Chapter 10 Two-way Analysis of Variance

Angelo LaCommare-Soto

2025-04-09

Size-dependent Temperature Constraints in Fiddler Crabs

Research Question 1: Does body size interact with hydration state to influence the thermal preferences of fiddler crabs?

Section 1 - Importing Data

```
# set working directory for all chunks in this file (default working directory is wherever Rmd file is)
getwd()
```

```
## [1] "C:/Users/Angelo L/Documents/GitHub/BIOL710/RCode710/RCode/working_directory"
```

```
library(tidyverse)
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
#commands header=TRUE in order to treat the first row of the data frame as a header and stringsAsFactor
crab <- read.csv("crab.csv",header=TRUE,stringsAsFactors = TRUE)
str(crab)
```

```
## 'data.frame':    80 obs. of  3 variables:
##  $ sand: Factor w/ 2 levels "dry","wet": 2 2 2 2 2 2 2 2 2 2 ...
##  $ size: Factor w/ 2 levels "large","small": 2 2 2 2 2 2 2 2 2 2 ...
##  $ tb  : num  26.6 35.7 25.5 27.1 27.3 25.7 25.7 30 27.7 28.3 ...
```

```r
head(crab)
```

```
##   sand  size   tb
## 1  wet small 26.6
## 2  wet small 35.7
## 3  wet small 25.5
## 4  wet small 27.1
## 5  wet small 27.3
## 6  wet small 25.7
```

Question Answers

a. The dataset "crab.csv" contains 80 total observations of 3 different variables.
b. Given the question 'Does body size interact with hydration state to influence the thermal preferences of fiddler crabs?', we are interested in all three of the variables: 'sand', 'size', 'tb'.
c. I will compare two factor variables ('sand' and 'size'), each with two factor levels, to each other in regards to their effects on the numerical variable body temperature ('tb').

Section 2 - Checking Model Assumptions of Equal Variances

```r
# equal variances test between tb and size
bartlett.test(tb~size, data=crab)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  tb by size
## Bartlett's K-squared = 2.1649, df = 1, p-value = 0.1412
```

```r
# equal variances test between tb and sand
bartlett.test(tb~sand, data=crab)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  tb by sand
## Bartlett's K-squared = 0.45685, df = 1, p-value = 0.4991
```

```r
# If the test results in a p-value > 0.05, then the data has equal variance (no difference in variance
```

Section 3 - Carrying Out a Two-way Analysis of Variance

```r
# ANOVA
crab_aov <- aov(tb ~ sand + size + sand*size, data=crab)
crab_aov
```

```
## Call:
##    aov(formula = tb ~ sand + size + sand * size, data = crab)
##
## Terms:
##                      sand      size sand:size Residuals
## Sum of Squares  1381.1220   14.2805   29.5245  413.9810
## Deg. of Freedom         1         1         1        76
##
## Residual standard error: 2.333906
## Estimated effects may be unbalanced
```

```r
# This is the long, written-out version of the same model below, small differences are due to rounding

crab_aov1 <- aov(tb ~ sand*size, data=crab)
crab_aov1
```

```
## Call:
##    aov(formula = tb ~ sand * size, data = crab)
##
## Terms:
##                      sand      size sand:size Residuals
## Sum of Squares  1381.1220   14.2805   29.5245  413.9810
## Deg. of Freedom         1         1         1        76
##
## Residual standard error: 2.333906
## Estimated effects may be unbalanced
```

```r
# ANOVA table
summary(crab_aov)
```

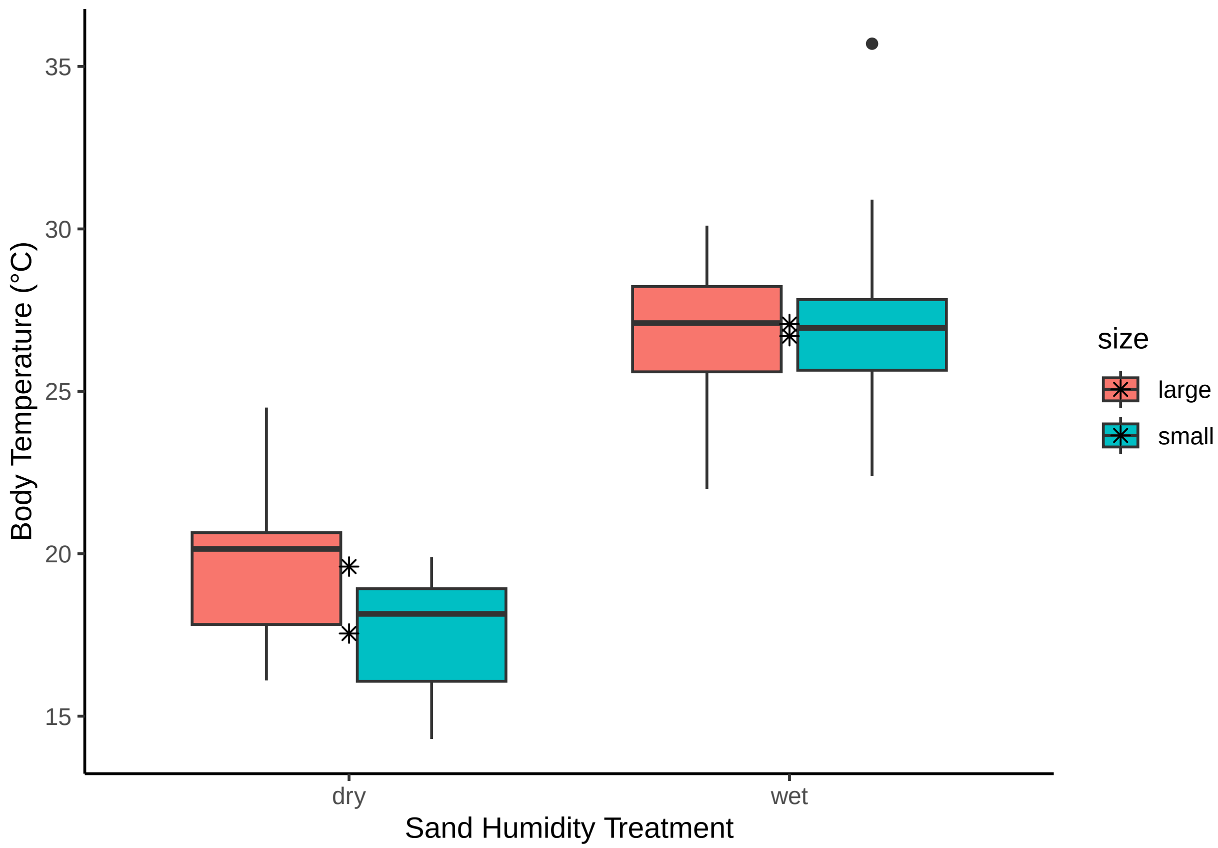
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## sand         1 1381.1  1381.1 253.551 <2e-16 ***
## size         1   14.3    14.3   2.622 0.1096
## sand:size    1   29.5    29.5   5.420 0.0226 *
## Residuals   76  414.0     5.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(crab_aov1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## sand         1 1381.1  1381.1 253.551 <2e-16 ***
## size         1   14.3    14.3   2.622 0.1096
## sand:size    1   29.5    29.5   5.420 0.0226 *
## Residuals   76  414.0     5.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Challenge: Plot the Data.

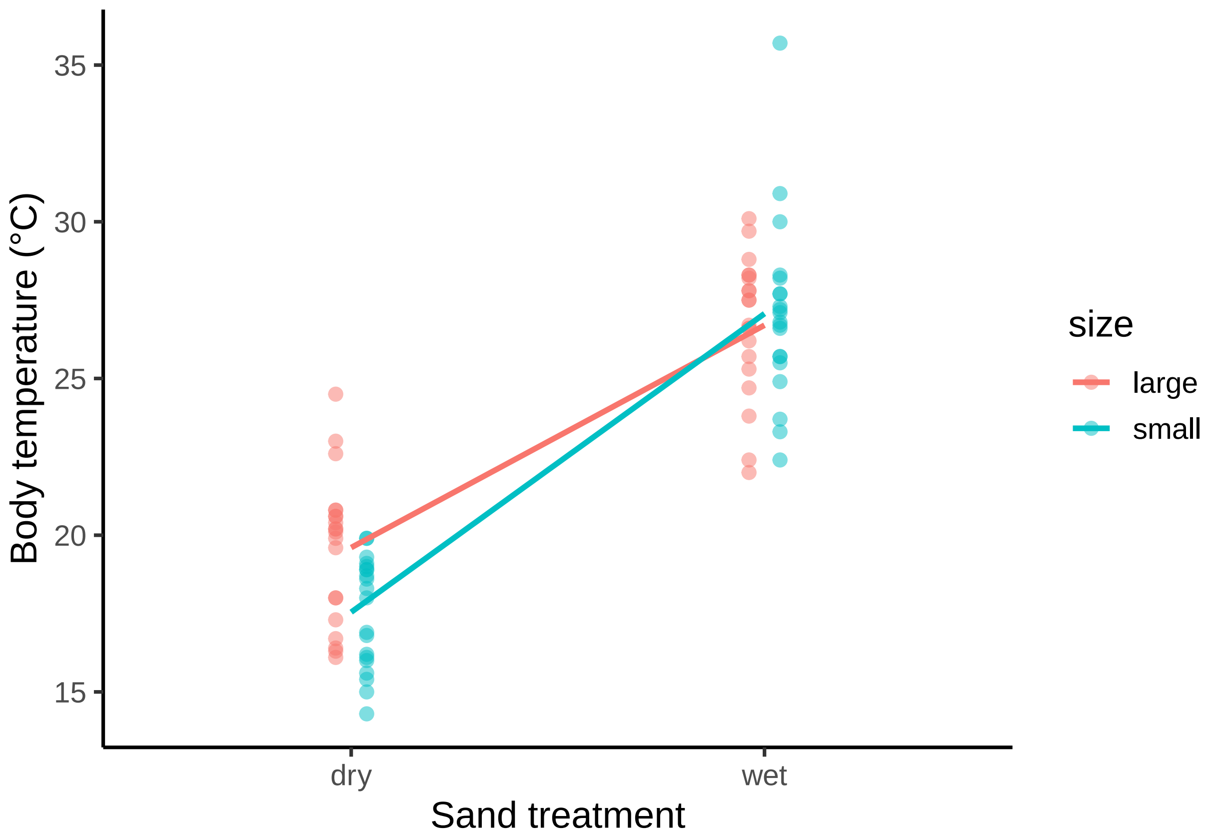
```r
# creating a boxplot with means
p1 <- ggplot(crab,aes(x=sand,y=tb, fill=size)) +
  geom_boxplot() +
  stat_summary(fun=mean, geom = 'point', shape = 8, size = 2)+
  ylab("Body Temperature (°C)") +
  xlab("Sand Humidity Treatment") +
  theme_classic()
p1
```



Section 4 - Plotting the Interaction

```r
# Interaction plot
p2 <- ggplot(crab,aes(sand,tb,group=size,color=size)) +
  geom_point(size=2,alpha=.5,position = position_dodge(width=.15)) +
  geom_smooth(method = "lm", se = FALSE) +
  ylab("Body temperature (°C)") +
  xlab("Sand treatment") +
  theme_classic(14)
p2
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Section 5 - Carrying Out a Multiple Comparision Using the Tukey Test

```
# Tukey test
t <- TukeyHSD(crab_aov)
t
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tb ~ sand + size + sand * size, data = crab)
##
## $sand
##         diff      lwr      upr p adj
## wet-dry 8.31 7.270591 9.349409     0
##
## $size
##               diff       lwr       upr     p adj
## small-large -0.845 -1.884409 0.1944087 0.1095567
##
## $`sand:size`
##                       diff       lwr        upr     p adj
## wet:large-dry:large  7.095  5.156303  9.0336967 0.0000000
## dry:small-dry:large -2.060 -3.998697 -0.1213033 0.0329093
## wet:small-dry:large  7.465  5.526303  9.4036967 0.0000000
```

```
## dry:small-wet:large -9.155 -11.093697 -7.2163033 0.0000000
## wet:small-wet:large  0.370  -1.568697  2.3086967 0.9585169
## wet:small-dry:small  9.525   7.586303 11.4636967 0.0000000
```

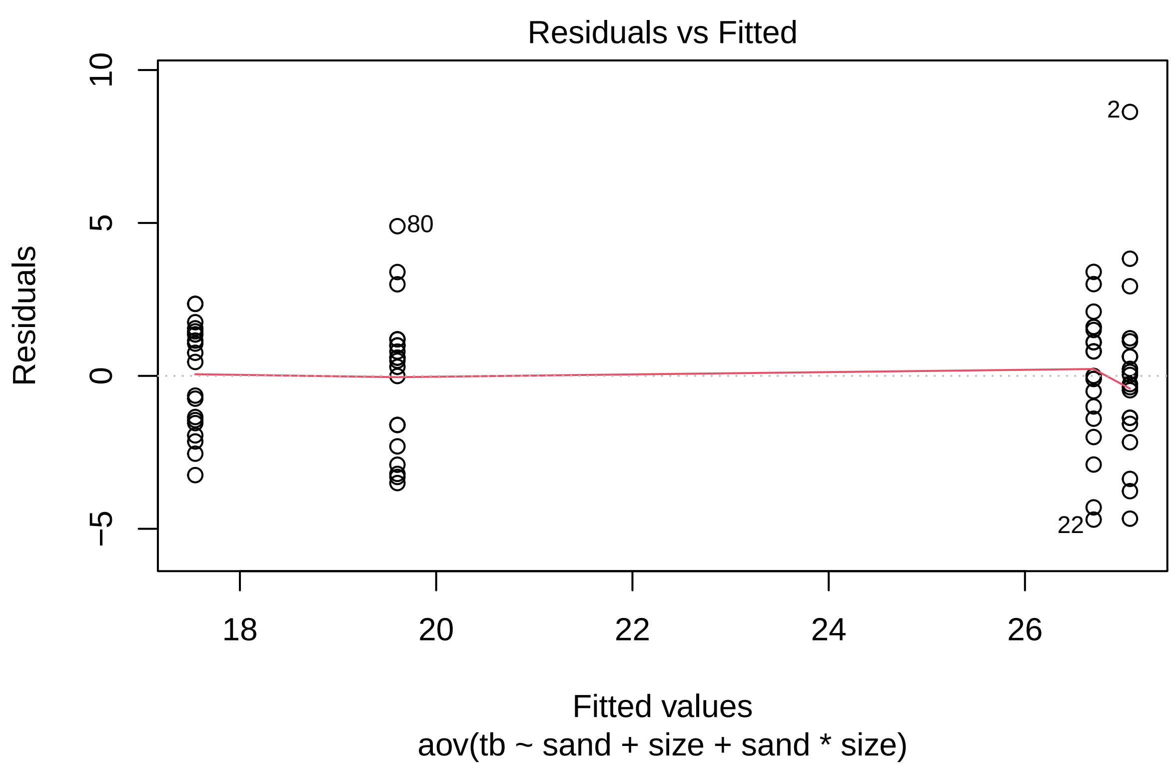
Question Answers

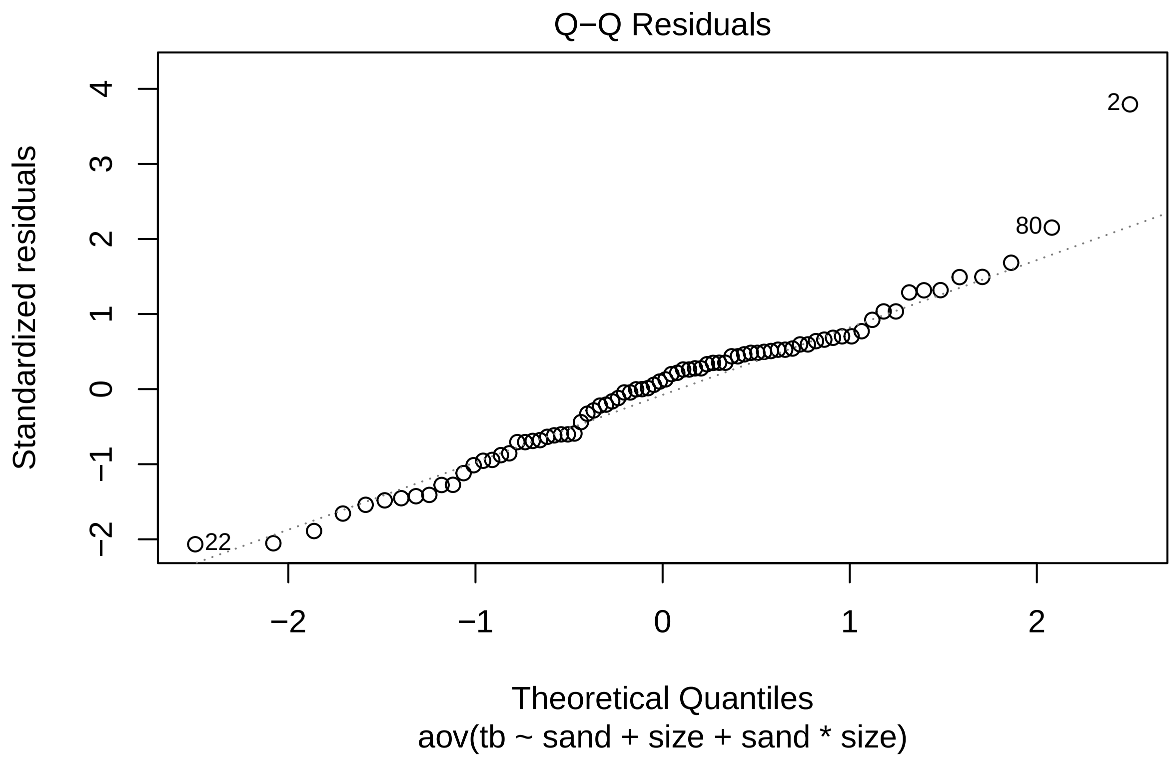
a. All group means are significantly different, except for between all small and large crab body temperature readings in this study (p-value: 0.109) because the mean body temperature difference between small and large crabs under wet sand conditions is not significantly different (p-value: 0.959).
b. Overall, fiddler crabs exposed to dry sand conditions exhibited a much lower mean body temperature than those exposed to wet sand conditions. Crab size was not associated with body temperature differences in wet sand conditions, but it was a factor in dry sand conditions, with larger crabs demonstrating higher body temperatures than their smaller counterparts. This could be due to larger crabs having lower surface area to volume ratios that allows them to more efficiently retain their body temperature.

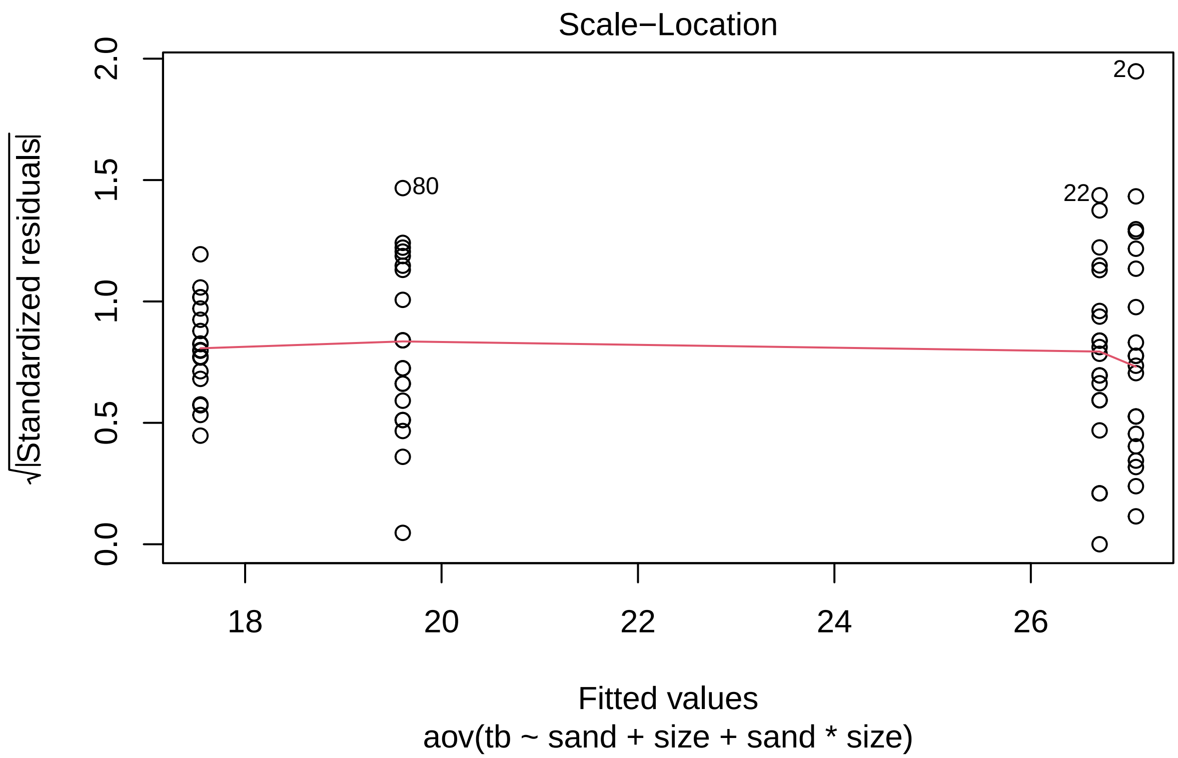
Section 6 - Checking Model Assumptions of Normality

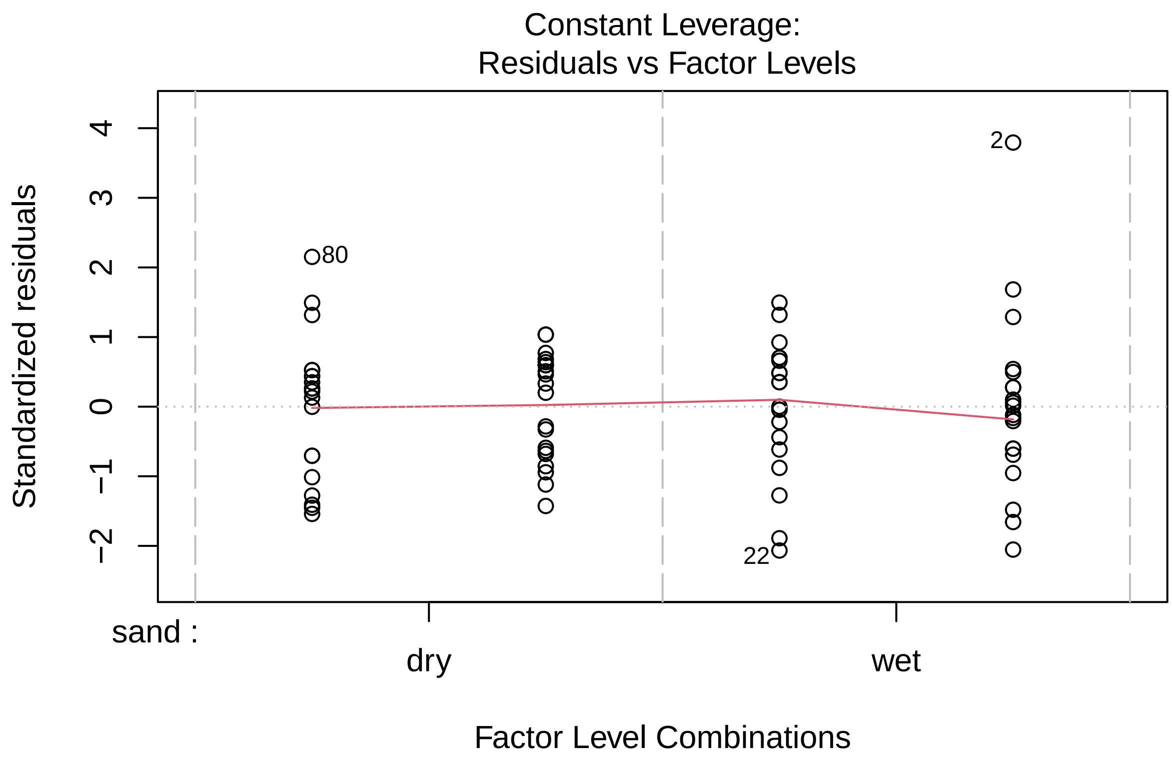
Challenge: Test the model assumptions of normality.

```
# The aov() function (done originally in the 4th chunk) prepares the data for model checking plots: a p
# model checking plots
plot(crab_aov)
```




Q−Q Residuals
Standardized residuals
Theoretical Quantiles
aov(tb ~ sand + size + sand * size)
4
3
2
1
0
−1
−2
−2
−1
0
1
2
2
80
22


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
18
20
22
24
26
Fitted values
aov(tb ~ sand + size + sand * size)
80
22
2




```r
# Another way we can test for normality in our data is by employing a Shapiro test. In this test, the n
crab_aov_res <- residuals(crab_aov)

shapiro.test(crab_aov_res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  crab_aov_res
## W = 0.96571, p-value = 0.03034
```

Question Answers

a. The assumptions of a two-way ANOVA are 1) independence of observations, 2) normality within groups, and 3) and homoscedasticity (variance equality between groups).
b. According to the QQ residuals plot and the Shapiro-Wilk normality test, my model does not follow the assumptions for an ANOVA.

Discussion Question Answers

a. A one-way ANOVA tests the influence of three or more groups within a single independent variable on a continuous numerical variable. A two-way ANOVA tests the influence of multiple groups within two independent variables and their interaction on a continuous numerical variable.

b. In the linear formula for the two-way ANOVA, we multiply the two factors to examine the interaction between them on body temperature in fiddler crabs.
c. The model output demonstrates three different p-values because it is testing for the effect of sand humidity and crab size separately, as well as their interaction on the body temperature of fiddler crabs.