Assignment 2

Kshitij Kavimandan, Pablo Alves, Pooja Mangal (Group 15)

11 March 2024

Exercise 1. Fruit flies

a)

```
# Load the data
fruitflies <- read.table("fruitflies.txt", header = TRUE)

# Add Logarithm of Longevity
fruitflies$loglongevity <- log(fruitflies$longevity)

attach(fruitflies)

# Prepare data for plotting
long1 <- fruitflies[fruitflies$activity == "isolated",]$loglongevity
size1 <- fruitflies[fruitflies$thorax == "isolated",]$loglongevity
```

```
# Do basic plots
#plot(long1,size1,col="red",pch=3);points(long2,size2,col="blue",pch=3);points(long3,size3,col=

#legend(4.2,0.75,legend=c('Isolated','Low','High'),pch=c(3,3,3),col=c('red','blue','green'))
```

```
attach(fruitflies)
```

```
## The following objects are masked from fruitflies (pos = 3):
##
##     activity, loglongevity, longevity, thorax
```

```
# Perform ANOVA to test for the effect of sexual activity on longevity
anova_model <- aov(loglongevity ~ activity, data = fruitflies)
summary(anova_model)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## activity     2  3.666  1.8332   19.42 1.8e-07 ***
## Residuals   72  6.797  0.0944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Estimated longevities for the three conditions
means <- aggregate(loglongevity ~ activity, data = fruitflies, FUN = mean)
print(means)
```

```
##   activity loglongevity
## 1     high     3.602124
## 2 isolated     4.119349
## 3      low     3.999836
```

```r
# Subset the data for 'high' and 'low' activity levels
high_activity <- loglongevity[activity == "high"]
low_activity <- loglongevity[activity == "isolated"]

# Perform the t-test
t_test_result <- t.test(high_activity, low_activity)

# Print the result
print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  high_activity and low_activity
## t = -5.8923, df = 44.433, p-value = 4.704e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6940857 -0.3403635
## sample estimates:
## mean of x mean of y
##  3.602124  4.119349
```

Our one-way ANOVA test shows a significant impact of sexual activity (p=1.8e-07). The estimated longevity for these conditions, with means of isolated, low, and high, are 4.12, 4, and 3.6, respectively.

The Welch Two Sample t-test results indicate a highly significant difference in mean log longevity between the "high" and "low" activity groups of fruit flies ($p < 0.0001$). This suggests that activity level significantly impacts fruit fly longevity, with flies exhibiting lower activity levels tending to live longer.

b)

```r
# Perform ANCOVA to include thorax length as an explanatory variable
ancova_model <- lm(loglongevity ~ activity + thorax, data = fruitflies)
summary(ancova_model)
```

```
##
## Call:
## lm(formula = loglongevity ~ activity + thorax, data = fruitflies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4858 -0.1612  0.0104  0.1510  0.3574
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.21893    0.24865   4.902 5.79e-06 ***
## activityisolated  0.40998    0.05839   7.021 1.07e-09 ***
## activitylow       0.28570    0.05849   4.885 6.18e-06 ***
## thorax            2.97899    0.30665   9.715 1.14e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2027 on 71 degrees of freedom
## Multiple R-squared:  0.7211, Adjusted R-squared:  0.7093
## F-statistic:  61.2 on 3 and 71 DF,  p-value: < 2.2e-16
```

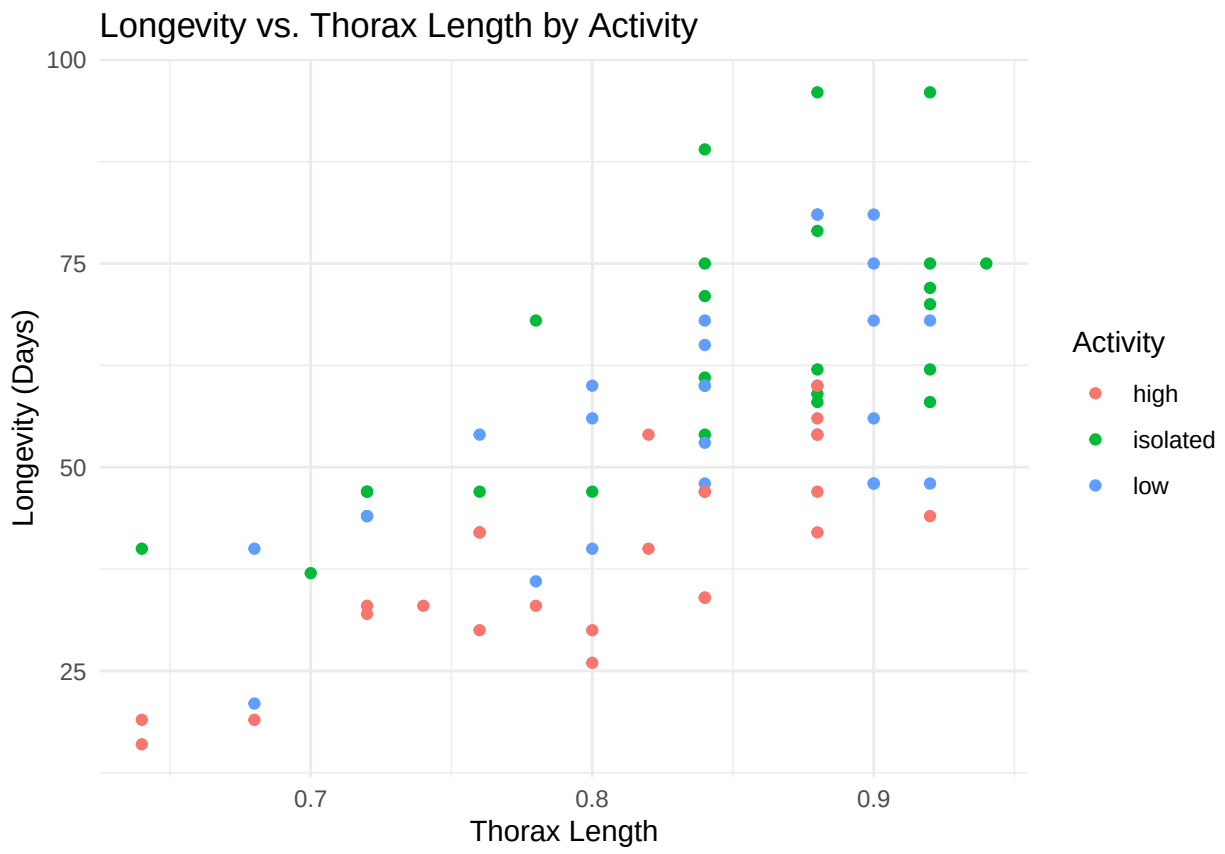
```r
# Calculate estimated longevities for the three groups with average thorax lengths
#avg_thorax <- mean(thorax)
#estimated_longevities <- predict(ancova_model, newdata = data.frame(activity = levels(activity
```

c)

```r
# Load necessary packages
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
# Scatterplot of longevity against thorax length, colored by activity
ggplot(fruitflies, aes(x = thorax, y = longevity, color = activity)) +
  geom_point() +
  labs(title = "Longevity vs. Thorax Length by Activity",
       x = "Thorax Length",
       y = "Longevity (Days)",
       color = "Activity") +
  theme_minimal()
```




```
# Test for the similarity of dependence using ANCOVA
model_c <- lm(longevity ~ activity * thorax, data = fruitflies)
summary(model_c)
```

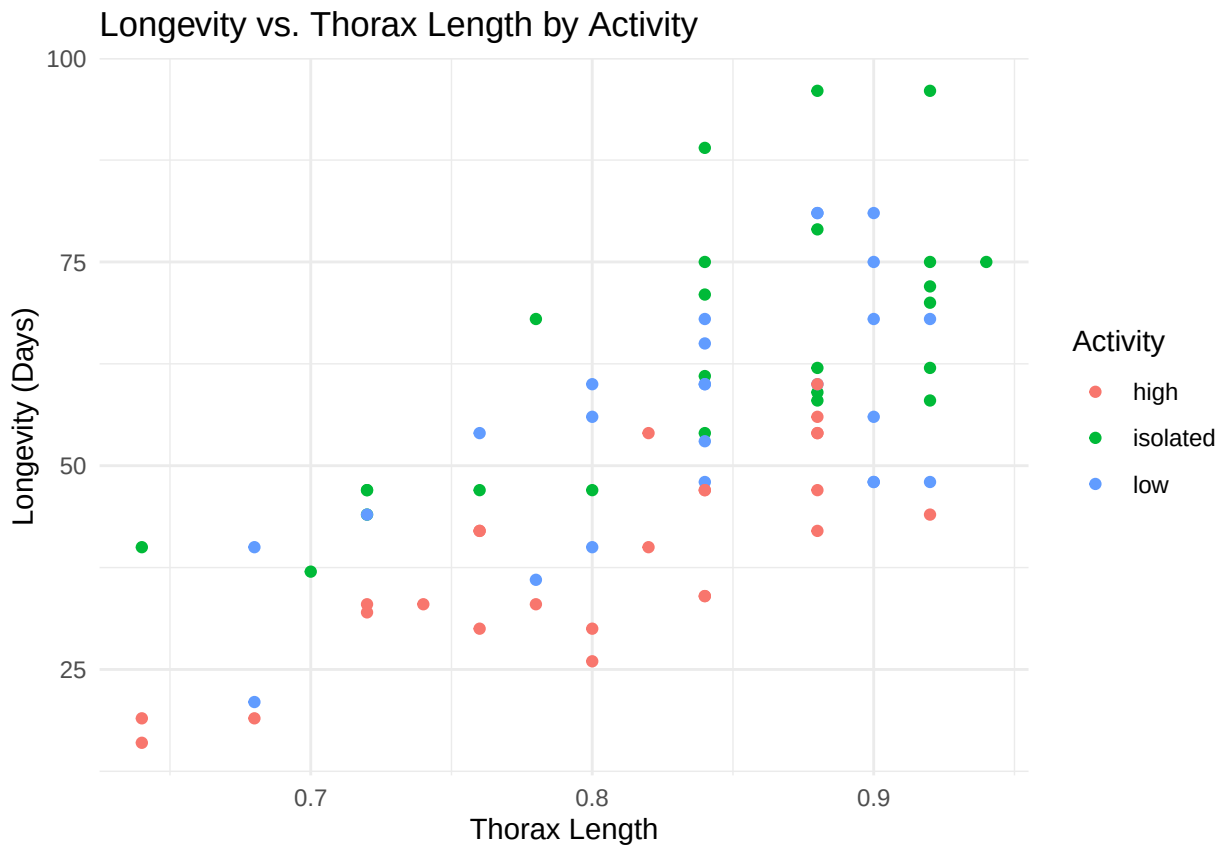
```
##
## Call:
## lm(formula = longevity ~ activity * thorax, data = fruitflies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.0489  -8.3834  -0.7693   7.0877  26.4504
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -61.280     22.071  -2.777  0.00707 **
## activityisolated          11.038     30.772   0.359  0.72091
## activitylow                3.288     33.817   0.097  0.92283
## thorax                   125.000     27.462   4.552 2.22e-05 ***
## activityisolated:thorax   11.127     37.492   0.297  0.76753
## activitylow:thorax        12.001     41.031   0.292  0.77079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 10.54 on 69 degrees of freedom
## Multiple R-squared:  0.6754, Adjusted R-squared:  0.6519
## F-statistic: 28.71 on 5 and 69 DF,  p-value: 1.261e-15
```

d)

```r
# Load necessary packages
library(ggplot2)

# Scatterplot of longevity against thorax length, colored by activity
ggplot(fruitflies, aes(x = thorax, y = longevity, color = activity)) +
  geom_point() +
  labs(title = "Longevity vs. Thorax Length by Activity",
       x = "Thorax Length",
       y = "Longevity (Days)",
       color = "Activity") +
  theme_minimal()
```




```r
# Test for the similarity of dependence using ANCOVA
model_c <- lm(longevity ~ activity * thorax, data = fruitflies)
summary(model_c)
```

```
##
## Call:
## lm(formula = longevity ~ activity * thorax, data = fruitflies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.0489  -8.3834  -0.7693   7.0877  26.4504
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -61.280     22.071  -2.777  0.00707 **
## activityisolated          11.038     30.772   0.359  0.72091
## activitylow                3.288     33.817   0.097  0.92283
## thorax                   125.000     27.462   4.552 2.22e-05 ***
## activityisolated:thorax   11.127     37.492   0.297  0.76753
## activitylow:thorax        12.001     41.031   0.292  0.77079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.54 on 69 degrees of freedom
## Multiple R-squared:  0.6754, Adjusted R-squared:  0.6519
## F-statistic: 28.71 on 5 and 69 DF,  p-value: 1.261e-15
```

e)

```r
# Fit ANCOVA model
ancova_model <- lm(longevity ~ activity + thorax, data = fruitflies)

# Summary of ANCOVA model
summary(ancova_model)
```

```
##
## Call:
## lm(formula = longevity ~ activity + thorax, data = fruitflies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.688  -8.622  -1.176   6.790  26.605
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -67.375     12.750  -5.284 1.33e-06 ***
## activityisolated   20.066      2.994   6.701 4.13e-09 ***
## activitylow        13.054      2.999   4.352 4.43e-05 ***
## thorax            132.618     15.725   8.434 2.62e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 10.4 on 71 degrees of freedom
## Multiple R-squared:  0.6749, Adjusted R-squared:  0.6611
## F-statistic: 49.12 on 3 and 71 DF,  p-value: < 2.2e-16
```

Exercise 2. Birthweights

a)

```r
# Load the data
birthweights <- read.table("Birthweight.csv", header = TRUE, sep=",")
```

Exercise 3. School awards

a)

```r
# Load the data
awards <- read.table("awards.txt", header = TRUE)
```