Report: Analysis of heart rate data

Coursework

1 Introduction

The data come from an experiment to study the effect of an artificial stimulant on heart rate, and how this varies as a function of underlying health, measured by the body mass index (BMI). The study was conducted over two days. On the first day, their resting heart rate was measured. On the second day, each subject was given a pill containing either 100mg of the stimulant or a placebo. Their heart rate was measured 30 minutes later.

Summary statistics are provided in Table:1. First of all, we observe that there are 108 observations for the placebo group '1' and 55 for the treatment group '0'. The mean stimulated heart rate is at about 77.5 beats/minute in the treatment group '0' and 71.8 beats/minute in the placebo group '1'. The standard deviation is also higher in group '0', with 8.6 b/min compared to 6.1 b/min in group '1'. The other statistics regarding 'rest_pulse' and 'BMI' are nearly identical between the two groups.

Next, Figure:1 shows some exploratory analysis of the variables. For the joint distribution of before and post-treatment heart rate, we notice a non-constant variance within each group. The distribution seems to follow a gamma distribution. Additionally, we observe a lower mean of the stimulated pulse in the placebo group than in the treatment one.

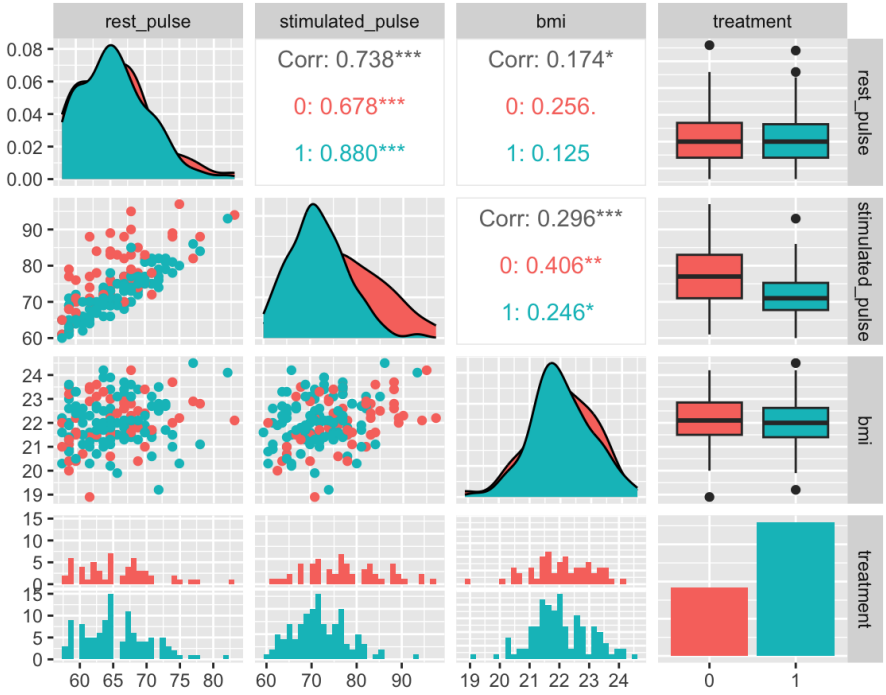


Figure 1: Joint distributions

Treatment group	Count	Mean Rest Pulse	sd Rest Pulse	Mean Stimulated Pulse	sd Stimulated Pulse	Mean bmi	sd bmi
0	55	65.76	5.48	77.53	8.59	22.07	1.06
1	108	65.63	4.87	71.79	6.14	22.04	0.98

Table 1: Summary statistics

2 Results

2.1 The clinicians' model

The clinicians' model (`fit0`) is a linear model allowing stimulated pulse to vary additively by treatment group and rest pulse. In the model, the mean stimulated pulse in treated individuals is about 5.6 b/min higher than in those in the placebo group. The standard error of this estimate is 0.72 b/min, so there is sufficient evidence for a treatment effect. Nevertheless, diagnostic plots Figure:2 show evidence of a non-normal distribution of the residuals (QQ plot). In addition, the 'Scale-Location' plot demonstrates non-constant variance of the residuals. With these results we understand that a linear model is not appropriate in our context.

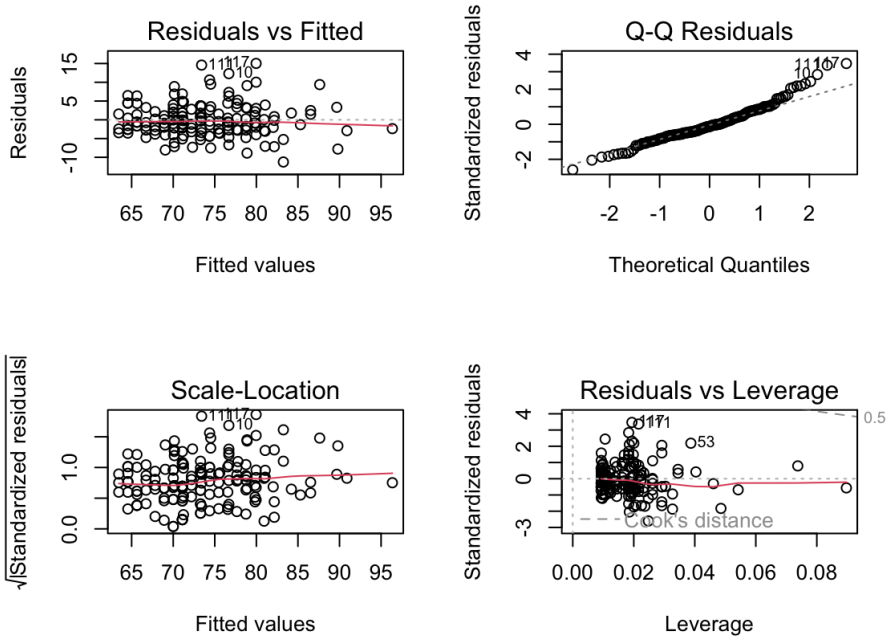


Figure 2: Diagnostic plots `fit0`

2.2 The statistician's model

The statistician's model (`fit1`) is a Gamma generalised linear model (with 'inverse' link function) using the difference between the stimulated and resting heart rates as a response variable, and BMI and treatment as covariates. We implement the IWLS algorithm (appendix) to fit the model and estimate the expected increase in heart rate μ for each individual. We observe better results on the diagnostic plots Figure:3; especially the 'QQ' and 'Scale-Location' plots. The model fits better.

As a first insight, we estimate the difference of the mean increase in heart rate between treatment and placebo groups: 6.2 b/min.
Having derived these results, we decide to pursue the analysis by comparing the difference in expected increase in heart rate between an individual who receives the treatment and an individual who receives the placebo, both having the same BMI of $21.6\,kg/m^2$ (the mode of the BMIs, i.e. the most recurrent BMI). Hence, we obtain a difference of 4.47 b/min (i.e. heart rate is 4.47 b/min faster for the treated individual) with a confidence interval of $(3.60, 5.86)$ b/min. This confidence interval does not contain 0, so there is evidence of a treatment effect.

Next, we fit an interaction model (`fit2`) which allows for the treatment effects to be different for different BMIs. In comparing the performance of both models, Table:2, we observe that they have similar performances: `fit2` has a slightly lower deviance than `fit1` (-0.2), while `fit1` has a slightly lower AIC than `fit2`. Based on the AIC, we decide to pursue our analysis with `fit1`.

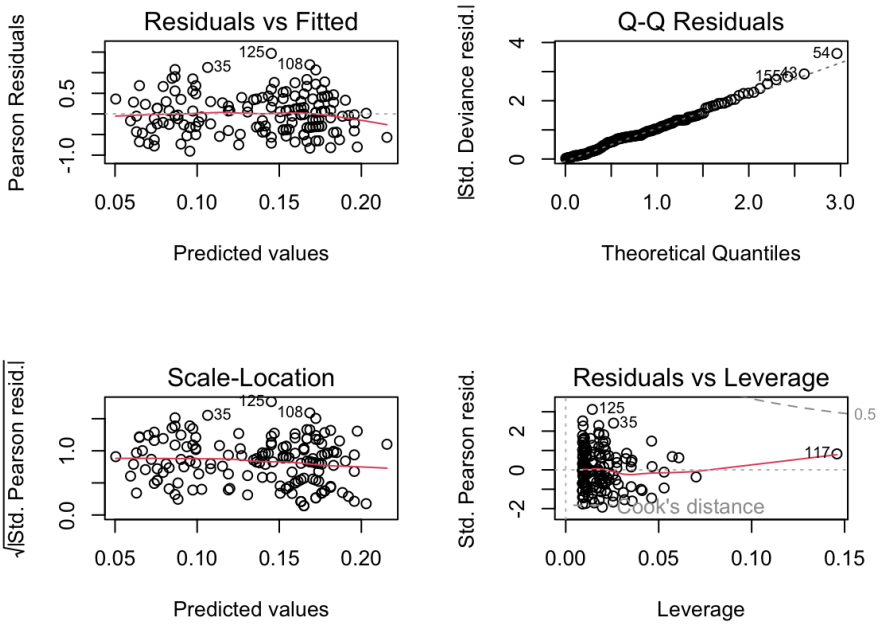


Figure 3: Diagnostic plots `fit1`

	Model	Deviance	AIC
1	fit1	43.83992	879.8940
2	fit2	43.64296	881.1275

Table 2: Comparison `fit1` and `fit2`

Furthermore, we are interested in analysing the dependence of the treatment effect on the BMI. We plot the expected increase in heart rate for individuals receiving stimulant and placebo as a function of BMI, Figure:4. An interesting fact to observe is that the treatment seems to have more effect on individuals with higher BMIs. Indeed, the gap between the two slopes tends to increase as the BMI increases. The red slope, corresponding to the treatment group, tends to grow 'exponentially', while the blue slope, corresponding to the placebo group, tends to grow 'linearly'. The higher the BMI is, the more effective the stimulant is.

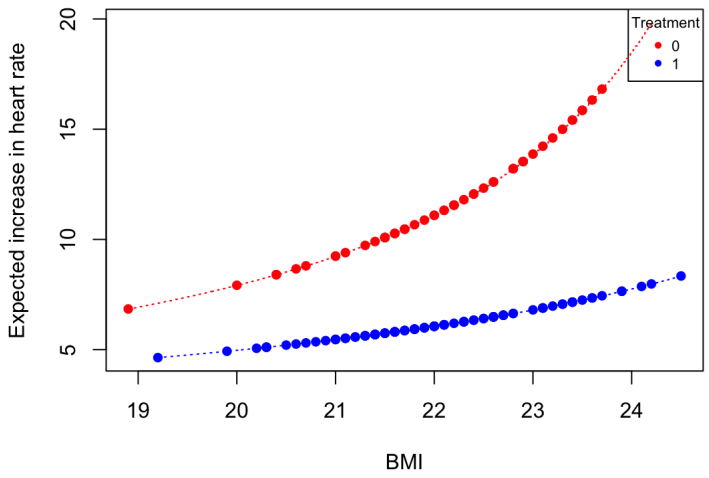


Figure 4: Expected increase in heart rate as a function of BMI

3 Evaluation

For the given sample size, we evaluate the power of model `fit1` to detect treatment effects in the range 0 to 10 b/min. We perform 1000 simulations with data simulated assuming `fit1` is correct. For simplicity, we assume a uniform treatment effect: the post-treatment increase in heart rate of all treated subjects is increased by the assumed treatment effect, independently of the covariate values. The results are plotted on Figure:5.
From the previous analysis, we know that for the typical BMI of $21.6\,kg/m^2$, there is evidence of a treatment effect of around 4.47 b/min, with a 95% confidence interval of $(3.60, 5.86)$ b/min. This means that between our two selected individuals, post-treatment, the treated one has a heart rate 4.47 b/min higher than the one in the placebo group. Hence, Figure:5 demonstrates that `fit1` has sufficient power to detect the size of treatment effect that might be expected here.

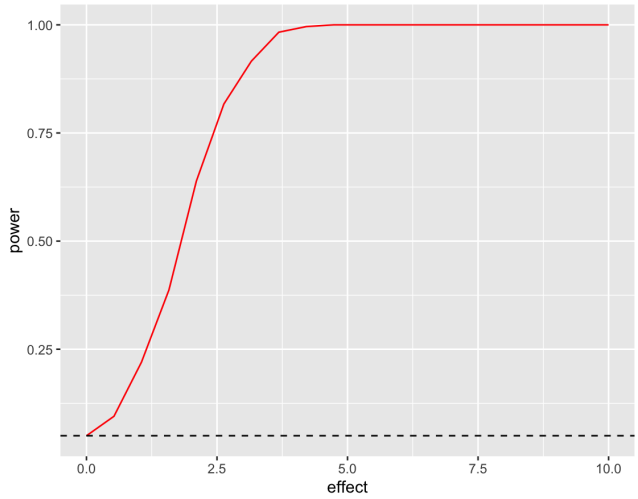


Figure 5: Power as a function of the strength of treatment effect for `fit1`

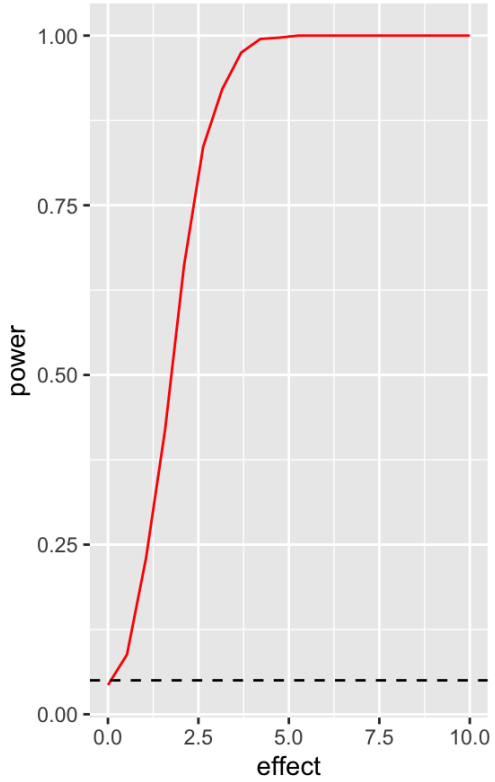
4 Summary

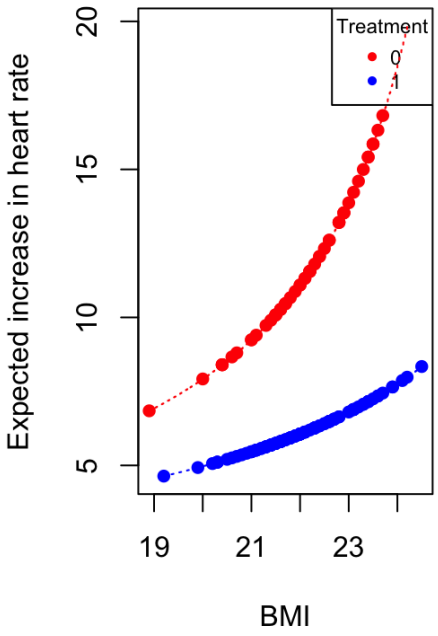
The initial linear model is not appropriate for the problem: the distributions of the observed variables being not well estimated. The statistician's model is more appropriate and fits the given data well. Using this model, for a typical BMI of $21.6\,kg/m^2$, there is evidence of a treatment effect of around 4.47 beats/min with a 95% confidence interval of $(3.60, 5.86)$ beats/min. This means that 'on average', for two individuals with a BMI of $21.6\,kg/m^2$, after treatment we would expect the heart rate to be 4.47 beats/min higher for a treated individual than an individual who received the placebo. As an indicative result, for general BMI, the difference would be on average around 6.2 beats/min. As observed on the first graph below, the statistician's model has sufficient power to detect the size of a treatment effect over about 3 beats/min, and, thus, sufficient power to detect the treatment effect that might be expected here.

Additionally, a noteworthy result, illustrated on the second graph below, is that the treatment effect appears to increase as the BMI of the individuals increases (exponential tendency for treatment slope and the gap between the slopes increases).

In conclusion, the stimulant treatment appears to perform well, with reasonable increases in heart rate. However, additional precautions should be taken for individual with a high BMI.

A suggestion for a more detailed dataset, would be to also take into account other features such as: age, sex, health (e.g. sporty, smoker). Finally, it would be recommended to have the same number of individuals in each group (placebo/treatment).





5 Appendix

```r
# Build X matrix
data <- dat[, c(3,4)]
X_bt <- data.matrix(data)

# Initialise beta vector parameters
epsilon = 1e-9
beta_4 <- rep(epsilon, 3)

# stimulated_pulse - rest_pulse
y = dat[, 2] - dat[, 1]

# IWLS
for (i in 1:250){
  eta_4 <- cbind(1,X_bt)%*%beta_4 # Estimated linear predictor
  mu_4 <- 1 / eta_4 # Estimated mean response
  z_4 <-  eta_4 + (mu_4 - y)/mu_4^2 # Form the adjusted dependent variable in terms of y and mu
  w_4 <-  mu_4^2 # Compute weights in terms of mu
  lmod <- lm(z_4 ~ X_bt, weights = w_4) # Regress z on x with weights w.
  beta_4 <- as.numeric(lmod$coeff) # New beta
}
```

Figure 6: IWLS algorithm for `fit1`

```r
data <- dat[, c(3,4)]
# Add predictors
data$prod <- data$bmi * as.numeric(as.character(data$treatment))
X_aug <- data.matrix(data)

# Initialise beta, vector parameters
epsilon = 1e-9
beta_5 <- rep(epsilon, 4)

# stimulated_pulse - rest_pulse
y = dat[, 2] - dat[, 1]

# IWLS
for (i in 1:250){
  eta_5 <- cbind(1,X_aug)%*%beta_5 # Estimated linear predictor
  mu_5 <- 1 / eta_5 # Estimated mean response
  z_5 <-  eta_5 + (mu_5 - y)/mu_5^2 # Form the adjusted dependent variable in terms of y and mu
  w_5 <-  mu_5^2 # Compute weights in terms of mu
  lmod <- lm(z_5 ~ X_aug, weights = w_5) # Regress z on x with weights w.
  beta_5 <- as.numeric(lmod$coeff) # New beta
}
```

Figure 7: IWLS algorithm for `fit2`

```r
  # 7. For the clinician's model
plot(dat$bmi, mu_4, xlab = "BMI", ylab = "Expected increase in heart rate",
     col = treatment_colors[as.numeric(as.character(dat$treatment)) + 1], pch=16)

# Add legend for treatment groups
legend("topright", legend = levels(dat$treatment), col = treatment_colors,
       pch = 16, cex=0.7, title = "Treatment")

# Add guide lines
library(KernSmooth)
LR.fit0 <- locpoly(dat$bmi[dat$treatment == 0], mu_4[dat$treatment == 0], degree = 3,
                   bandwidth = 0.5,
                   kernel = "normal")
LR.fit1 <- locpoly(dat$bmi[dat$treatment == 1], mu_4[dat$treatment == 1], degree = 3,
                   bandwidth = 0.5,
                   kernel = "normal")

lines(LR.fit0$x, LR.fit0$y, type = "s", col = "red", lwd = 1, lty = 3)
lines(LR.fit1$x, LR.fit1$y, type = "s", col = "blue", lwd = 1, lty = 3)
```

Figure 8: Expected increase in heart rate according to BMI for `fit1`

```r
  # 8. Simulation study
n_sim <- 1000
treat_effs <- seq(0, 10, length.out = 20)
power <- matrix(0, nrow = length(treat_effs), ncol = 1)
j <- 1

for(treat_eff in treat_effs){
  p_store <- matrix(0, nrow = n_sim, ncol = 1)
  for(i in 1:n_sim){
    dat$diff_sim <- sample(unlist(simulate(fit1)))
    wh_treat <- dat$treatment == 0
    dat$diff_sim[wh_treat] <- dat$diff_sim[wh_treat] + treat_eff
    fit_sim1 <- glm(diff_sim ~ bmi + treatment, data=dat, family = Gamma(link = "inverse"))
    sum1 <- summary(fit_sim1)
    p_store[i,1] <- sum1$coefficients[3,4]
  }

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


power_long <- data.frame(
  effect = treat_effs,
  power = power
)

power_plot <- ggplot(data = power_long, mapping = aes(x = effect,
                                                      y = power) ) +
                    geom_line(col = 'red') +
                    geom_hline(yintercept = 0.05, lty = 2)

power_plot
```

Figure 9: Power as a function of the strength of treatment effect for `fit1`