

ADA2: Class 09, Ch 05b Paired Experiments and Randomized Block Experiments: Two-way Factor design

Tim Farkas

February 15, 2022

Kangaroos skull measurements: crest width

What effect does sex and species have on the crest width of a kangaroo skull?

The data to be analyzed here are selected skull measurements on 148 kangaroos of known sex and species. There are 11 columns of data, corresponding to the following features. The measurements are in meters/10000 (mm/10).

column	Variable name	Description
1 *	sex	sex (1=M, 2=F)
2 *	species	species (0=M. giganteus, 1=M.f. melanops, 2=M.f. fuliginosus)
3	pow	post orbit width
4	rw	rostral width
5	sopd	supra-occipital - paroccipital depth
6 *	cw	crest width
7	ifl	incisive foramina length
8	ml	mandible length
9	mw	mandible width
10	md	mandible depth
11	arh	ascending ramus height

Some of the observations in the data set are missing (not available). These are represented by a period ., which in the `read_csv()` function is specified by the `na = "."` option.

```
setwd("~/Dropbox/3_Education/Courses/stat_528_ada2/")
library(erikmisc)
library(tidyverse)
```

```
# First, download the data to your computer,
#   save in the same folder as this Rmd file.
```

```
dat_kang <-
  read_csv(
    "ADA2_CL_09_kang.csv"
  , na = c("", ".")
  ) %>%
  # subset only our columns of interest
  select(
    sex, species, cw
```

```

) %>%
# make dose a factor variable and label the levels
mutate(
  sex      = factor(sex      , labels = c("M","F"))
, species = factor(species, labels = c("Mg", "Mfm", "Mff"))
)

# remove observations with missing values
n_start  <- nrow(dat_kang)
dat_kang <- na.omit(dat_kang)
n_keep   <- nrow(dat_kang)
n_drop   <- n_start - n_keep
cat("Removed", n_start, "-", n_keep, "=", n_drop, "observations with missing values.")

```

Removed 148 - 148 = 0 observations with missing values.

```

# The first few observations
head(dat_kang)

```

sex	species	cw
M	Mg	153
M	Mg	141
M	Mg	144
M	Mg	116
M	Mg	120
M	Mg	188

(1 p) Interpret plots of the data, distributional centers and shapes

The side-by-side boxplots of the data compare the crest widths across the 6 combinations of sex and species. Comment on the distributional shapes and compare the typical crest widths across groups.

```

# Calculate the cell means for each (sex, species) combination
# Group means
kang_mean      <- dat_kang %>% summarise(m = mean(cw))
kang_mean_x    <- dat_kang %>% group_by(sex)      %>% summarise(m = mean(cw)) %>% ungroup()
kang_mean_s    <- dat_kang %>% group_by(species)   %>% summarise(m = mean(cw)) %>% ungroup()
kang_mean_xs   <- dat_kang %>% group_by(sex, species) %>% summarise(m = mean(cw)) %>% ungroup()

```

```
kang_mean
```

m
123.4865

```
kang_mean_x
```

sex	m
M	111.0959
F	135.5467

```
kang_mean_s
```

species	m
Mg	110.120
Mfm	115.625
Mff	144.400

```
kang_mean_xs
```

sex	species	m
M	Mg	103.0800
M	Mfm	101.6522
M	Mff	127.8000
F	Mg	117.1600
F	Mfm	128.4800
F	Mff	161.0000

```
# Interaction plots, ggplot
```

```
library(ggplot2)
```

```
p1 <- ggplot(dat_kang, aes(x = sex, y = cw, colour = species))
```

```
p1 <- p1 + geom_hline(aes(yintercept = 0), colour = "black"  
                        , linetype = "solid", size = 0.2, alpha = 0.3)
```

```
p1 <- p1 + geom_boxplot(alpha = 0.5, outlier.size=0.1)
```

```
p1 <- p1 + geom_point(data = kang_mean_xs, aes(y = m), size = 4)
```

```
p1 <- p1 + geom_line(data = kang_mean_xs, aes(y = m, group = species), size = 1.5)
```

```
p1 <- p1 + labs(title = "Kangaroo interaction plot, species by sex")
```

```
#print(p1)
```

```
p2 <- ggplot(dat_kang, aes(x = species, y = cw, colour = sex))
```

```
p2 <- p2 + geom_hline(aes(yintercept = 0), colour = "black"  
                        , linetype = "solid", size = 0.2, alpha = 0.3)
```

```
p2 <- p2 + geom_boxplot(alpha = 0.5, outlier.size=0.1)
```

```
p2 <- p2 + geom_point(data = kang_mean_xs, aes(y = m), size = 4)
```

```
p2 <- p2 + geom_line(data = kang_mean_xs, aes(y = m, group = sex), size = 1.5)
```

```
p2 <- p2 + labs(title = "Kangaroo interaction plot, sex by species")
```

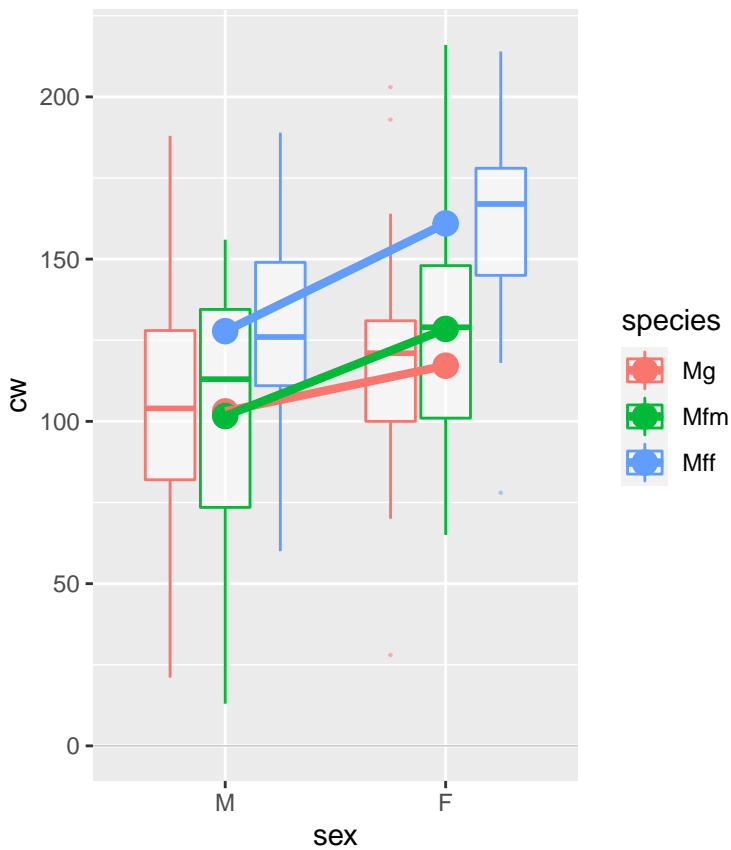
```
#print(p2)
```

```
library(gridExtra)
```

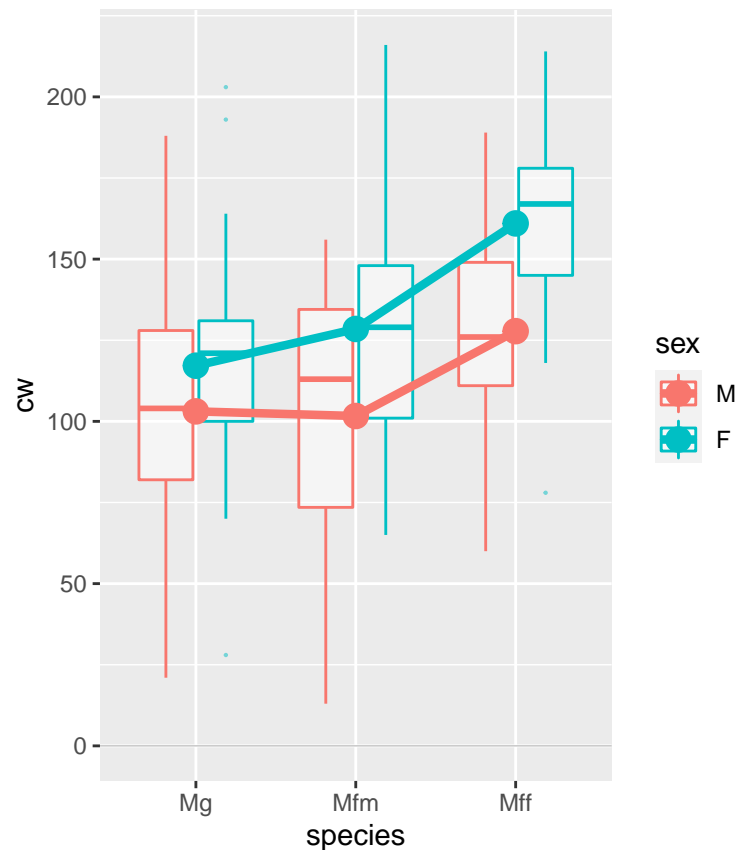
```
grid.arrange(grobs = list(p1, p2), nrow=1, top="Kangaroo crestwidth plots")
```

Kangaroo crestwidth plots

Kangaroo interaction plot, species by sex



Kangaroo interaction plot, sex by species



Solution

These graphs demonstrate that, if differences exist, it's that females have larger crests than males, and the species Mff has larger crests than the other two species. Some other differences are apparent, but they are smaller.

(1 p) Do the plots above suggest there is an interaction?

Do the lines for each group seem to be very different from parallel?

Solution

It looks as though female kangaroos have larger crests than males, but this effect is smaller for the Mg species than the other two species.

Fit the two-way interaction model

Here it is.

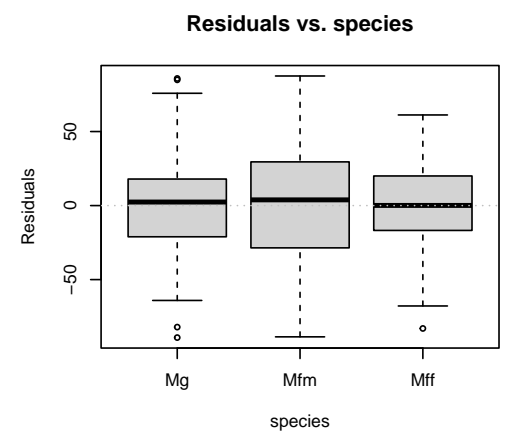
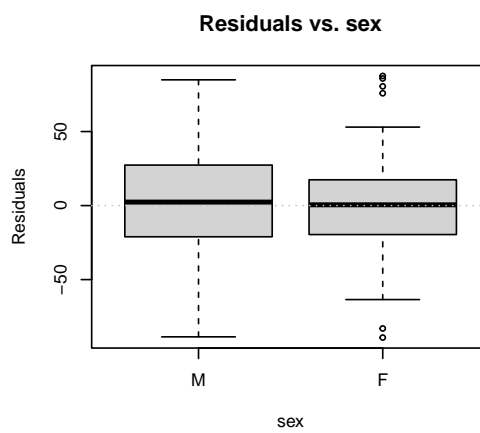
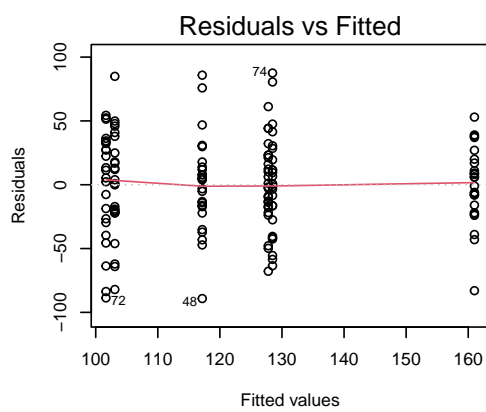
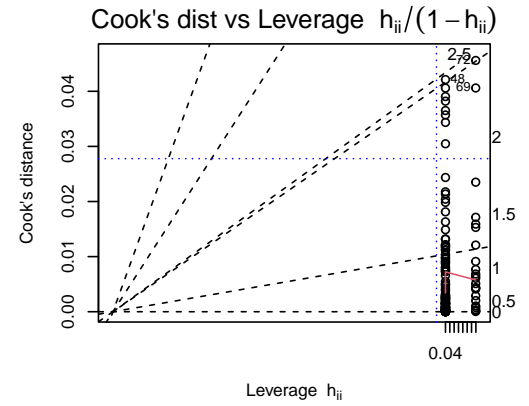
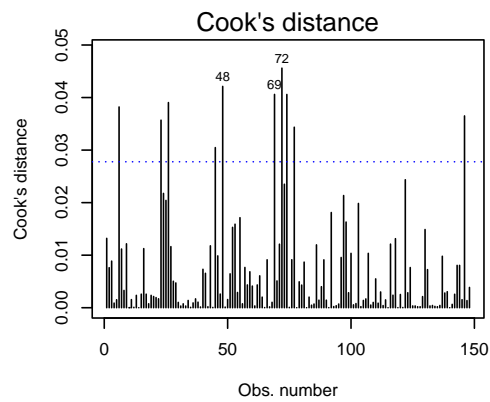
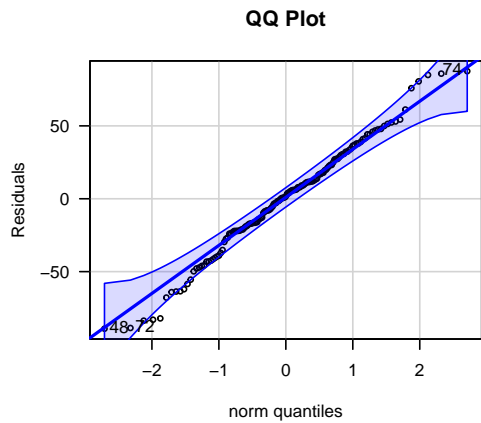
```
lm_cw_x_s_xs <-
  lm(
    cw ~ sex * species
    , data = dat_kang
    , contrasts = list(sex = contr.sum, species = contr.sum)
  )
```

(1 p) Check model assumptions for full model

Recall that we assume that the full model is correct before we perform model reduction by backward selection.

```
# plot diagnostics
```

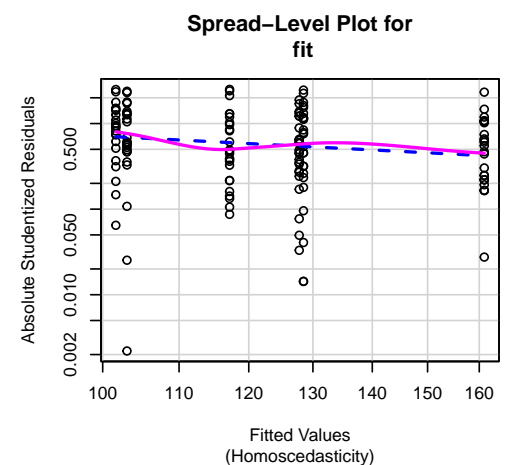
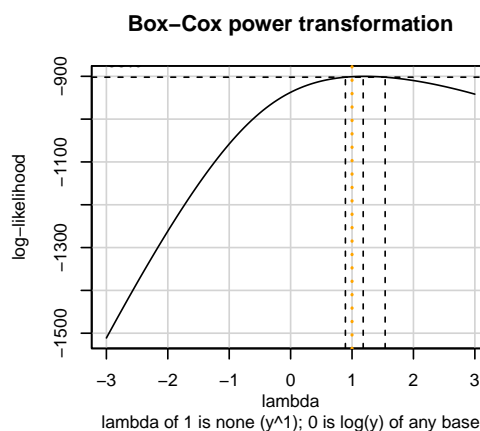
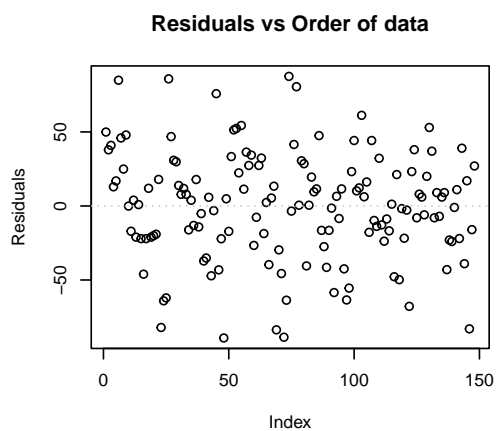
```
e_plot_lm_diagnostics(lm_cw_x_s_xs)
```



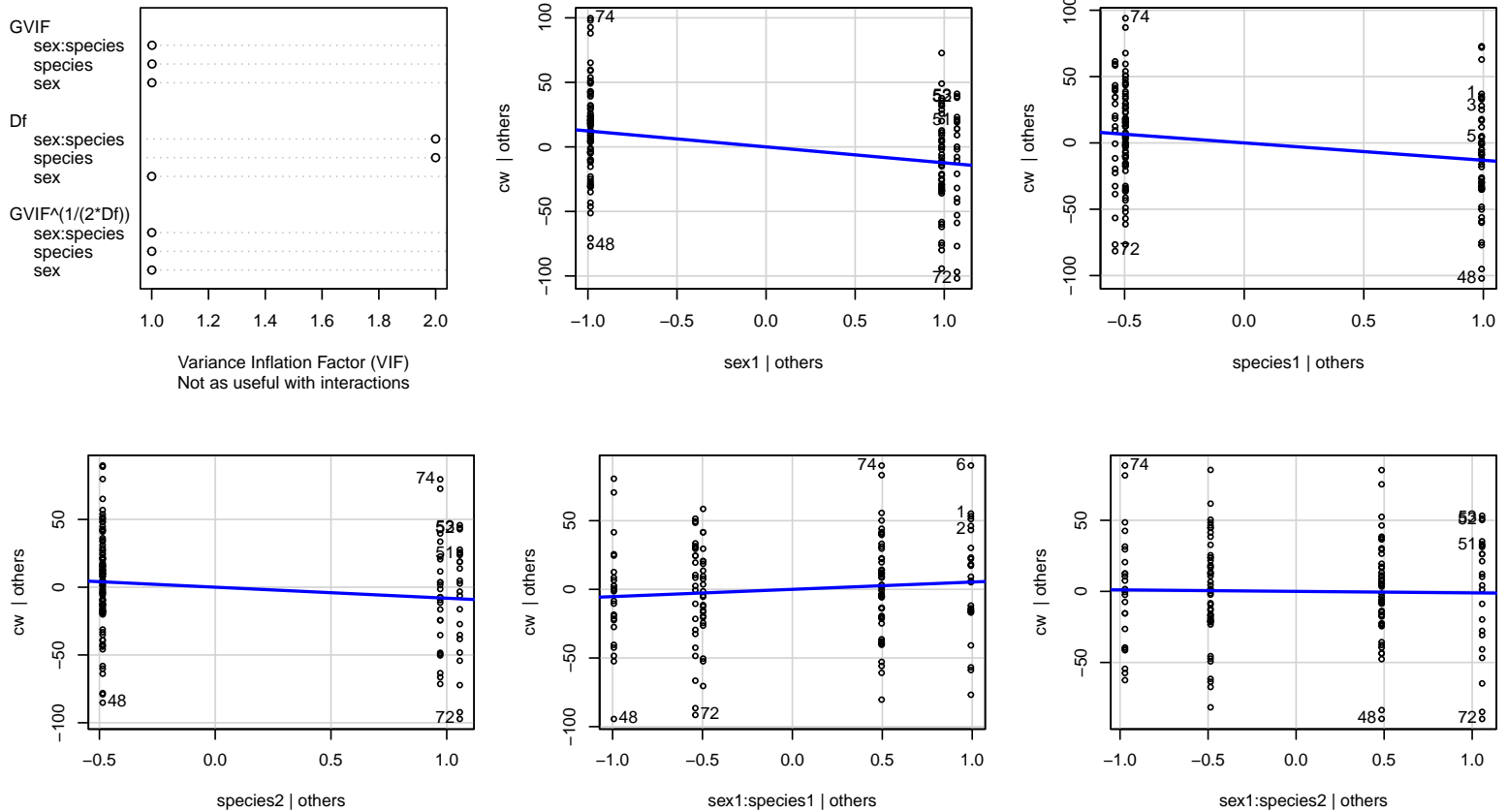
Non-constant Variance Score Test

Variance formula: \sim fitted.values

Chisquare = 3.094663, Df = 1, p = 0.078549



Collinearity



Solution

Yeah this model looks pretty good. The residuals are normally distributed, variance appears homogenous across both sexes and species, there is no relationship between residual variance and fitted values, and Box-Cox indicates no transformation is necessary.

(1 p) ANOVA table, test for interaction

Provide your conclusion for the test for interaction.

```
library(car)
Anova(lm_cw_x_s_xs, type=3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2244042.24	1	1643.7948431	0.0000000
sex	22556.30	1	16.5228341	0.0000793
species	34194.73	2	12.5240772	0.0000098
sex:species	2367.14	2	0.8669828	0.4224297
Residuals	193852.66	142	NA	NA

Solution

Based on the figures, it's not surprising that we see no significant interaction between species and sex, with $p = 0.4224$.

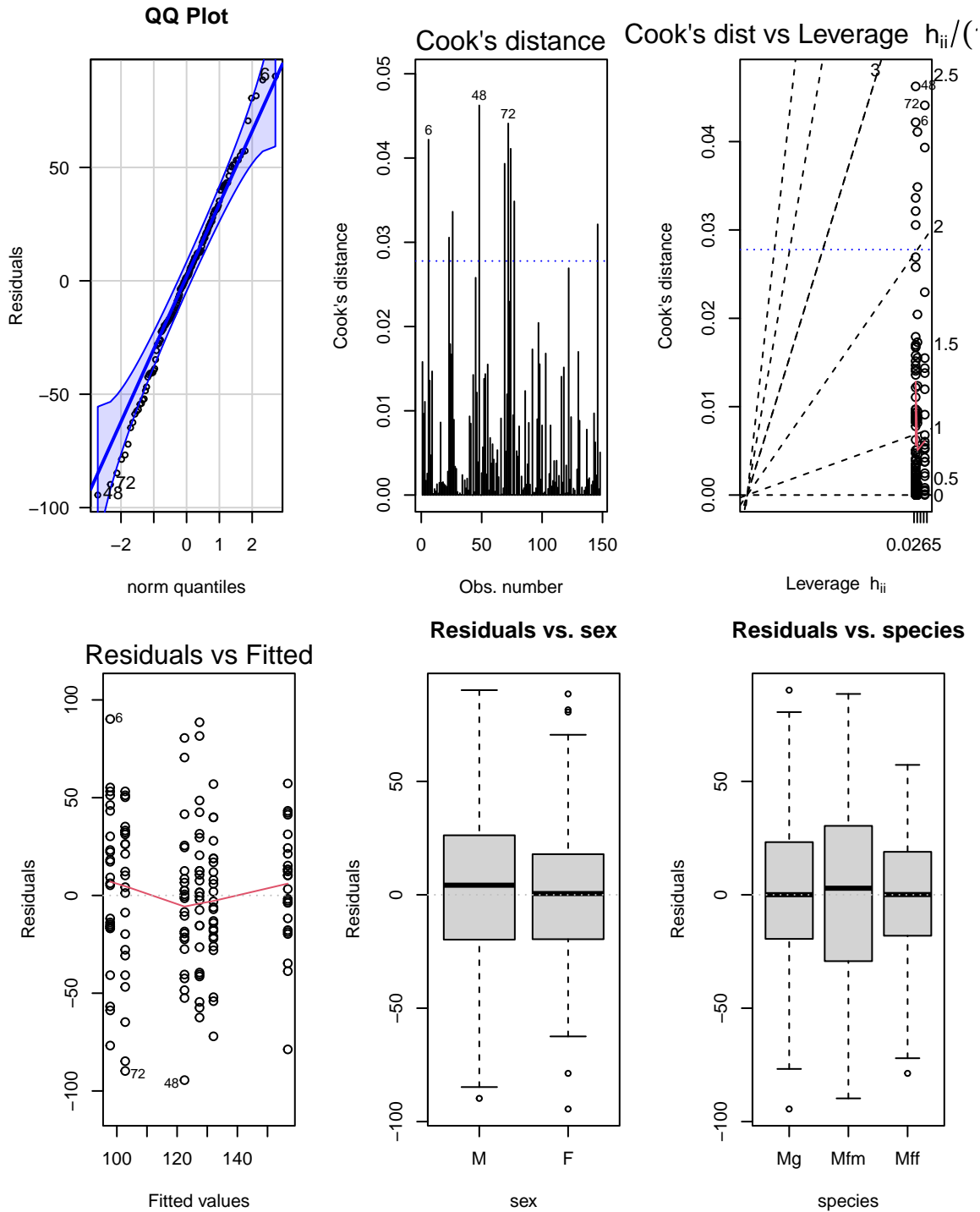
(4 p) Reduce to final model, test assumptions

If the model can be simplified (because interaction is not significant), then refit the model with only the main effects. Test whether the main effects are significant, reduce further if sensible. Test model assumptions of your final model.

Solution

```
lm_cw_x_s_xs <-
  lm(
    cw ~ sex + species
    , data = dat_kang
    , contrasts = list(sex = contr.sum, species = contr.sum)
  )
```

```
e_plot_lm_diagnostics(lm_cw_x_s_xs)
```

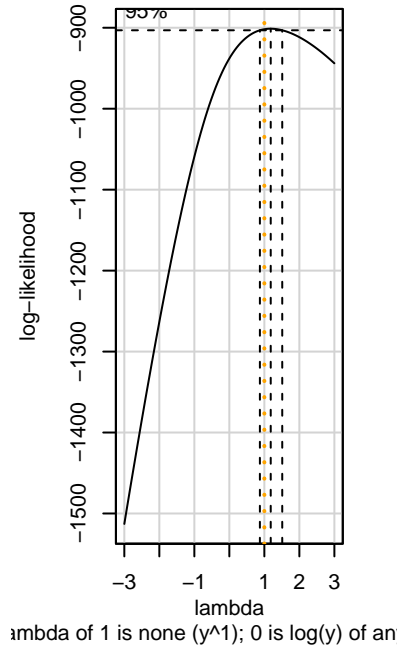
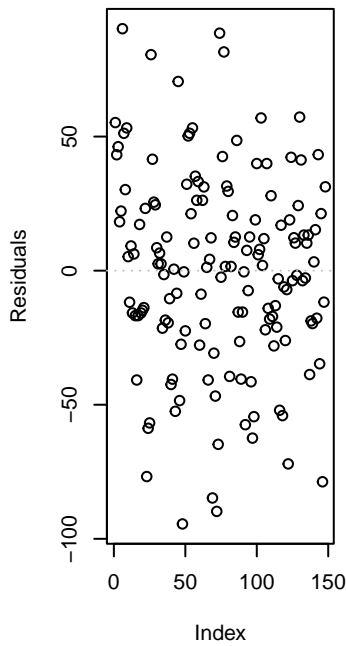


Non-constant Variance Score Test

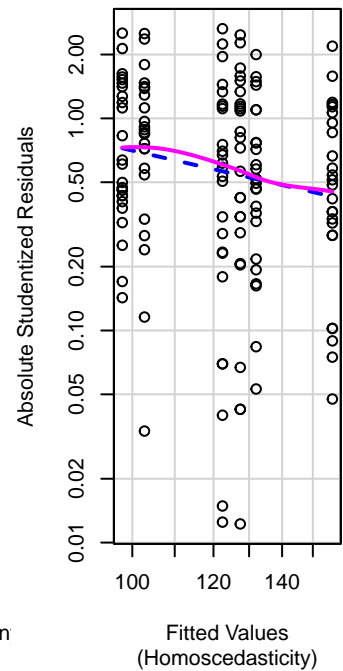
Variance formula: \sim fitted.values

Chisquare = 3.255567, Df = 1, p = 0.071181

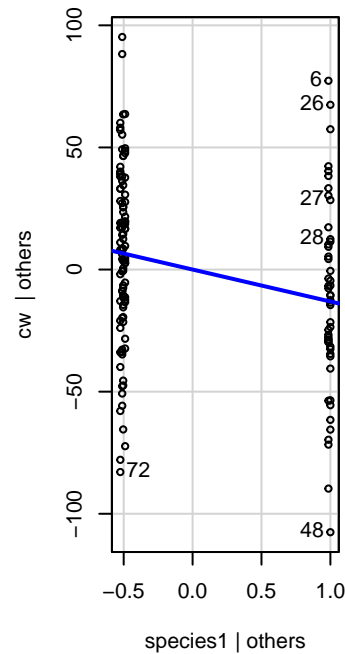
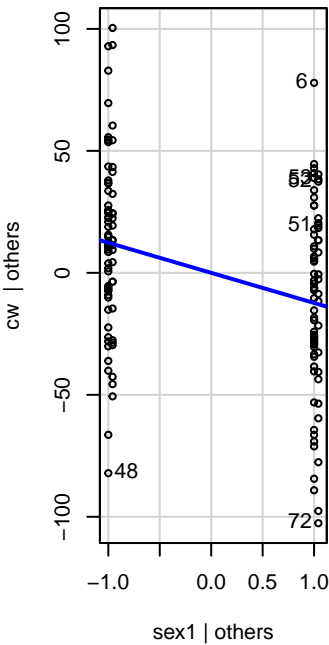
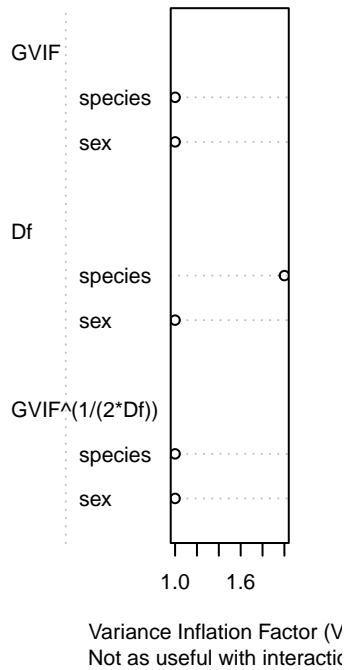
Residuals vs Order of data Box-Cox power transformati

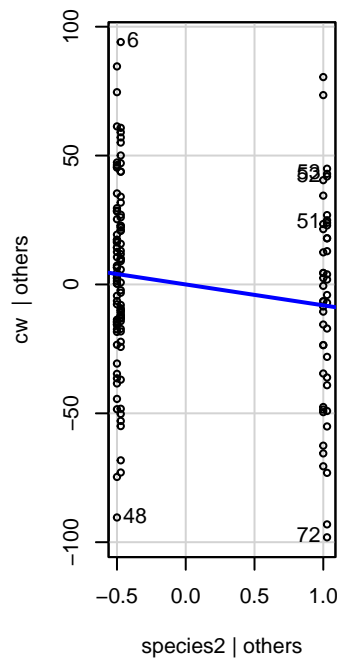


Spread-Level Plot for fit



Collinearity





```
Anova(lm_cw_x_s_xs, type=3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2245491.13	1	1647.90060	0.00e+00
sex	22510.73	1	16.51997	7.89e-05
species	34163.12	2	12.53566	9.60e-06
Residuals	196219.80	144	NA	NA

The model diagnostics are equally good for a model excluding the interaction between species and sex, and the ANOVA shows significant differences between sexes and among species.

(2 p) Summarize the differences

Summarize differences, if any, in sexes and species using relevant multiple comparisons. Give clear interpretations of any significant effects.

This code is here to get you started. Determine which comparisons you plan to make and modify the appropriate code. Make the code chunk active by moving the {R} to the end of the initial code chunk line.

```
library(emmeans)
lm_object <- lm_cw_x_s_xs
# Contrasts to perform pairwise comparisons
cont_kang1 <- emmeans(lm_object, specs = "sex")
cont_kang2 <- emmeans(lm_object, specs = "species")
#cont_kang <- emmeans(lm_object, specs = "sex", by = c("species"))
#cont_kang <- emmeans(lm_object, specs = "species", by = c("sex"))
# Means and CIs
#cont_kang
# Pairwise comparisons
cont_kang1 %>% pairs()
```

```
contrast estimate SE df t.ratio p.value
M - F          -24.7 6.07 144  -4.064  0.0001
```

Results are averaged over the levels of: species

```
cont_kang2 %>% pairs()
```

contrast	estimate	SE	df	t.ratio	p.value
Mg - Mfm	-4.99	7.46	144	-0.669	0.7818
Mg - Mff	-34.28	7.38	144	-4.643	<.0001
Mfm - Mff	-29.29	7.46	144	-3.926	0.0004

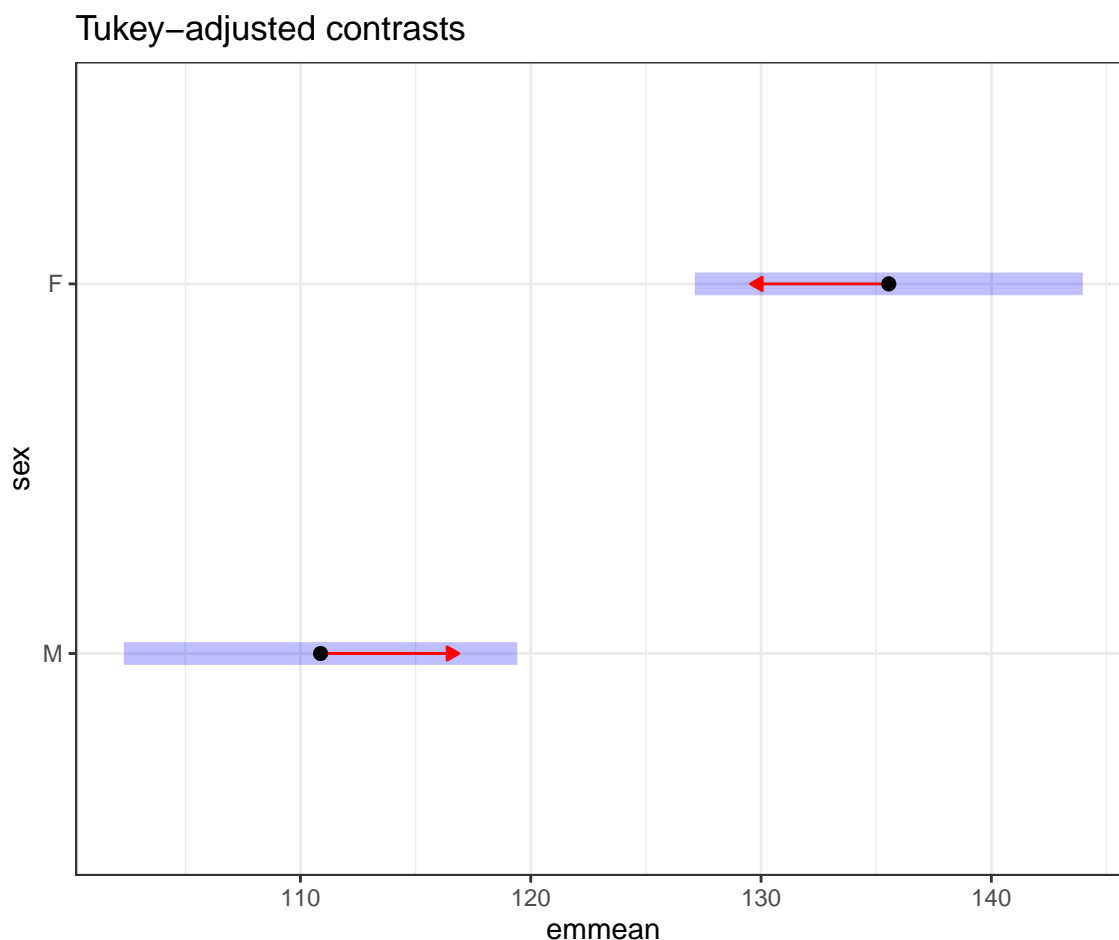
Results are averaged over the levels of: sex

P value adjustment: tukey method for comparing a family of 3 estimates

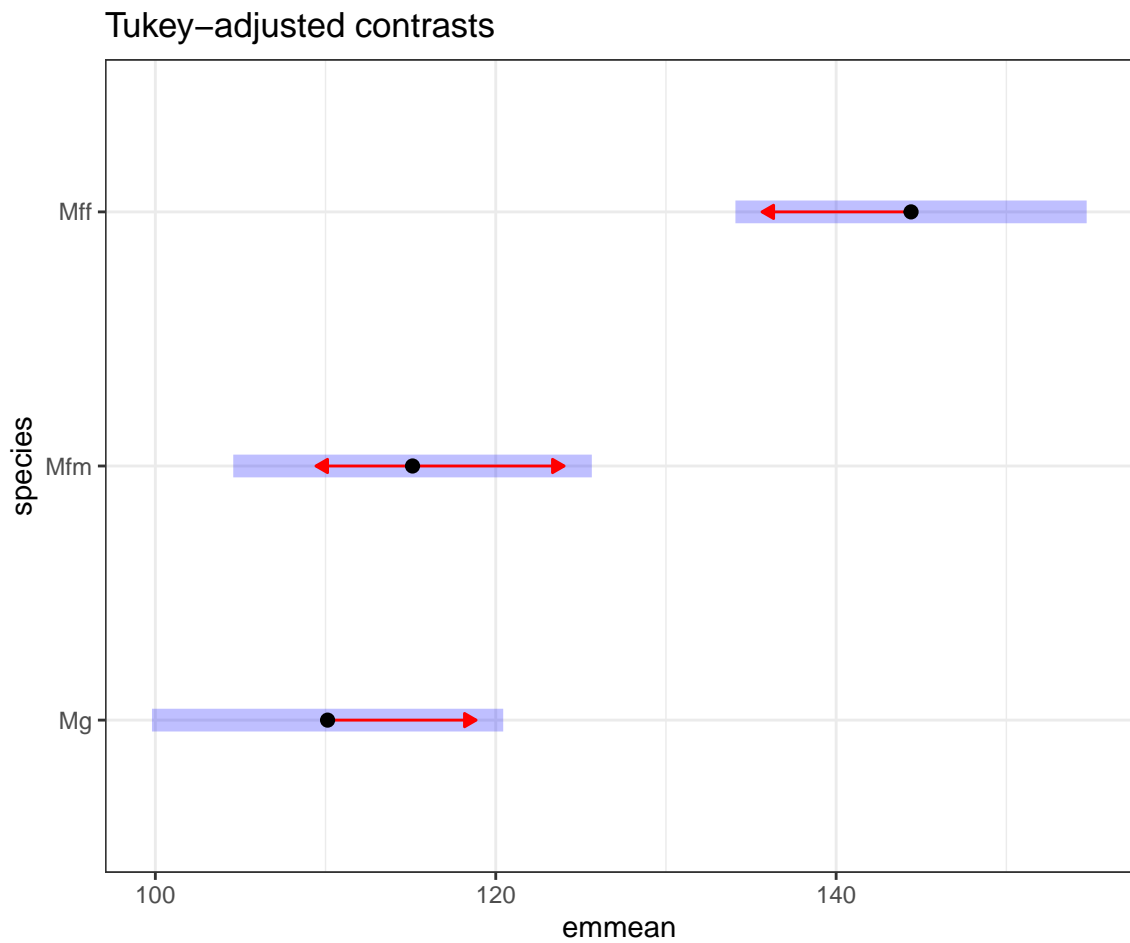
EMM plot interpretation

This **EMM plot (Estimated Marginal Means, aka Least-Squares Means)** is only available when conditioning on one variable. The **blue bars** are confidence intervals for the EMMs; don't ever use confidence intervals for EMMs to perform comparisons – they can be very misleading. The **red arrows** are for the comparisons among means; the degree to which the “comparison arrows” overlap reflects as much as possible the significance of the comparison of the two estimates. If an arrow from one mean overlaps an arrow from another group, the difference is not significant, based on the adjust setting (which defaults to “tukey”).

```
# Plot means and contrasts
p <- plot(cont_kang1, comparisons = TRUE)
p <- p + labs(title = "Tukey-adjusted contrasts")
p <- p + theme_bw()
print(p)
```



```
p <- plot(cont_kang2, comparisons = TRUE)
p <- p + labs(title = "Tukey-adjusted contrasts")
p <- p + theme_bw()
print(p)
```



Solution

Since there is no evidence of an interaction between sex and species, we perform pairwise comparisons for sex and species independently.

The comparison of sexes is really unnecessary, since there are only two groups. Males have smaller crests than females.

The comparison of species shows, as anticipated, that Mff has larger crests than both Mfm and Mg, but Mfm and Mg do not differ in crest size from one another.