

# Homework 1

Güney Berkay Ateş  
120200047

The aim was to model the relationship between the dependent variable 'heart.disease' and the independent variable 'biking'. The `lm` function was used to fit the regression model, and the summary function was employed to obtain detailed statistics, including coefficients and p-values. Predictions based on the model were generated using the prediction function. Additionally, a scatter plot was created to visualize the relationship between 'biking' and 'heart.disease', and the regression line was overlaid for a clearer understanding of the trend. This analysis serves as a basic exploration of the association between biking habits and the occurrence of heart disease in the given dataset.

The model equation is expressed as  $\text{heart.disease} = 17.70 - 0.20 \times \text{biking}$ , indicating that for every unit increase in 'biking', the 'heart.disease' is expected to decrease by 0.20 units. The coefficient for 'biking' is highly significant (p-value <  $2e-16$ ), suggesting a robust relationship. The residuals, representing the differences between observed and predicted values, exhibit a relatively small spread, with a residual standard error of 1.618. The multiple R-squared value of 0.8751 indicates that approximately 87.5% of the variability in 'heart.disease' can be explained by the linear relationship with 'biking'. The F-statistic (3474) is highly significant (p-value <  $2.2e-16$ ), supporting the overall significance of the model. In summary, the results suggest a strong and statistically significant inverse association between biking and the incidence of heart disease in the given dataset.

Code:

```
# Assuming your dataset is named 'heart'

# Fit the simple regression model
simple_model <- lm(heart.disease ~ biking, data = heart)

# Display the summary of the regression model
summary(simple_model)

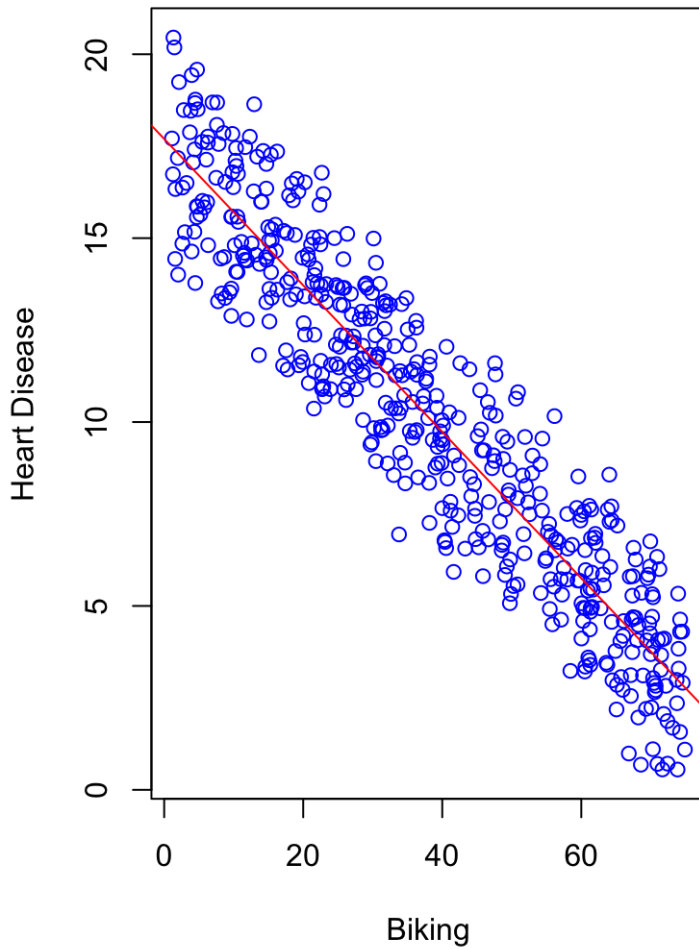
# Make predictions
predictions <- predict(simple_model, newdata = heart)
```

```
# Visualize the relationship
plot(heart$biking, heart$heart.disease, main = "Simple Regression",
     xlab = "Biking", ylab = "Heart Disease", col = "blue")
abline(simple_model, col = "red")
```

Results:

Plots:

## Simple Regression



:

Console:

Call:

```
lm(formula = heart.disease ~ biking, data = heart)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.028	-1.206	-0.004	1.151	3.643

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	17.697884	0.146780	120.57	<2e-16 ***
biking	-0.199091	0.003378	-58.94	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.618 on 496 degrees of freedom

Multiple R-squared: 0.8751, Adjusted R-squared: 0.8748

F-statistic: 3474 on 1 and 496 DF, p-value: < 2.2e-16