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ , если  $\alpha_1 = 0$  примет уже знакомую нам форму

$$y_{ij} = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i + \varepsilon_{ij}$$

Коэффициенты линейной модели обозначают отклонения от базового уровня

Группа	Среднее на базовом уровне	Эффект для данного уровня	Случайная изменчивость
A1	$\beta_0$		$\varepsilon_{1j}$
A2	$\beta_0$	$\beta_1$	$\varepsilon_{2j}$
...	...	...	...
Ai	$\beta_0$	$\beta_i$	$\varepsilon_{ij}$

## Задание:

- ▶ Сколько переменных-болванок нужно, чтобы записать модель зависимости длины яиц кукушек от вида птиц-хозяев?

## Решение:

Сколько переменных-болванок нужно, чтобы записать модель зависимости длины яиц кукушек от вида птиц-хозяев?

- ▶ 5 переменных-болванок, т.к. 6 уровней у фактора species

```
levels(cu$species)
```

```
# [1] "лес_зав" "лес_кон" "бел_тряс" "малин" "луг_кон" "крапив"
```

Уровень лес\_зав будет базовым, и для его кодирования не нужна отдельная переменная).



## Дисперсионный анализ в матричном виде

Уравнение линейной модели для этого примера (в параметризации фиктивных переменных, `contr.treatment`):

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_5 x_{5i} + \varepsilon_i$$

- ▶ Здесь  $i = 1, \dots, n$ , т.е. порядковый номер наблюдения,
- ▶  $x_{1i}, \dots, x_{5i}$  — переменные-болванки для фактора `species`

Если расписать эту формулу, получится по отдельному уравнению для каждого из наблюдений:

$$y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_5 x_{51} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{12} + \cdots + \beta_5 x_{52} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{1n} + \cdots + \beta_5 x_{5n} + \varepsilon_n$$

Эту систему уравнений

$$y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_5 x_{51} + \varepsilon_i$$

$$y_2 = \beta_0 + \beta_1 x_{12} + \cdots + \beta_5 x_{52} + \varepsilon_i$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{1n} + \cdots + \beta_5 x_{5n} + \varepsilon_n$$

можно переписать в виде матриц:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{51} \\ 1 & x_{12} & \cdots & x_{52} \\ \vdots & & & \\ 1 & x_{1n} & \cdots & x_{5n} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Для такой длинной формы записи матриц

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{51} \\ 1 & x_{12} & \cdots & x_{52} \\ \vdots & & & \\ 1 & x_{1n} & \cdots & x_{5n} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

есть сокращенная форма записи:

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

- ▶  $\mathbf{Y}$  - матрица предсказанных значений, 1 столбец в  $n$  строк ( $n$  - число наблюдений)
- ▶  $\mathbf{X}$  - матрица независимых переменных с  $n$  строк, ее первый столбец содержит единицы, далее по столбцу для каждой из переменных в модели
- ▶  $\beta$  - матрица коэффициентов линейной модели, столбец
- ▶  $\varepsilon$  - матрица остатков, 1 столбец в  $n$  строк

# Модельная матрица **X** в дисперсионном анализе

Посмотреть своими глазами на эти переменные-болванки можно так:

```
X <- model.matrix(~ species, data = cu)
head(X)
```

```
# (Intercept) speciesлес_кон speciesбел_тряс speciesмалин
# 1           1             0             0             0
# 2           1             0             0             0
# 3           1             0             0             0
# 4           1             0             0             0
# 5           1             0             0             0
# 6           1             0             0             0
# speciesлуг_кон speciesкрапив
# 1           1             0
# 2           1             0
# 3           1             0
# 4           1             0
# 5           1             0
# 6           1             0
```



## Условия применимости дисперсионного анализа



## Условия применимости дисперсионного анализа:

- ▶ Случайность и независимость наблюдений внутри групп
- ▶ Нормальное распределение остатков
- ▶ Гомогенность дисперсий остатков
- ▶ Отсутствие коллинеарности факторов (независимость групп)

### Другие ограничения:

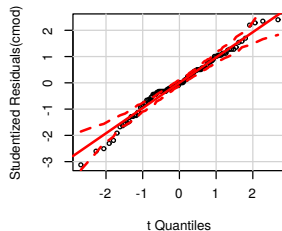
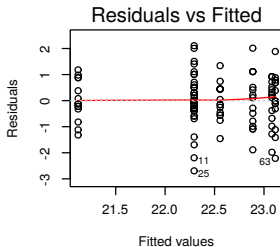
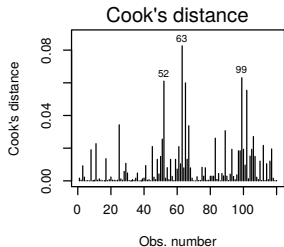
- ▶ Лучше работает, если размеры групп примерно одинаковы (т.наз. сбалансированный дисперсионный комплекс)
- ▶ Устойчив к отклонениям от нормального распределения (при равных объемах групп или при больших выборках)

## Задание:

- ▶ Подберите линейную модель зависимости длины яиц кукушек в гнездах от вида птиц-хозяев
- ▶ Проверьте условия применимости

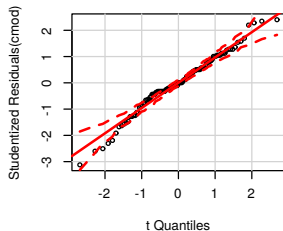
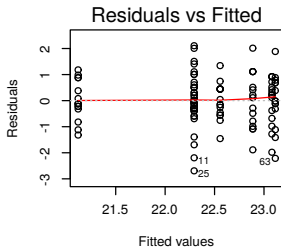
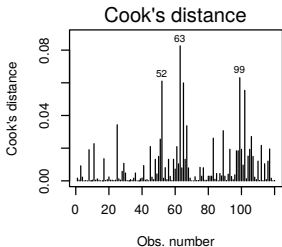
## Решение:

```
cmod <- lm(length ~ species, data = cu)
library(car)
op <- par(mfrow = c(1, 3))
plot(cmod, which = 4)
plot(cmod, which = 1)
qqPlot(cmod)
par(op)
```



## Решение:

```
cmod <- lm(length ~ species, data = cu)
library(car)
op <- par(mfrow = c(1, 3))
plot(cmod, which = 4)
plot(cmod, which = 1)
qqPlot(cmod)
par(op)
```

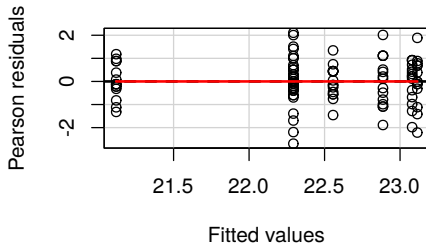
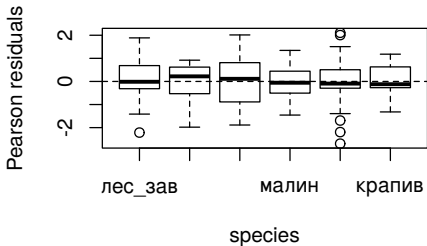


- ▶ Выбросов нет
- ▶ Дисперсии почти одинаковые. Может быть, в одной из групп чуть меньше
- ▶ Распределение остатков отличается от нормального

# Полезно строить графики остатков от переменных в модели

...и не в модели, если такие есть

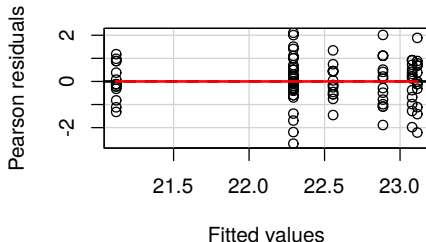
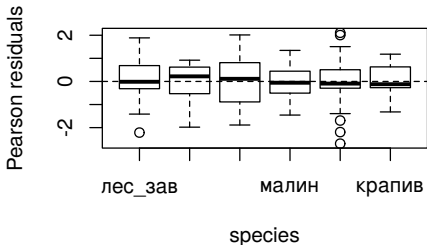
`residualPlots(cmod)`



# Полезно строить графики остатков от переменных в модели

...и не в модели, если такие есть

`residualPlots(cmod)`



- Разброс остатков практически одинаков для разных видов птиц-хозяев

# Коэффициенты линейной модели

Влияет ли вид птиц-хозяев на длину яиц кукушек?

```
coef(summary(cmod))
```

#	Estimate	Std. Error	t value	Pr(> t )
# (Intercept)	23.1143	0.242	95.543	1.17e-110
# speciesлес_кон	-0.0343	0.336	-0.102	9.19e-01
# speciesбел_тряс	-0.2276	0.336	-0.677	5.00e-01
# speciesмалин	-0.5580	0.331	-1.685	9.48e-02
# speciesлуг_кон	-0.8210	0.277	-2.964	3.70e-03
# speciesкрапив	-1.9943	0.336	-5.929	3.33e-08



# Коэффициенты линейной модели

Влияет ли вид птиц-хозяев на длину яиц кукушек?

```
coef(summary(cmod))
```

#	Estimate	Std. Error	t value	Pr(> t )
# (Intercept)	23.1143	0.242	95.543	1.17e-110
# speciesлес_кон	-0.0343	0.336	-0.102	9.19e-01
# speciesбел_тряс	-0.2276	0.336	-0.677	5.00e-01
# speciesмалин	-0.5580	0.331	-1.685	9.48e-02
# speciesлуг_кон	-0.8210	0.277	-2.964	3.70e-03
# speciesкрапив	-1.9943	0.336	-5.929	3.33e-08

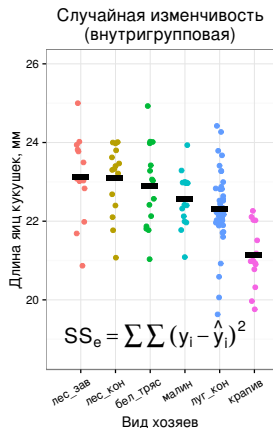
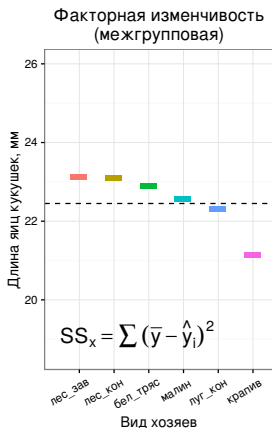
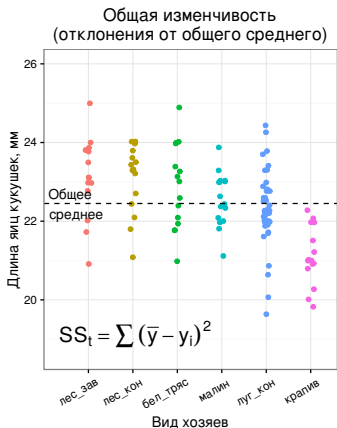
По коэффициентам это сложно сказать. Было бы удобнее, если бы ответ был из одного числа, вместо пачки коэффициентов. Дисперсионный анализ позволит оценить влияние каждого из факторов.





# Структура общей изменчивости

Общая изменчивость ( $SS_t$ ) = Факторная ( $SS_x$ ) + Случайная ( $SS_e$ )



Если выборки из одной совокупности, то Факторная изменчивость = Случайная изменчивость

## Таблица дисперсионного анализа

Источник изменчивости	SS	df	MS	F
Название фактора	$SS_x = \sum (\bar{y} - \hat{y}_i)^2$	$df_x = a - 1$	$MS_x = \frac{SS_x}{df_x}$	$F_{df_x, df_e} = \frac{MS_x}{MS_e}$
Случайная	$SS_e = \sum (y_i - \hat{y}_i)^2$	$df_e = N - a$	$MS_e = \frac{SS_e}{df_e}$	
Общая	$SS_t = \sum (\bar{y} - y_i)^2$	$df_t = N - 1$		

Гипотезы:  $H_0 : MS_x = MS_e$ ,  $H_A : MS_x \neq MS_e$

Минимальное упоминание результатов в тексте должно содержать  $F_{df_x, df_e}$  и  $p$ .

## Делаем дисперсионный анализ в R

В R есть много функций для дисперсионного анализа. Мы рекомендуем `Anova()` из пакета `car`. Зачем? Когда факторов будет больше одного, эта функция сможет правильно оценить достоверность каждого из них независимо от других.

```
cu_anova <- Anova(cmod)
cu_anova
```

```
# Anova Table (Type II tests)
#
# Response: length
#           Sum Sq  Df F value    Pr(>F)
# species      42.8   5    10.4 0.000000029 ***
# Residuals     93.4 114
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Результаты дисперсионного анализа

Результаты дисперсионного анализа можно описать в тексте:

- ▶ Длина яиц кукушек в гнездах разных птиц-хозяев достоверно различается ( $F_{5,114} = 10.45, p < 0.01$ ).



# Результаты дисперсионного анализа

Результаты дисперсионного анализа можно представить в виде таблицы

- ▶ Длина яиц кукушек достоверно различалась в гнездах разных птиц-хозяев (Табл. 1).

Table 1: Результаты дисперсионного анализа длины яиц кукушек в гнездах разных птиц-хозяев. SS — суммы квадратов отклонений, df — число степеней свободы, F — значение F-критерия, P — доверительная вероятность.

	SS	df	F	P
Влажность	42.8	5	10.4	<0.01
Остаточная	93.4	114		

## Как понять, какие именно группы различаются

Дисперсионный анализ говорит нам только, есть ли влияние фактора, но не говорит, какие именно группы различаются.

Коэффициенты линейной модели в `summary(cmod)` содержат лишь часть ответа — сравнение средних значений всех групп со средним на базовом уровне.

Если нас интересуют другие возможные попарные сравнения, нужно сделать пост хок тест.

## Пост хок тесты

# Post hoc тесты

Пост хок тесты — попарные сравнения средних **после того, как дисперсионный анализ показал, что влияние фактора достоверно**

Свойства post hoc тестов:

- ▶ **Применяются, только если влияние фактора значимо**
- ▶ Делают поправку для снижения вероятности ошибки I рода  $\alpha$ , (но не слишком большую, чтобы не снизилась мощность, и чтобы не возросла вероятность ошибки II рода  $\beta$ )
- ▶ Учитывают величину различий между средними
- ▶ Учитывают количество сравниваемых пар
- ▶ Различаются по степени консервативности (тест Тьюки — разумный компромисс)
- ▶ Работают лучше при равных объемах групп, при гомогенности дисперсий





## Пост хок тест Тьюки в R

- ▶ `glht()` — “general linear hypotheses testing”
- ▶ `linfct` — аргумент, задающий гипотезу для тестирования
- ▶ `mcp()` — функция, чтобы задавать множественные сравнения (обычные пост хоки)
- ▶ `species = “Tukey”` — тест Тьюки по фактору `species`

```
library(multcomp)  
cu_ph <- glht(cmod, linfct = mcp(species = "Tukey"))
```



# Результаты попарных сравнений (тест Тьюки)

Таблица результатов пост хок теста практически нечитабельна.

`summary(cu_ph)`

```
#
# Simultaneous Tests for General Linear Hypotheses
#
# Multiple Comparisons of Means: Tukey Contrasts
#
# Fit: lm(formula = length ~ species, data = cu)
#
# Linear Hypotheses:
```

	Estimate	Std. Error	t value	Pr(> t )
# лес_кон - лес_зав == 0	-0.0343	0.3364	-0.10	1.000
# бел_тряс - лес_зав == 0	-0.2276	0.3364	-0.68	0.984
# малин - лес_зав == 0	-0.5580	0.3313	-1.68	0.538
# луг_кон - лес_зав == 0	-0.8210	0.2770	-2.96	0.041 *
# крапив - лес_зав == 0	-1.9943	0.3364	-5.93	<0.001 ***
# бел_тряс - лес_кон == 0	-0.1933	0.3305	-0.58	0.992
# малин - лес_кон == 0	-0.5237	0.3253	-1.61	0.587
# луг_кон - лес_кон == 0	-0.7867	0.2699	-2.91	0.047 *
# крапив - лес_кон == 0	-1.9600	0.3305	-5.93	<0.001 ***
# малин - бел_тряс == 0	-0.3304	0.3253	-1.02	0.909
# луг_кон - бел_тряс == 0	-0.5933	0.2699	-2.20	0.241
# крапив - бел_тряс == 0	-1.7667	0.3305	-5.34	<0.001 ***
# луг_кон - малин == 0	-0.2629	0.2635	-1.00	0.915



# Результаты пост хок теста

Результаты пост хок теста можно привести в виде текста...

- ▶ Размер яиц кукушек в гнездах крапивника достоверно меньше, чем в гнездах лугового конька (тест Тьюки,  $p < 0.01$ ). Размер яиц кукушек в гнездах лесной завирушки, белой трясогузки, малиновки и лесного конька не различается, но яйца в гнездах этих видов крупнее, чем у лугового конька или крапивника (тест Тьюки, от  $p < 0.01$  до 0.05).

...или построить график



## Данные для графика при помощи predict()

```
MyData <- data.frame(
  species = factor(levels(cu$species),
                    levels = levels(cu$species)))
MyData <- data.frame(MyData, predict(cmod, newdata = MyData,
                                   interval = "confidence"))
MyData
```

```
#   species  fit  lwr  upr
# 1 лес_зав 23.1 22.6 23.6
# 2 лес_кон 23.1 22.6 23.5
# 3 бел_тряс 22.9 22.4 23.3
# 4   малин 22.6 22.1 23.0
# 5 луг_кон 22.3 22.0 22.6
# 6  крапив 21.1 20.7 21.6
```

## Задание:

Создайте MyData вручную:

- ▶ предсказанные значения
- ▶ стандартные ошибки
- ▶ верхнюю и нижнюю границы доверительных интервалов

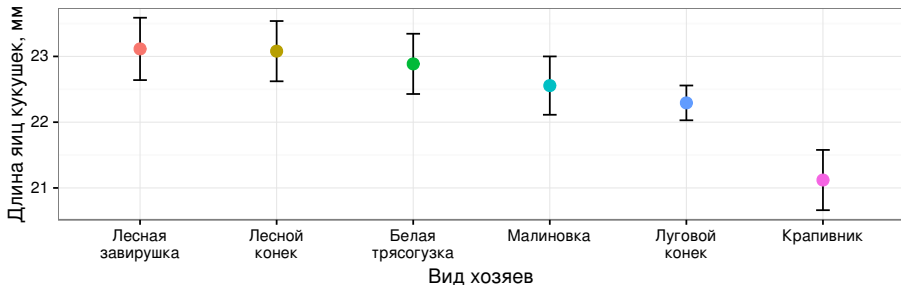
## Решение:

```
MyData <- data.frame(  
  species = factor(levels(cu$species),  
                    levels = levels(cu$species)))  
X <- model.matrix(~species, data = MyData)  
betas <- coef(cmod)  
MyData$fit <- X %*% betas  
MyData$se <- sqrt(diag(X %*% vcov(cmod) %*% t(X)))  
MyData$lwr <- MyData$fit - qnorm(0.975) * MyData$se  
MyData$upr <- MyData$fit + qnorm(0.975) * MyData$se  
MyData
```

```
#   species  fit    se  lwr  upr  
# 1 лес_зав 23.1 0.242 22.6 23.6  
# 2 лес_кон 23.1 0.234 22.6 23.5  
# 3 бел_тряс 22.9 0.234 22.4 23.3  
# 4   малин 22.6 0.226 22.1 23.0  
# 5 луг_кон 22.3 0.135 22.0 22.6  
# 6  крапив 21.1 0.234 20.7 21.6
```

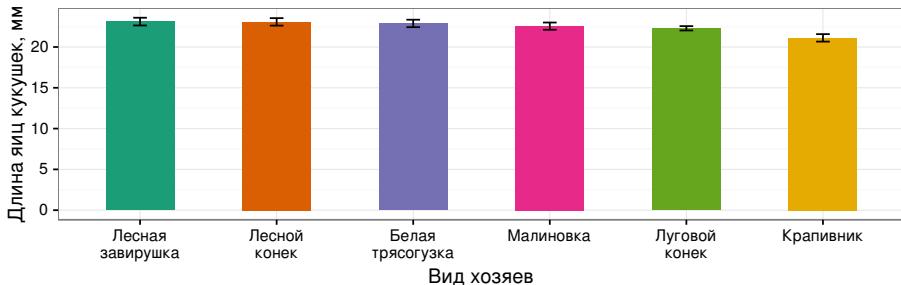
# Точечный график

```
gg_points <- ggplot(data = MyData, aes(x = species, y = fit)) +  
  geom_errorbar(aes(ymin = lwr, ymax = upr), width = 0.1) +  
  geom_point(aes(colour = species), size = 3) +  
  labs(x = "Вид хозяев", y = "Длина яиц кукушек, мм") +  
  scale_fill_brewer(name = "Вид \nхозяев", palette = "Dark2") +  
  scale_x_discrete(labels = c("Лесная\nзавирушка", "Лесной\nконек",  
                              "Белая\nтрясогузка", "Малиновка", "Луговой\nконек",  
                              "Крапивник")) +  
  theme(legend.position = "none")  
gg_points
```



# Столбчатый график

```
gg_bars <- ggplot(data = MyData, aes(x = species, y = fit)) +  
  geom_bar(stat = "identity", aes(fill = species), width = 0.5) +  
  geom_errorbar(aes(ymin = lwr, ymax = upr), width = 0.1) +  
  labs(x = "Вид хозяев", y = "Длина яиц кукушек, мм") +  
  scale_fill_brewer(name = "Вид \nхозяев", palette = "Dark2") +  
  scale_x_discrete(labels = c("Лесная\nзавирушка", "Лесной\nконек",  
                             "Белая\nтрясогузка", "Малиновка", "Луговой\nконек",  
                             "Крапивник")) +  
  theme(legend.position = "none")  
gg_bars
```

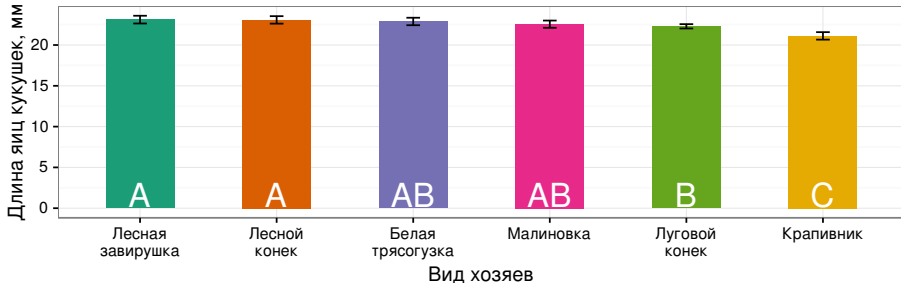




# Можно привести результаты пост хок теста на столбчатом графике

Достоверно различающиеся группы обозначим разными буквами

```
gg_bars_coded <- gg_bars +  
  geom_text(aes(y = 1.6, label = c("A", "A", "AB", "AB", "B", "C")),  
            colour = "white", size = 7)  
gg_bars_coded
```



# Take home messages

- ▶ Дисперсионный анализ — линейная модель с дискретными предикторами, существует в нескольких параметризациях, которые отличаются трактовками коэффициентов
- ▶ При помощи дисперсионного анализа можно проверить гипотезу о равенстве средних значений в группах
- ▶ Условия применимости дисперсионного анализа
  - ▶ Случайность и независимость групп и наблюдений внутри групп
  - ▶ Нормальное распределение в группах
  - ▶ Гомогенность дисперсий в группах
- ▶ При множественных попарных сравнениях увеличивается вероятность ошибки первого рода, поэтому нужно вносить поправку для уровня значимости
- ▶ Post hoc тесты — это попарные сравнения после дисперсионного анализа, которые позволяют сказать, какие именно средние различаются

- ▶ Quinn, Keough, 2002, pp. 173–207
- ▶ Logan, 2010, pp. 254–282
- ▶ [Open Intro to Statistics](#), pp.236–246
- ▶ Sokal, Rohlf, 1995, pp. 179–260
- ▶ Zar, 2010, pp. 189–207