Analysis of Heart Rate Data

CID: 02099078

Introduction

The data provided contains measurements of heart rate in beats per minute, taken before and after treatment. Patients were randomly distributed between being given the treatment and the placebo. Each patient's BMI was measured and on the first day, their resting heart rate was measured. On the second day, each patient took a pill containing either the stimulant or a placebo, and their heart rate was measured 30 minutes later.

Summary statistics are given in Table 1 below. There are 110 observations in the placebo group and 48 observations in the stimulant group. The table shows that the mean increase in heart rate after the pill is taken is 6.04 bpm for the placebo group and 12.58 bpm for the stimulant group. This gives a difference between the two treatment groups of 6.54 bpm. The table also shows the standard deviation of each treatment group for the heart rates before and after the pill is taken. We observe that the standard deviation for the stimulant group post-treatment is especially high, which suggests there was a lot of variation in how the heart rate of patients reacted to the stimulant.

type	time	mean	sd	n
placebo	0	64.95	5.02	110
placebo	1	70.99	6.22	110
treatment	0	64.94	4.97	48
treatment	1	77.52	9.19	48

Table 1: Summary statistics for heart rate in beats per minute.

Figure 1 shows the joint distribution of heart rate measurements before and after the pill was taken, for each treatment group. For each group, the distribution appears to be bivariate normal, with the stimulated heart rate generally being higher for patients with higher rest heart rates.

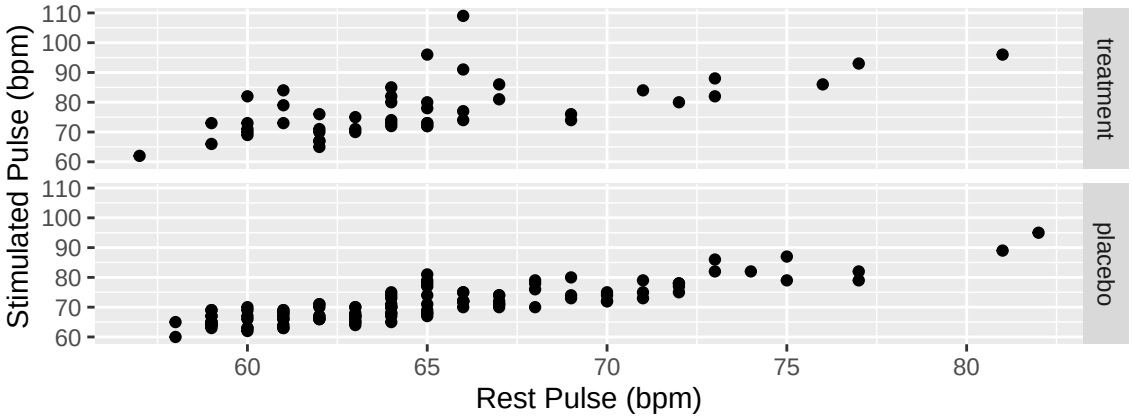


Figure 1: Joint distribution of heart rate measurements by treatment group.

Figure 2 shows a box plot of the the heart rate measurements before and after the pill was taken, for each treatment group. This helps us to visualise the effect of the pill for the stimulant group and the placebo group comparatively.

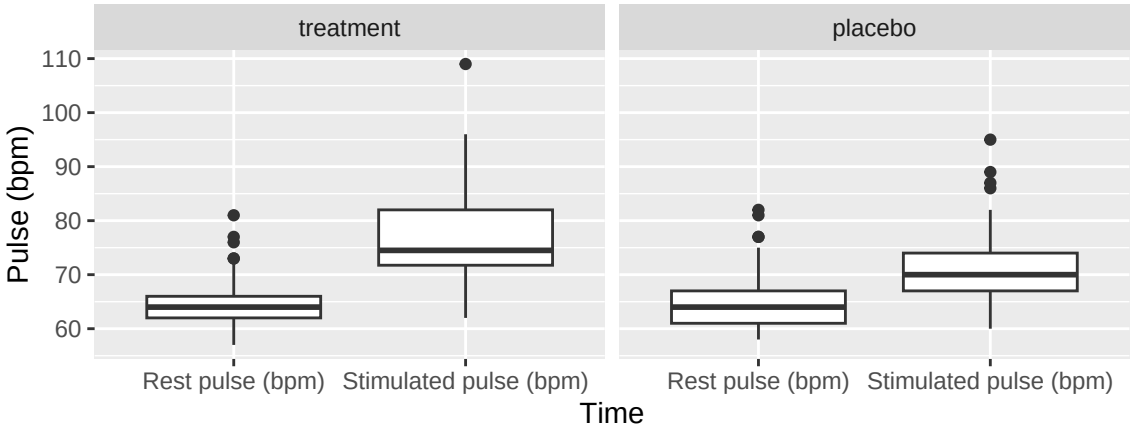
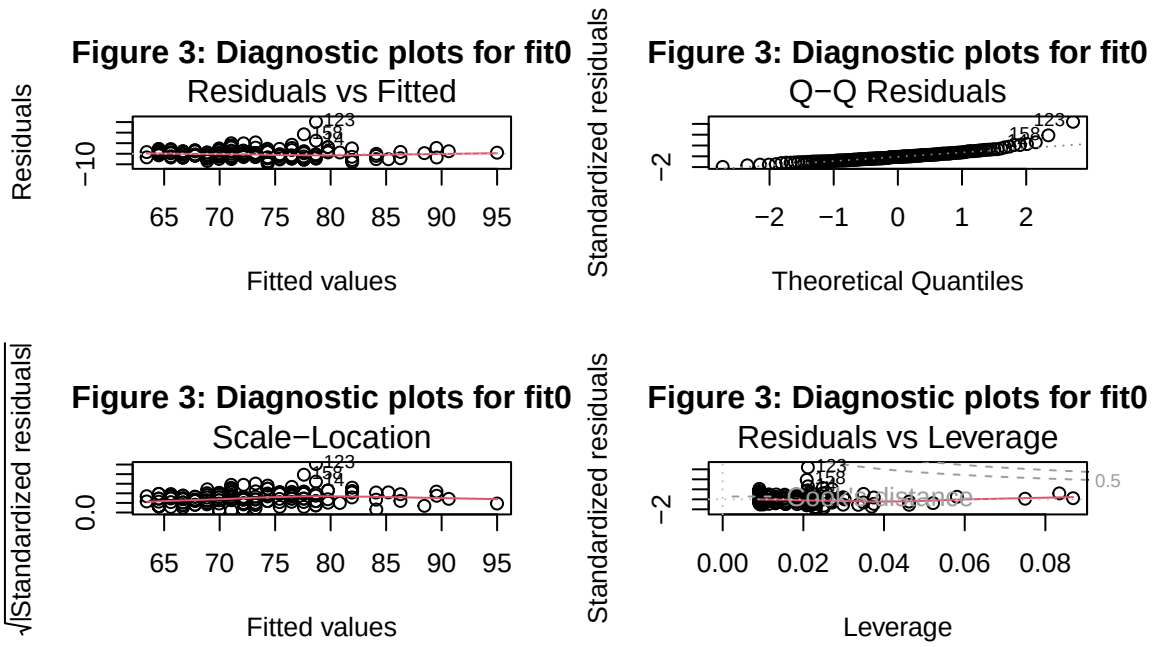


Figure 2: Box plot of heart rate measurements by treatment group.

Initial Model

We first fit the initial model suggested by the clinicians (fit0). For this model, the stimulated pulse varies by adding the rest pulse and the treatment type. However, it ignores the BMI of each patient, therefore this may be an unreasonable model to use. The summary of the model tells us that post-treatment, the mean heart rate for the stimulated group is 6.55 bpm higher than it is for the placebo group. The standard error of this estimate is 0.08 bpm, and as the p-value is extremely small, there is sufficient evidence to suggest that there is a treatment effect at the 5% significance level.

We notice in figure 3 that for the Q-Q residuals plot, the tails are heavier. This suggests that the underlying distribution of the residuals is not normal. There are more extreme values than we would usually expect under a normal distribution. This suggests that the underlying distribution of the heart rates post-treatment is also not normal. In addition, the scale-location plot does not appear to be constant, which implies that the residuals have non-constant variance. Adding the BMI as a covariate may lead to a better model for the data.



Statistician's Model

The statistician advised that we use the difference between the stimulated and resting heart rates as a response variable, and fit a Gamma GLM using the inverse link function, with BMI and treatment as covariates (fit1). In order to achieve this, we use the iterated weighted least squares algorithm. To find a sensible starting point, we fitted the linear model, as implemented below, and used the coefficients as the initial beta values.

```
difference <- data$stimulated_pulse - data$rest_pulse

# fit linear model to obtain starting beta values
initial_lm <- lm(1/difference ~ bmi + treatment, data = dat)

beta <- c(0.6423, -0.0244, 0.1210) # initial guess using coefficients of initial_lm

for (i in 1:25){
  eta <- cbind(1, data$bmi, as.numeric(data$treatment) - 1)%*%beta # estimated linear predictor
  mu <- 1/eta # estimated mean response
  z <- eta + (mu - difference)/(mu^2)  # form the adjusted variate
  w <- mu^2 # weights
  lmod <- lm(z ~ data$bmi + data$treatment, weights = w) # regress z on x with weights w.
  beta <- as.numeric(lmod$coeff) # new beta
}
```

Using our numerical algorithm for the statistician's model, we plot the fitted values to obtain figure 4. This plot shows the differences in heart rates (before and after taking the pill) against the BMI values for each patient, separated by treatment group. It appears that patients with higher BMI values experience a slight increase in the effect of the pill on their heart rate. The increase seems to be more significant for the stimulant group than for the placebo group.

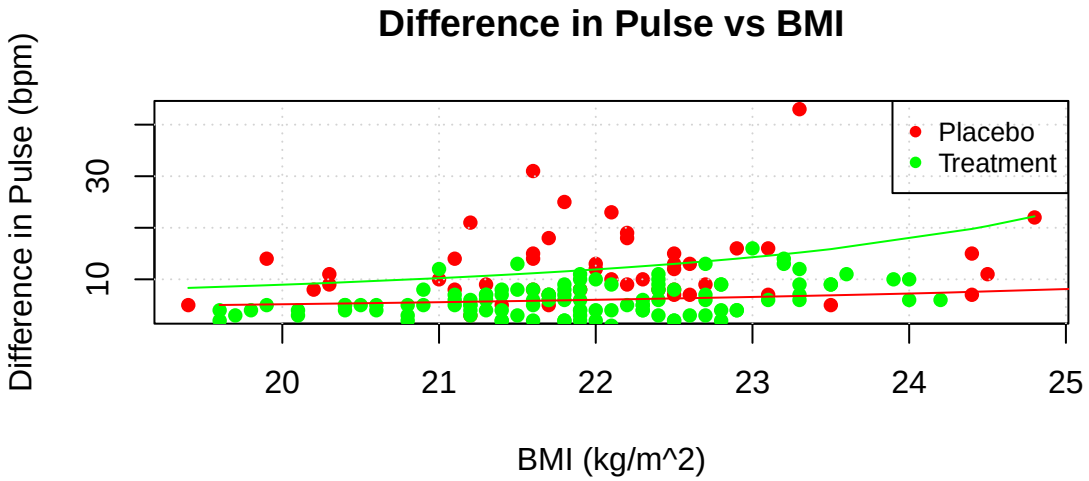


Figure 4: Difference in pulse against BMI.

We now look to obtain an estimate of the size of the treatment effect, for a typical BMI from the sample. Using the coefficients of our model and the mean BMI from the data, we obtain 5.83 as our treatment effect estimate. This means that for an average BMI from the sample, we would expect the stimulant to increase a patient's heart rate by 5.83 bpm.

Next, we use bootstrap resampling to construct a 95% confidence interval for the size of the treatment effect. We iteratively fit the statistician's model using shuffled subsets of the data, computing treatment effect

estimates each time. This gives us a 95% confidence interval of (4.05, 7.33), meaning that there is a 95% chance of the treatment effect lying in this range for an average BMI from the sample.

Improved Model

The statistician's model is clearly an improvement on the initial model suggested by the clinicians. However, adding a cross-term between treatment and BMI may be even more effective, and so we investigate whether this change leads to a better-fitting model (fit2). Table 2 shows the AIC scores for all three models. The improved model is 4 AIC units lower than the statistician's model, which suggests there is a significant improvement by adding the cross-term to the model. To further analyse whether the improvement is significant, we check the p-value of the cross-term and the p-value of the F test between fit1 and fit2. We get 0.014 and 0.016 respectively. These are both significant at the 5% significance level and suggest that this new model is a significant improvement on the statistician's model.

model	AIC
fit0	948.8
fit1	847.8
fit2	843.8

Table 2: AIC scores for each model.

As before, we obtain an estimate of the treatment effect, this time for the improved model. Our value for the treatment effect is 6.43, suggesting that for an average BMI from the sample, the stimulant increases heart rate by 6.43 bpm compared to the placebo.

Also as before, we use bootstrap resampling to construct a 95% confidence interval for the size of the treatment effect. This gives us a 95% confidence interval of (4.41, 8.47), meaning that for the improved model, there is a 95% chance of the treatment effect lying in this range for an average BMI from the sample. Figure 5 plots the distribution of the bootstrapped treatment effects for both the statistician's model and the improved model.

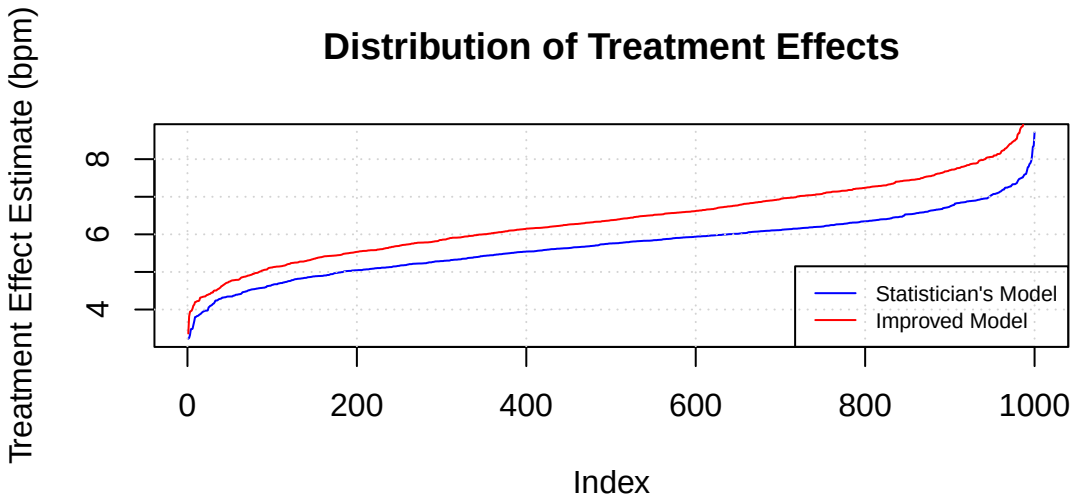


Figure 5: Treatment effects for statistician's and improved models.

Figure 6 displays a plot of the expected increase in heart rate for individuals receiving the stimulant and the placebo as a function of BMI. It can be seen that for this model, patients with a higher BMI have a significantly higher expected increase in heart rate for both treatment groups. The treatment effect itself, which is the difference between the two curves, doesn't appear to change too much as BMI varies.

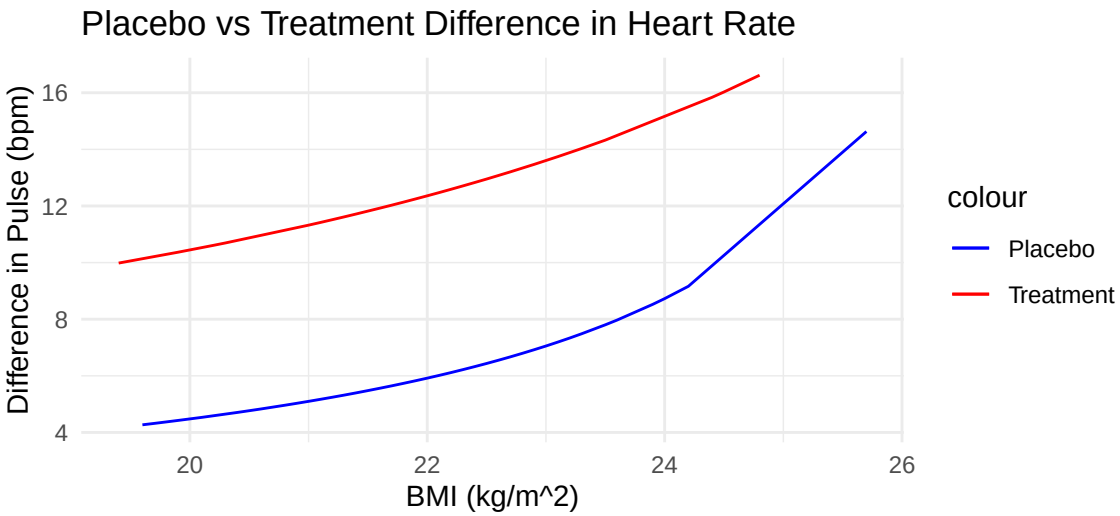


Figure 6: Expected increase in heart rate as a function of BMI.

Finally, we perform a simulation study to determine the size of the treatment effect that could be reliably detected with the current sample size. For each simulated dataset, we determine whether the treatment effect is significant by calculating the power, which is the proportion of times the null hypothesis (no treatment effect) is rejected across all simulations. We find that the proportion of simulations in which the treatment coefficient is significantly different from zero at the 5% level is 81.5%. Figure 7 visualises the relationship between the size of the treatment effect and the power. As the size of the treatment effect increases, the power also increases, meaning there is a higher chance of detecting the treatment effect when it exists. When the size of the treatment effect is small, the power is closer to 0, resulting in there being a small chance of detecting the treatment effect regardless of whether it exists.

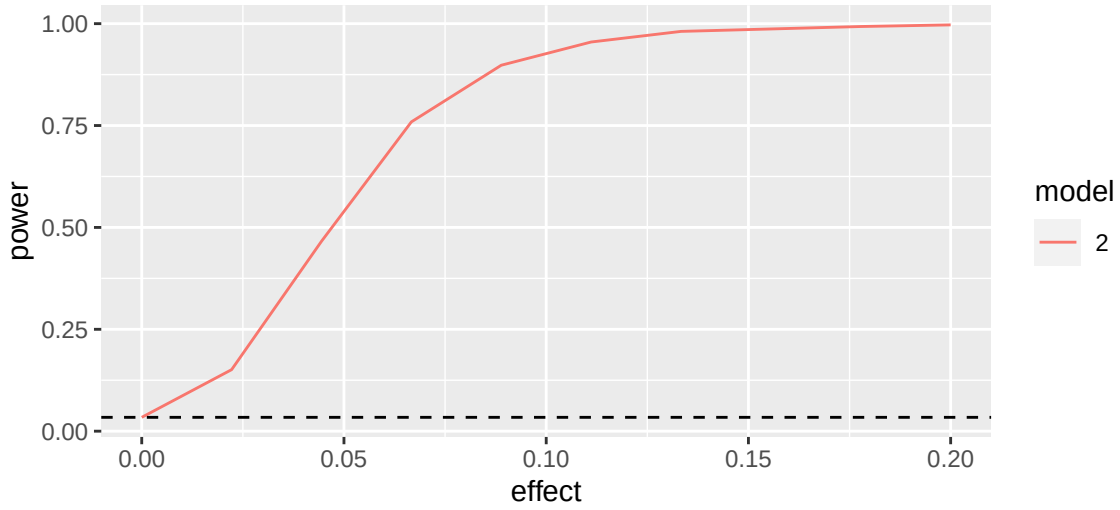


Figure 7: Power as a function of the strength of the treatment effect for fit2.

The simulation study shows that the trial is overpowered, since very small treatment effects can be detected with high probability. This means there is a risk of Type I errors, as we may conclude that there is a treatment effect when there isn't one. To reduce this risk, we could use a lower significance level. Overall, our improved model does allow the treatment effect to be estimated precisely. Increasing the sample size of the study would lead to an even more precise estimate.

Accessible Conclusions

The main difficulty with the initial model provided by the clinicians is that it uses a linear model when the response variable does not follow a normal distribution. As a result, generalised linear models are more useful for modelling this data, as they only assume that the reponse variable follows a specific probability distribution. It is also more important to look at the difference between the rest and stimulated heart rates instead of just the stimulated heart rate, as we would expect a patient with a higher rest heart rate to have a higher stimulated heart rate. Another flaw is that the initial model ignores the BMI of each patient, even though it has an effect on the rest and stimulated heart rates of the patients. A limitation of the final improved model is that adding extra parameters can increase the risk of overfitting, meaning the model may capture noise in the data.

The treatment effect represents the true effect of taking the pill with the stimulant on a patient's heart rate. Our results found that on average, 30 minutes after taking the pill, a patient's heart rate will have increased by 6.43 beats per minute. We are 95% sure that a patient's heart rate will increase by between 4.41 beats per minute and 8.47 beats per minute. At the 5% significance level, there is sufficient evidence to reject the null hypothesis, which is that there is no treatment effect. There is sufficient evidence to suggest that the artificial stimulant does increase heart rate. Patients with higher BMI's experience larger increases in heart rate for both treatment groups, however, the treatment effect itself, which is the difference between the two treatment groups, stays relatively constant across the BMI values.

Appendix

```
# import libraries
library(tidyverse)
library(ggplot2)
library(EnvStats)
library(knitr)
```

```
# Download the data and read it into R.
load("2099078.RData")

# create copy of data which will have renamed treatment categories for better plots
data <- dat

# treatment is categorical
data$treatment <- as.factor(data$treatment)

# rename treatment categories for better plots
levels(data$treatment) <- c("treatment", "placebo")

# add a subject ID for individual change plots
data$id <- 1:nrow(data)

#summary statistics
summary(dat[dat$treatment == 1,])
summary(dat[dat$treatment == 0,])
sd(dat[dat$treatment == 1,]$rest_pulse)
sd(dat[dat$treatment == 1,]$stimulated_pulse)
sd(dat[dat$treatment == 0,]$rest_pulse)
sd(dat[dat$treatment == 0,]$stimulated_pulse)
```

```
# Create a data frame with the summary statistics
summary_stats <- data.frame(
  type = c("placebo", "placebo", "treatment", "treatment"),
  time = c(0, 1, 0, 1),
  mean = c(64.95, 70.99, 64.94, 77.52),
  sd = c(5.02, 6.22, 4.97, 9.19),
  n = c(110, 110, 48, 48)
)

# Print the data frame in LaTeX table format
kable(summary_stats, format = "latex", caption = "Summary Statistics for heart rate in
      beats per minute")
```

```
# plot of stimulated pulse against rest pulse
stim_rest <- ggplot(data = data, mapping = aes(x = rest_pulse, y = stimulated_pulse)) +
  geom_point() + facet_grid(data$treatment) + xlab("Rest Pulse (bpm)") + ylab("Stimulated Pulse (bpm)")
stim_rest
```

```
# plot of rest pulse against BMI
rest_bmi <- ggplot(data = data, mapping = aes(x = bmi, y = stimulated_pulse - rest_pulse)) + geom_point
rest_bmi
```

```r
# box plot of rest and stimulated pulse for the treatment and placebo
df_long <- pivot_longer(data, cols = c(rest_pulse, stimulated_pulse),
                        names_to = "Time", names_prefix = "Pulse", values_to = "Pulse")
box_plots <- ggplot(data = df_long, mapping = aes(x = Time, y = Pulse)) + geom_boxplot() +
  facet_wrap(~treatment) + labs(y = "Pulse (bpm)", caption = "Figure 3: Box plot of heart rate measureme

box_plots
```

```r
fit0 <- lm(stimulated_pulse ~ rest_pulse + treatment, data = data)

# summary
summary(fit0)

# AIC score
aic_fit0 <- AIC(fit0)
print(paste("AIC:", aic_fit0))
```

```r
# diagnostic plots
par(mfrow = c(2, 2))
plot(fit0, main = "Figure 4: Diagnostic plots for fit0")
```

```r
difference <- data$stimulated_pulse - data$rest_pulse

# fit a linear model to obtain starting beta values
initial_lm <- lm(1/difference ~ bmi + treatment, data = dat)

beta <- c(0.6423, -0.0244, 0.1210) # initial guess

for (i in 1:25){
  eta <- cbind(1, data$bmi, as.numeric(data$treatment) - 1)%*%beta # estimated linear predictor
  mu <- 1/eta # estimated mean response
  z <- eta + (mu - difference)/(mu^2)  # form the adjusted variate
  w <- mu^2 # weights
  lmod <- lm(z ~ data$bmi + data$treatment, weights = w) # regress z on x with weights w.
  beta <- as.numeric(lmod$coeff) # new beta
}
```

```r
# obtain initial beta values
initial_lm

# beta values
print(beta)
```

```r
# check the numerical algorithm using the glm function.
fit1 <- glm(difference ~ bmi + treatment, data = dat, family = Gamma())
fit1

# AIC score
AIC(fit1)
```

```r
# plot the fitted function.
placebo_data <- dat[dat$treatment==1,]
placebo_data <- placebo_data[order(placebo_data$bmi),]
placebo_difference <- placebo_data$stimulated_pulse - placebo_data$rest_pulse
# Plotting the data
plot(placebo_data$bmi, placebo_difference, xlab = "BMI (kg/m^2)",
     ylab = "Difference in Pulse (bpm)", main = "Difference in Pulse vs BMI (Placebo)", col = "blue", p
mu_hat_placebo = rep(0,times = length(placebo_data))
for (i in 1:110){
  mu_hat_placebo[i] <- 1/(beta[1] + beta[2] * placebo_data$bmi[i] + beta[3])
  }
# Adding regression line
lines(x = placebo_data$bmi, y = mu_hat_placebo, col = "red")

# Adding caption
title("Figure 5: Difference in pulse against BMI for placebo group.", line = -18,
      cex.main = 0.9)
```

```r
treatment_data <- dat[dat$treatment==0,]
treatment_data <- treatment_data[order(treatment_data$bmi),]
treatment_difference <- treatment_data$stimulated_pulse - treatment_data$rest_pulse
# Plotting the data
plot(treatment_data$bmi, treatment_difference, xlab = "BMI (kg/m^2)",
     ylab = "Difference in Pulse (bpm)", main = "Difference in Pulse vs BMI (Treatment)", col = "blue",
mu_hat_stimulated = rep(0,times = length(treatment_data))
for (i in 1:48){
  mu_hat_stimulated[i] <- 1/(beta[1] + beta[2] * treatment_data$bmi[i])
  }

# Adding regression line
lines(x = treatment_data$bmi, y = mu_hat_stimulated, col = "red")

# Adding caption
title("Figure 6: Difference in pulse against BMI for treatment group.", line = -18,
      cex.main = 0.9)
```

```r
# Combine data
combined_data <- rbind(placebo_data, treatment_data)
combined_difference <- c(placebo_difference, treatment_difference)
combined_mu_hat <- c(mu_hat_placebo, mu_hat_stimulated)

# Plotting the data for placebo group
plot(combined_data$bmi[combined_data$treatment == 0], combined_difference[combined_data$treatment == 0]
     xlab = "BMI (kg/m^2)", ylab = "Difference in Pulse (bpm)",
     main = "Difference in Pulse vs BMI", col = "red", pch = 16)
grid()

# Adding points for treatment group
points(combined_data$bmi[combined_data$treatment == 1], combined_difference[combined_data$treatment ==
       col = "green", pch = 16)

# Adding regression lines
lines(x = combined_data$bmi[combined_data$treatment == 1], y = combined_mu_hat[combined_data$treatment
```

```r
lines(x = combined_data$bmi[combined_data$treatment == 0], y = combined_mu_hat[combined_data$treatment =

# Adding legend
legend("topright", legend = c("Placebo", "Treatment"), col = c("red", "green"), pch = 16, cex = 0.8)

# Adding caption
title("Figure 4: Difference in pulse against BMI.", line = -10.5, cex.main = 0.9)
```

```r
# obtain an estimate of the size of the treatment effect.

# set seed for reproducability
set.seed(25)

beta = (coef(fit1))

# estimate of the treatment effect for the placebo
mu_hat_placebo <- 1/(beta[1] + beta[2] * mean(dat$bmi) + beta[3])

# estimate of the treatment effect for the stimulated
mu_hat_stimulated <- 1/(beta[1] + beta[2] * mean(dat$bmi))

# estimate of the size of the treatment effect
estimate <- mu_hat_stimulated - mu_hat_placebo

print(estimate)
```

```r
# use bootstrapping to find 95% confidence interval for size of the treatment effect.

# set seed for reproducability
set.seed(25)

ind<- 1:length(dat)
estimates = rep(0,times = length(dat))

for (i in 1:1000){
  shuffled_index <- sample(ind,158, replace = T)
  shuffled_data <- dat[shuffled_index,]
  fitboot <- glm(stimulated_pulse - rest_pulse ~ bmi + treatment, data = shuffled_data,
                 family = Gamma())
  beta = (coef(fitboot))

  # estimate of the treatment effect for the placebo
  mu_hat_placebo <- 1/(beta[1] + beta[2] * mean(shuffled_data$bmi) + beta[3])

  # estimate of the treatment effect for the stimulated
  mu_hat_stimulated <- 1/(beta[1] + beta[2] * mean(shuffled_data$bmi))

  # estimate of the size of the bootstrapped treatment effects
  estimates[i] <- mu_hat_stimulated - mu_hat_placebo
  }

#sorted list of bootstrapped treatment effects
estimates<-sort(estimates)
```

```r
# print confidence interval
print(c(estimates[25], estimates[975]))
```

```r
# plot the distribution of these estimates for the statistician's model.

# plot distribution
plot(estimates, col = "blue", type = "l", xlab = "Index",
     ylab = "Treatment Effect Estimate", main = "Treatment Effect Estimates for Statistician's Model")

# Adding caption
title("Figure 7: Distribution of treatment effects for statistician's model.",
      line = -18, cex.main = 0.9)
```

```r
# add cross-term between BMI and treatment
fit2 <- glm(difference ~ treatment + bmi + treatment*bmi, data = dat, family = Gamma())

#summary
summary(fit2)

#AIC score
AIC(fit2)
```

```r
par(mfrow = c(2, 2))
plot(fit2, main = "Figure 8: Diagnostic plots for fit2")
```

```r
# F test
anova(fit1, fit2, test="F")
```

```r
# obtain an estimate of the size of the treatment effect for this improved model.

# set seed for reproducability
set.seed(26)

beta = (coef(fit2))

# estimate of the treatment effect for the placebo
mu_hat_placebo_improved <- 1/(beta[1] + beta[2] + beta[3] * mean(dat$bmi) + beta[4] * mean(dat$bmi))


# estimate of the treatment effect for the stimulated
mu_hat_stimulated_improved <- 1/(beta[1] + beta[3] * mean(dat$bmi))

# estimate of the size of the treatment effect
estimate_improved <- mu_hat_stimulated_improved - mu_hat_placebo_improved

print(estimate_improved)
```

```r
# again use bootstrapping to find 95% confidence interval for size of the treatment effect for the impr

# set seed for reproducability
set.seed(26)
```

```r
index = 1:as.numeric(nrow(dat))
estimates_improved = rep(0,times = length(1000))

for (i in 1:1000){
  shuffled_index <- sample(index, replace = T)
  shuffled_data <- dat[shuffled_index,]
  fitboot2 <- glm(stimulated_pulse - rest_pulse ~ treatment + bmi + treatment * bmi, data = shuffled_da

  beta = (coef(fitboot2))

  # estimate of the treatment effect for the placebo
  mu_hat_placebo_improved <- 1/(beta[1] + beta[2] +  beta[3] * mean(shuffled_data$bmi) + beta[4] *
                                  mean(shuffled_data$bmi))

  # estimate of the treatment effect for the stimulated
  mu_hat_stimulated_improved <- 1/(beta[1] + beta[3] * mean(shuffled_data$bmi))

  # estimate of the size of the bootstrapped treatment effects
  estimate_improved <- mu_hat_stimulated_improved - mu_hat_placebo_improved

  estimates_improved[i] = (estimate_improved)
  }

#sorted list of bootstrapped treatment effects
estimates_improved <- sort(estimates_improved)

# print confidence interval
print(c(estimates_improved[25], estimates_improved[975]))
```

```r
# plot the distribution of these estimates for the improved model.

# plot distribution
plot(estimates_improved, col = "red", type = "l", xlab = "Index", ylab = "Treatment Effect Estimate", ma

# Adding caption
title("Figure 9: Distribution of treatment effects for the new improved model.", line = -18, cex.main =
```

```r
# Plotting the distribution for the statistician's model
plot(estimates, col = "blue", type = "l", xlab = "Index", ylab = "Treatment Effect Estimate (bpm)", mai
grid()

# Adding the distribution for the improved model
lines(estimates_improved, col = "red")

# Adding a legend
legend("bottomright", legend = c("Statistician's Model", "Improved Model"), col = c("blue", "red"), lty

# Adding caption
title("Figure 5: Treatment effects for statistician's and improved models.", line = -10.5, cex.main = 0
```

```r
# produce a plot of the expected increase in heart rate for individuals receiving stimulant and placebo
```

```r
fit3 <- glm(stimulated_pulse[treatment == 1] - rest_pulse[treatment == 1] ~ bmi[treatment == 1], data =
fit4 <- glm(stimulated_pulse[treatment == 0] - rest_pulse[treatment == 0] ~ bmi[treatment == 0], data =

# placebo
x_placebo = sort(dat$bmi[dat$treatment == 1])
mu_hat_placebo = 1/(coef(fit3)[1] + coef(fit3)[2] * x_placebo)
plot_placebo = as.data.frame(data.frame(x_placebo, mu_hat_placebo))
plot(plot_placebo)
```

```r
# stimulated
x_stimulated = sort(dat$bmi[dat$treatment == 0])
mu_hat_stimulated = 1/(coef(fit4)[1] + coef(fit4)[2] * x_stimulated)
plot_stimulated = as.data.frame(data.frame(x_stimulated, mu_hat_stimulated))
plot(plot_stimulated)
```

```r
# produce a plot of the expected increase in heart rate for individuals receiving stimulant and placebo

fit3 <- glm(stimulated_pulse[treatment == 1] - rest_pulse[treatment == 1] ~ bmi[treatment == 1], data =
fit4 <- glm(stimulated_pulse[treatment == 0] - rest_pulse[treatment == 0] ~ bmi[treatment == 0], data =

# combine plots
ggplot() +
  geom_line(data = plot_placebo, aes(x = x_placebo, y = mu_hat_placebo, color = "Placebo")) +
  geom_line(data = plot_stimulated, aes(x = x_stimulated, y = mu_hat_stimulated, color = "Treatment")) +
  labs(x = "BMI", y = "Difference in Pulse", title = "Placebo vs Treatment Difference in Heart Rate") +
  scale_color_manual(values = c("Placebo" = "blue", "Treatment" = "red"),
                     labels = c("Placebo", "Treatment")) +
  theme_minimal() + labs(caption = "Figure 10: Expected increase in heart rate as a function of BMI.") +
```

```r
# compute the proportion of simulations in which the treatment coefficient is significantly different f

# set seed for reproducability
set.seed(26)

n_sim <- 1000
coef_store <- rep(0, nrow = n_sim)
p_store <- rep(0, nrow = n_sim)

for (i in 1:n_sim){
  #simulate data
  dat$pulse_sim <- unlist(simulate(fit2))

  # fit model
  fit2_sim <- glm(pulse_sim ~ treatment + bmi + treatment * bmi, data = dat, family = Gamma())
  sum <- summary(fit2_sim)
  coef_store[i] <- sum$coefficients[2,1] # check [2,1]
  p_store[i] <- sum$coefficients[2,4] # check [2,4]
}

hist(p_store)
```

```r
kable(mean(p_store < 0.05), format="simple")
```

```r
# simulation study to determine the size of treatment effect that could reliably be detected with the c

# set seed for reproducability
set.seed(26)

# 1000 simulated datasets
n_sim <- 1000
p_store <- rep(0,length(n_sim))
treat_effs <- seq(1, 1.2, length.out = 10)
power <- rep(0, nrow = length(treat_effs))
j <- 1
placebo_df = dat[dat$treatment == 1,]
stimulated_df = dat[dat$treatment == 0,]
treatment_effect = rep(0, times = length(treat_effs))

for(treat_eff in treat_effs){

  # estimate of the treatment effect for the placebo
  mu_hat_placebo_improved <- 1/(beta[1] + treat_eff + beta[3] * placebo_df$bmi + beta[4] * placebo_df$br

  # estimate of the treatment effect for the stimulated
  mu_hat_stimulated_improved <- 1/(beta[1] + beta[3] * stimulated_df$bmi)

  # shape and scale parameters
  alpha_placebo <- egamma(mu_hat_placebo_improved)$parameters[1]
  beta_placebo <- egamma(mu_hat_placebo_improved)$parameters[2]

  alpha_stimulated <- egamma(mu_hat_stimulated_improved)$parameters[1]
  beta_stimulated <- egamma(mu_hat_stimulated_improved)$parameters[2]

  for (i in 1:n_sim){
    # regenerate response variable data
    placebo_sim = rgamma(nrow(placebo_df), shape = alpha_placebo, scale = beta_placebo)
    stimulated_sim = rgamma(nrow(stimulated_df), shape = alpha_stimulated, scale = beta_stimulated)

    # create bootstrap samples
    placebo_sample = placebo_df[sample(1:nrow(placebo_df), replace = T),]
    stimulated_sample = stimulated_df[sample(1:nrow(stimulated_df), replace = T),]

    new_response = c(stimulated_sim, placebo_sim)

    new_bmi = c(stimulated_sample$bmi, placebo_sample$bmi)
    new_treatment = c(rep(0, times = nrow(stimulated_sample)), rep(1, times = nrow(placebo_sample)))
    new_treatment_bmi = c(rep(0, times = nrow(stimulated_sample)), placebo_sample$bmi)

    fitboot_new = glm(new_response ~ new_treatment + new_bmi + new_treatment_bmi)

    sum_new = summary(fitboot_new)
    p_store[i] <- sum_new$coefficients[2,4]
  }
```

```
  power[j] <- mean(p_store < 0.05)
  j <- j + 1
}
```

```
power_long <- data.frame(
  effect = c(treat_effs) - 1,
  power = c(power),
  model = as.factor(rep(2, length(treat_effs)))
)

# plot power against effect
power_plot <- ggplot(data = power_long, mapping = aes(x = effect, y = power, group = model, colour = mo

power_plot
```