Lab_Assignment_8

2024-11-22

```r
#Excersise 2
library(tidyr)
data_wide <- read.csv("Lab 8 penguin - Lab 8 penguin.csv")
data_long <- gather(data = data_wide, key = "group", value = "bill_depth")
data_long <- separate(data_long, col = group, into = c("region", "sex"), sep = "\\.")
print(data_long)
```

```
##    region     sex bill_depth
## 1     Arg   males       25.9
## 2     Arg   males       25.0
## 3     Arg   males       25.8
## 4     Arg   males       24.1
## 5     Arg   males       23.9
## 6     Arg   males       23.4
## 7     Arg females       22.4
## 8     Arg females       20.2
## 9     Arg females       21.4
## 10    Arg females       21.9
## 11    Arg females       21.8
## 12    Arg females       21.5
## 13  Galap   males       18.1
## 14  Galap   males       21.0
## 15  Galap   males       18.2
## 16  Galap   males       19.3
## 17  Galap   males       19.0
## 18  Galap   males       19.1
## 19  Galap females       17.3
## 20  Galap females       16.5
## 21  Galap females       15.5
## 22  Galap females       17.0
## 23  Galap females       16.8
## 24  Galap females       16.6
```

```r
#Excersise 3
aov.res <- aov(formula = data_long$bill_depth ~ data_long$region * data_long$sex, data=data_long, strin
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'stringsAsFactors' will be disregarded
```

```r
print(aov.res)
```

```
## Call:
##    aov(formula = data_long$bill_depth ~ data_long$region * data_long$sex,
##     data = data_long, stringsAsFactors = TRUE)
##
## Terms:
##                  data_long$region data_long$sex data_long$region:data_long$sex
```

```
## Sum of Squares         164.85042      47.88375                        0.63375
## Deg. of Freedom                1             1                              1
##                  Residuals
## Sum of Squares   15.55833
## Deg. of Freedom        20
##
## Residual standard error: 0.8819958
## Estimated effects may be unbalanced
```

```
summary(aov.res)
```

```
##                                 Df Sum Sq Mean Sq F value   Pr(>F)
## data_long$region                 1 164.85  164.85 211.913 4.18e-12 ***
## data_long$sex                    1  47.88   47.88  61.554 1.57e-07 ***
## data_long$region:data_long$sex  1   0.63    0.63   0.815    0.377
## Residuals                       20  15.56    0.78
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

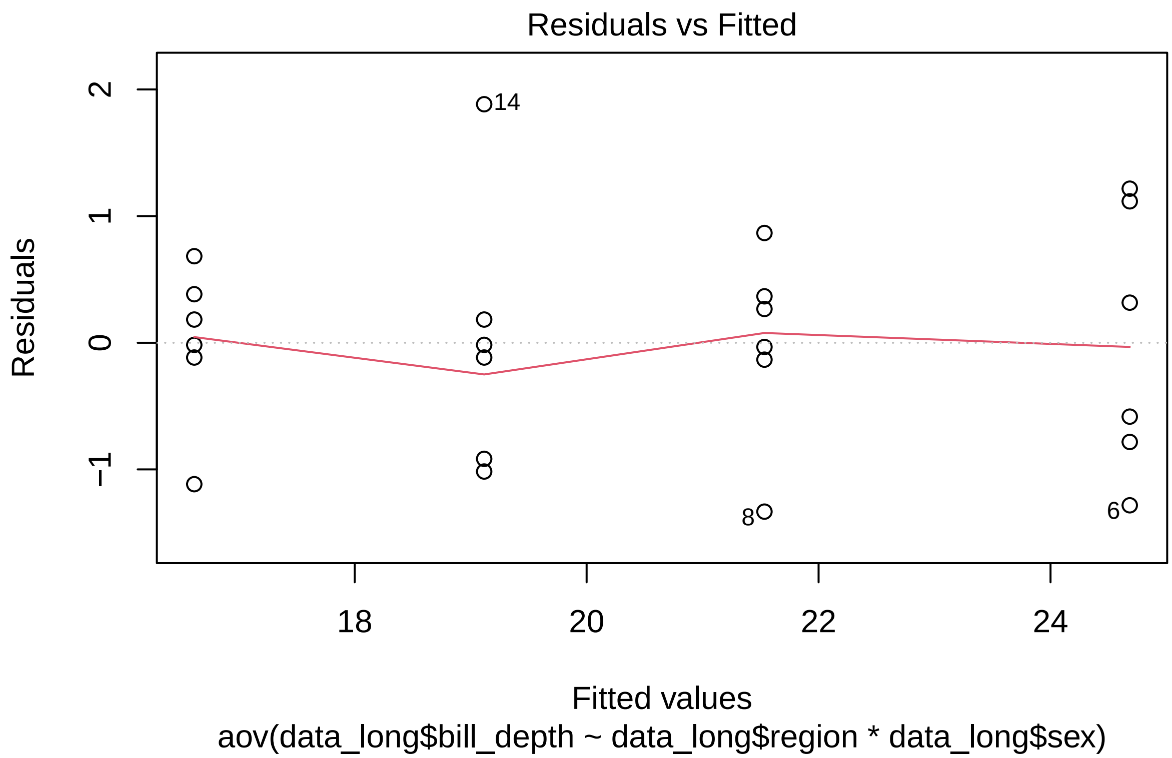
```
#Excersise 4
library(car)
```

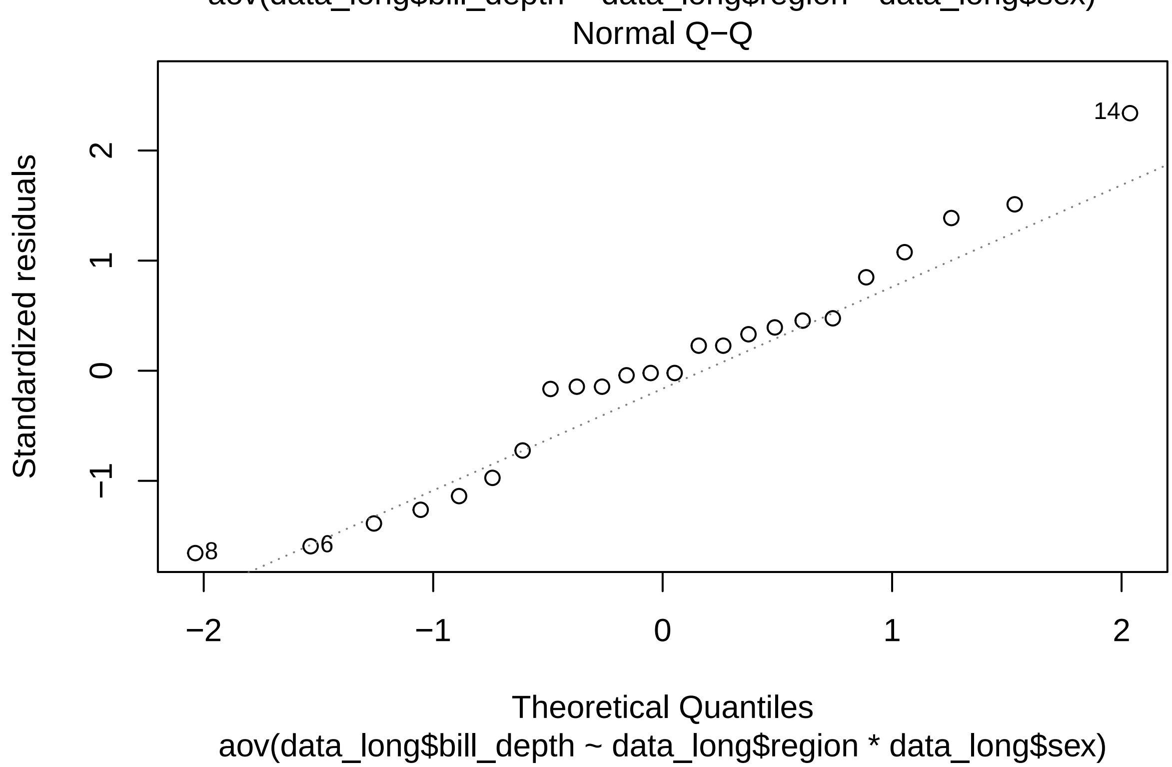
```
## Loading required package: carData
```

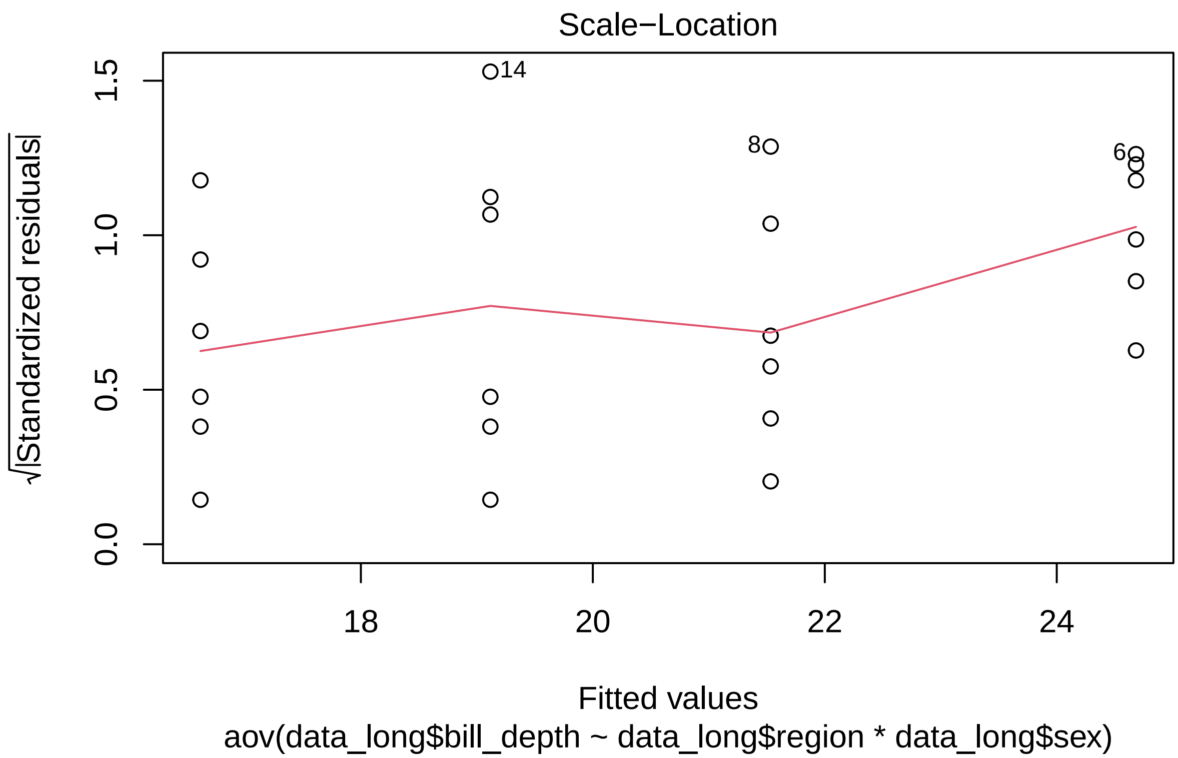
```
levene_test <- leveneTest(bill_depth ~ region * sex, data = data_long)
print(levene_test)
```

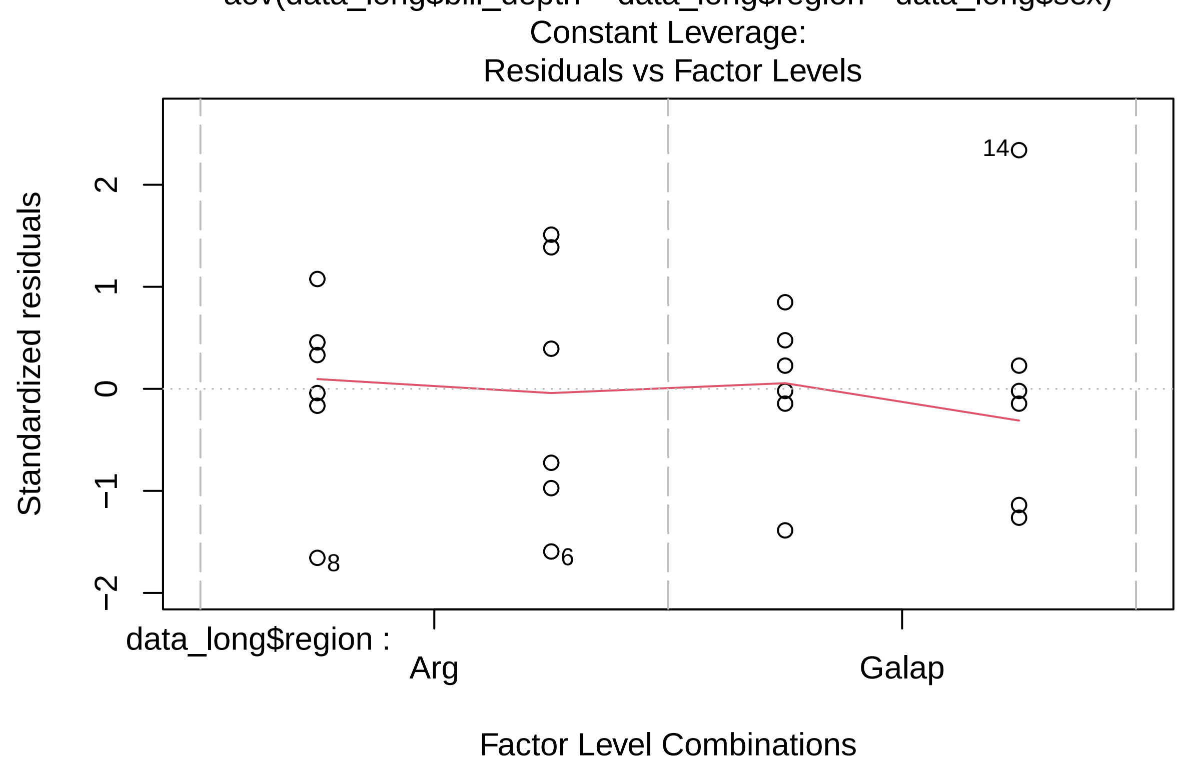
```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.8919 0.4624
##       20
```

```
# Excersise 5
plot(aov.res)
```


Residuals vs Fitted
Residuals
Fitted values
aov(data_long$bill_depth ~ data_long$region * data_long$sex)
14
8
6
18
20
22
24
-1
0
1
2


Normal Q−Q
Standardized residuals
Theoretical Quantiles
aov(data_long$bill_depth ~ data_long$region * data_long$sex)
14
8
6
−2
−1
0
1
2



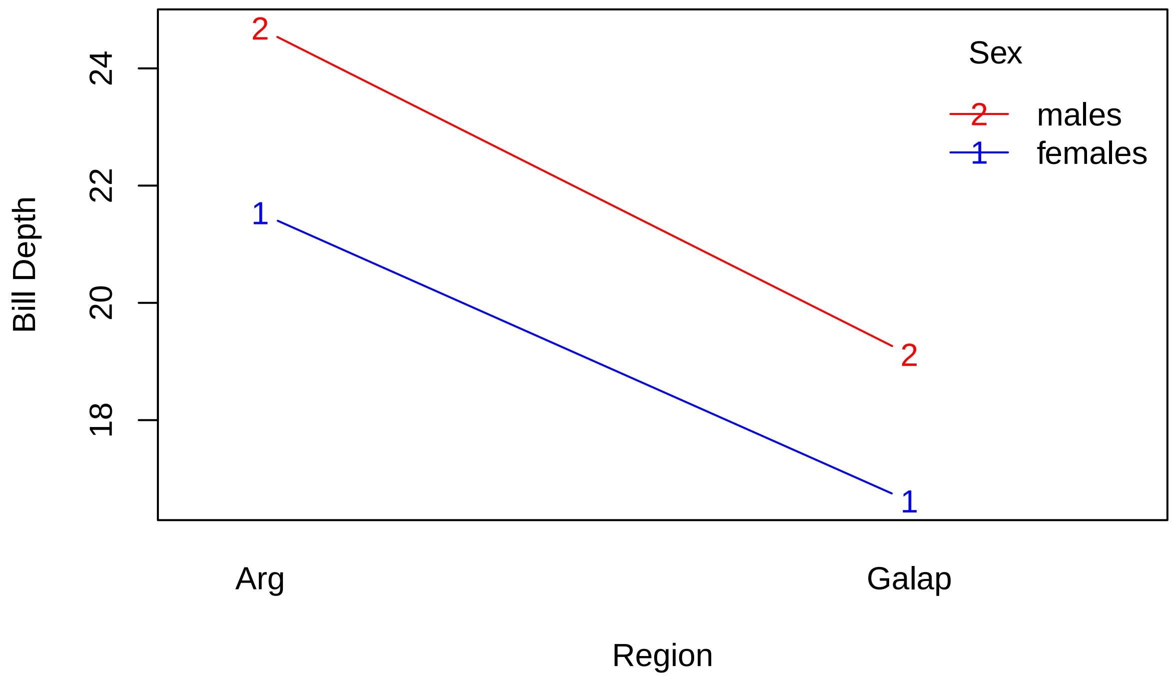



```
residuals_anova <- residuals(aov.res)
shapiro_test <- shapiro.test(residuals_anova)
print(shapiro_test)
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuals_anova
## W = 0.96564, p-value = 0.5616
```

```
#Excersise 6
interaction.plot(
  x.factor = data_long$region,  # x-axis factor
  trace.factor = data_long$sex,  # Grouping factor (trace)
  response = data_long$bill_depth,  # Response variable
  col = c("blue", "red"),  # Colors for lines
  lty = 1,  # Line type
  type = "b",  # Both lines and points
  xlab = "Region",
  ylab = "Bill Depth",
  trace.label = "Sex"
)
```




```
#Excersise 7
#Anova Long way
#intermediate calculations
grandmean <- mean(data_long$bill_depth)
Ameans <- c()
Ameans[1] <- mean(data_long[data_long$region=="Arg", 3])
Ameans[2] <- mean(data_long[data_long$region=="Galap", 3])

Bmeans <- c()
Bmeans[1] <- mean(data_long[data_long$sex=="males", 3])
Bmeans[2] <- mean(data_long[data_long$sex=="females", 3])

ABmeans <- c()
ABmeans[1] <- mean(data_long[data_long$region=="Arg" & data_long$sex=="males", 3])
ABmeans[2] <- mean(data_long[data_long$region=="Galap" & data_long$sex=="males", 3])
ABmeans[3] <- mean(data_long[data_long$region=="Arg" & data_long$sex=="females", 3])
```

```r
ABmeans[4] <- mean(data_long[data_long$region=="Galap" & data_long$sex=="females", 3])
```

```r
totalss <- sum((data_long$bill_depth - grandmean)^2)
cellss <- (nrow(data_long)/length(ABmeans)) * ((ABmeans - grandmean)^2)
errorss <- totalss -cellss
Afactorss <- (nrow(data_long)/length(Ameans)) * sum((Ameans - grandmean)^2)
Bfactorss <- (nrow(data_long)/length(Bmeans)) * sum((Bmeans - grandmean)^2)

grandmean <- mean(data_long$bill_depth)
Ameans <- c()
Ameans[1] <- mean(data_long[data_long$region=="Arg", 3])
Ameans[2] <- mean(data_long[data_long$region=="Galap", 3])

Bmeans <- c()
Bmeans[1] <- mean(data_long[data_long$sex=="males", 3])
Bmeans[2] <- mean(data_long[data_long$sex=="females", 3])

ABmeans <- c()
ABmeans[1] <- mean(data_long[data_long$region=="Arg" & data_long$sex=="males", 3])
ABmeans[2] <- mean(data_long[data_long$region=="Galap" & data_long$sex=="males", 3])
ABmeans[3] <- mean(data_long[data_long$region=="Arg" & data_long$sex=="females", 3])
ABmeans[4] <- mean(data_long[data_long$region=="Galap" & data_long$sex=="females", 3])
```

```r
totalss <- sum((data_long$bill_depth - grandmean)^2)
cellss <- (nrow(data_long)/length(ABmeans)) * ((ABmeans - grandmean)^2)
errorss <- totalss -cellss
Afactorss <- (nrow(data_long)/length(Ameans)) * sum((Ameans - grandmean)^2)
Bfactorss <- (nrow(data_long)/length(Bmeans)) * sum((Bmeans - grandmean)^2)

ABinteractionss <- cellss - Afactorss - Bfactorss

df_total <- nrow(data_long) - 1
df_A <- length(Ameans) - 1
df_B <- length(Bmeans) - 1
df_AB <- (length(Ameans) - 1) * (length(Bmeans) - 1)
df_error <- df_total - (df_A + df_B + df_AB)

MS_A <- Afactorss / df_A
MS_B <- Bfactorss / df_B
MS_AB <- ABinteractionss / df_AB
MS_error <- errorss / df_error

F_A <- MS_A / MS_error
F_B <- MS_B / MS_error
F_AB <- MS_AB / MS_error

p_A <- pf(F_A, df_A, df_error, lower.tail = FALSE)
p_B <- pf(F_B, df_B, df_error, lower.tail = FALSE)
p_AB <- pf(F_AB, df_AB, df_error, lower.tail = FALSE)

print(p_A)
```

```
## [1] 4.643577e-05 9.054110e-04 9.974784e-04 9.276743e-05
```

```r
print(p_B)
```

```
## [1] 0.01137977 0.04884869 0.05107560 0.01623867
```

```r
print(p_AB)
```

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