# Math 133 Group Work 1

Pranav Jayakumar

January 27, 2025

**Abstract**

In this assignment, we analyze the relationship between the number of hot wings a person buys vs how much beer they drink

## 1 Data Analysis

### 1.1 Data Visualization

We will create a scatterplot with `Hotwings` as the y-axis and `Beer` as the x-axis.

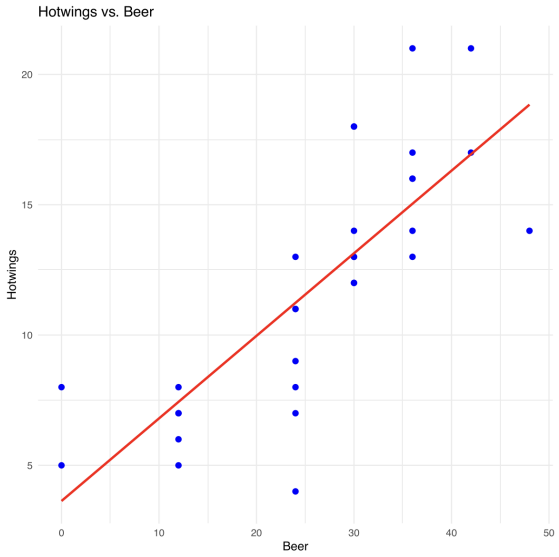


Figure 1: Scatterplot of Hotwings vs. Beer

## 1.2 Fitting a Linear Model

We will now fit a linear model `Hotwings ~ Beer`. We will use an 80-20 train-test split.

```r
fit_linear_model <- function(y, x, raw_data) {

  # train test split
  n <- nrow(raw_data)
  trainIndex <- sample(n, round(0.8 * n, 0))
  train <- raw_data[trainIndex, ]
  test <- raw_data[-trainIndex, ]

  # construct formula
  formula <- as.formula(paste(y, "~", x))

  # fit model
  model <- lm(formula, data = train)

  # Store model summary
  model_summary <- summary(model)
  print(model_summary)  # This will print the summary when the
      function runs

  # predict on testing data
  y_test <- test[[y]]
  y_hat <- predict(model, newdata = test)

  # analyze accuracy
  SSE <- sum((y_test - y_hat)^2)
  MSE <- SSE / nrow(test)
  RMSE <- sqrt(MSE)
  SST <- sum((y_test - mean(y_test))^2)
  R2 <- 1 - SSE / SST

  return(list(
    summary = model_summary,  # Include summary in return value
    SSE = SSE,
    MSE = MSE,
    RMSE = RMSE,
    SST = SST,
    R2 = R2
  ))
}
```

We will now show our linear model in the form $\hat{y} = \beta_0 + \beta_1 x$

$$\hat{y} = 3.41405 + 0.31241x$$

We find that the linear model returns a root mean squared error of RMSE $= 2.8019$.

### 1.3 Scatterplot of Hotwings vs. Beer with Gender

We will now add a color aesthetic "Gender" to our scatterplot. We observe that the trendlines

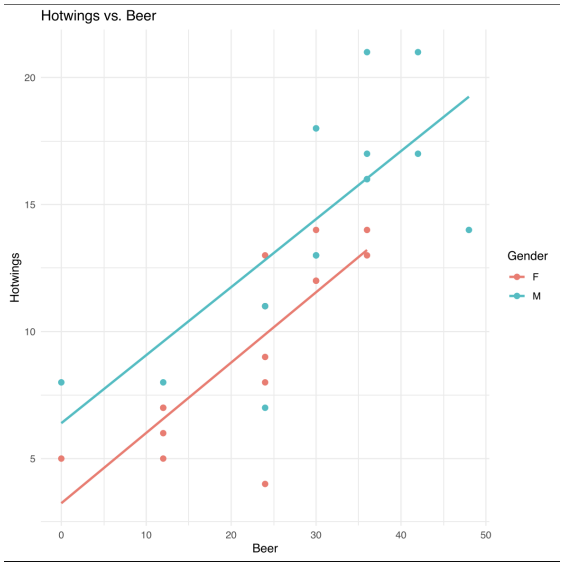


Figure 2: Scatterplot of Hotwings vs. Beer grouped by Gender

for men and women appear to have similar slopes, while the y-intercept for men is higher than that of women