

101b_project

Daren Sathasivam

2024-05-27

Load packages needed:

```
library(ggplot2)
library(readxl)
library(DescTools)
library(agricolae)
library(pwr)
```

Read in Data(make sure working directory is correct):

```
data <- read_excel("project.xlsx")
# data <- read.csv('project.csv')
head(data)

## # A tibble: 6 x 13
##   `Participant Name`     Age Sex  Location (Island & House Number) Date
##   <chr>      <dbl> <chr> <chr>                <dttm>
## 1 Sophia Carlsen        63   F    Hayarano - 95      2024-05-21 00:00:00
## 2 Kyosuke Endo         23   M    Hayarano - 396     2024-05-21 00:00:00
## 3 Husaam Ahmad          78   M    Hayarano - 418     2024-05-21 00:00:00
## 4 RODGER AITKEN         71   M    Akkeshi - 31      2024-05-21 00:00:00
## 5 DR KRISTINA BLOMGREN 51   F    Akkeshi - 18      2024-05-21 00:00:00
## 6 Alexander Wilson      26   M    Akkeshi - 407     2024-05-21 00:00:00
## # i abbreviated name: 1: `Location (Island & House Number)`
## # i 8 more variables: `Time Pre-Jump` <dttm>,
## #   `Pre-Jump Adrenaline (pg/mL)` <dbl>, `Time Jumped (sec)` <dbl>,
## #   `Time Post-Jumping` <dttm>, `Post-Jump Adrenaline (pg/mL)` <dbl>,
## #   `Change in Blood Adrenaline Levels` <dbl>,
## #   `Percent Change of Blood Adrenaline Levels` <dbl>, `Additional Notes` <chr>
```

Create Model:

```
# Define response variable, treatments, and blocks
response <- data$`Change in Blood Adrenaline Levels`
treatments <- as.factor(data$`Time Jumped (sec)`)
age <- (data$Age)
gender <- (data$Sex)
head(treatments)

## [1] 30 30 30 30 30 30
```

```

## Levels: 30 60 90 120
# Simple model
model1 <- lm(response ~ treatments + age + gender)
summary(model1)

##
## Call:
## lm(formula = response ~ treatments + age + gender)
##
## Residuals:
##       Min     1Q Median     3Q    Max 
## -3153.01 -334.62   43.96  440.58  953.68 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3161.970   189.864  16.654 <2e-16 ***
## treatments60 2098.330   185.094  11.337 <2e-16 ***
## treatments90 4067.552   187.381  21.707 <2e-16 ***
## treatments120 5581.360   186.472  29.931 <2e-16 ***
## age          1.816     2.907   0.625   0.533    
## genderM      -1796.291   132.883 -13.518 <2e-16 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 679.2 on 102 degrees of freedom
## Multiple R-squared:  0.9298, Adjusted R-squared:  0.9264 
## F-statistic: 270.3 on 5 and 102 DF,  p-value: < 2.2e-16

# ANOVA
model1_aov <- aov(response ~ treatments + age + gender)
summary(model1_aov)

##           Df Sum Sq Mean Sq F value Pr(>F)    
## treatments   3 539006855 179668952 389.508 <2e-16 ***
## age          1  20067   20067   0.044  0.835  
## gender       1  84288805  84288805 182.731 <2e-16 *** 
## Residuals  102 47049720   461272                        
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
coef(model1_aov)

## (Intercept) treatments60 treatments90 treatments120      age
## 3161.96990  2098.32996  4067.55243  5581.36023  1.81645
## genderM
## -1796.29070

# Model Equation Model without factor
treatments_nofac <- data$`Time Jumped (sec)`
model1_nofac <- lm(response ~ treatments_nofac + age + gender)
model1_nofac_aov <- aov(response ~ treatments_nofac + age + gender)
summary(model1_nofac)

##
## Call:
## lm(formula = response ~ treatments_nofac + age + gender)

```

```

## 
## Residuals:
##   Min     1Q Median     3Q    Max 
## -3311.3 -301.0  162.8  511.0  889.4 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1397.295   215.318   6.489 2.98e-09 ***
## treatments_nofac  62.350    2.004  31.116 < 2e-16 ***
## age          2.453    2.908   0.844   0.401    
## genderM      -1800.411   134.825 -13.354 < 2e-16 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 689.7 on 104 degrees of freedom 
## Multiple R-squared:  0.9262, Adjusted R-squared:  0.9241 
## F-statistic:  435 on 3 and 104 DF,  p-value: < 2.2e-16 

summary(model1_nofac_aov)

##           Df  Sum Sq Mean Sq F value Pr(>F)    
## treatments_nofac 1 536049452 536049452 1126.775 <2e-16 ***
## age            1  5199   5199   0.011  0.917    
## gender          1 84834063 84834063  178.321 <2e-16 *** 
## Residuals      104 49476733 475738                
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

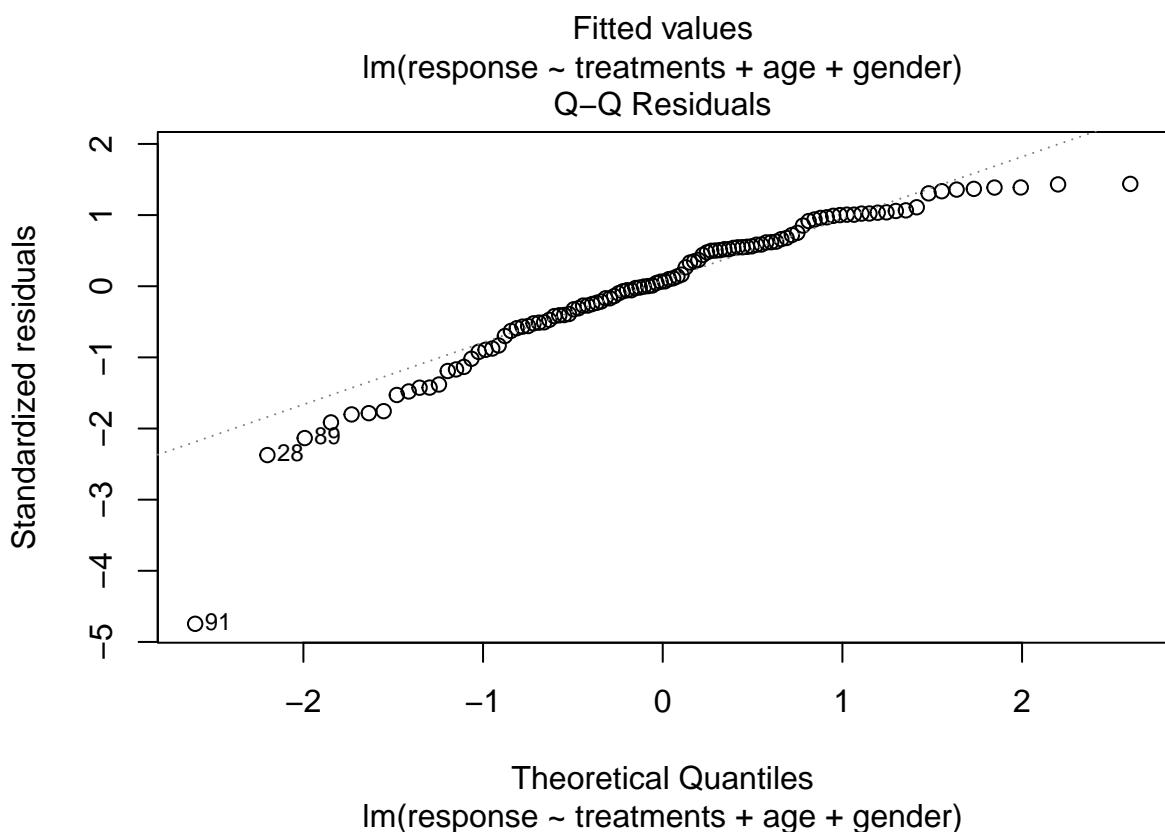
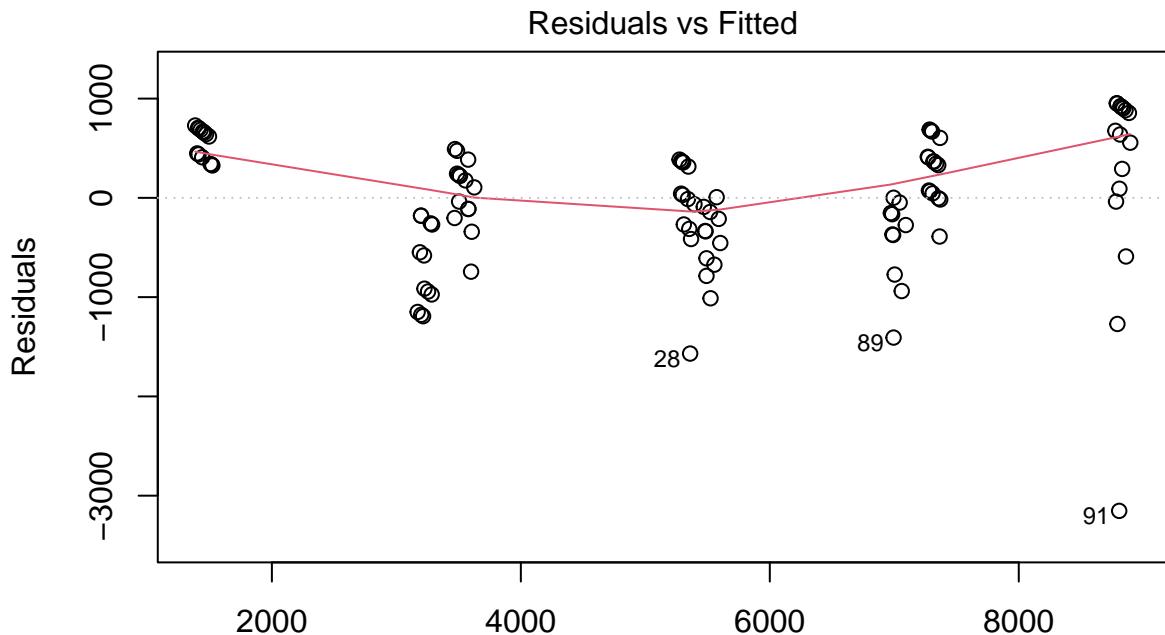
model_no_baseline <- lm(response ~ 0 + treatments_nofac + age + 
  gender, data = data)
# Calculate the mean coefficient for treatments
treatment_coeffs <- coef(model_no_baseline)[grep("treatments",
  names(coef(model_no_baseline)))]
mean_treatment_effect <- mean(treatment_coeffs)
# Construct the model equation including the average
# treatment effect
model_coeffs <- coef(model_no_baseline)
coeff_names <- names(model_coeffs)
# Initialize the equation with the intercept (manually
# calculated or assumed here)
equation <- "y = 1397.295"
equation <- sprintf("%s + %.2f*Treatment", equation, mean_treatment_effect)
for (name in coeff_names[!grepl("treatments", coeff_names)]) {
  sign <- ifelse(model_coeffs[name] >= 0, "+", "-")
  equation <- sprintf("%s %s %.2f*%s", equation, sign, abs(model_coeffs[name]),
    name)
}
equation <- sprintf("%s + e", equation)
equation

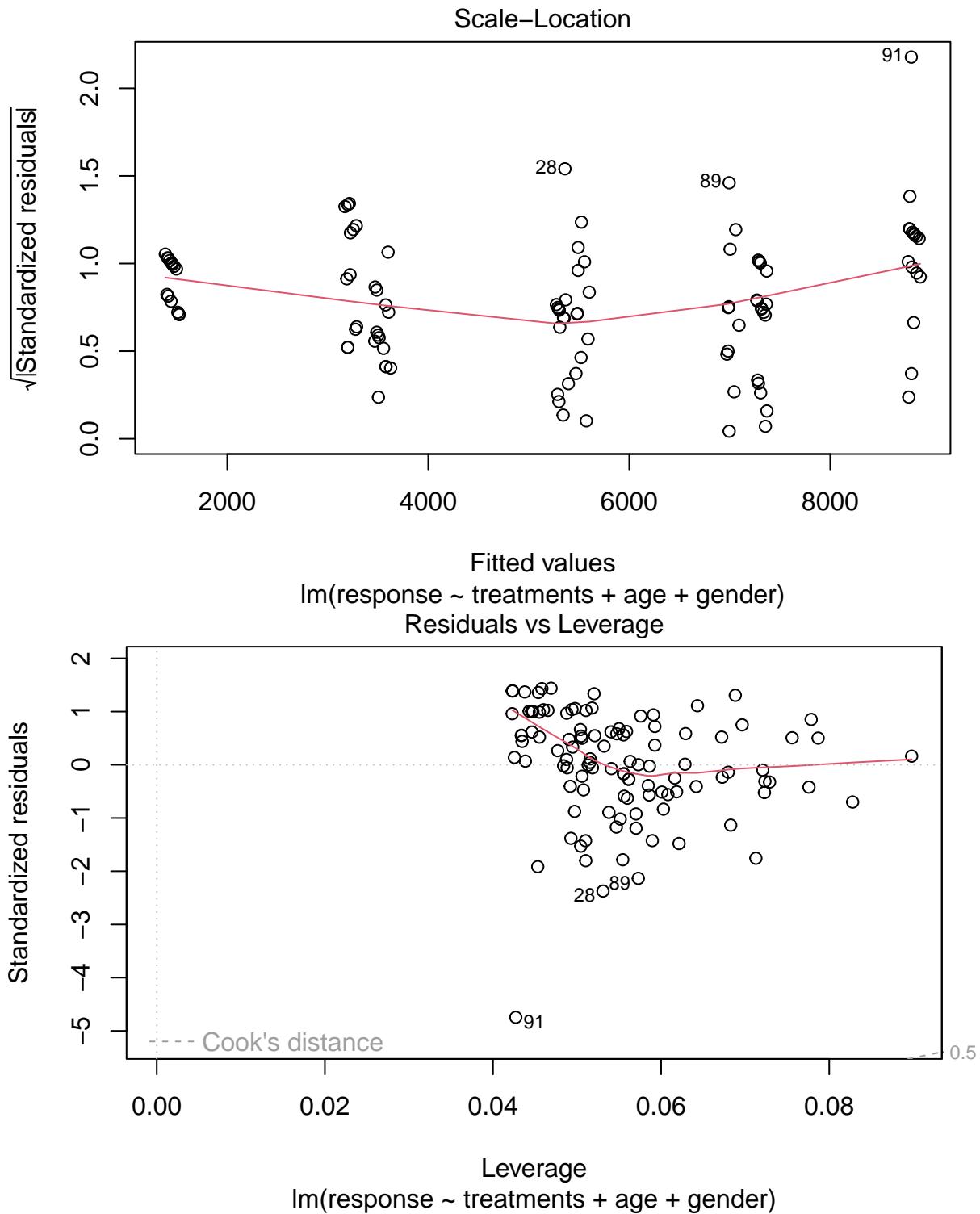
## [1] "y = 1397.295 + 62.35*Treatment + 2.45*age + 1397.29*genderF - 403.12*genderM + e"

```

Plot Models:

```
# Basic plots  
plot(model1)
```





```
# Post-hoc test
tukey1 <- TukeyHSD(model1_aov, "treatments")

## Warning in replications(paste("~", xx), data = mf): non-factors ignored: age
```

```

tukey1

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = response ~ treatments + age + gender)
##
## $treatments
##      diff     lwr      upr p adj
## 60-30 2159.544 1676.745 2642.344    0
## 90-30 4347.663 3864.863 4830.463    0
## 120-30 5912.863 5430.063 6395.663    0
## 90-60 2188.119 1705.319 2670.918    0
## 120-60 3753.319 3270.519 4236.118    0
## 120-90 1565.200 1082.400 2048.000    0

posthoc_lsd <- LSD.test(model1_aov, "treatments", p.adj = "none")
posthoc_lsd

## $statistics
##   MSerror Df      Mean       CV t.value      LSD
##   461271.8 102 5344.599 12.70759 1.983495 366.6424
##
## $parameters
##      test p.adjusted      name.t ntr alpha
## Fisher-LSD      none treatments     4  0.05
##
## $means
##      response      std r      se      LCL      UCL      Min      Max      Q25
## 120 8152.444 1530.2463 27 130.7063 7893.189 8411.700 5586.9 9741.4 6818.60
## 30  2239.581 390.7401 27 130.7063 1980.326 2498.837 1848.2 3019.2 2022.05
## 60  4399.126 946.2791 27 130.7063 4139.871 4658.381 2856.9 5660.9 3731.05
## 90  6587.244 1289.3710 27 130.7063 6327.989 6846.500 4512.4 7971.9 5262.35
##      Q50      Q75
## 120 8744.1 9740.35
## 30  2113.1 2310.90
## 60  3961.9 5329.95
## 90  7352.1 7682.70
##
## $comparison
## NULL
##
## $groups
##      response groups
## 120 8152.444     a
## 90  6587.244     b
## 60  4399.126     c
## 30  2239.581     d
##
## attr(,"class")
## [1] "group"

```

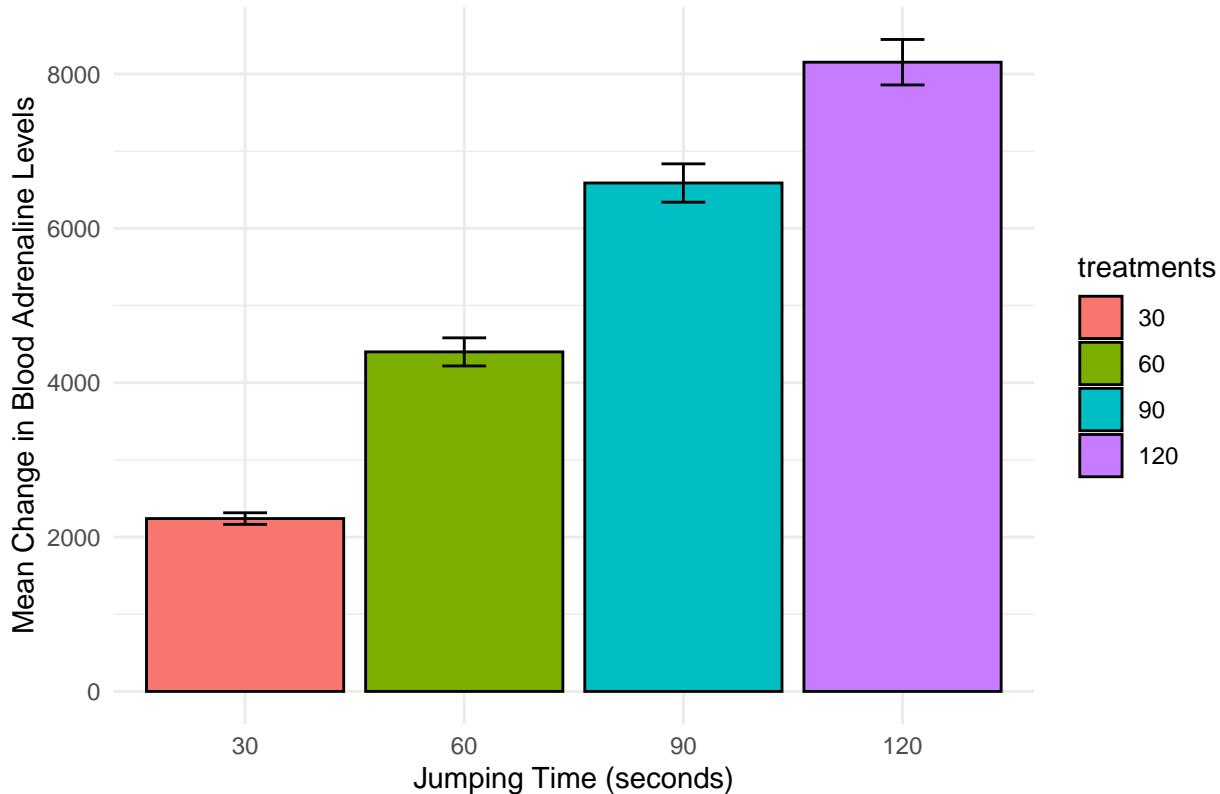
Other Plots:

```
# Calculate means and confidence intervals for plotting
treatment_means <- aggregate(response ~ treatments, data = data,
                                mean)
treatment_se <- aggregate(response ~ treatments, data = data,
                           function(x) sd(x)/sqrt(length(x)))

# Merge means and standard errors for plotting
plot_data <- merge(treatment_means, treatment_se, by = "treatments")
names(plot_data) <- c("treatments", "mean", "se")

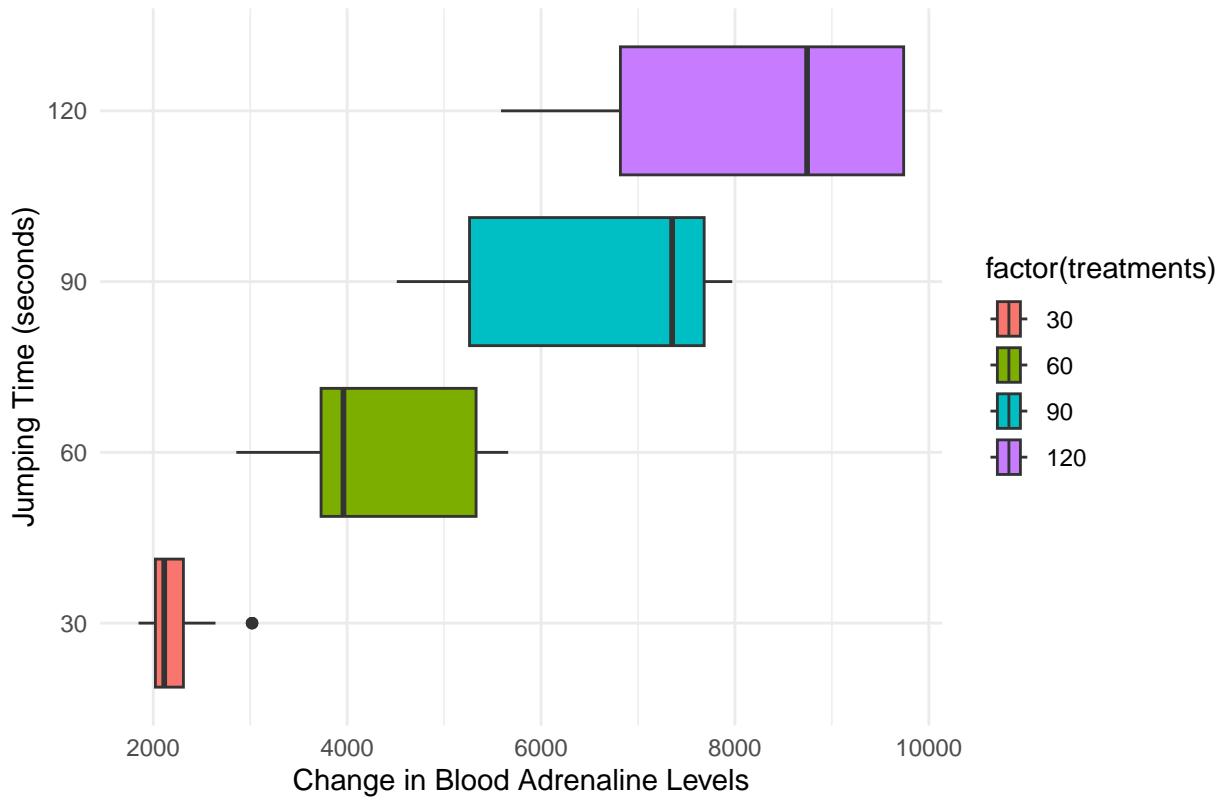
# Bar Graph
ggplot(plot_data, aes(x = treatments, y = mean, fill = treatments)) +
  geom_bar(stat = "identity", position = position_dodge(),
           color = "black") + geom_errorbar(aes(ymax = mean + se,
                                               ymin = mean - se),
                                         width = 0.2, position = position_dodge(0.9)) +
  labs(title = "Effect of Jumping Time on Blood Adrenaline Levels",
       x = "Jumping Time (seconds)", y = "Mean Change in Blood Adrenaline Levels") +
  theme_minimal()
```

Effect of Jumping Time on Blood Adrenaline Levels



```
# Boxplot
ggplot(data, aes(y = factor(treatments), x = response, fill = factor(treatments))) +
  geom_boxplot() + labs(title = "Adrenaline Level Changes Across Jumping Times",
                        y = "Jumping Time (seconds)", x = "Change in Blood Adrenaline Levels") +
  theme_minimal()
```

Adrenaline Level Changes Across Jumping Times



Power Analysis:

```
summary(model1_aov)

##                               Df   Sum Sq   Mean Sq F value Pr(>F)
## treatments            3 539006855 179668952 389.508 <2e-16 ***
## age                   1    20067    20067   0.044  0.835
## gender                1  84288805  84288805 182.731 <2e-16 ***
## Residuals          102  47049720   461272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model1)

##
## Call:
## lm(formula = response ~ treatments + age + gender)
##
## Residuals:
##      Min       1Q     Median       3Q      Max 
## -3153.01  -334.62     43.96   440.58  953.68 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3161.970   189.864  16.654 <2e-16 ***
## treatments60 2098.330   185.094  11.337 <2e-16 ***
```

```

## treatments90   4067.552    187.381   21.707   <2e-16 ***
## treatments120  5581.360    186.472   29.931   <2e-16 ***
## age            1.816      2.907     0.625     0.533
## genderM        -1796.291   132.883  -13.518   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 679.2 on 102 degrees of freedom
## Multiple R-squared:  0.9298, Adjusted R-squared:  0.9264
## F-statistic: 270.3 on 5 and 102 DF,  p-value: < 2.2e-16

# R^2 from model summary
R_squared <- 0.9298
k <- 4 # Number of predictors
n <- length(response) / k # 27
# Effect size f:
group_means <- c(8152.444, 6587.244, 4399.126, 2239.581) # Group means obtained from Fisher LSD
d <- max(group_means) - min(group_means) # 5912.863
sigma <- sqrt(461272) # MSE from ANOVA for model1_aov (679.1701)
f0 <- d / sigma # From CH3 Slide 43/44
f0 # 8.706012

## [1] 8.706012
# Adjusted d so effect size is not too large
d_set <- 1500
f <- d_set / sigma # 2.208578 with max distance set as 1500

# Power analysis for ANOVA
power_analysis <- pwr.anova.test(k = k, # Number of predictors
                                    n = n, # Degrees of freedom of residuals
                                    f = f0,
                                    sig.level = 0.05,
                                    power = NULL)
power_analysis

##
##      Balanced one-way analysis of variance power calculation
##
##      k = 4
##      n = 27
##      f = 8.706012
##      sig.level = 0.05
##      power = 1
##
## NOTE: n is number in each group
# Calculate required sample size for a power of 0.95
sample_size_analysis <- pwr.anova.test(k = k,
                                         f = f,
                                         power = 0.95)
sample_size_analysis

##
##      Balanced one-way analysis of variance power calculation
##
##      k = 4

```

```

##          n = 2.182995
##          f = 2.208578
##      sig.level = 0.05
##      power = 0.95
##
## NOTE: n is number in each group
a <- read.csv("project.csv")
head(a)

##      Participant.Name Age Sex Location..Island...House.Number.           Date
## 1      Sophia Carlsen  63   F          Hayarano - 95 May 21, 2024
## 2      Kyosuke Endo   23   M          Hayarano - 396 May 21, 2024
## 3      Husaam Ahmad   78   M          Hayarano - 418 May 21, 2024
## 4      RODGER AITKEN  71   M          Akkeshi - 31 May 21, 2024
## 5 DR KRISTINA BLOMGREN 51   F          Akkeshi - 18 May 21, 2024
## 6      Alexander Wilson 26   M          Akkeshi - 407 May 21, 2024
##      Time.Pre.Jump Pre.Jump.Adrenaline..pg.mL. Time.Jumped..sec. Time.Post.Jumping
## 1        11:38                  32.4            30             11:41
## 2        11:43                  41.0            30             11:48
## 3        11:51                  38.5            30             11:53
## 4        11:58                  33.9            30             11:59
## 5        12:04                  41.1            30             12:07
## 6        12:10                  38.4            30             12:15
##      Post.Jump.Adrenaline..pg.mL. Change.in.Blood.Adrenaline.Levels
## 1                    3051.6            3019.2
## 2                    2153.7            2112.7
## 3                    1887.2            1848.7
## 4                    2146.8            2112.9
## 5                    2352.1            2311.0
## 6                    2151.5            2113.1
##      Percent.Change.of.Blood.Adrenaline.Levels
## 1                    9318.519
## 2                    5152.927
## 3                    4801.818
## 4                    6232.743
## 5                    5622.871
## 6                    5502.865
##      Additional.Notes
## 1      smoked a couple times throughout her life
## 2      mostly healthy individual with almost no smoking
## 3      light - moderate smoker for most of life, recently got heart disease
## 4 alternative light & moderate smoker, had and recovered from bowel cancer towards end of life
## 5      light-non smoker for most of life, got liver cancer in recent years
## 6      light - non smoker for most of life

age <- as.factor(a$Age)
gender <- as.factor(a$Sex)
response <- a$Post.Jump.Adrenaline..pg.mL.
treatments <- as.factor(a$Time.Jumped..sec.)

model2 <- aov(response ~ treatments + age + gender)
summary(model2)

##          Df     Sum Sq   Mean Sq F value    Pr(>F)

```

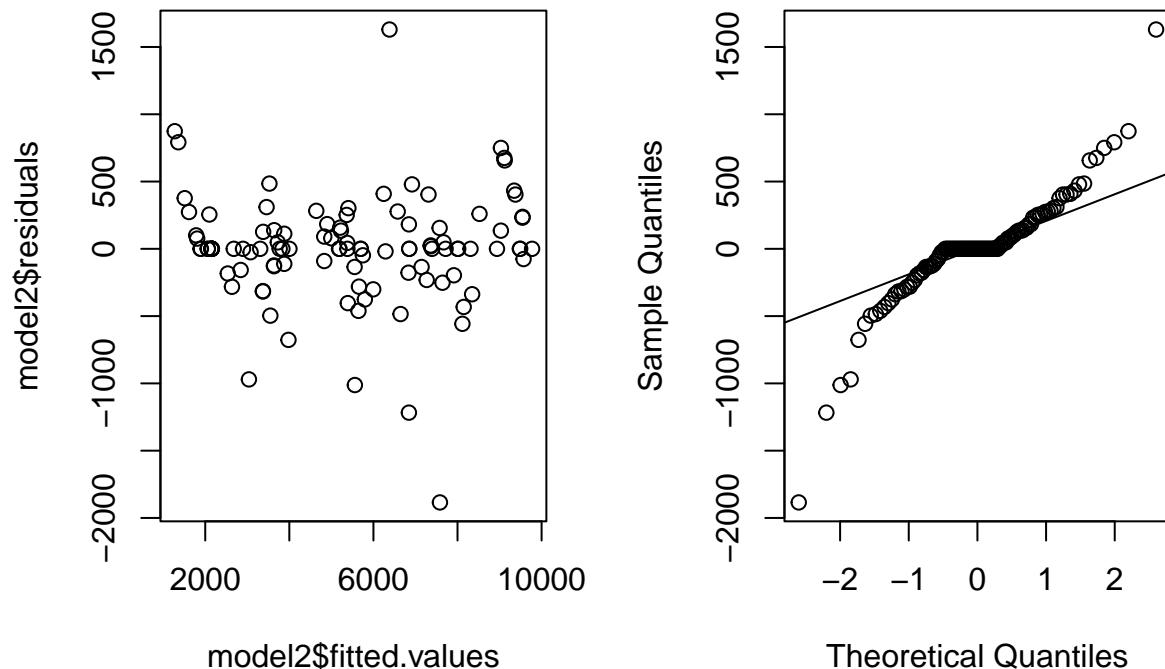
```

## treatments     3 538880120 179626707 438.732 < 2e-16 ***
## age          60 66174542   1102909   2.694 0.000446 ***
## gender        1 47584442  47584442 116.223 8.39e-14 ***
## Residuals    43 17605172   409423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(mfrow = c(1, 2))
plot(model2$fitted.values, model2$residuals)

qqnorm(model2$residuals)
qqline(model2$residuals)

```

Normal Q-Q Plot



```

# PostHocTest(model2, method = 'lsd') TukeyHSD(model2)

treatment_means <- aggregate(response ~ treatments, data = a,
                               mean)
treatment_se <- aggregate(response ~ treatments, data = a, function(x) sd(x)/sqrt(length(x)))

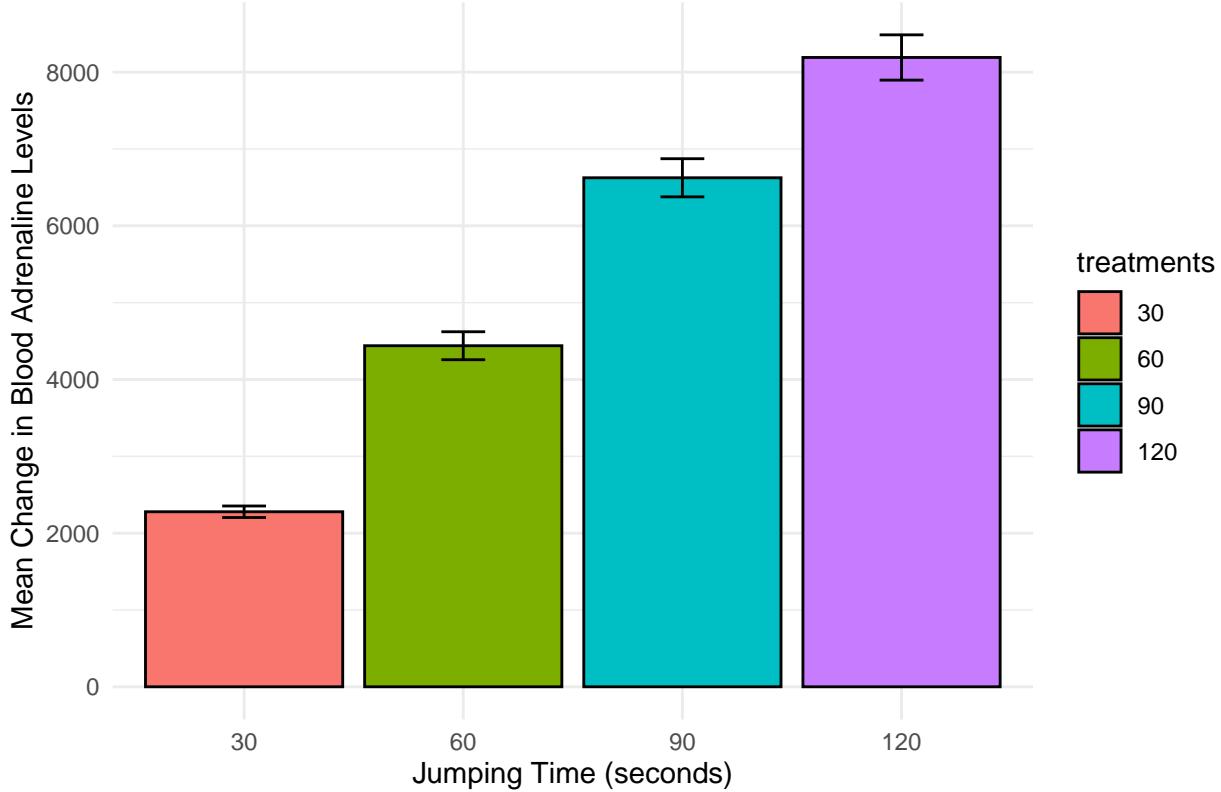
# Merge means and standard errors for plotting
plot_data <- merge(treatment_means, treatment_se, by = "treatments")
names(plot_data) <- c("treatments", "mean", "se")

# Bar Graph Plot for each level of treatment
ggplot(plot_data, aes(x = treatments, y = mean, fill = treatments)) +
  geom_bar(stat = "identity", position = position_dodge(),
           color = "black") + geom_errorbar(aes(ymax = mean + se,
                                               ymin = mean - se),
                                         width = 0.2, position = position_dodge(0.9)) +
  labs(title = "Effect of Jumping Time on Blood Adrenaline Levels",
       x = "Jumping Time (seconds)", y = "Mean Change in Blood Adrenaline Levels") +

```

```
theme_minimal()
```

Effect of Jumping Time on Blood Adrenaline Levels



```
# Boxplot
ggplot(data, aes(y = factor(treatments), x = response, fill = factor(treatments))) +
  geom_boxplot() + labs(title = "Adrenaline Level Changes Across Jumping Times",
  y = "Jumping Time (seconds)", x = "Change in Blood Adrenaline Levels") +
  theme_minimal()
```

Adrenaline Level Changes Across Jumping Times

