PS03 Answers

kg

11/9/2021

## QUESTION 1

**We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent’s vote share.**

**a. Run a regression where the outcome variable is voteshare and the explanatory variable is difflog.**

outcome variable = voteshare

explanatory variable = difflog

#x explains y, and x = explanator. y~x  
reg1=lm(data = data, voteshare ~ difflog)

1. **Make a scatterplot of the two variables and add the regression line.**

library(ggplot2)  
ggplot(data, aes(difflog, voteshare)) +  
 geom\_point(alpha = 0.5) + #add a scatterplot  
 geom\_smooth(method = "lm")

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

**3. Save the residuals of the module in a separate object.**

summary(data$voteshare)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.3476 0.5846 0.6569 0.6552 0.7233 0.9930

#voteshare is a value between 0 and 1 so assuming percentage  
  
plot1=ggplot(data, aes(difflog, voteshare)) +  
 geom\_point(alpha = 0.5) + #add a scatterplot  
 geom\_smooth(method = "lm")  
  
residuals1=summary(reg1)$residuals

1. **Write the prediction equation.**

summary(reg1)

##   
## Call:  
## lm(formula = voteshare ~ difflog, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.26832 -0.05345 -0.00377 0.04780 0.32749   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.579031 0.002251 257.19 <2e-16 \*\*\*  
## difflog 0.041666 0.000968 43.04 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07867 on 3191 degrees of freedom  
## Multiple R-squared: 0.3673, Adjusted R-squared: 0.3671   
## F-statistic: 1853 on 1 and 3191 DF, p-value: < 2.2e-16

= 0.579031

= 0.041666

= 0.579031 + 0.041666

## QUESTION 2

(## makes title)

**We are interested in knowing how the difference between incumbent and challenger’s spending and the vote share of the presidential candidate of the incumbent’s party are related.**

1. **Run a regression where the outcome variable is presvote and the explanatory variable is difflog.**

outcome = presvote

explanatory = difflog

reg2=lm(data = data, presvote ~ difflog)

1. **Make a scatterplot of the two variables and add the regression line.**

plot2= ggplot(data, aes(difflog, presvote)) +  
 geom\_point(alpha = 0.5) + #add a scatterplot  
 geom\_smooth(method = "lm")  
plot2

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

1. **Save the residuals of the model in a separate object.**

residuals2=summary(reg2)$residuals

1. **Write the prediction equation.**

summary(reg2)

##   
## Call:  
## lm(formula = presvote ~ difflog, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32196 -0.07407 -0.00102 0.07151 0.42743   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.507583 0.003161 160.60 <2e-16 \*\*\*  
## difflog 0.023837 0.001359 17.54 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1104 on 3191 degrees of freedom  
## Multiple R-squared: 0.08795, Adjusted R-squared: 0.08767   
## F-statistic: 307.7 on 1 and 3191 DF, p-value: < 2.2e-16

*=*

=

## QUESTION 3

**We are interested in knowing how the vote share of the presidential candidate of the incum- bent’s party is associated with the incumbent’s electoral success.**

**1. Run a regression where the outcome variable is voteshare and the explanatory variable is presvote.**

reg3<- lm(data = data, voteshare ~ presvote)

1. **Make a scatterplot of the two variables and add the regression line.**

plot3<- ggplot(data, aes(presvote, voteshare)) +  
 geom\_point(alpha = 0.5) + #add a scatterplot  
 geom\_smooth(method = "lm") #add line  
plot3

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

1. **Write the prediction equation.**

summary(reg3)

##   
## Call:  
## lm(formula = voteshare ~ presvote, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.27330 -0.05888 0.00394 0.06148 0.41365   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.441330 0.007599 58.08 <2e-16 \*\*\*  
## presvote 0.388018 0.013493 28.76 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08815 on 3191 degrees of freedom  
## Multiple R-squared: 0.2058, Adjusted R-squared: 0.2056   
## F-statistic: 827 on 1 and 3191 DF, p-value: < 2.2e-16

*=*

=

## 

## QUESTION 4

**The residuals from part (a) tell us how much of the variation in voteshare is not explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in presvote is not explained by the difference in spending between incumbent and challenger in the district.**

1. **Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.**

outcome = residuals1

explanatory = residuals2

reg4<-lm(residuals1 ~ residuals2)  
reg4

##   
## Call:  
## lm(formula = residuals1 ~ residuals2)  
##   
## Coefficients:  
## (Intercept) residuals2   
## -5.207e-18 2.569e-01

1. **Make a scatterplot of the two residuals and add the regression line.**

plot4<- ggplot(data, aes(residuals2, residuals1)) +  
 geom\_point(alpha = 0.5) + #add a scatterplot  
 geom\_smooth(method = "lm") #add line  
plot4

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

1. **Write the prediction equation.**

summary(reg4)

##   
## Call:  
## lm(formula = residuals1 ~ residuals2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.25928 -0.04737 -0.00121 0.04618 0.33126   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.207e-18 1.299e-03 0.00 1   
## residuals2 2.569e-01 1.176e-02 21.84 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07338 on 3191 degrees of freedom  
## Multiple R-squared: 0.13, Adjusted R-squared: 0.1298   
## F-statistic: 477 on 1 and 3191 DF, p-value: < 2.2e-16

*=*

=

## 

## QUESTION 5

**What if the incumbent’s vote share is affected by both the president’s popularity and the difference in spending between incumbent and challenger?**

1. **Run a regression where the outcome variable is the incumbent’s voteshare and the explanatory variables are difflog and presvote.**

outcome = voteshare

explanatory = difflog , presvote

reg5= lm(data = data, voteshare ~ difflog+presvote)

1. **Write the prediction equation.**

reg5$coefficients

## (Intercept) difflog presvote   
## 0.44864422 0.03554309 0.25687701

*=*

=

=

1. **What is it in this output that is identical to the output in Question 4? Why do you think this is the case? Look at the summary to get information about the lines.**

summary(reg4)

##   
## Call:  
## lm(formula = residuals1 ~ residuals2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.25928 -0.04737 -0.00121 0.04618 0.33126   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.207e-18 1.299e-03 0.00 1   
## residuals2 2.569e-01 1.176e-02 21.84 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07338 on 3191 degrees of freedom  
## Multiple R-squared: 0.13, Adjusted R-squared: 0.1298   
## F-statistic: 477 on 1 and 3191 DF, p-value: < 2.2e-16

summary(reg5)

##   
## Call:  
## lm(formula = voteshare ~ difflog + presvote, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.25928 -0.04737 -0.00121 0.04618 0.33126   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.4486442 0.0063297 70.88 <2e-16 \*\*\*  
## difflog 0.0355431 0.0009455 37.59 <2e-16 \*\*\*  
## presvote 0.2568770 0.0117637 21.84 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07339 on 3190 degrees of freedom  
## Multiple R-squared: 0.4496, Adjusted R-squared: 0.4493   
## F-statistic: 1303 on 2 and 3190 DF, p-value: < 2.2e-16

I tried plotting to see what information visualization would give me.

par(mfrow=c(1,2))  
  
plot(reg4)

Chart

Description automatically generatedChart, scatter chart

Description automatically generated

plot(reg5)

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

The residuals of reg4 and reg5 are the same.

I think the correlation is related to the sum of least squares rule in statistical analysis, which says that the line of best fit in the data minimizes the sum of the squared residuals.

sum(resid(reg4)^2)

## [1] 17.18224

sum(resid(reg5)^2)

## [1] 17.18224

Since they both are equal then they are equally well-fitted to the line. This suggests that the residuals. I think this is because a multivariate analysis of voteshare, presvote, and difflog is a faster way of getting a line of best fit, and the slower way is doing the lines individually as was done in Question 1 and Question 2.