

Penguin Assignment

Anonymous

06/12/2022

Organising and cleaning data:

Setting working directory as “PenguinProjects” folder (please adjust to suit your working directory):

```
setwd("~/RWorkingDirectory/PenguinProjects")
```

Loading in libraries, data and functions:

```
source("functions/libraries.r")
source("functions/cleaning.r")
source("functions/plotting.r")
```

Creating a new folder for the raw data and saving the raw data:

```
#Creating a new data folder
dir.create("data")

#Creating a new raw data folder
dir.create("data/data_raw")

#Saving the raw data
write.csv(penguins_raw, "data/data_raw/penguins_raw_06-12-2022.csv")
```

Cleaning the data, creating a new folder for the clean data and saving the clean data:

```
#Using function from "cleaning.r" to clean the data
penguins_clean <- cleaning(penguins_raw)

#Creating a new clean data folder
dir.create("data/data_clean")

#Saving the clean data
write.csv(penguins_clean, "data/data_clean/penguins_clean_06-12-2022.csv")
```

Looking at the data:

```
head(penguins_clean)
```

Preparing data for analysis:

Removing rows without sex or body mass recorded:

```
#Using function from "cleaning.r" to prepare data
penguins_mass <- remove_empty_body_mass(penguins_clean)

#Defining factors and numerical variables
penguins_mass$species <- as.factor(penguins_mass$species)
penguins_mass$sex <- as.factor(penguins_mass$sex)
penguins_mass$body_mass_g <- as.numeric(penguins_mass$body_mass_g)

#Saving mass data
write.csv(penguins_mass, "data/data_clean/penguins_mass_06-12-2022.csv")
```

Looking at the data:

```
head(penguins_mass, 3)
```

```
## # A tibble: 3 x 3
##   species                sex  body_mass_g
##   <fct>                <fct>      <dbl>
## 1 Adelie Penguin (Pygoscelis adeliae) MALE      3750
## 2 Adelie Penguin (Pygoscelis adeliae) FEMALE     3800
## 3 Adelie Penguin (Pygoscelis adeliae) FEMALE     3250
```

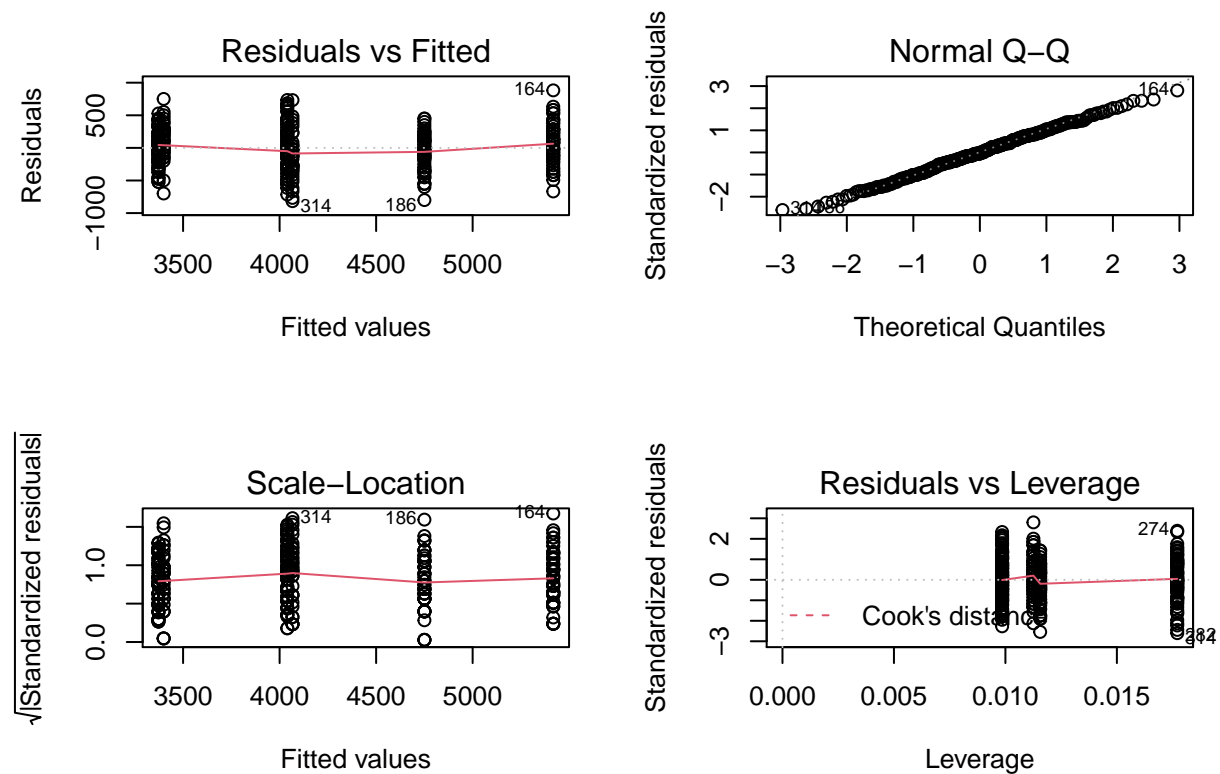
Analysing the data:

Q: To what extent does the species and sex of a penguin influence body mass?

Creating a model with no interaction and testing assumptions of statistical analysis:

```
#Creating the model
lmSpecies_Sex_Mass_No_Interaction <- aov(body_mass_g ~ species + sex, data = penguins_mass)

#Plotting to test assumptions
par(mfrow = c(2,2))
plot(lmSpecies_Sex_Mass_No_Interaction)
```



Overall, the data appears to meet assumptions of ANOVA.

Running ANOVA:

```
summary(lmSpecies_Sex_Mass_No_Interaction)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|-----|-----------|----------|---------|------------|
| species | 2 | 145190219 | 72595110 | 724.2 | <2e-16 *** |
| sex | 1 | 37090262 | 37090262 | 370.0 | <2e-16 *** |
| Residuals | 329 | 32979185 | 100241 | | |

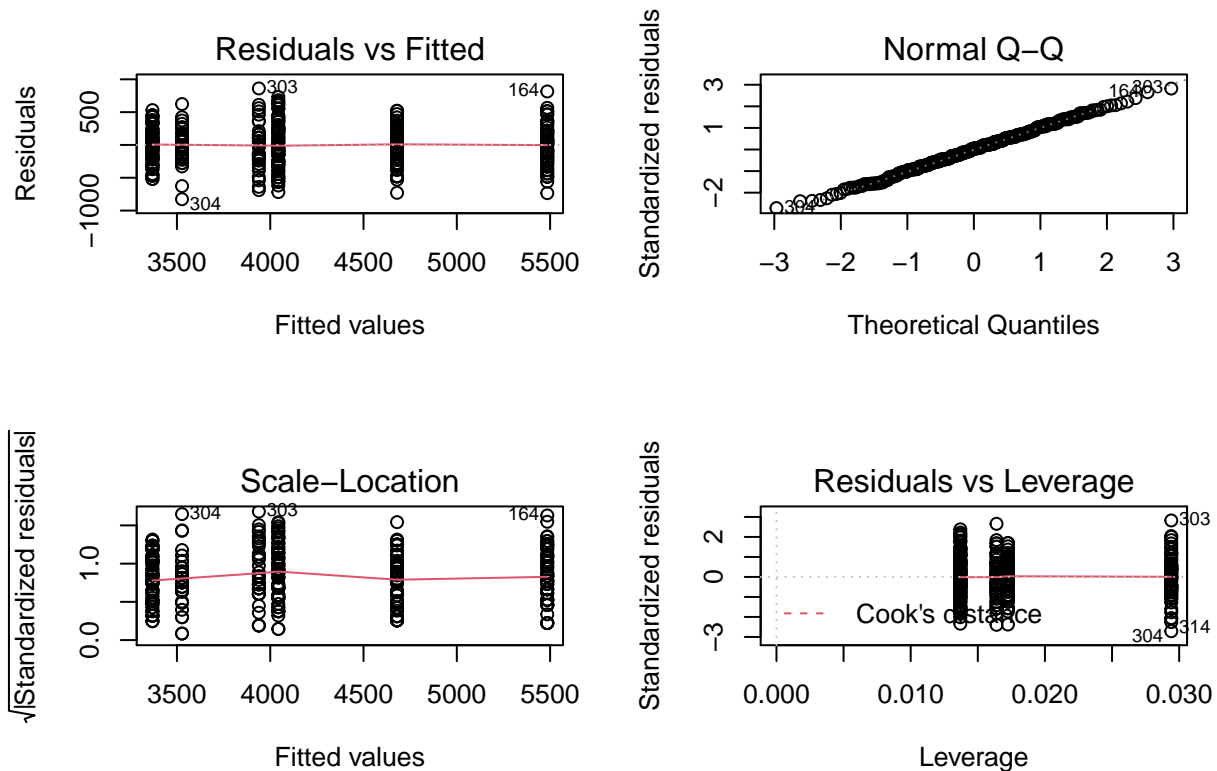
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Both species and sex appear to have a significant effect on body mass.

Building a model that takes into account interaction between species and sex:

```
#Creating the model
lmSpecies_Sex_Mass_Interaction <- aov(body_mass_g ~ species * sex, data = penguins_mass)

#Plotting to test assumptions
par(mfrow = c(2,2))
plot(lmSpecies_Sex_Mass_Interaction)
```



Overall, the data appears to fit the assumptions of ANCOVA.

Running ANCOVA:

```
summary(lmSpecies_Sex_Mass_Interaction)
```

| | ## | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------|----------------|-----|-----------|----------|---------|--------------|
| species | ## species | 2 | 145190219 | 72595110 | 758.358 | < 2e-16 *** |
| sex | ## sex | 1 | 37090262 | 37090262 | 387.460 | < 2e-16 *** |
| species:sex | ## species:sex | 2 | 1676557 | 838278 | 8.757 | 0.000197 *** |
| Residuals | ## Residuals | 327 | 31302628 | 95727 | | |

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

There is a significant effect of interaction between species and sex on body mass.

Comparing the two models:

```
anova(lmSpecies_Sex_Mass_No_Interaction, lmSpecies_Sex_Mass_Interaction)
```

```
## Analysis of Variance Table
##
## Model 1: body_mass_g ~ species + sex
## Model 2: body_mass_g ~ species * sex
##   Res.Df      RSS Df Sum of Sq    F    Pr(>F)
## 1      329 32979185
## 2      327 31302628  2    1676557 8.757 0.0001973 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding an interaction between species and sex significantly improves the model.

Running a Tukey HSD test:

```
#Using TukeyHSD test to compare between groups (results hidden here to keep document tidy)
TukeyHSD(lmSpecies_Sex_Mass_Interaction)
```

The Tukey HSD test suggests significant differences between body mass of penguins grouped by species and sex.

The comparisons between groups which are not significantly different are:

- 1) Chinstrap females and Adelie females
- 2) Chinstrap males and Adelie males

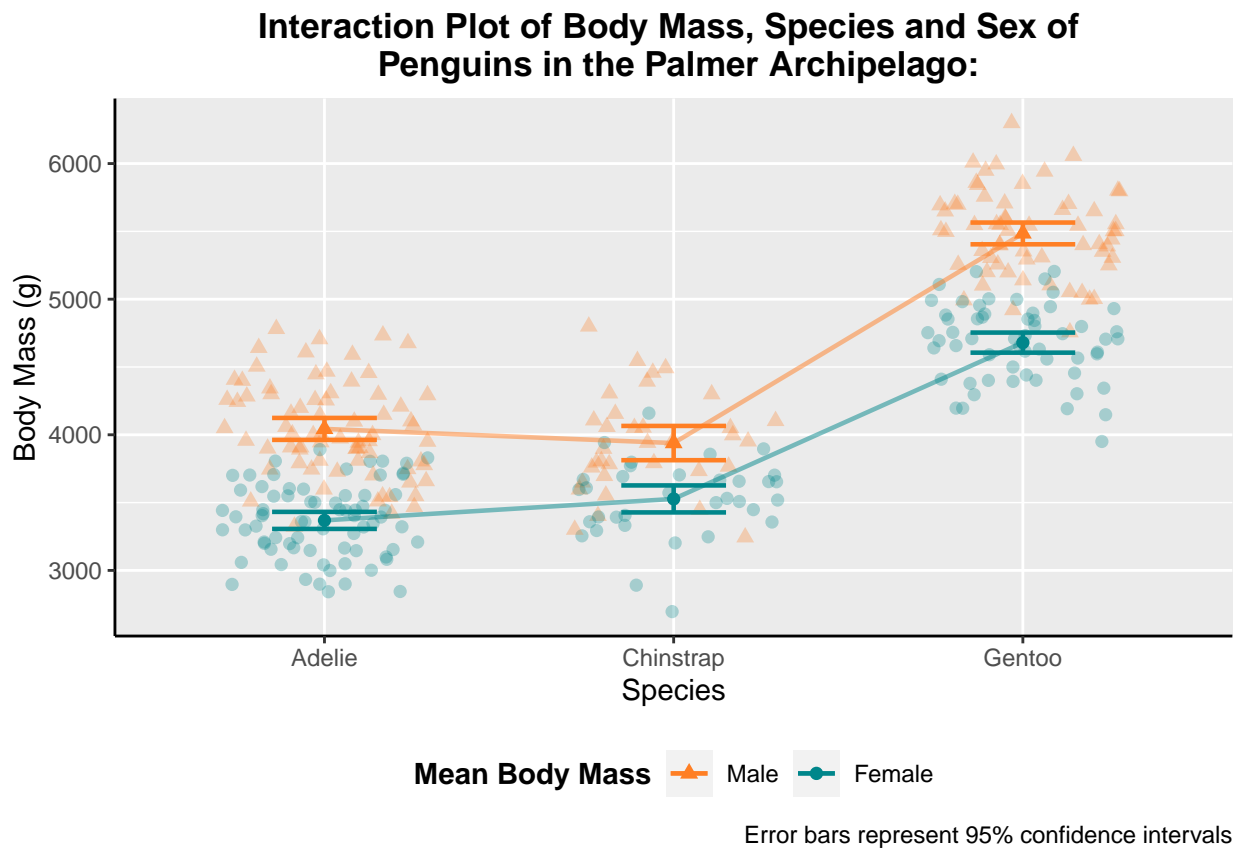
Producing figures to display the data:

Creating a new folder to store figures in:

```
dir.create("figures")
```

Producing a dot plot with lines for each species:

```
#Using function from "plotting.R" to produce graph  
plot_mass_figure(penguins_mass)
```



Saving the figure as a PNG file:

```
#Using function from "plotting.R" to save figure as PNG  
save_mass_plot_png(penguins_mass, "figures/fig_01_06-12-2022.png", 20, 600, 1.4)
```

Saving the figure as an SVG file:

```
#Using function from "plotting.R" to save figure as SVG  
save_mass_plot_svg(penguins_mass, "figures/fig_01_06-12-2022.svg", 20, 1.4)
```

Session info:

```
sessionInfo()
```

```
## R version 4.0.5 (2021-03-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 22000)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
## [2] LC_CTYPE=English_United Kingdom.1252
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] svglite_2.1.0      ragg_1.2.2          dplyr_1.0.7
## [4] janitor_2.1.0      ggplot2_3.3.3       palmerpenguins_0.1.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.8.3      lubridate_1.8.0     lattice_0.20-41
## [4] deldir_1.0-6      png_0.1-7           assertthat_0.2.1
## [7] digest_0.6.27     utf8_1.2.1          R6_2.5.1
## [10] backports_1.4.1   evaluate_0.18       highr_0.9
## [13] pillar_1.6.0      rlang_0.4.11        data.table_1.14.2
## [16] rstudioapi_0.14   rpart_4.1-15        Matrix_1.4-1
## [19] checkmate_2.0.0   rmarkdown_2.18      labeling_0.4.2
## [22] textshaping_0.3.6 splines_4.0.5        stringr_1.4.0
## [25] foreign_0.8-81    htmlwidgets_1.5.4   munsell_0.5.0
## [28] compiler_4.0.5    xfun_0.22           pkgconfig_2.0.3
## [31] systemfonts_1.0.4 base64enc_0.1-3      htmltools_0.5.2
## [34] nnet_7.3-15       tidyselect_1.1.1    htmlTable_2.4.1
## [37] gridExtra_2.3     tibble_3.1.1        Hmisc_4.7-0
## [40] fansi_0.4.2       crayon_1.5.2         withr_2.5.0
## [43] grid_4.0.5        gtable_0.3.1        lifecycle_1.0.0
## [46] DBI_1.1.3         magrittr_2.0.1       scales_1.1.1
## [49] cli_3.0.1         stringi_1.7.6        farver_2.1.0
## [52] latticeExtra_0.6-30 snakecase_0.11.0     ellipsis_0.3.2
## [55] generics_0.1.3    vctrs_0.3.8         Formula_1.2-4
## [58] RColorBrewer_1.1-3 tools_4.0.5          interp_1.0-33
## [61] glue_1.4.2        purrr_0.3.4          jpeg_0.1-9
## [64] fastmap_1.1.0     survival_3.2-10     yaml_2.3.5
## [67] colorspace_2.0-0  cluster_2.1.1        knitr_1.33
```