ADS_21

2045

5 30 10:58:42 2024

1. Neural activity patterning and osmotic challenge

1.1 Import and organise the data set.

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## -- 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 errors
```

```
vaso_data<-read.csv("D:/ /ADS2_mockexam/vasotypes.csv")
str(vaso_data)
```

```
## 'data.frame':    166 obs. of  3 variables:
##  $ Cell    : chr  "n200706" "n200706-2" "n200706-3" "n200706-4" ...
##  $ Protocol: chr  "control" "control" "control" "control" ...
##  $ Fit.Type: chr  "All" "HAP" "All" "DAP" ...
```

```
anyNA(vaso_data)
```

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

```
anyDuplicated(vaso_data)
```

```
## [1] 0
```

```
# Create a contingency table
#
formatted_vaso <- vaso_data %>%
  #
```

```
  select(Protocol, Fit.Type) %>%
  #
  group_by(Protocol, Fit.Type) %>%
  summarize(count = n()) %>%
  #
  spread(key = Fit.Type, value = count)
```

```
## `summarise()` has grouped output by 'Protocol'. You can override using the
## `.groups` argument.
```

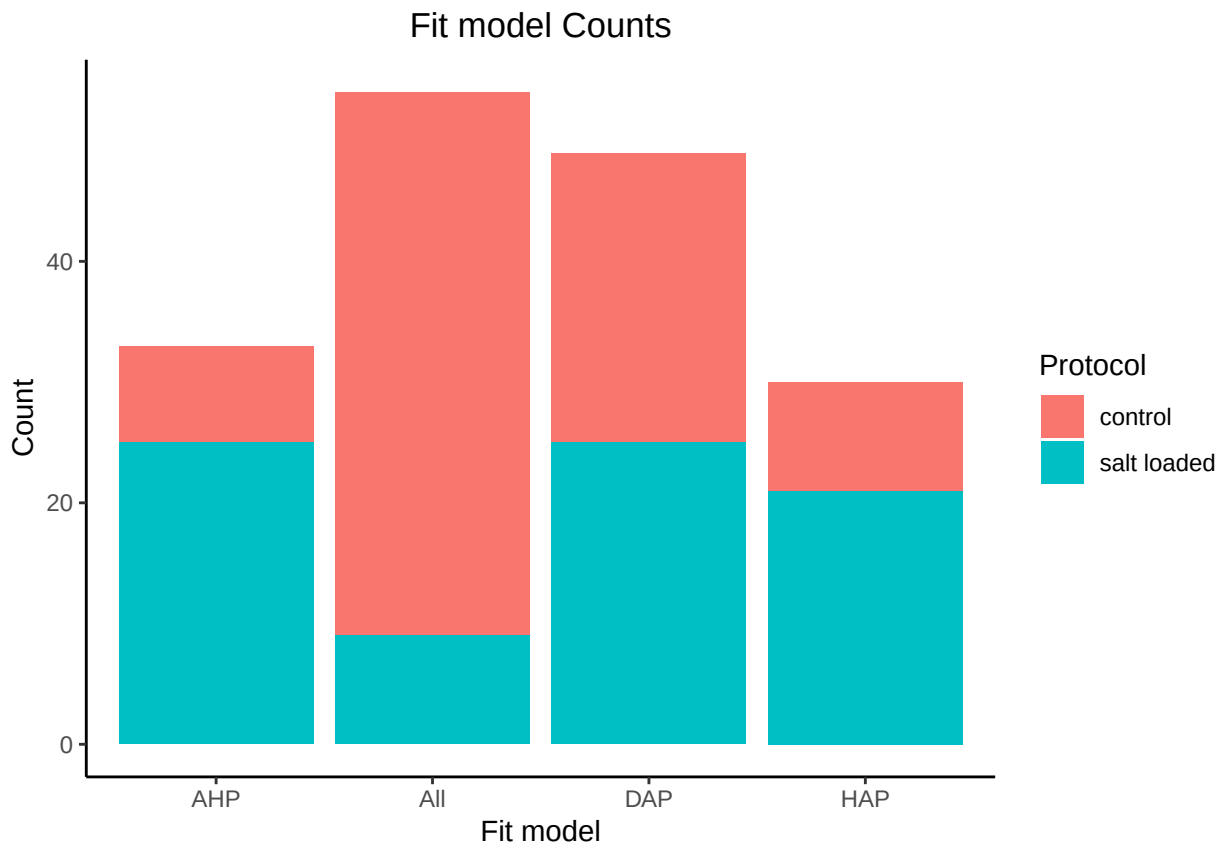
```
formatted_vaso
```

```
## # A tibble: 2 x 5
## # Groups:   Protocol [2]
##   Protocol      AHP   All   DAP   HAP
##   <chr>       <int> <int> <int> <int>
## 1 control         8    45    24     9
## 2 salt loaded    25     9    25    21
```

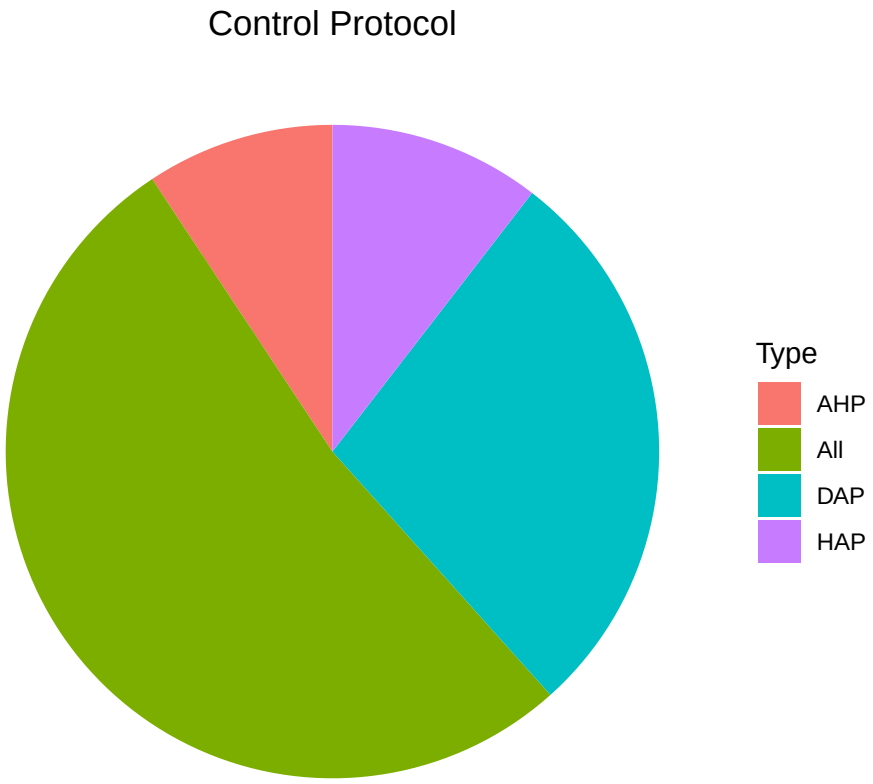
1.2 Including Plots

You can also embed plots, for example:

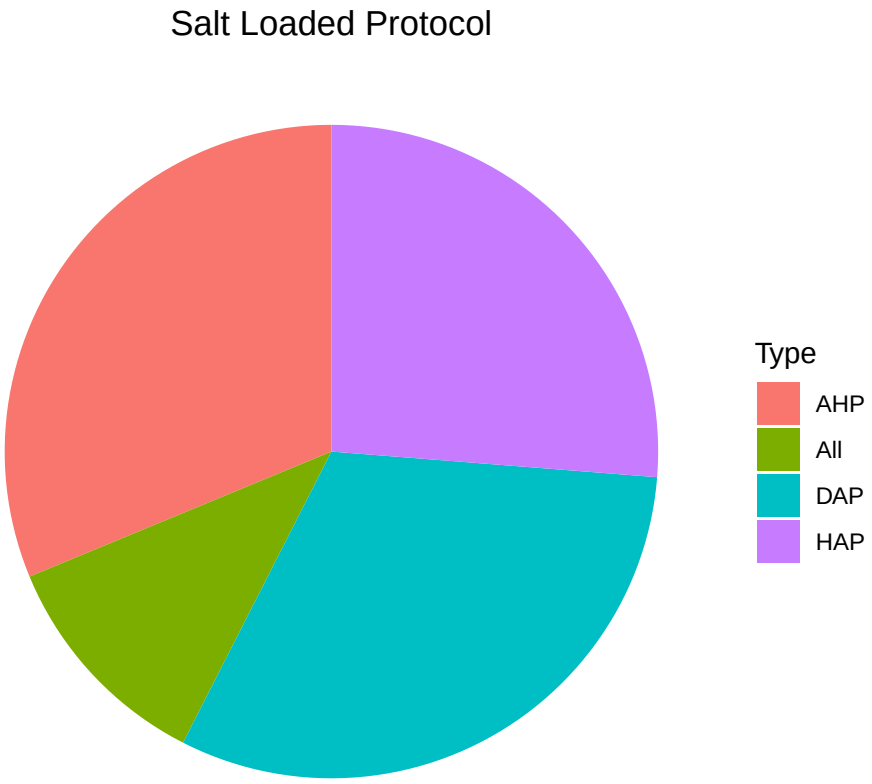
```
# Plotting the data
ggplot(vaso_data, aes(x = Fit.Type, fill = Protocol)) +
  geom_bar(position = "stack") +
  labs(title = "Fit model Counts",
       x = "Fit model",
       y = "Count",
       fill = "Protocol") +
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))
```




```
formatted_vaso_long <- formatted_vaso %>%
  pivot_longer(cols = -Protocol, names_to = "Type", values_to = "Count")

#   control
control_data <- formatted_vaso_long %>% filter(Protocol == "control")

ggplot(control_data, aes(x = "", y = Count, fill = Type)) +
  geom_bar(stat = "identity",width = 1) +
  coord_polar("y") +
  labs(title = "Control Protocol", fill = "Type") +
  theme_void()+
  theme(plot.title = element_text(hjust = 0.5))
```




```
#   salt loaded
salt_loaded_data <- formatted_vaso_long %>% filter(Protocol == "salt loaded")

ggplot(salt_loaded_data, aes(x = "", y = Count, fill = Type)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Salt Loaded Protocol", fill = "Type") +
  theme_void()+
  theme(plot.title = element_text(hjust = 0.5))
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

1.3 Choose, justify, state the statistical hypotheses, and carry out an appropriate test.

We'll use a chi-squared test to determine if there is a significant difference in the distribution of best fit models between control and salt-loaded neurons.

Hypotheses:

Null Hypothesis (H0): There is no difference in the distribution of best fit models between control and salt-loaded neurons. Alternative Hypothesis (H1): There is a difference in the distribution of best fit models between control and salt-loaded neurons.

```
# Create a contingency table
contingency_table <- table(vaso_data$Protocol, vaso_data$Fit.Type)

# Perform the chi-squared test
chi_squared_test <- chisq.test(contingency_table)

# Display the results
chi_squared_test
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  contingency_table
## X-squared = 37.41, df = 3, p-value = 3.768e-08
```

1.4 Discuss the result, have the neurons' intrinsic properties changed? Are there any problems or limitations with the current study?

Chi-squared statistic: 37.41 P-value: 3.768e-08 There is a significant difference in the distribution of best fit models between control and salt-loaded neurons. This suggests that the osmotic challenge affects the spiking properties of the neurons. Discussion

Limitations: Potential limitations include the sample size and variability in experimental conditions. Further studies could include larger sample sizes and more controlled conditions to verify these findings. 2. The interpretation of the model fitting results requires careful consideration. Each combination of afterpotentials (HAP, DAP, AHP) may represent different aspects of neuronal physiology, and their relative contributions to the observed spiking patterns need to be interpreted cautiously. 3. The study's focus on a prolonged osmotic challenge raises questions about the long-term effects of salt loading on neuronal activity. It's crucial to consider whether any observed changes are transient or persist over time. 3. Covariates and Confounding Factors: It's essential to consider whether any covariates or confounding factors might influence the results. Factors such as the age of the animals, their health status, or other experimental conditions could affect the neurons' spiking properties independently of the osmotic challenge.

EXAM1: covariates 1. there is no control group (it would not be possible to add sham-operated persons anyway); 2. the information about the patient is incomplete as it does not include details about their clinical picture and possible important clinical covariates (age, the location of the trauma, their detailed case description upon admission, etc). You may comment about the need to add other clinical covariates to predict the clinical outcome. You may mention that the condition of several persons not only did not improve but even deteriorated further, and this issue must be addressed.

2. Myoinositol and gestational diabetes mellitus

Your colleagues wonder whether the suggested treatments improve insulin resistance and prevent excessive weight gain in pregnant women. You may compare the difference in the analyzed characteristics between both time points.

2.1 Import the data and "clean" it.

```
# Import data from all sites
files <- list.files(path = "D:/ /ADS2_mockexam/", pattern = "site_.*.csv",
                    full.names = TRUE)
data_list <- lapply(files, read.csv)#

# Combine data into one dataframe
combined_data <- bind_rows(data_list)


anyNA(combined_data)
```

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

```r
str(combined_data)
```

```
## 'data.frame':    80 obs. of  7 variables:
##  $ X                : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ID               : int  41 68 44 14 37 7 10 46 12 72 ...
##  $ Site             : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Patients         : chr  "Myoinositol" "Chiroinositol" "control" "Myo/Chiroinositol" ...
##  $ Insulin_free_days: int  207 209 201 233 218 195 198 246 195 223 ...
##  $ Weight_beginning : num  68.9 76 87.3 70.3 103.1 ...
##  $ Weight_delivery  : num  79.6 86.4 101.3 79.3 113.2 ...
```

2.2 Choose the appropriate method of statistical analysis, check assumptions for this analysis, and explain your choice briefly.

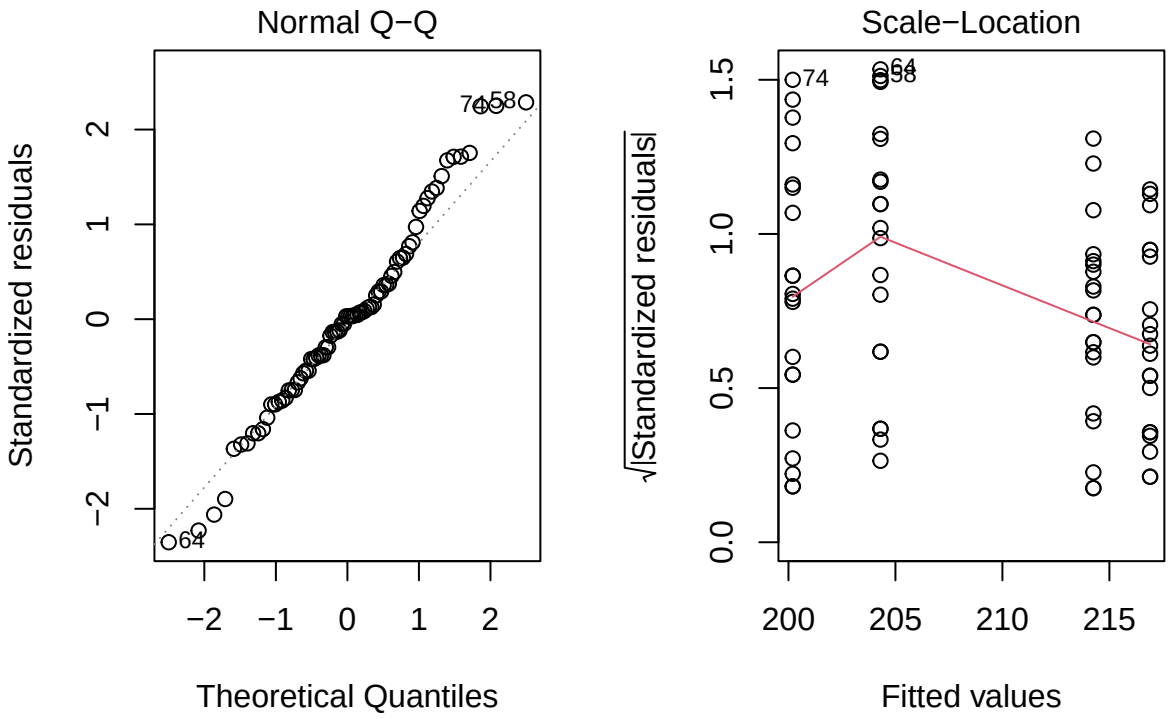
Appropriate Statistical Test: Use ANOVA to compare the means across multiple groups.

Hypothesis Formulation: Null Hypothesis: There is no difference in insulin resistance among the different treatment groups. Alternative Hypothesis: There is a significant difference in insulin resistance among the different treatment groups.

```r
# Convert treatment to factor
combined_data$Patients <- as.factor(combined_data$Patients)

# Perform ANOVA for insulin-free days
anova_insulin <- aov(Insulin_free_days ~ Patients, data = combined_data)
summary(anova_insulin)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Patients     3   3789  1263.1   2.023  0.118
## Residuals   76  47447   624.3
```

```r
par(mfrow = c(1, 2)) # combine plots
# 1. QQ-plot
plot(anova_insulin, which = 2)
# 2. Homogeneity of variances
plot(anova_insulin, which = 3)
```



The left QQ plot shows the residuals are close to normality, since the dots do not offset the dashed line. But the right plot shows variances not homogeneous, since the red line indicating the mean of residuals is not horizontal.

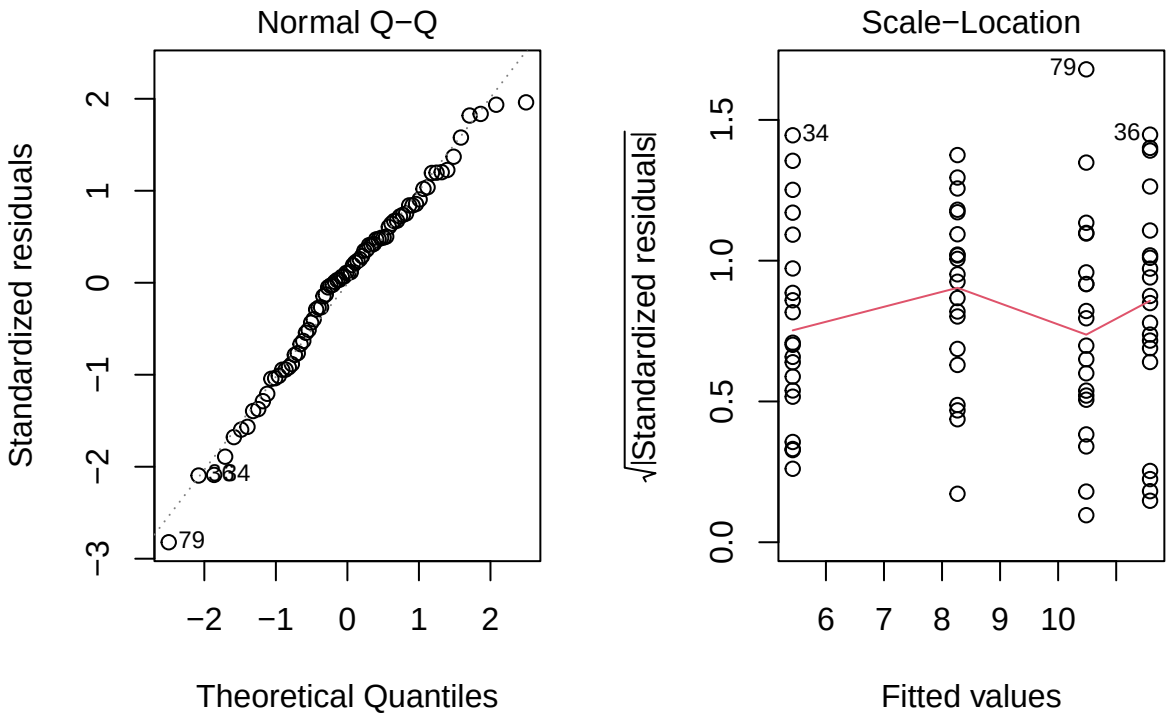
Null Hypothesis: There is no difference in weight gain among the different treatment groups. Alternative Hypothesis: There is a significant difference in weight gain among the different treatment groups.

```
# Perform ANOVA for weight change
combined_data <- combined_data %>%
  mutate(weight_change = Weight_delivery - Weight_beginning)

anova_weight <- aov(weight_change ~ Patients, data = combined_data)
summary(anova_weight)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Patients     3  443.2  147.75    9.48 2.15e-05 ***
## Residuals   76 1184.5   15.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The normality of residuals and the homogeneity of variances are tested

```
par(mfrow = c(1, 2)) # combine plots
# 1. QQ-plot
plot(anova_weight, which = 2)
# 2. Homogeneity of variances
plot(anova_weight, which = 3)
```



The left QQ plot shows the residuals are close to normality, since the dots do not offset the dashed line. But the right plot shows variances not homogeneous, since the red line indicating the mean of residuals is not horizontal.

2.3 Formulate the correct statistical hypotheses, conduct the analysis, identify the effect of treatment, and show which factor has a higher effect on the tested characteristics. Plot your data.

We find that the residuals are normal, while the variances are not equal in both insulin resistance and weight gain datasets. So the Welch'one way ANOVA should be used.

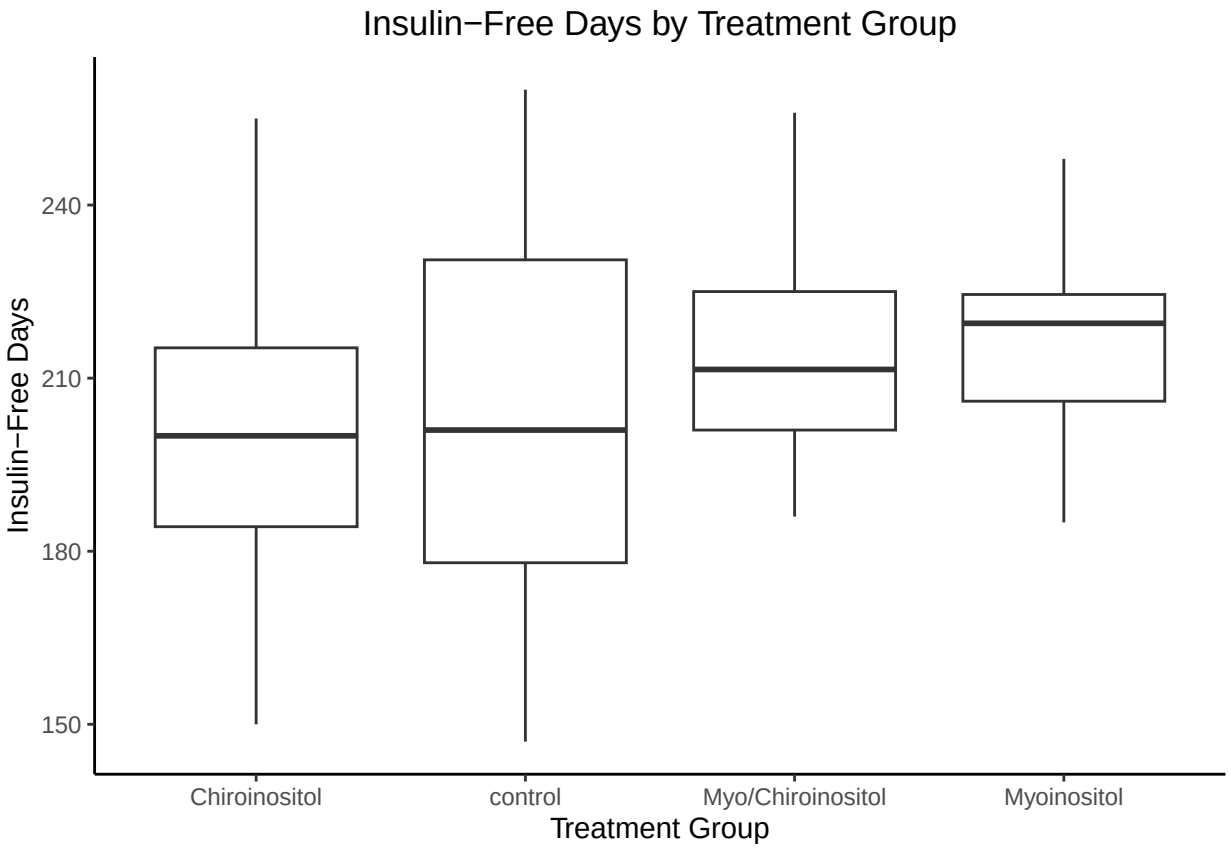
For insulin resistance: Null Hypothesis: There is no difference in insulin resistance among the different treatment groups. Alternative Hypothesis: There is a significant difference in insulin resistance among the different treatment groups.

For weight gain: Null Hypothesis: There is no difference in weight gain among the different treatment groups. Alternative Hypothesis: There is a significant difference in weight gain among the different treatment groups.

```
oneway.test(Insulin_free_days ~ Patients,data = combined_data,var.equal = F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  Insulin_free_days and Patients
## F = 2.2469, num df = 3.00, denom df = 40.99, p-value = 0.09726
```

```r
ggplot(combined_data, aes(x = Patients, y = Insulin_free_days)) +
  geom_boxplot() +
  labs(title = "Insulin-Free Days by Treatment Group", x = "Treatment Group",
       y = "Insulin-Free Days")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))
```

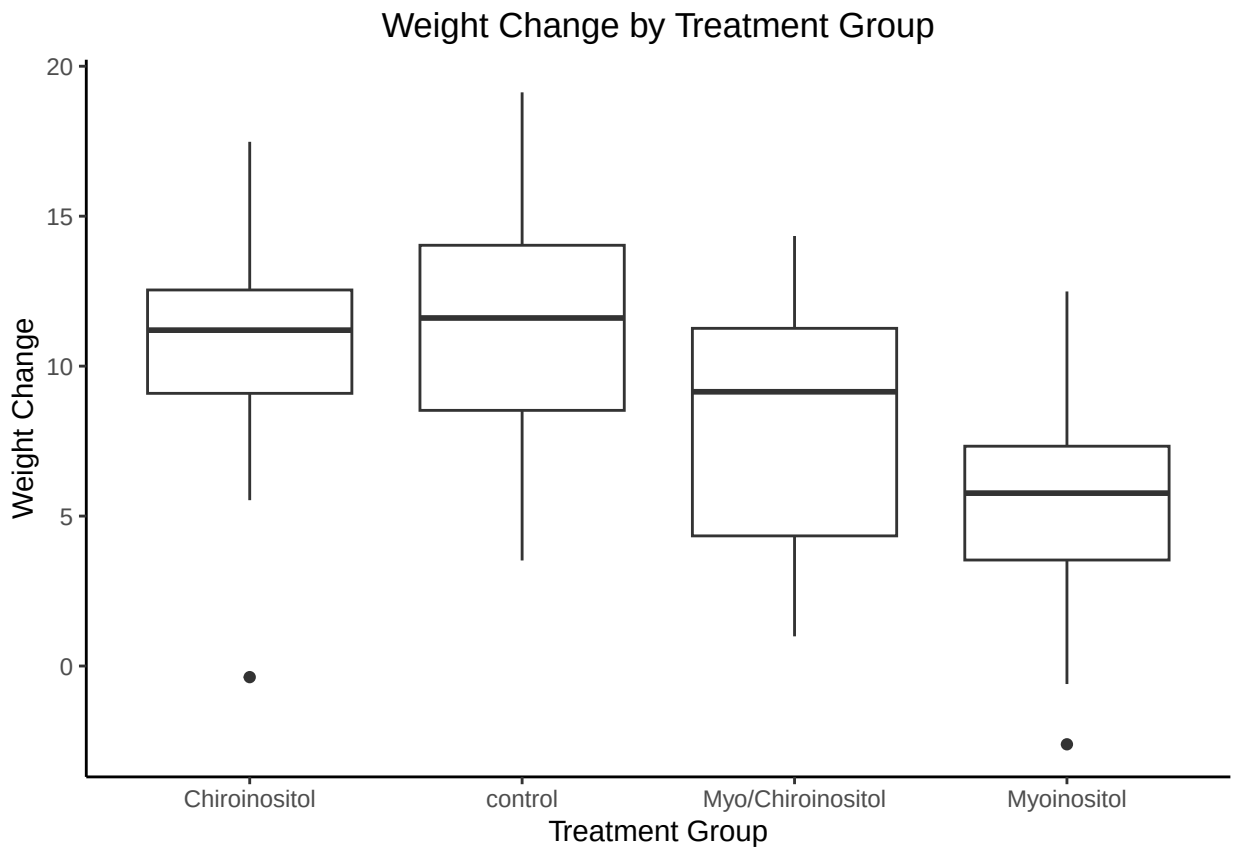


The Welch's ANOVA shows p-value = 0.09726, which indicates no difference in insulin resistance among the different treatment groups. And the Post-hoc test is not required.

```r
oneway.test(weight_change ~ Patients,data = combined_data,var.equal = F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  weight_change and Patients
## F = 9.6821, num df = 3.000, denom df = 42.177, p-value = 5.525e-05
```

```r
ggplot(combined_data, aes(x = Patients, y = weight_change)) +
  geom_boxplot() +
  labs(title = "Weight Change by Treatment Group", x = "Treatment Group",
       y = "Weight Change")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5))
```



The Welch's ANOVA shows p-value = 5.525e-05, which indicates significant difference in weight gain among the different treatment groups. And the Post-hoc test (Tukey HSD tests) is required.

```
library(multcomp)
```

```
##      mvtnorm

##      survival

##      TH.data

##      MASS

##
##    'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
##    'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

```r
post_test_weight <- glht(anova_weight,linfct = mcp(Patients = "Tukey"))
summary(post_test_weight)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = weight_change ~ Patients, data = combined_data)
##
## Linear Hypotheses:
##                                        Estimate Std. Error t value Pr(>|t|)
## control - Chiroinositol == 0              1.099      1.248   0.880    0.815
## Myo/Chiroinositol - Chiroinositol == 0   -2.219      1.248  -1.777    0.292
## Myoinositol - Chiroinositol == 0         -5.057      1.248  -4.050   <0.001 ***
## Myo/Chiroinositol - control == 0         -3.318      1.248  -2.657    0.046 *
## Myoinositol - control == 0               -6.156      1.248  -4.931   <0.001 ***
## Myoinositol - Myo/Chiroinositol == 0     -2.838      1.248  -2.273    0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

The Tukey HSD test shows p value <0.001 when comparing Myoinositol group with control and Chiroinositol group and p value $=0.0464 < 0.05$ when comparing Myo/Chiroinositol group with control. And considering the boxplots, this indicates inhibitory effect on weight gain of Myoinositol and Myo/Chiroinositol treatment. The Myoinositol treatment itself has the highest level of inhibitory effects on the weight gain.

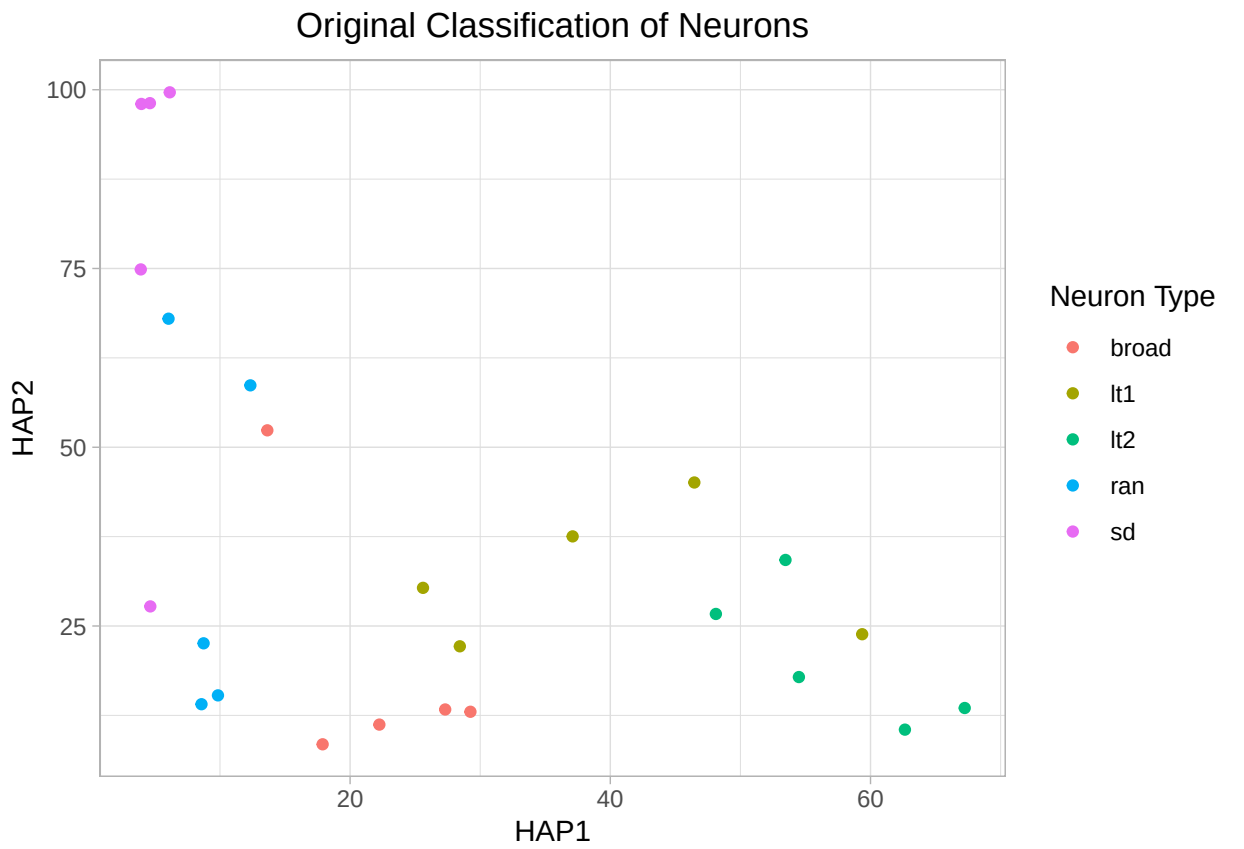
2.4 What could be done in the future to follow up on this study? Give some suggestions to the researchers.

Additional measurements such as blood glucose levels and HbA1c to better understand the impact of treatments on gestational diabetes. More information about the patients like the ages can be further collected and researched.

3. Classifying neuron types from electrophysiological recordings

3.1 Import the original data and plot it in a useful format to show the original classifications (type)

```r
vmndata<-read.csv("D:/ /ADS2_mockexam/vmndata.csv")

# Plot original data
ggplot(vmndata, aes(x = hap1, y = hap2, color = as.factor(type))) +
  geom_point() +
  labs(title = "Original Classification of Neurons", x = "HAP1", y = "HAP2",
       color = "Neuron Type")+
  theme_light()+
  theme(plot.title = element_text(hjust = 0.5))
```



3.2 Use clustering of the model fit data to make your own classification of the recordings (You are allowed to use the kmeans function in R)

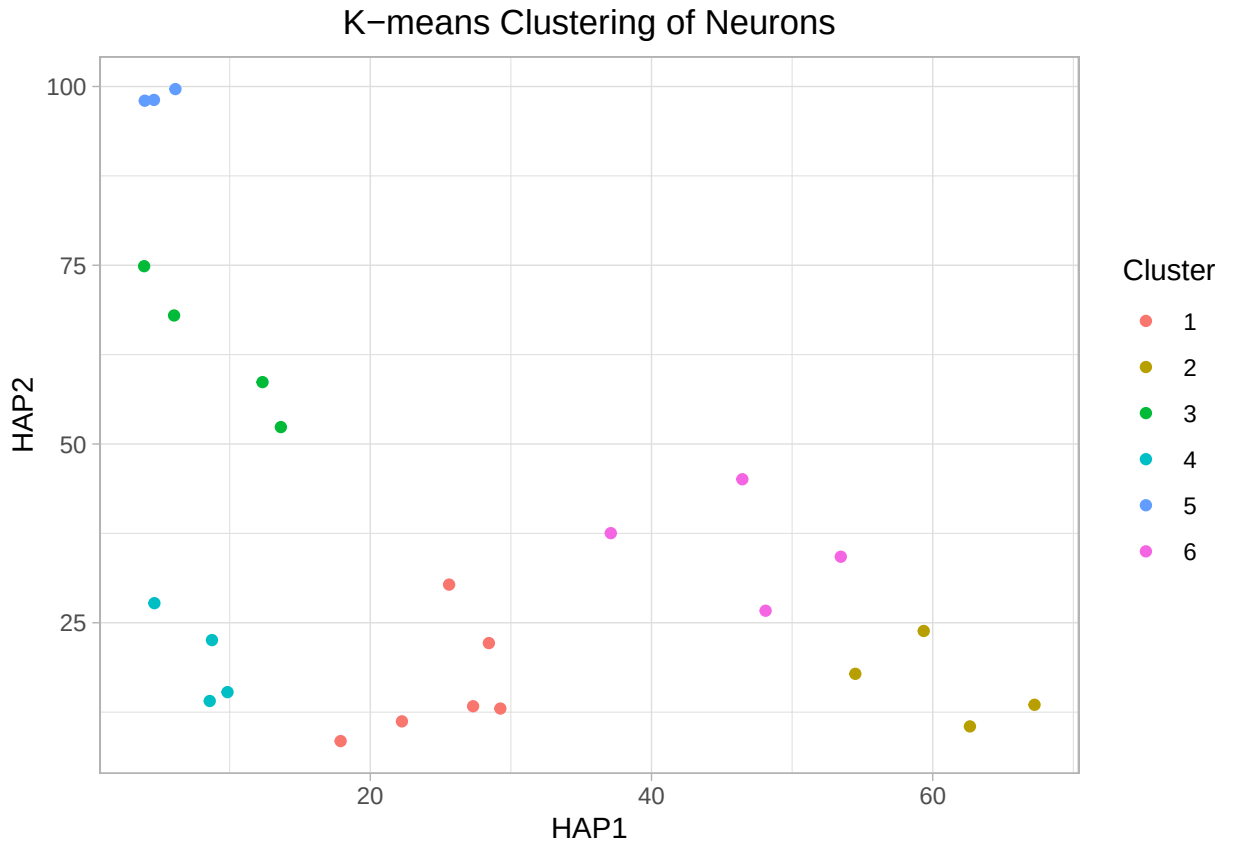
```
# Perform k-means clustering
set.seed(123)
kmeans_result <- kmeans(vmndata[, c("hap1", "hap2")], centers = 6)

# Add cluster result to the data
vmndata$cluster <- as.factor(kmeans_result$cluster)
str(vmndata)
```

```
## 'data.frame':    25 obs. of  4 variables:
##  $ type   : chr  "ran" "ran" "ran" "ran" ...
##  $ hap1   : num  8.58 6.04 8.74 9.85 12.34 ...
##  $ hap2   : num  14.1 68 22.6 15.3 58.7 ...
##  $ cluster: Factor w/ 6 levels "1","2","3","4",..: 4 3 4 4 3 5 5 5 3 4 ...
```

3.3 Plot the outcome of this clustering (e.g. by assigning colours by cluster).

```
# Plot clustering result
ggplot(vmndata, aes(x = hap1, y = hap2, color = cluster)) +
  geom_point() +
  labs(title = "K-means Clustering of Neurons", x = "HAP1", y = "HAP2",
       color = "Cluster")+
```

```
  theme_light()+
  theme(plot.title = element_text(hjust = 0.5))
```



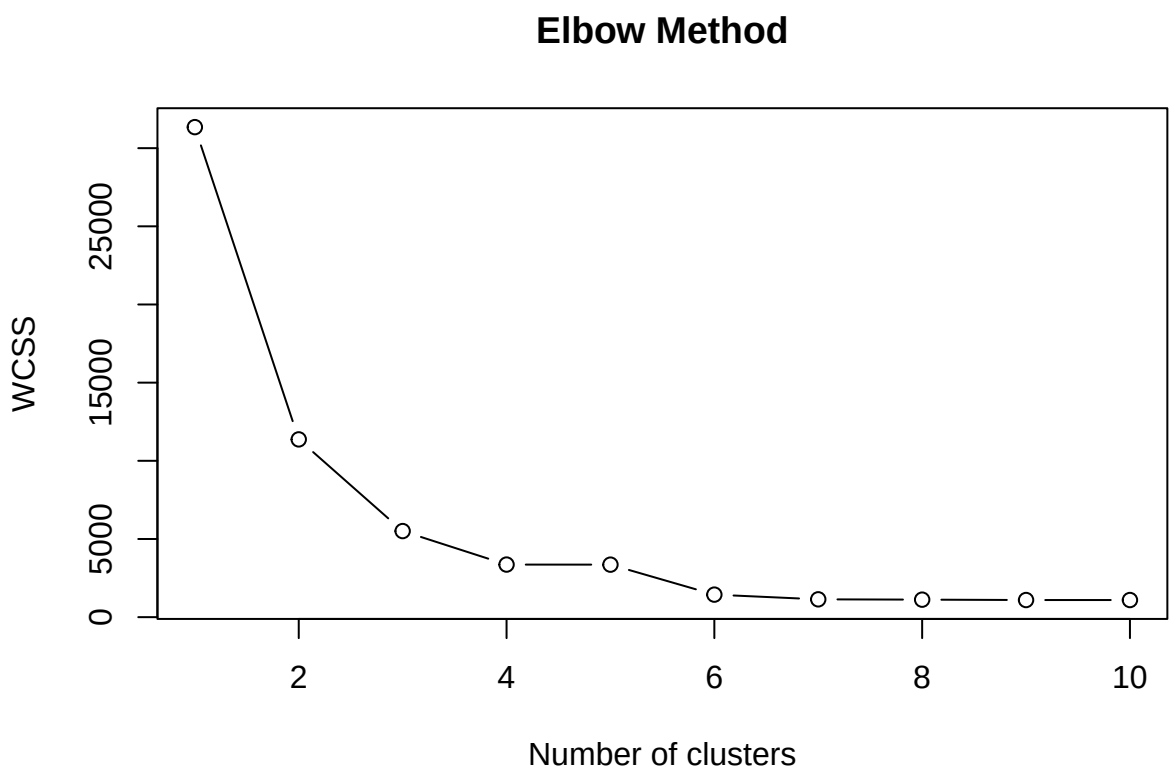

```
# # cluster the data
# h_cluster <- hclust(dist(data[, 2:3]))
# # plot the data
# plot(h_cluster, xlab = "Guest names", labels = data$names)
```

3.4 Test the clustering using different subsets of the fit parameters. Describe in words how your clustering compares to the original classifications

Elbow Method : k K k Within-Cluster Sum of Squares, WCSS WCSS WCSS WCSS is the sum of squares of the distance between the samples in the cluster and the cluster center, and a smaller WCSS value usually indicates a better clustering effect. WCSS The cluster number is selected at the point where the WCSS decline rate decreases sharply

```
#
data_matrix <- vmndata[, c("hap1", "hap2")]

#     WCSS
wcss <- sapply(1:10, function(k) {
  km <- kmeans(data_matrix, centers = k)
  km$tot.withinss
})
```

```
#  WCSS
plot(1:10, wcss, type = "b", xlab = "Number of clusters",
     ylab = "WCSS", main = "Elbow Method")
```




```
drug_df<-read.csv("D:/ /ADS_provide/short/t1d_drug.csv")
which(is.na(drug_df$Glucose))
```

```
## [1] 32 37 38 39 40 47 48 52 53
```

```
drug_df_clean<-drug_df[-c(which(is.na(drug_df$Glucose))-30,which(is.na(drug_df$Glucose))),]
drug_df_clean$Measurement<-factor(drug_df_clean$Measurement,levels=c("Glucose_before","Glucose_after"))
drug_df_clean$Treatment<-factor(drug_df_clean$Treatment,levels=c("Vehicle", "1 mg/ml", "5 mg/ml"))


#library(ggplot2)
# create plot using ggplot() and geom_boxplot() functions
ggplot(drug_df_clean, aes(Treatment, Glucose, fill=Measurement)) +
geom_boxplot()+
# geom_point() is used to make points at data values
geom_jitter(aes(y=Glucose),
              size = 2, shape = 21,
              color="black",
              stroke = 0.15, show.legend = FALSE,
              position = position_jitterdodge(jitter.height=0.5,
```

```
                                          jitter.width = 0.1,
                                          dodge.width = 0.8))+
  #geom_line(aes(group = ID), color = "black", alpha = 0.5) +
theme_classic()
```