ST552 HOMEWORK 1

Brian Cervantes Alvarez
January 19, 2024

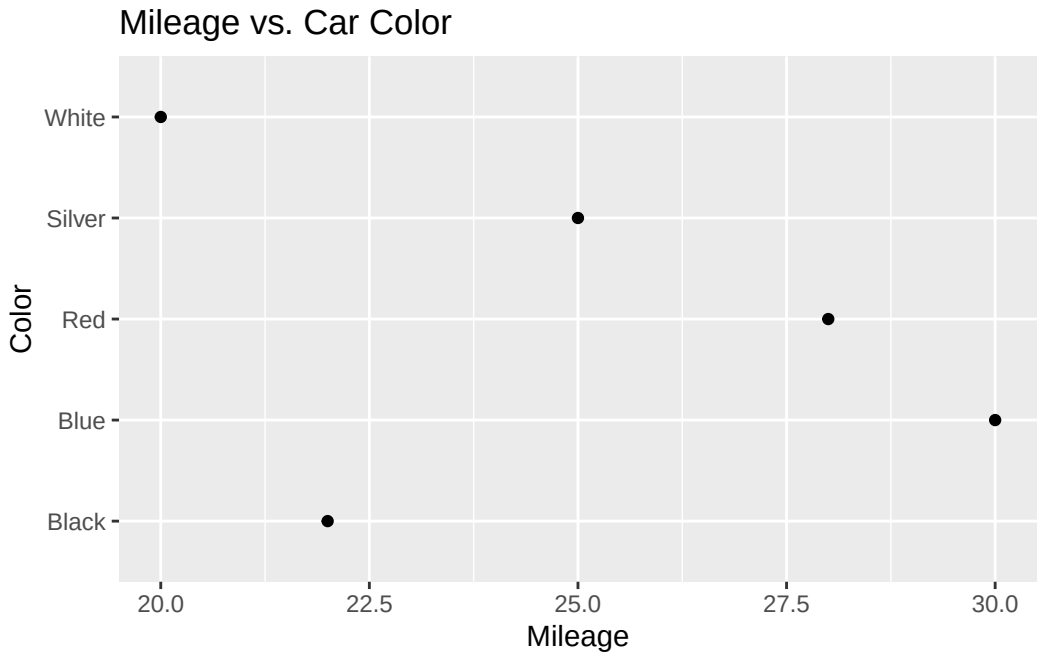
Problem 1:

Part A: Writing reproducible code

```
# Create a random dataset
set.seed(123)
carDs <- data.frame(
  name = c("Toyota", "Honda", "Ford", "Chevrolet", "BMW"),
  mileage = c(30, 28, 25, 22, 20),
  color = c("Blue", "Red", "Silver", "Black", "White"),
  transmission = factor(c("Automatic", "Manual", "Automatic", "Automatic", "Manual"))
)
```

Part B: Make A Plot

```
library(ggplot2)
ggplot(carDs, aes(x = mileage, y = color)) +
  geom_point() +
  labs(title = "Mileage vs. Car Color",
       x = "Mileage",
       y = "Color")
```



Part C: Load Mammals Data and Fit Linear Regression Model

```r
library(MASS)
data("mammals")
model <- lm(body ~ brain, data = mammals)
summary(model)
```

```
Call:
lm(formula = body ~ brain, data = mammals)

Residuals:
     Min       1Q   Median       3Q      Max
-1552.25    -8.00    47.36    55.10  1553.42

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -56.85555   42.97805  -1.323    0.191
brain         0.90291    0.04453  20.278   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 323.5 on 60 degrees of freedom
Multiple R-squared:  0.8727,    Adjusted R-squared:  0.8705
F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

Part D: Extract The Estimate of The Intercept

```r
interceptEstimate <- coef(model)[1]
print(paste0("The estimate of  0 is: ", interceptEstimate))
```

```
[1] "The estimate of  0 is: -56.8555454285966"
```

Part E: Extract the estimate of sigma

```r
stdEstimate <- summary(model)$sigma
print(paste0("The estimate of  is: ", stdEstimate))
```

```
[1] "The estimate of  is: 323.522268021538"
```

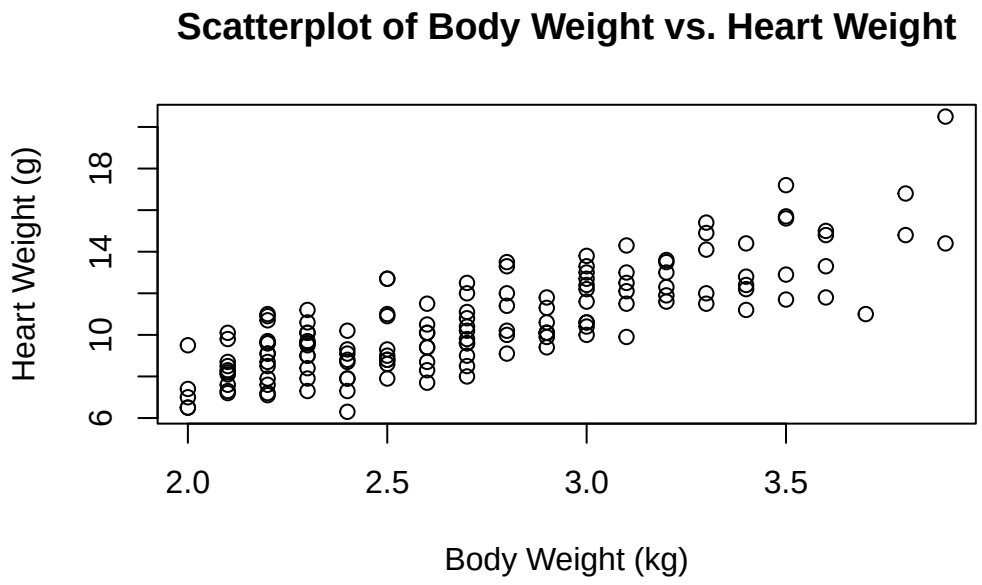
Problem 2: Simple Linear Regression

Part A: Explanatory Variable and Response

The explanatory variable is Body weight (Bwt), and the response variable is Heart weight (Hwt).

Part B: Scatterplot and Description

```
data("cats")
plot(cats$Bwt, cats$Hwt, main = "Scatterplot of Body Weight vs. Heart Weight",
     xlab = "Body Weight (kg)", ylab = "Heart Weight (g)")
```



Part C: Appropriateness of Linear Regression

The plot demonstrates that a linear regression model could be appropriate as there appears to be a linear trend between body weight and heart weight.

Part D: Fit Linear Regression Model and Add Regression Line

```r
# Fit linear regression model
catModel <- lm(Hwt ~ Bwt, data = cats)

summary(catModel)
```

```
Call:
lm(formula = Hwt ~ Bwt, data = cats)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5694 -0.9634 -0.0921  1.0426  5.1238

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3567     0.6923  -0.515    0.607
Bwt           4.0341     0.2503  16.119   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.452 on 142 degrees of freedom
Multiple R-squared:  0.6466,    Adjusted R-squared:  0.6441
F-statistic: 259.8 on 1 and 142 DF,  p-value: < 2.2e-16
```

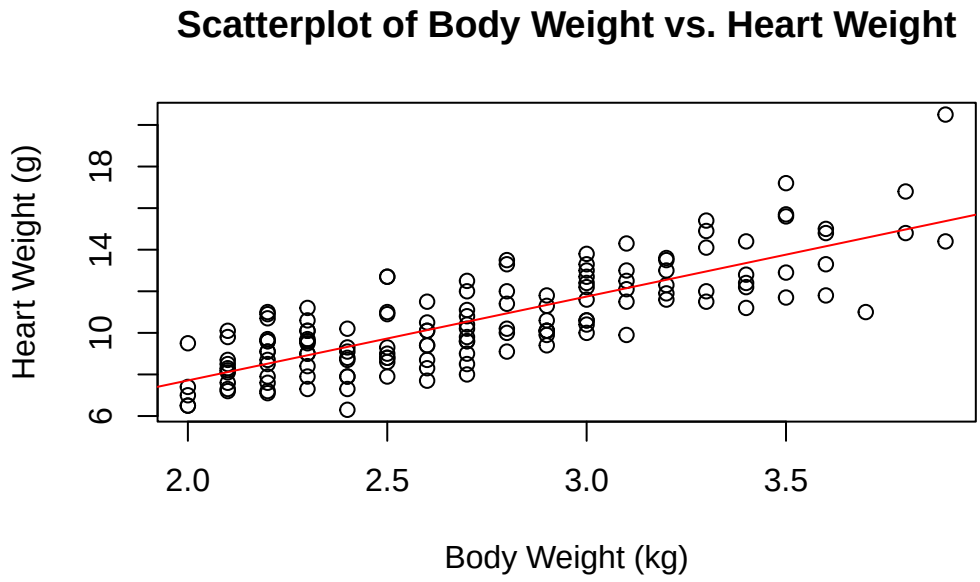
```r
# Extract parameter estimates
intEstimate <- coef(catModel)[1]
slopeEstimate <- coef(catModel)[2]

print(paste0("Intercept estimate: ", round(intEstimate, 4)))
```

```
[1] "Intercept estimate: -0.3567"
```

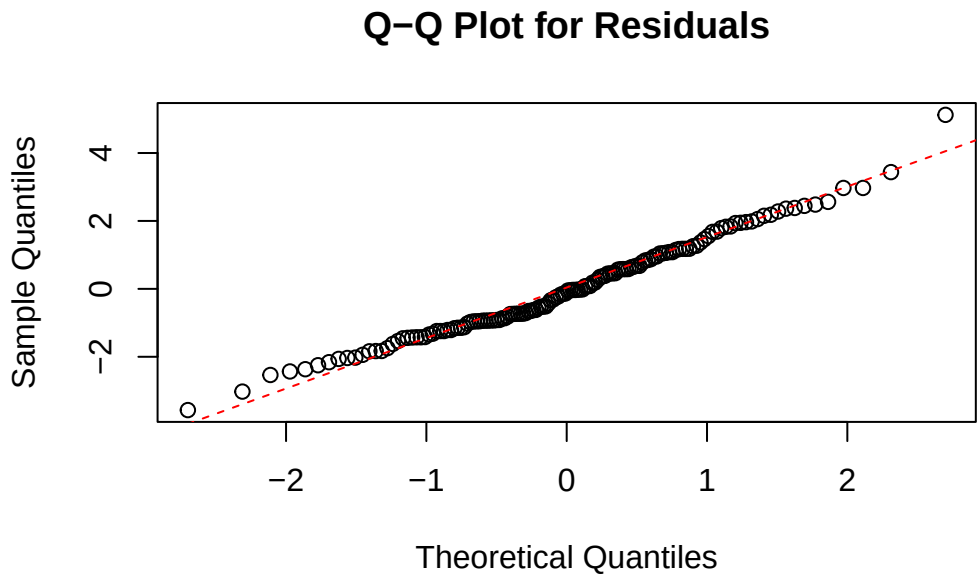
```r
print(paste0("Slope estimate: ", round(slopeEstimate, 4)))
```

```
[1] "Slope estimate: 4.0341"
```

```r
# Add regression line to the scatterplot
plot(cats$Bwt, cats$Hwt, main = "Scatterplot of Body Weight vs. Heart Weight",
     xlab = "Body Weight (kg)", ylab = "Heart Weight (g)")
abline(catModel, col = "red")
```




```
# Create Q-Q plot for the residuals
qqnorm(residuals(catModel), main = "Q-Q Plot for Residuals")
qqline(residuals(catModel), col = "red", lty = 2) # Reference Line
```



Part E: Conclusion About The Relationship

The linear regression analysis indicates a statistically significant positive relationship between cat body weight and heart weight, with a slope estimate of 4.0341. Approximately 64.41% of the variability in heart weight can be explained by the linear relationship with body weight.

Part F: Hypothesis Test for Slope Coefficient

The highly significant slope estimate of 4.0341 in the linear regression model with the $pval < 2.2e-16$ indicates a substantial and practically meaningful positive relationship between cat body weight and heart weight.

```
# Test hypothesis
test <- summary(catModel)$coefficients[2, c("Estimate",
                                            "Std. Error",
                                            "t value",
                                            "Pr(>|t|)")]
results <- data.frame(
  Estimate = test["Estimate"],
  `Std. Error` = test["Std. Error"],
  `t value` = test["t value"],
  `Pr(>|t|)` = test["Pr(>|t|)"]
)
results
```

```
         Estimate Std..Error  t.value     Pr...t..
Estimate 4.034063  0.2502615 16.11939 6.969045e-34
```

Part G: Predicted Mean Heart Weight for a Cat with Body Weight 3.5 Kilograms

```
# Predict mean heart weight
predHeartWeight <- predict(catModel,
                           data.frame(Bwt = 3.5),
                           interval = "confidence")
print(paste0("Predicted mean heart weight for a cat with body weight 3.5kg: ",
    round(predHeartWeight[1],2),
    "g"))
```

```
[1] "Predicted mean heart weight for a cat with body weight 3.5kg: 13.76g"
```