# lab10.R

Feipeng Huang

2022-11-16

```r
#Q1
rm(list = ls())

rope =
read.csv("/Users/stonehuang/Documents/environmental_data/data/rope.csv")
rope$rope.type = factor(rope$rope.type)

levels(rope$rope.type)

## [1] "BLAZE" "BS"    "PI"    "SB"    "VEL"   "XTC"

n_obs = nrow(rope)
n_groups = length(levels(rope$rope.type))

grandmean = mean(rope$p.cut)
residuals = grandmean - rope$p.cut

ss_tot = sum(residuals ^ 2)
df_tot = n_obs - 1

agg_resids =
 aggregate(
    x = rope$p.cut,
    by = list(rope$rope.type),
    FUN = function(x) x - mean(x))

str(agg_resids)

## 'data.frame':    6 obs. of  2 variables:
##  $ Group.1: Factor w/ 6 levels "BLAZE","BS","PI",..: 1 2 3 4 5 6
##  $ x      :List of 6
##   ..$ : num  0.633 0.633 0.623 0.173 0.143 ...
##   ..$ : num  0.303 0.223 0.193 0.183 0.093 ...
##   ..$ : num  0.363 0.133 0.113 0.103 0.083 0.063 0.053 0.053 0.033 -0.007
## ...
##   ..$ : num  0.398 0.238 0.178 0.168 0.168 0.138 0.118 0.118 0.048 0.038
## ...
##   ..$ : num  0.65 0.36 0.3 0.22 0.16 ...
##   ..$ : num  0.3545 0.3145 0.2745 0.2545 0.0745 ...

#sums of squared residuals within each group
agg_sum_sq_resids =
```

```
 aggregate(
    x = rope$p.cut,
    by = list(rope$rope.type),
    FUN = function(x) sum((x - mean(x))^2))

str(agg_sum_sq_resids)

## 'data.frame':    6 obs. of  2 variables:
##  $ Group.1: Factor w/ 6 levels "BLAZE","BS","PI",..: 1 2 3 4 5 6
##  $ x      : num  1.808 0.405 0.312 0.633 1.129 ...

ss_within = sum(agg_sum_sq_resids$x)
df_within = n_obs - n_groups

ss_among = ss_tot - ss_within
df_among = n_groups - 1

ms_within = ss_within / df_within
ms_among  = ss_among / df_among

f_ratio = ms_among / ms_within
f_pval = 1 - pf(f_ratio, df_among, df_within)
```

*#Q2 Based on the figure, I think the variances are not equal among the groups because the boxes are in different sizes.*
*#Q3*

```
bartlett.test(p.cut ~ rope.type, data = rope)

##
##  Bartlett test of homogeneity of variances
##
## data:  p.cut by rope.type
## Bartlett's K-squared = 19.687, df = 5, p-value = 0.00143
```

*#Q4 I think an ANOVA-type analysis is not appropriate for the raw data because the raw data does not have equal variances of the residuals within groups, which is an assumption of ANOVA.*

```
fit_rope_1 = lm(p.cut ~ rope.type, data = rope)
summary(fit_rope_1)

##
## Call:
## lm(formula = p.cut ~ rope.type, data = rope)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2800 -0.1500 -0.0355  0.1030  0.6500
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.36714    0.04493   8.172 4.45e-13 ***
## rope.typeBS  -0.13014    0.06433  -2.023  0.04538 *
## rope.typePI  -0.18014    0.06433  -2.800  0.00599 **
## rope.typeSB  -0.09514    0.06433  -1.479  0.14186
## rope.typeVEL -0.01714    0.06433  -0.266  0.79033
## rope.typeXTC -0.10164    0.06433  -1.580  0.11683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2059 on 115 degrees of freedom
## Multiple R-squared:  0.08843,    Adjusted R-squared:  0.0488
## F-statistic: 2.231 on 5 and 115 DF,  p-value: 0.05582
```

#Q5 BLAZE
#Q6 0.36714, it is the intercept from the model coefficient table.
#Q7 0.2655

```r
0.36714 + (-0.10164)
```

```
## [1] 0.2655
```

#Q8 p-value = 7.238e-07

```r
shapiro.test(residuals(fit_1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit_1)
## W = 0.91144, p-value = 7.238e-07
```

#Q9 The model residuals do not meet the normality assumption because the p-value is lower than 0.05 so we reject the null hypothesis of the normality test.
#Q10 3 groups meet the normality assumption.

```r
sapply(X = agg_resids$x, FUN = function(x) shapiro.test(x))
```

```
##           [,1]                              [,2]
## statistic 0.7948103                         0.9229202
## p.value   0.0005471159                      0.1127864
## method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
## data.name "x"                               "x"
##           [,3]                              [,4]
## statistic 0.9259378                         0.9135802
## p.value   0.1289522                         0.07460241
## method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
## data.name "x"                               "x"
##           [,5]                              [,6]
## statistic 0.9017058                         0.8863599
## p.value   0.04439502                        0.02306614
## method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
## data.name "x"                               "x"
```
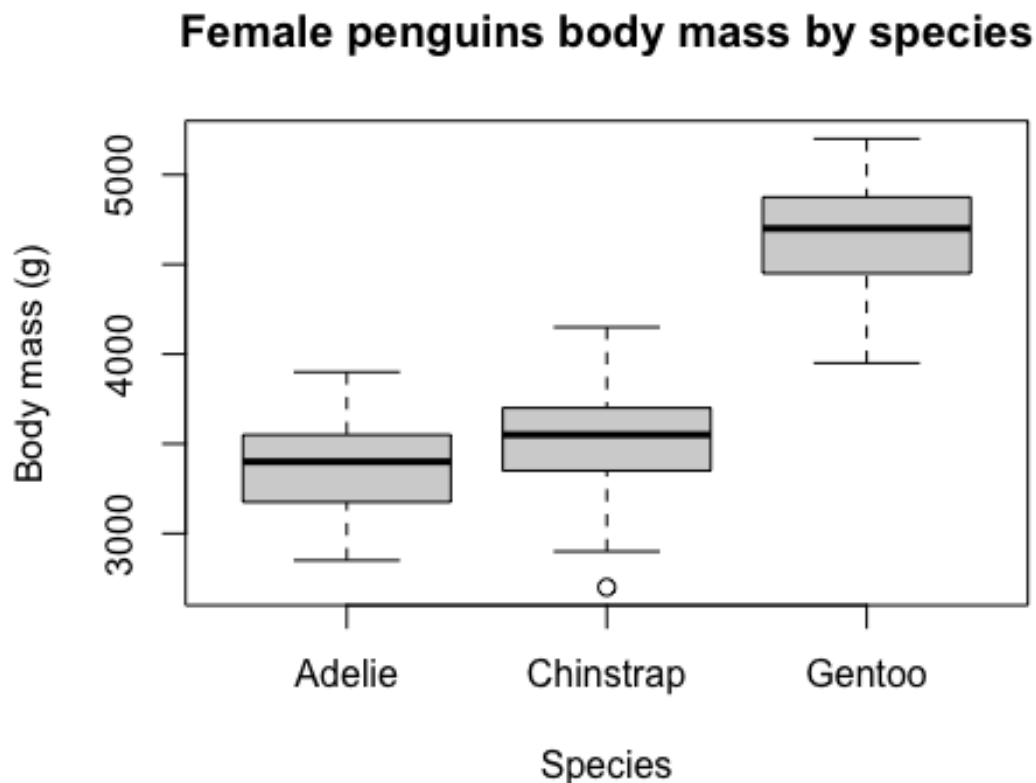
```
require(palmerpenguins)

## Loading required package: palmerpenguins

pen_fem = subset(penguins, sex == "female")
#Q12
boxplot(body_mass_g ~ species, data = pen_fem, main = "Female penguins body
mass by species", xlab = "Species", ylab = "Body mass (g)")
```

## Female penguins body mass by species

```
bartlett.test(body_mass_g ~ species, data = pen_fem)

##
##  Bartlett test of homogeneity of variances
##
```

```
## data:  body_mass_g by species
## Bartlett's K-squared = 0.19828, df = 2, p-value = 0.9056
```

```
fit_penguin = lm(body_mass_g ~ species, data=pen_fem)
shapiro.test(residuals(fit_penguin))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(fit_penguin)
## W = 0.99077, p-value = 0.3639
```

```
penguin_hsd = TukeyHSD(aov(fit_penguin))
round(penguin_hsd$species, digits = 4)

##                       diff       lwr       upr   p adj
## Chinstrap-Adelie   158.3703   22.3208  294.4197 0.0179
## Gentoo-Adelie     1310.9058 1195.6491 1426.1624 0.0000
## Gentoo-Chinstrap  1152.5355 1011.0062 1294.0648 0.0000
```