

Assignment03

Chakrit Jiarasatit

2024-11-10

The ensuing case study deals with the relation of ADP-induced platelet aggregation levels and Clopidogrel resistance, according to three genetic markers: rs4244285 (CYP2C192), rs4986893 (CYP2C193), and rs662 (PON1 192Q>R). These analyses are adjusted for age and sex as covariates. Given that ADP levels are continuous, linear regression was used, whereas logistic regression models were applied to drug resistance given its binary nature. PlateletHW.tsv: The following are 11 variables in the dataset: IID, ADP, Resistance, rs4244285, rs4986893, rs662, AGE, SEX, PON1 192Q>R, CYP2C192, CYP2C193. Of the above variables in the dataset, the sex and ADP are continuous variables, while the rest are categorical. Rs4244285 SNP is coded as 0 = GG, 1 = AG, and 2 = AA and is related to CYP2C192. SNPs rs4986893 (CYP2C193) and rs662 (PON1 192Q>R) are coded 0 = AA, 1 = AG and 2 = GG. Sex is coded 0 = male and 1 = female, whereas Drug Resistance is 0 = not resistant, and 1 = resistant.

1. Load necessary libraries

```
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

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

library(e1071)
```

2. Load the data

Load the data from a TSV (Tab-Separated Values) file into a data frame named data

```
data <- read_tsv("Raw_data/PlateletHW.tsv")

## Rows: 211 Columns: 11
## -- Column specification -----
## Delimiter: "\t"
## chr (3): PON1.192Q>R, CYP2C19*2, CYP2C19*3
## dbl (8): IID, ADP, Resistance, rs4244285, rs4986893, rs662, AGE, SEX
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3. Data Cleaning

Clean the data by taking the absolute value of ADP

```
data <- data %>% mutate(ADP = abs(ADP))
```

4. Checking for Skewness and Outliers

- Calculate skewness to check the asymmetry of the data distribution.
- Calculate quartiles and Interquartile Range (IQR) to measure data spread.
- Determine outlier boundaries using IQR method and filter out the outliers.

Calculate Skewness

```
adp_skewness <- skewness(data$ADP)
```

Calculate Quartiles and IQR

```
Q1 <- quantile(data$ADP, 0.25)
Q3 <- quantile(data$ADP, 0.75)
IQR <- Q3 - Q1
```

Determine Outlier Boundaries

```
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR
```

Filter Outliers

```
outliers <- data %>% filter(ADP < lower_bound | ADP > upper_bound)
cat("Skewness of ADP:", adp_skewness, "\n")
```

```
## Skewness of ADP: 0.6072203
```

```
cat("Number of outliers:", nrow(outliers), "\n")
```

```
## Number of outliers: 0
```

- The skewness of the ADP data is 0.6072203, indicating a slight right skew. This means that the data distribution has a longer tail on the right side. Additionally, there are no outliers in the dataset, as indicated by an outlier count of 0.

Write a new clean data

```
write_tsv(data, "Clean_data/PlateletHW_clean.tsv")
```

5. Normalizing ADP by Taking the Logarithm

- Normalize the ADP values by taking their logarithms to make the data distribution more symmetric.

```
data <- data %>% mutate(ADP_log = log(ADP))
```

6. Creating Linear Regression Models

- Create linear regression models to analyze the relationship between $\log(\text{ADP})$ and various SNPs (rs4244285, rs4986893, and rs662).

```
model1 <- lm(ADP_log ~ rs4244285, data = data)
model2 <- lm(ADP_log ~ rs4986893, data = data)
model3 <- lm(ADP_log ~ rs662, data = data)
```

7. Summarizing the Models

- Display the summary of each model, including coefficients, standard errors, t-values, p-values, R-squared values, and residuals.

Summary of Linear Regression Model 1: $\text{ADP_log} \sim \text{rs4244285}$

```
summary(model1)

##
## Call:
## lm(formula = ADP_log ~ rs4244285, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4317 -0.5467 -0.0235  0.8128  1.4120
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.18940    0.07873   40.51  < 2e-16 ***
## rs4244285     0.35644    0.09530    3.74 0.000238 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8651 on 209 degrees of freedom
## Multiple R-squared:  0.06273,    Adjusted R-squared:  0.05825
## F-statistic: 13.99 on 1 and 209 DF,  p-value: 0.0002375
```

- This linear regression model shows a significant positive relationship between the log of ADP values and the SNP rs4244285. Specifically, for each unit increase in rs4244285, log(ADP) increases by 0.356 units. The model is R-squared value indicates the proportion of variance in log(ADP) explained by rs4244285.

Summary of Linear Regression Model 2: ADP_log ~ rs4986893

```
summary(model2)
```

```
##
## Call:
## lm(formula = ADP_log ~ rs4986893, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5803 -0.6227 -0.0348  0.8844  1.2972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.33804    0.06257  53.353  < 2e-16 ***
## rs4986893    0.66218    0.24289   2.726  0.00695 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8781 on 209 degrees of freedom
## Multiple R-squared:  0.03434,    Adjusted R-squared:  0.02972
## F-statistic: 7.433 on 1 and 209 DF,  p-value: 0.006949
```

- This linear regression model shows a significant positive relationship between the log of ADP values and the SNP rs4986893. Specifically, for each unit increase in rs4986893, log(ADP) increases by 0.662 units. The model is R-squared value would indicate the proportion of variance in log(ADP) explained by rs4986893.

Summary of Linear Regression Model 3: ADP_log ~ rs662

```
summary(model3)
```

```
##
## Call:
## lm(formula = ADP_log ~ rs662, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6073 -0.6565 -0.0787  0.9437  1.2492
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.32192    0.14211  23.376  <2e-16 ***
## rs662        0.04310    0.09195   0.469    0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8932 on 209 degrees of freedom
## Multiple R-squared:  0.00105,    Adjusted R-squared:  -0.003729
```

```
## F-statistic: 0.2197 on 1 and 209 DF, p-value: 0.6397
```

- This linear regression model shows a non-significant relationship between the log of ADP values and the SNP rs662. Specifically, for each unit increase in rs662, log(ADP) increases by 0.0431 units. The model is R-squared value would indicate the proportion of variance in log(ADP) explained by rs662.

8. Creating Logistic Regression Models

- Create logistic regression models to analyze the relationship between drug resistance (Resistance) and various SNPs.

```
logit_model1 <- glm(Resistance ~ rs4244285, data = data, family = binomial)
logit_model2 <- glm(Resistance ~ rs4986893, data = data, family = binomial)
logit_model3 <- glm(Resistance ~ rs662, data = data, family = binomial)
```

9. Summarizing the Logistic Regression Models

- Display the summary of each logistic regression model, including coefficients, standard errors, z-values, p-values, and deviance.

Summary of Logistic Regression Model 1: Resistance ~ rs4244285

```
summary(logit_model1)

##
## Call:
## glm(formula = Resistance ~ rs4244285, family = binomial, data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2583     0.2165  -5.813 6.13e-09 ***
## rs4244285      0.9433     0.2436   3.872 0.000108 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 252.16  on 209  degrees of freedom
## AIC: 256.16
##
## Number of Fisher Scoring iterations: 4
```

- This model shows that there is a strong positive link between drug resistance and the SNP rs4244285. In simple terms, as rs4244285 increases, the likelihood of drug resistance also increases. For each unit increase in rs4244285, the odds of drug resistance go up by 0.943 units.

Summary of Logistic Regression Model 2: Resistance ~ rs4986893

```
summary(logit_model2)
```

```
##
```

```
## Call:
## glm(formula = Resistance ~ rs4986893, family = binomial, data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8256      0.1548  -5.333 9.65e-08 ***
## rs4986893      1.7419      0.6115   2.848 0.00439 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 258.94  on 209  degrees of freedom
## AIC: 262.94
##
## Number of Fisher Scoring iterations: 4
```

- This model shows that there is a strong positive link between drug resistance and the SNP rs4986893. In simple terms, as rs4986893 increases, the likelihood of drug resistance also increases. For each unit increase in rs4986893, the odds of drug resistance go up by 1.742 units.

Summary of Logistic Regression Model 3: Resistance ~ rs662

```
summary(logit_model3)
```

```
##
## Call:
## glm(formula = Resistance ~ rs662, family = binomial, data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.66470    0.33669  -1.974  0.0484 *
## rs662        -0.02556    0.21826  -0.117  0.9068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 268.13  on 209  degrees of freedom
## AIC: 272.13
##
## Number of Fisher Scoring iterations: 4
```

- This model shows that there is a very weak negative link between drug resistance and the SNP rs662. In simpler terms, as rs662 increases, the likelihood of drug resistance slightly decreases. For each unit increase in rs662, the odds of drug resistance go down by 0.0256 units.

10. Creating Adjusted Models with Confounding Factors (AGE and SEX)

- Create adjusted linear and logistic regression models including AGE and SEX as additional predictors to account for confounding factors.

```
model11_adj <- lm(ADP_log ~ rs4244285 + AGE + SEX, data = data)
model12_adj <- lm(ADP_log ~ rs4986893 + AGE + SEX, data = data)
model13_adj <- lm(ADP_log ~ rs662 + AGE + SEX, data = data)

logit_model11_adj <- glm(Resistance ~ rs4244285 + AGE + SEX, data = data, family = binomial)
logit_model12_adj <- glm(Resistance ~ rs4986893 + AGE + SEX, data = data, family = binomial)
logit_model13_adj <- glm(Resistance ~ rs662 + AGE + SEX, data = data, family = binomial)
```

11. Summarizing the Adjusted Models

- Display the summary of each adjusted model.

Summary of Adjusted Linear Regression Model 1: ADP_log ~ rs4244285 + AGE + SEX

```
summary(model11_adj)

##
## Call:
## lm(formula = ADP_log ~ rs4244285 + AGE + SEX, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4329 -0.5847  0.0234  0.7691  1.3790
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.642427   0.369440   9.859  < 2e-16 ***
## rs4244285    0.360830   0.095387   3.783 0.000203 ***
## AGE         -0.006644   0.005629  -1.180 0.239184
## SEX         -0.047234   0.134027  -0.352 0.724883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8652 on 207 degrees of freedom
## Multiple R-squared:  0.07152,    Adjusted R-squared:  0.05806
## F-statistic: 5.315 on 3 and 207 DF,  p-value: 0.001507
```

- This adjusted linear regression model shows a significant positive relationship between log(ADP) and the SNP rs4244285 after controlling for AGE and SEX. Specifically, for each unit increase in rs4244285, log(ADP) increases by 0.36083 units. Additionally, age has a slight negative effect on log(ADP), while being male slightly decreases log(ADP).

Summary of Adjusted Linear Regression Model 2: ADP_log ~ rs4986893 + AGE + SEX

```
summary(model2_adj)

##
## Call:
## lm(formula = ADP_log ~ rs4986893 + AGE + SEX, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5707 -0.5926 -0.0460  0.8435  1.3307
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.787300   0.372706  10.162 < 2e-16 ***
## rs4986893    0.663689   0.243691   2.723  0.00701 **
## AGE         -0.006749   0.005720  -1.180  0.23941
## SEX         -0.006994   0.136411  -0.051  0.95916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.879 on 207 degrees of freedom
## Multiple R-squared:  0.04167,    Adjusted R-squared:  0.02778
## F-statistic:      3 on 3 and 207 DF,  p-value: 0.03161
```

- This adjusted linear regression model shows a significant positive relationship between log(ADP) and the SNP rs4986893 after controlling for AGE and SEX. Specifically, for each unit increase in rs4986893, log(ADP) increases by 0.66369 units. Additionally, age has a slight negative effect on log(ADP), while being male slightly decreases log(ADP).

Summary of Adjusted Linear Regression Model 3: ADP_log ~ rs662 + AGE + SEX

```
summary(model3_adj)

##
## Call:
## lm(formula = ADP_log ~ rs662 + AGE + SEX, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6018 -0.6330 -0.0758  0.9179  1.2493
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.758708   0.392094   9.586 <2e-16 ***
## rs662        0.047704   0.092747   0.514  0.608
```



```
## AGE          -0.006572   0.005838  -1.126    0.262
## SEX          -0.024108   0.139273  -0.173    0.863
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8941 on 207 degrees of freedom
## Multiple R-squared:  0.008599,    Adjusted R-squared:  -0.005769
## F-statistic: 0.5985 on 3 and 207 DF,  p-value: 0.6167
```

- This adjusted linear regression model shows a small, non-significant relationship between $\log(\text{ADP})$ and the SNP rs662 after controlling for AGE and SEX. Specifically, for each unit increase in rs662, $\log(\text{ADP})$ increases by 0.04770 units. Additionally, age has a slight negative effect on $\log(\text{ADP})$, while being male slightly decreases $\log(\text{ADP})$.

Summary of Adjusted Logistic Regression Model 1: Resistance ~ rs4244285 + AGE + SEX

```
summary(logit_model1_adj)

##
## Call:
## glm(formula = Resistance ~ rs4244285 + AGE + SEX, family = binomial,
##      data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.35306    0.94235   0.375  0.7079
## rs4244285    0.97318    0.24693   3.941 8.11e-05 ***
## AGE         -0.02451    0.01463  -1.675  0.0938 .
## SEX         -0.05410    0.34977  -0.155  0.8771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 248.82  on 207  degrees of freedom
## AIC: 256.82
##
## Number of Fisher Scoring iterations: 4
```

- This adjusted logistic regression model shows a significant positive relationship between drug resistance and the SNP rs4244285 after controlling for AGE and SEX. Specifically, for each unit increase in rs4244285, the log-odds of drug resistance increase by 0.9732 units. Age has a slight negative effect on the log-odds of drug resistance, while being male slightly decreases the log-odds of drug resistance.

Summary of Adjusted Logistic Regression Model 2: Resistance ~ rs4986893 + AGE + SEX

```
summary(logit_model2_adj)
```

```
##
## Call:
## glm(formula = Resistance ~ rs4986893 + AGE + SEX, family = binomial,
##      data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.73079    0.92024   0.794  0.42712
## rs4986893    1.78749    0.61864   2.889  0.00386 **
## AGE         -0.02411    0.01437  -1.678  0.09335 .
## SEX          0.08354    0.34338   0.243  0.80779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 255.98  on 207  degrees of freedom
## AIC: 263.98
##
## Number of Fisher Scoring iterations: 4
```

- This adjusted logistic regression model shows a significant positive relationship between drug resistance and the SNP rs4986893 after controlling for AGE and SEX. Specifically, for each unit increase in rs4986893, the log-odds of drug resistance increase by 1.7875 units. Age has a slight negative effect on the log-odds of drug resistance, while being male slightly increases the log-odds of drug resistance.

Summary of Adjusted Logistic Regression Model 3: Resistance ~ rs662 + AGE + SEX

```
summary(logit_model3_adj)

##
## Call:
## glm(formula = Resistance ~ rs662 + AGE + SEX, family = binomial,
##      data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.727160    0.927671   0.784   0.433
## rs662        -0.004631    0.222195  -0.021   0.983
## AGE         -0.021617    0.014006  -1.543   0.123
## SEX          0.005561    0.336527   0.017   0.987
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 268.14  on 210  degrees of freedom
## Residual deviance: 265.46  on 207  degrees of freedom
## AIC: 273.46
##
## Number of Fisher Scoring iterations: 4
```

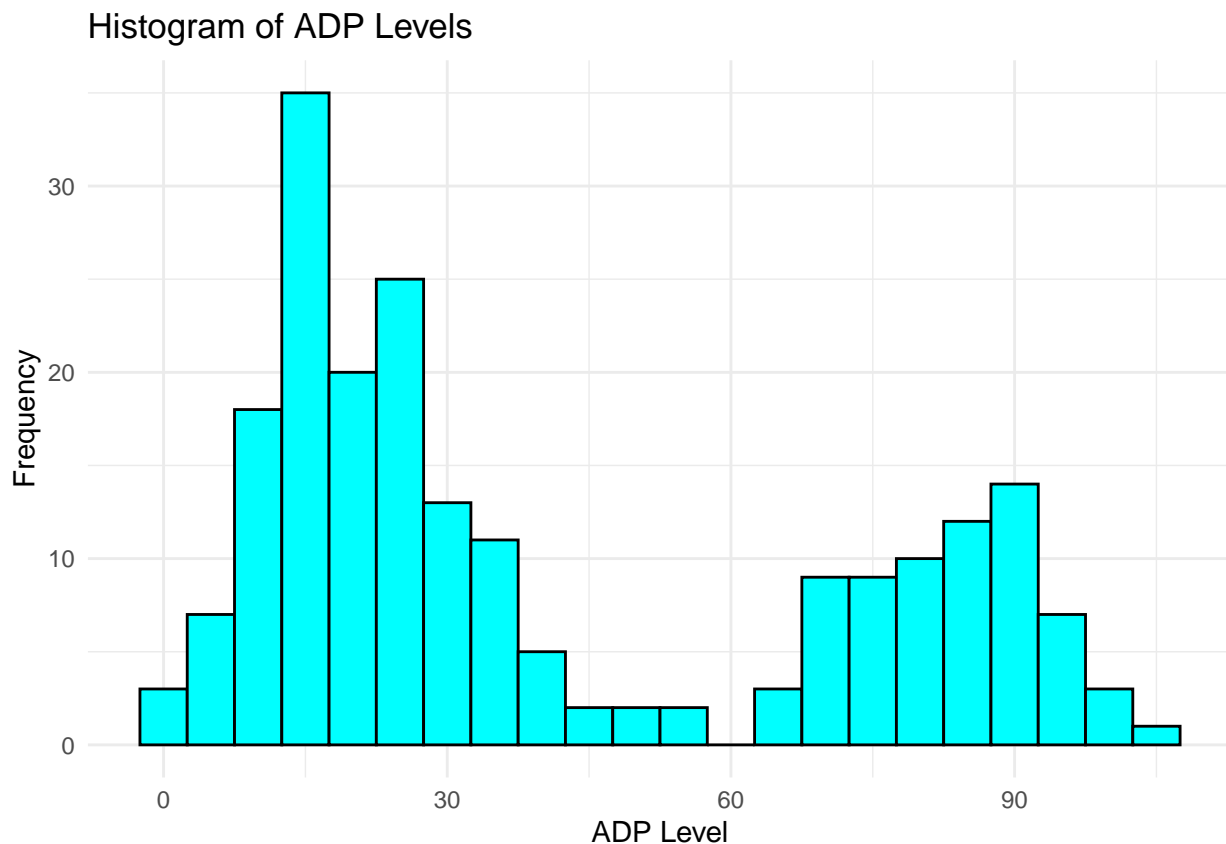
- This adjusted logistic regression model shows a very weak negative relationship between drug resistance and the SNP rs662 after controlling for AGE and SEX. Specifically, for each unit increase in rs662, the log-odds of drug resistance decrease by 0.0046 units. Age has a slight negative effect on the log-odds of drug resistance, while being male slightly increases the log-odds of drug resistance.

12. Creating Visualization

Create visualizations to show the distribution of ADP and the relationships between $\log(\text{ADP})$ and various SNPs.

Histogram of ADP Levels

```
ggplot(data, aes(x = ADP)) +  
  geom_histogram(binwidth = 5, fill = "cyan", color = "black") +  
  theme_minimal() +  
  labs(title = "Histogram of ADP Levels", x = "ADP Level", y = "Frequency")
```

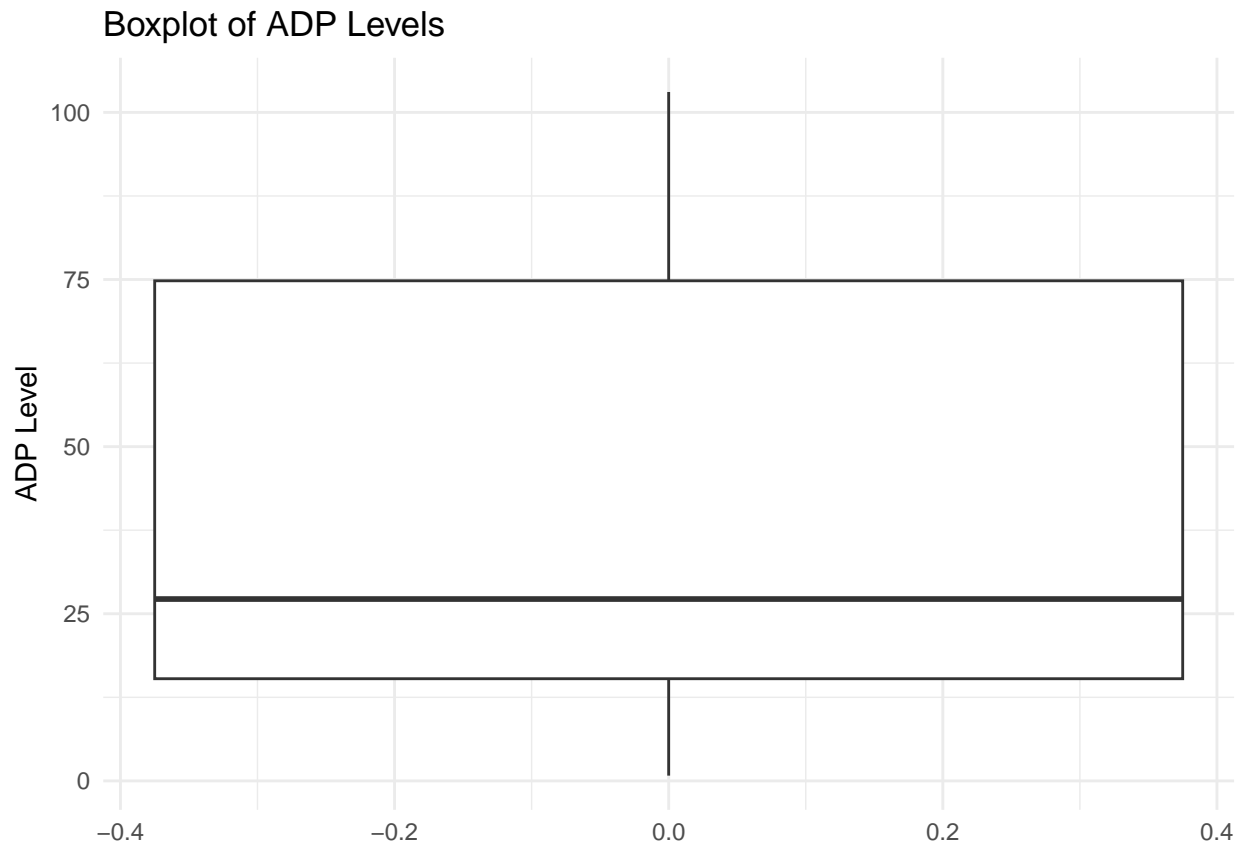


- There is a noticeable gap in the distribution between ADP levels 50 and 60, where the frequency drops to nearly zero. This distribution indicates a bimodal pattern, suggesting that there are two distinct groups or populations with different ADP levels.

Boxplot of ADP Levels

```
ggplot(data, aes(y = ADP)) +  
  geom_boxplot()
```

```
theme_minimal() +
labs(title = "Boxplot of ADP Levels", y = "ADP Level")
```

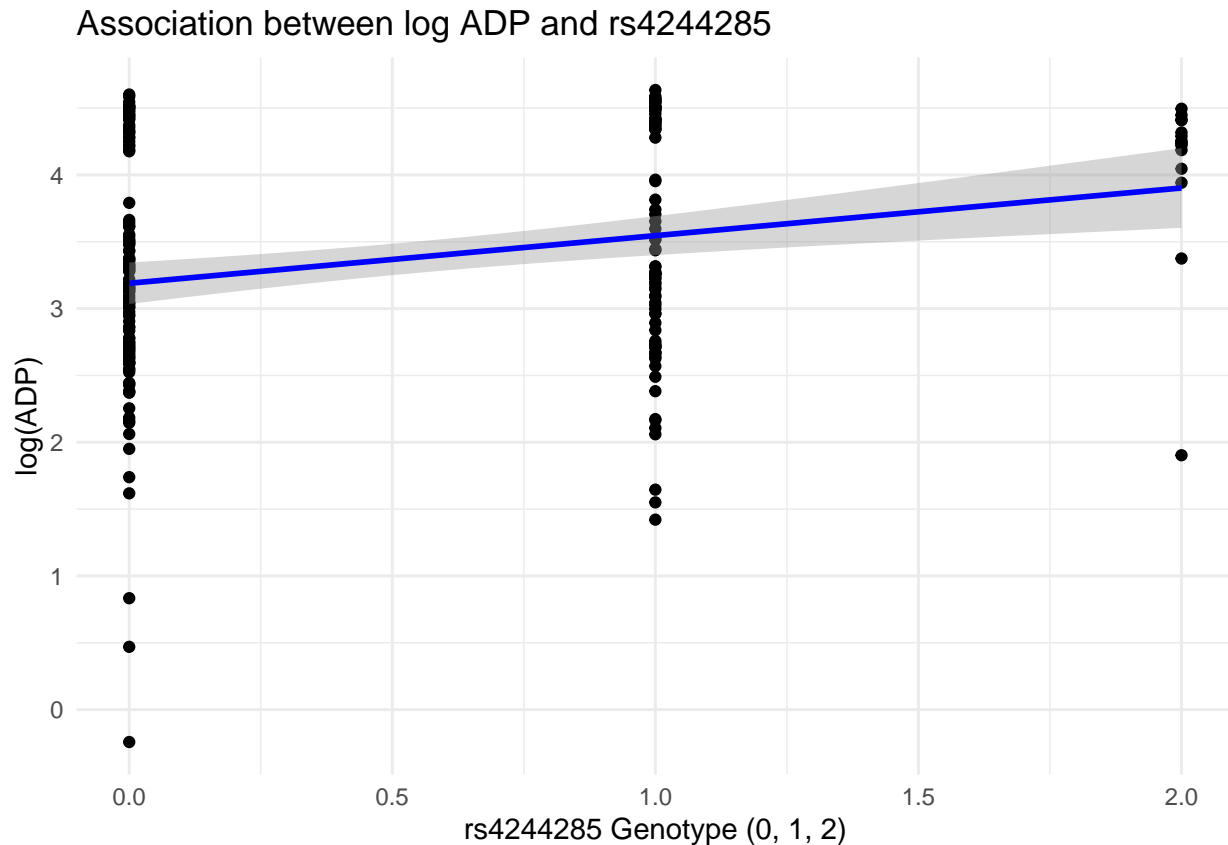


- From the boxplot there is no significant outliers affecting the data.

Association between log ADP and rs4244285

```
ggplot(data, aes(x = rs4244285, y = ADP_log)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue") +
  theme_minimal() +
  labs(title = "Association between log ADP and rs4244285", x = "rs4244285 Genotype (0, 1, 2)", y = "log ADP")

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

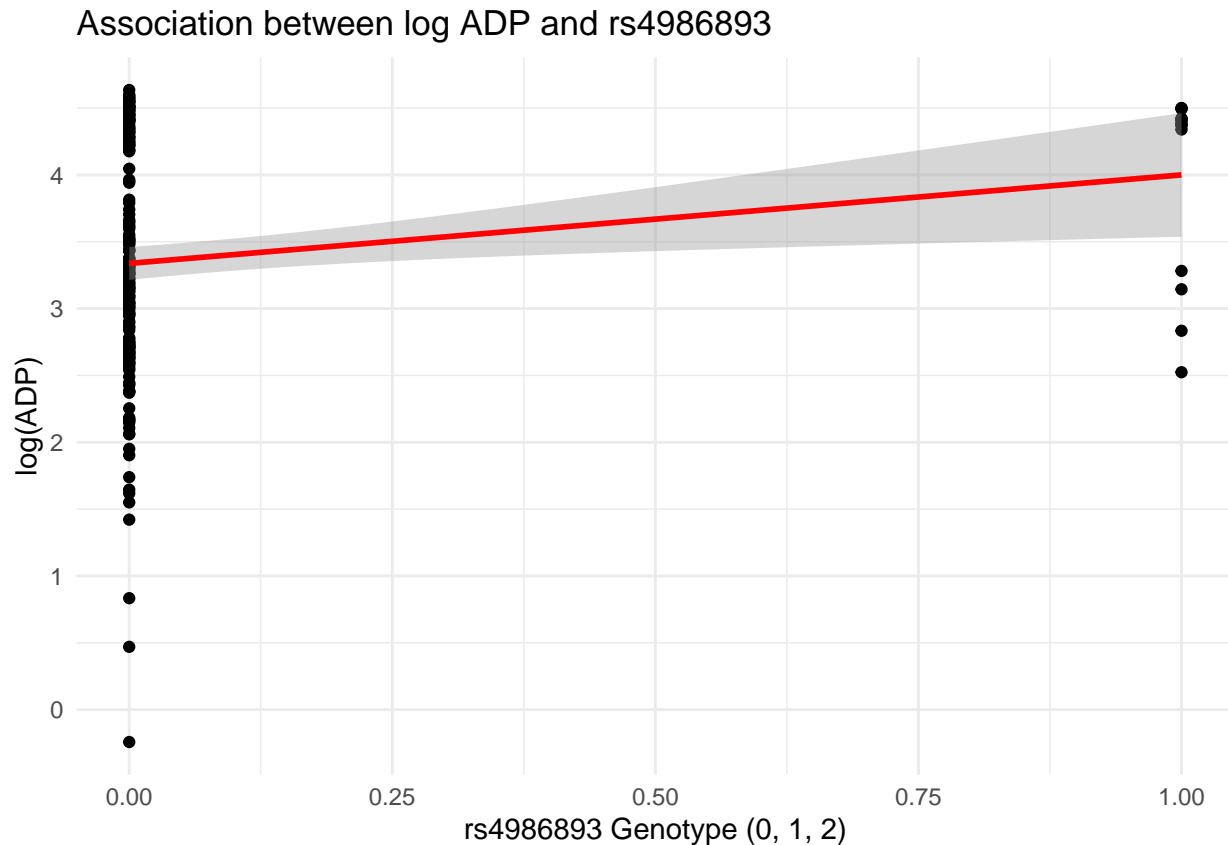


- The plot indicates a slight positive association between the rs4244285 genotype and log-transformed ADP levels, with the log(ADP) levels tending to increase slightly as the rs4244285 genotype value increases. The data points and the fitted regression line with its confidence interval provide a clear visualization of this relationship.

Association between log ADP and rs4986893

```
ggplot(data, aes(x = rs4986893, y = ADP_log)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red") +
  theme_minimal() +
  labs(title = "Association between log ADP and rs4986893", x = "rs4986893 Genotype (0, 1, 2)", y = "log(ADP)")

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

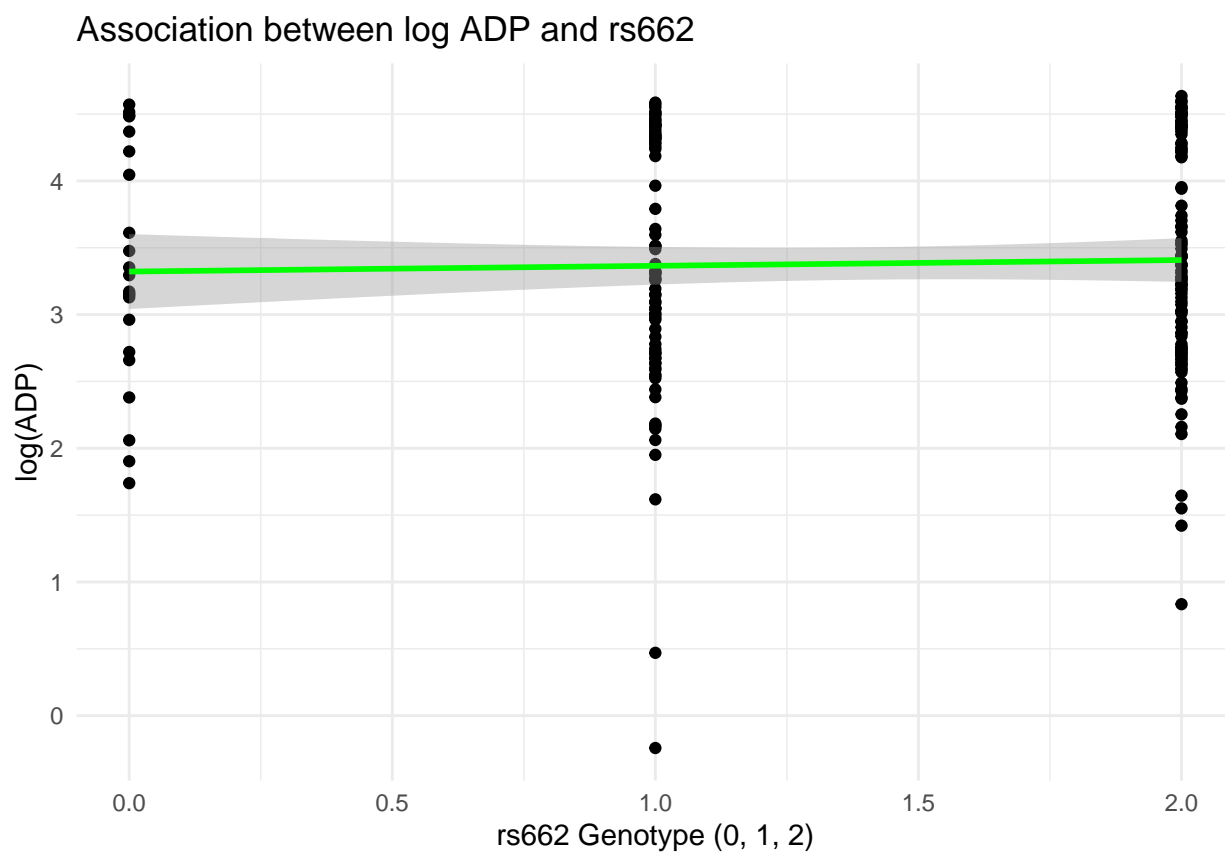


- The plot indicates a positive association between the rs4986893 genotype and log-transformed ADP levels, with log(ADP) levels tending to increase as the rs4986893 genotype value increases. The data points are scattered around the regression line, showing some variability in log(ADP) levels for each genotype.

Association between log ADP and rs662

```
ggplot(data, aes(x = rs662, y = ADP_log)) +
  geom_point() +
  geom_smooth(method = "lm", color = "green") +
  theme_minimal() +
  labs(title = "Association between log ADP and rs662", x = "rs662 Genotype (0, 1, 2)", y = "log(ADP)")

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



- The plot indicates a non-significant relationship between the rs662 genotype and log-transformed ADP levels, with log(ADP) levels tending to stay the same as the rs662 genotype value increases.

Discussion of experimental results

This study will identify the relationship between the level of ADP-induced platelet aggregation and Clopidogrel resistance using three genetic markers: rs4244285 (CYP2C192), rs4986893 (CYP2C193), and rs662 (PON1 192Q>R), adjusting for age and sex in refining the analysis. Linear regression will be applied to ADP levels as a continuous variable, logistic regression will be used for the binary outcome of drug resistance. The data set PlateletHW.tsv includes the following 11 variables: identifiers, genetic markers, age, sex, and two continuous measurements each of ADP and drug resistance. Genetic markers and drug resistance are categorically coded, while the sex is coded 0 for males and 1 for females. After initial cleaning, which consisted of taking absolute values of ADP and examining distribution, the slight right-skewed distribution of 0.607 indicated no outliers and was thus reliable for analysis. The linear regression of log-transformed ADP on rs4244285 revealed a significant positive association, with an increase of 0.356 per unit of rs4244285, this was modestly improved to 0.36083 after adjustment for age and sex. Rs4986893 also showed a very strong positive association with log(ADP), with a growth of 0.662 per unit of rs4986893, heavily replicated as 0.66369 after adjustment. Rs662 did not exhibit any significant association with ADP levels. The logistic regression models showed that rs4244285 and rs4986893 are strongly positively correlated with drug resistance, where rs4244285 increases the odds by 0.943 units and rs4986893 increases the odds by 1.742 units. These associations maintained after adjustment. Rs662 showed a very weak negative correlation against drug resistance, lowering the odds by 0.0256 units, and was further diminished to 0.0046 when adjusted. While adjusting for age and sex, older age and male sex slightly reduced log(ADP) levels and the probability of resistance to a medication. Thus, rs4244285 and rs4986893 seem to be the strong predictors of both ADP aggregation levels and Clopidogrel resistance, whereas rs662 doesn't have any significant effect in these contexts.