R_Activity_Assignment_5

Dahee Ahn

2024-09-27

```
rm(list=ls())
```

1. Download the birdies data set from Canvas and load it into R using read.table(). We will be analyzing the weight gain of chicks (grams) as a function of different feeds

```
birdies_df_txt <- read.table(file="C:/Users/chemk/OneDrive/Desktop/Classes/ENT6707_DataAnalysis/
week6/birdies.txt", header=TRUE)
head(birdies_df_txt)
```

```
##   weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
```

```
tail(birdies_df_txt)
```

```
##    weight   feed
## 66    352 casein
## 67    359 casein
## 68    216 casein
## 69    222 casein
## 70    283 casein
## 71    332 casein
```

```
str(birdies_df_txt)
```

```
## 'data.frame':    71 obs. of  2 variables:
##  $ weight: int  179 160 136 227 217 168 108 124 143 140 ...
##  $ feed  : chr  "horsebean" "horsebean" "horsebean" "horsebean" ...
```

```
nrow(birdies_df_txt)
```

```
## [1] 71
```

```
View(birdies_df_txt)
birdies_df_txt$feed <- as.factor(birdies_df_txt$feed)
summary(birdies_df_txt)
```

```
##      weight             feed   
##  Min.   :108.0   casein   :12  
##  1st Qu.:204.5   horsebean:10  
##  Median :258.0   linseed  :12  
##  Mean   :261.3   meatmeal :11  
##  3rd Qu.:323.5   soybean  :14  
##  Max.   :423.0   sunflower:12  
```

```
str(birdies_df_txt)
```

```
## 'data.frame':    71 obs. of  2 variables:
##  $ weight: int  179 160 136 227 217 168 108 124 143 140 ...
##  $ feed  : Factor w/ 6 levels "casein","horsebean",..: 2 2 2 2 2 2 2 2 2 2 ...
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2     
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
birdies_df_txt %>% group_by(feed) %>% summarise(Means= mean(weight), SD=sd(weight), max_birdies=
max(weight))
```

```
## # A tibble: 6 × 4
##   feed      Means    SD max_birdies
##   <fct>     <dbl> <dbl>       <int>
## 1 casein     324.  64.4         404
## 2 horsebean  160.  38.6         227
## 3 linseed    219.  52.2         309
## 4 meatmeal   277.  64.9         380
## 5 soybean    246.  54.1         329
## 6 sunflower  329.  48.8         423
```

```
birdies_df_txt %>% group_by(feed) %>% summarise(Means= mean(weight), SD=sd(weight), max_birdies=
max(weight)) %>% arrange(Means)
```

```
## # A tibble: 6 × 4
##   feed      Means    SD max_birdies
##   <fct>     <dbl> <dbl>       <int>
## 1 horsebean  160.  38.6         227
## 2 linseed    219.  52.2         309
## 3 soybean    246.  54.1         329
## 4 meatmeal   277.  64.9         380
## 5 casein     324.  64.4         404
## 6 sunflower  329.  48.8         423
```

2. Fit an ANOVA of weight as a function of feed. Note that when I write blah1 as a function of blah2, blah1 should be the response variable and/or appear on the y-axis in any graphs whereas blah2 would be the predictor(s). Provide the R output for the ANOVA table and briefly explain how the degrees of freedom were calculated for each line of the table.

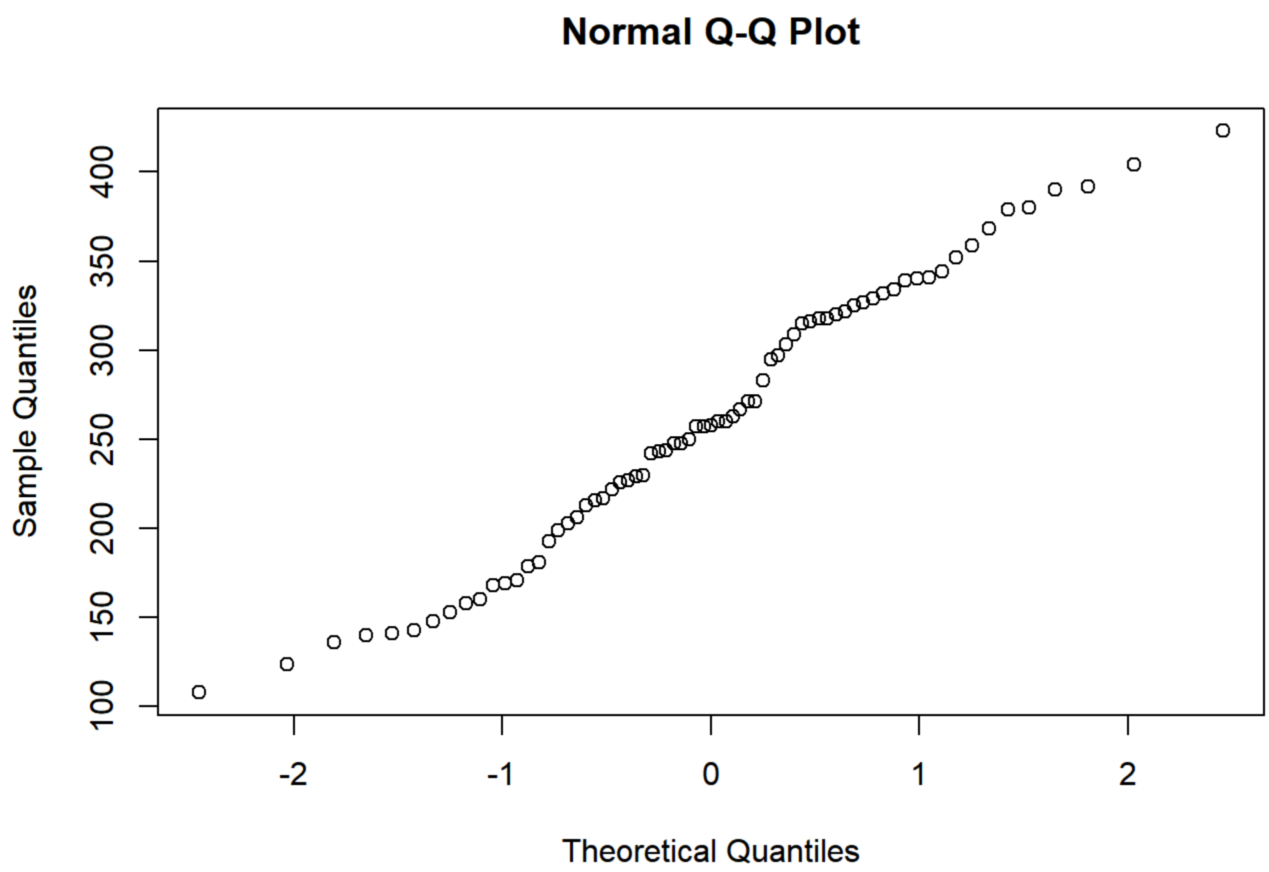
```
windows(width=3, height=2)
library(datasets)
library(ggplot2)
head(birdies_df_txt)
```

```
##   weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
```

```
plot(weight~feed, data=birdies_df_txt)
ggplot(birdies_df_txt, mapping=aes(x=feed, y=weight, fill=feed))+geom_boxplot()+theme_classic()+
xlab("Feed")+ylab("Weight")
```

```
hist(birdies_df_txt$weight)
```


```
qqnorm(birdies_df_txt$weight)
```




```
var_check <- birdies_df_txt %>% group_by(feed) %>% summarize(Variance=var(weight))
var_check$Variance[2]/var_check$Variance[1]
```

```
## [1] 0.3593585
```

```
birdies_df_txt$sqrt_weight <- sqrt(birdies_df_txt$weight)
var_check_sqrt <- birdies_df_txt %>% group_by(feed) %>% summarize(Variance=var(sqrt_weight))
variance_ratio_sqrt <- var_check_sqrt$Variance[2]/var_check_sqrt$Variance[1]
print(variance_ratio_sqrt)
```

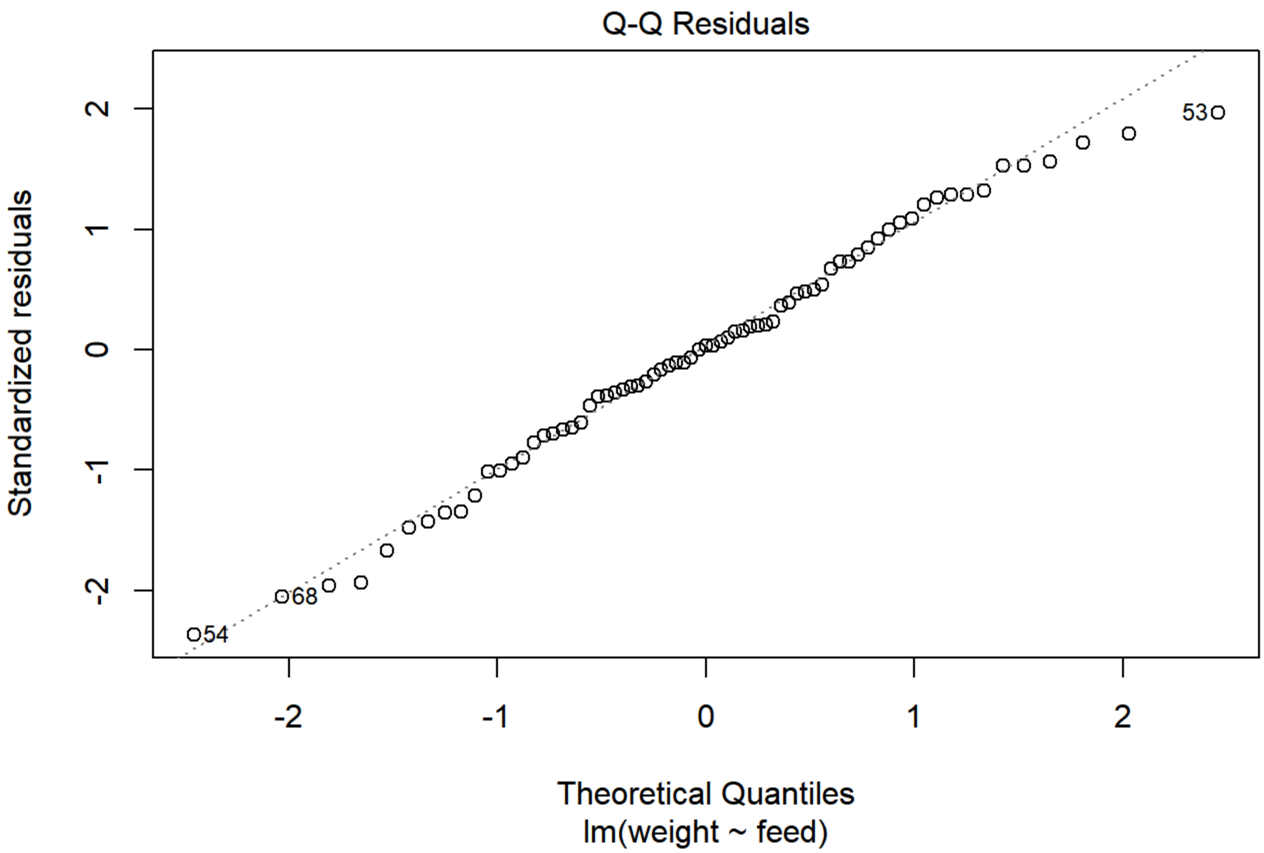
```
## [1] 0.6536348
```

```
fit_1 <- lm(weight~feed, data=birdies_df_txt)
anova(fit_1)
```

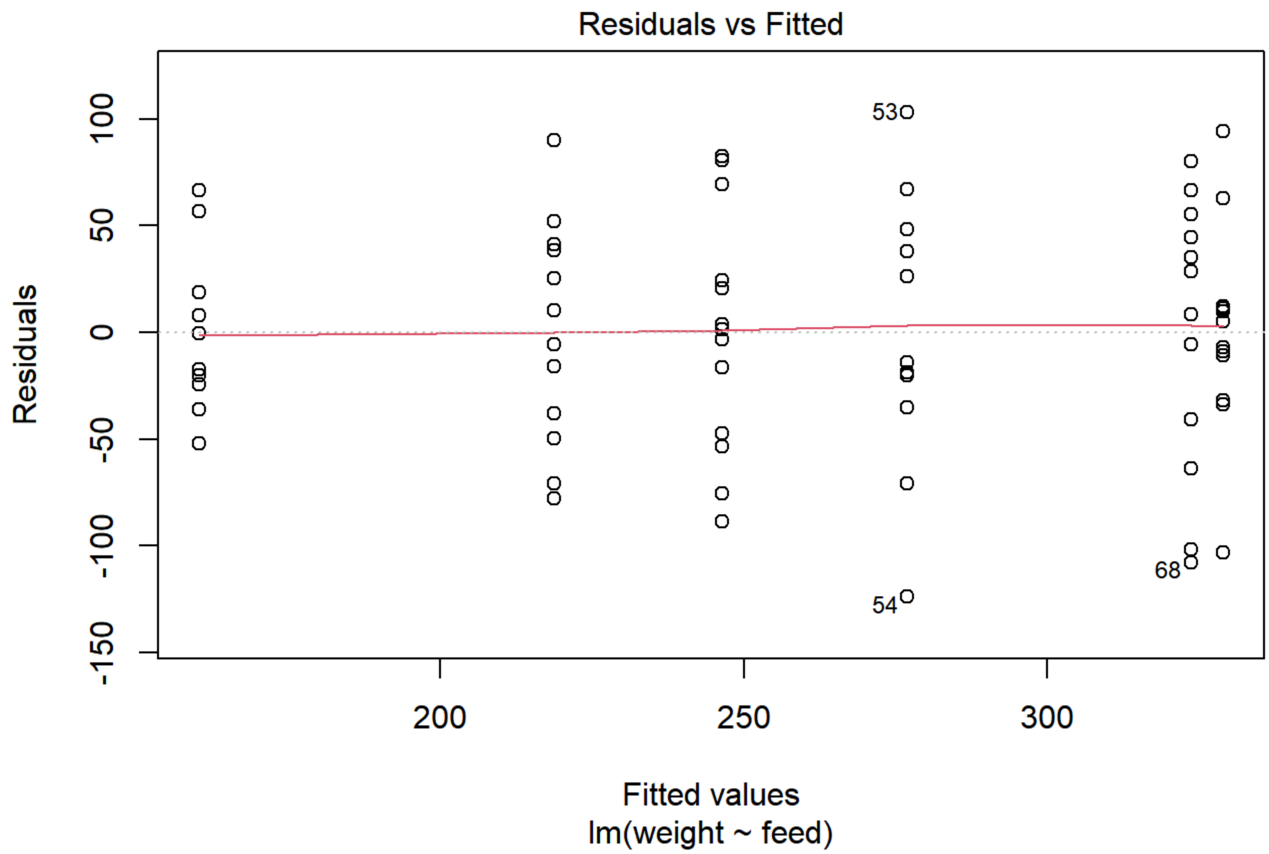
```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## feed       5 231129   46226  15.365 5.936e-10 ***
## Residuals 65 195556    3009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Degrees of freedom of feed: k-1 (K = the number of groups) = 6-1 = 5
# Degrees of freedom of residauls: N-k (N = total number of observations) = 71-6 = 65
```

```
plot(fit_1, which=c(2))
```




```
plot(fit_1, which=c(1))
```



3. The above ANOVA table you just created tells us if the variable weight varies across the levels of feed, but not HOW chick weight gain differs between feeds. Conduct pairwise comparisons using the emmeans package to identify any potential differences between groups (i.e., just report the R output for this question).

```
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
emmeans(fit_1, pairwise~feed)
```

```
## $emmeans
##  feed      emmean   SE df lower.CL upper.CL
##  casein       324 15.8 65      292      355
##  horsebean    160 17.3 65      126      195
##  linseed      219 15.8 65      187      250
##  meatmeal     277 16.5 65      244      310
##  soybean      246 14.7 65      217      276
##  sunflower    329 15.8 65      297      361
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast              estimate   SE df t.ratio p.value
##  casein - horsebean      163.38 23.5 65   6.957  <.0001
##  casein - linseed        104.83 22.4 65   4.682  0.0002
##  casein - meatmeal        46.67 22.9 65   2.039  0.3325
##  casein - soybean         77.15 21.6 65   3.576  0.0084
##  casein - sunflower       -5.33 22.4 65  -0.238  0.9999
##  horsebean - linseed     -58.55 23.5 65  -2.493  0.1413
##  horsebean - meatmeal   -116.71 24.0 65  -4.870  0.0001
##  horsebean - soybean     -86.23 22.7 65  -3.797  0.0042
##  horsebean - sunflower  -168.72 23.5 65  -7.184  <.0001
##  linseed - meatmeal      -58.16 22.9 65  -2.540  0.1277
##  linseed - soybean       -27.68 21.6 65  -1.283  0.7933
##  linseed - sunflower    -110.17 22.4 65  -4.920  0.0001
##  meatmeal - soybean       30.48 22.1 65   1.379  0.7391
##  meatmeal - sunflower    -52.01 22.9 65  -2.271  0.2207
##  soybean - sunflower     -82.49 21.6 65  -3.823  0.0039
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

4. Write a BRIEF summary of your analysis by answering only the following: (i) which feed was associated with the largest chick weight gain? (ii) was it statistically clear that a single feed was the best for weight gain? Report the necessary statistics (t-values, degrees of freedom, and p-values) to justify your conclusions.

Q1. Which feed was associated with the largest chick weight gain?

A1. The feed associated with the largest chick weight gain was Sunflower, with an estimated mean weight of 329 gram.

Q2. Was it statistically clear that a single feed was the best for weight gain?

A2. There were some significant differences in weight gain among the different types of feed, while others were not. For instance: Comparing casein and horsebean, the weight(mean ± SE) of casein was 163.38g greater than that of horsebean (Tukey's test: $t_{65} = 6.96, p < 0.0001$), indicating a statistically significant difference. However, the weight(mean ± SE) of sunflower was only 5.33g greater than that of casein (Tukey's test: $t_{65} = 0.24, p=0.9999$), indicating no significant difference. This suggests that while casein perfomed significantly better than horsebean, there was no clear evidence that sunflower was statistically superior to casein in promoting weight gain.

Q3. Report the necessary statistics (t-values, degrees of freedom, and p-values) to justify your conclusions.

A3. Comparing casein and horsebean: t-ratio: 6.96, df: 65, p-value: <0.0001 / Comparing casein and sunflower: t-ratio: 0.24, df: 65, p-value: 0.9999 / Comparing horsebean and sunflower: t-ratio: 7.18, df: 65, p-value: < 0.0001