Homework 2

Code

2023-05-02

Hide

```
library(dplyr)
```

```

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

Hide

```
setwd("/Users/victoriaritorto/Documents/SODA496")
Senate_Results <- read.csv("senate_results.csv")
```

Hide

```
# Split the data into training and testing sets
set.seed(1234)
train_index <- sample(nrow(Senate_Results), 0.8 * nrow(Senate_Results))
train_data <- Senate_Results[train_index, ]
test_data <- Senate_Results[-train_index, ]
```

Hide

```
Senate_Results$candidatevotes <- as.factor(Senate_Results$candidatevotes)
train_data$party_simplified <- as.numeric(train_data$party_simplified)
```

Hide

```
# Fit a linear regression model to the training data
model <- lm(candidatevotes ~ party_simplified, data = train_data)
summary(model)
```

```
Call:
lm(formula = candidatevotes ~ party_simplified, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1716.98  -969.90   -75.55   803.99  2134.88

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       1908.27      98.04  19.465  < 2e-16 ***
party_simplified  -172.29      34.01  -5.065 4.89e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1042 on 958 degrees of freedom
Multiple R-squared:  0.02608,   Adjusted R-squared:  0.02507
F-statistic: 25.66 on 1 and 958 DF,  p-value: 4.891e-07
```

Hide

```
# Predict the target variable for the test data using the model
predictions <- predict(model, type = 'response')
```