Assignment 3

1. A)

Table 1

|  |  |
| --- | --- |
| Dependent Variable: Turnout | |
| Voting compulsory  (reference = no) | 15.78819(2.79)\*\*\* |
| Population in million | 0.02(0.045) |
| Constant | 65.066(1.18)\*\*\* |
| N | 194 |
| Adjusted R-Squared | 0.135 |
| Significance | 3.465e-07 |

\**p*<0.05, \*\**p*<0.01, \*\*\**p*<0.001

Table 1 shows the predicted values from model 1, in this model compulsory voting was predicted to have a statistically significant relationship with voter turnout where as the population did not. Therefore, compulsory voting has a much larger effect on electoral turnout than population does.

The overall model was statistically significant as the p-value was <0.001 showing that the model was overall very significant. The constant was 65.066 meaning that if the both the independent variables were at 0 the turnout would be at 65.066, however this is likely skewed as the “no” in the compulsory voting variable was coded to 0.

For compulsory voting, for every 1 unit it increases (in this case shifting from no to yes) turnout percentage is predicted to increase by 15.78819 and for population for every 1 unit it increases the turnout percentage is predicted to increase by 0.02.

The adjusted R-squared of model 1 is 0.135 meaning that it accounts for 1.35% of the actual data.

B)A white rectangular object with black text

Description automatically generated

The population in millions is not a significant relationship with voter turnout as it’s coefficient is on the zero line. Compulsory voting does have a statistically significant relationship as the confidence intervals do not cross the zero line.

A graph with text on it

Description automatically generated

Figure 1.2 shows the relationship between each category of compulsory voting and electoral turnout in percent. It displays that the model predicts on average countries with compulsory voting have around 15% greater turnout than those that do not have compulsory voting.

A graph with a line and numbers

Description automatically generated

Figure 1.3 displays the model’s predicted change in electoral turnout as population in millions increases. It displays that as the population increases the predicted turnout has a much larger variance. The predicted turnout doesn’t have a large enough change as the coefficient line only changes marginally with a far greater variance therefore it is not possible to make concrete predictions and state whether population has an effect on turnout.

1. A) A graph of a graph

   Description automatically generated

Figure 2.1 is a QQ plot of the residuals from model 1. Figure 2.1 is normally distributed as the points are roughly along a straight line while the ends slightly deviate from the line which is common for normal distribution.

B)

A diagram of a graph

Description automatically generated

Figure 2.2 displays large variance along the y-axis and the residuals are not clustered around the zero line. There is not constant variance as the values clustered to the left have much greater variance than those clustered to the right. However, the mean of the values will be around zero meaning there is no bias so the assumption that there is zero bias is not violated.

C) Breusch-Pagan test

data: model1

BP = 1.5976, df = 2, p-value = 0.4499

The Breusch-Pagan test above is not statistically significant as the p-value is > 0.05, this means that the variance is constant and that there is homoscedasticity.

A graph of a person with a number of numbers

Description automatically generated with medium confidence

Figure 3.1 shows a histogram of the outliers from model 1 which has normal distribution with no outliers.

A graph of population

Description automatically generated

Figure 3.2 is a histogram of population from the dataset, again there are no major outliers and the data is skewed to the left.

A graph of a number of dots

Description automatically generated with medium confidence

Figure 3.3 shows a scatter plot of the residuals from Model 1 and the population independent variable. The data is clustered to the left but there are no major outliers in the plot.

A graph of a graph of a voting

Description automatically generated with medium confidence

Figure 3.4 displays a histogram of compulsory voting from the dataset. This distribution is odd due to the compulsory voting independent variable being binary therefore a figure 3.4 has a bimodal distribution with slight skewing to the left as there were more countries without compulsory voting than with it.

Overall, from figures 3.1, 3.2, 3.3 and 3.4 there are no major outliers in the independent variables or the residuals.

Appendix:

rm(list=ls())

setwd("/Users/kieranmcavoy/Documents/Data Science/Quant 1/Assignment 3")

Idea <- read.csv("idea.csv")

#Question 1a

table(Idea$Compulsory.voting)

table(Idea$Population\_in\_million)

class(Idea$Compulsory.voting)

class(Idea$Population\_in\_million)

#Change the compulsory voting category numeric

Idea$Compulsory <- NA

Idea$Compulsory[(Idea$Compulsory.voting == "No")] <- 0

Idea$Compulsory[(Idea$Compulsory.voting == "Yes")] <- 1

class(Idea$Compulsory.voting)

Idea$factorcomp <- as.factor(Idea$Compulsory.voting)

model1 <- lm(formula = turnout ~ factorcomp + Population\_in\_million, data = Idea)

summary(model1)

nrow(model.frame(model1))

model2 <- lm(formula = turnout ~ Compulsory + Population\_in\_million, data = Idea)

summary(model2)

#Question 1b

install.packages("margins")

library(margins)

#Coefficient Plot

margins.model2 <- margins(model2)

plot(margins.model2, xlab = "Independent Variables", main = "Figure 1.1 Coefficient Plot for Model 1")

#Fitted Values

cplot(model1, "factorcomp",

xlab = "Compulsory Voting",

ylab = "Turnout(%)",

main = "Figure 1.2 Fitted Values of Effect of Compulosory Voting on Turnout Percentage")

cplot(model1,"Population\_in\_million",

xlab = "Population in Million",

ylab = "Turnout(%)",

main = "Figure 1.3 Fitted Values Effects of Population on Turnout")

#Question 2A The residuals are (roughly) normally distributed

#QQ Plot

qqnorm(model1$residuals, main = "Figure 2.1: QQ Plot of The Residuals in Model 1")

qqline(model1$residuals,col="red")

#Question 2B The error term has zero bias

plot(y=model1$residuals, x=model1$fitted.values,

xlab="Fitted Values", ylab="Residuals", main = "Figure 2.2: Model 1 Residuals Against Fitted Values")

abline(a=0, b=0, col="red")

#The error variance is uniform (no heteroscedasticity)

#Breusch Pagan Test

install.packages("lmtest")

library(lmtest)

bptest(model1, studentize=FALSE)

#Question 3

hist(model1$residuals, xlab = "Model 1 Residuals", main = "Figure 3.1 Histogram of Model 1 Residuals") #Check outliers of the residuals

hist(Idea$Population\_in\_million, xlab = "Population in Millions", main = "Figure 3.2 Histogram of Population")

hist(Idea$Compulsory, xlab = "Compulsory Voting", main = "Figure 3.4 Histogram of Compulsory Voting")

plot(Idea$Population\_in\_million, model1$residuals, xlab = "Population in Millions", ylab = "Model 1 Residuals", main = "Figure 3.3 plot of Model 1 Residuals and Population")