Data Exercise 1 (PS813)

Lotem Bassan

1/27/2019

Uploading dataset in STATA

I begin by uploading the dataset using the net install command in STATA:

```
#net install PS813_EX1, from(https://faculty.polisci.wisc.edu/weimer)

##checking PS813_EX1 consistency and verifying not already installed...
##installing into u:\ado\plus\...
##installation complete.

#PS813_EX1 3451

#saveold "\\Desktop\Grad School\courses\PS813\Rwork\PS1.dta"
##(saving in Stata 12 format, which can be read by Stata 11 or 12)
##file \\Desktop\Grad School\courses\PS813\Rwork\PS1.dta saved
```
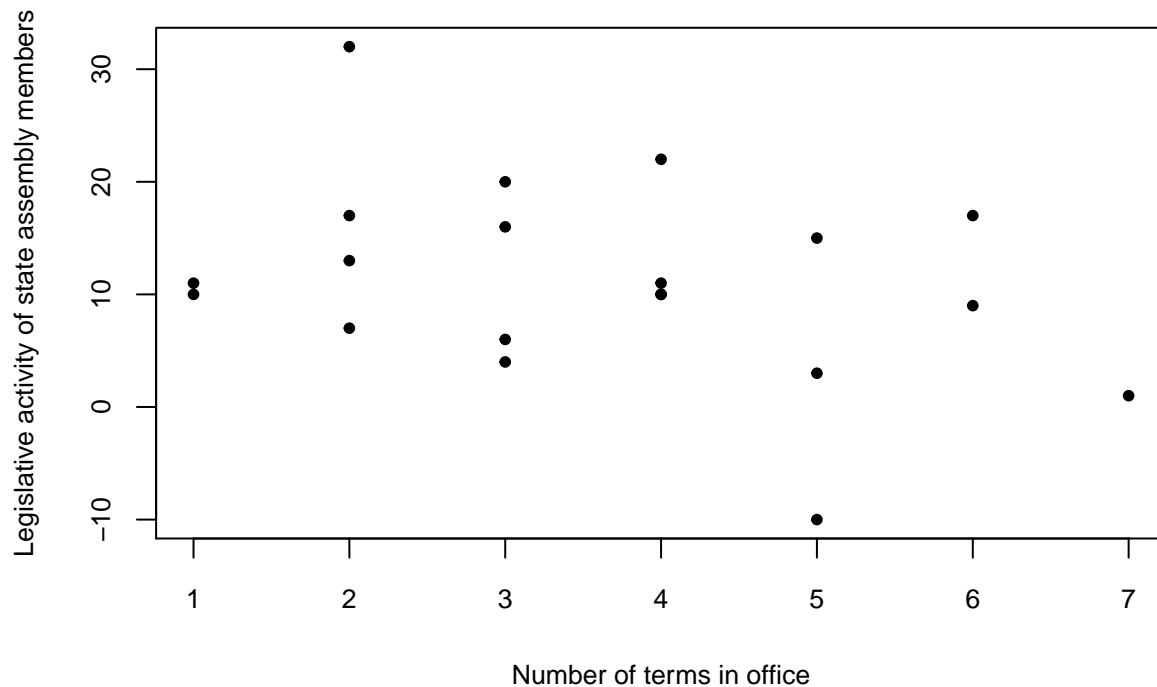
Uploading the file in R

I set the working directory in R and upload the .dta file:

```
setwd("~/Desktop/Grad School/courses/PS813/Rwork")
library(foreign)
PS1Data <- read.dta("PS1.dta")
```

Rough Plot of the data

```
plot(PS1Data$terms, PS1Data$Leg_Act,
     xlab = "Number of terms in office",
     ylab= "Legislative activity of state assembly members",
     cex.lab = 0.8, cex.axis= 0.8, pch=20)
```

Correlation

Compute correlation between variables:

```
cor(PS1Data$terms, PS1Data$Leg_Act)
```

```
## [1] -0.3336006
```

Regression

Regress Leg_Act on terms:

```
fit <- lm(Leg_Act ~ terms, data = PS1Data)
summary(fit)
```

```
##
## Call:
## lm(formula = Leg_Act ~ terms, data = PS1Data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.7883  -5.7062  -0.5109   4.3777  18.0438
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.401      4.546   3.828  0.00123 **
## terms         -1.723      1.147  -1.501  0.15060
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 8.494 on 18 degrees of freedom
## Multiple R-squared:  0.1113, Adjusted R-squared:  0.06192
## F-statistic: 2.254 on 1 and 18 DF,  p-value: 0.1506
```

Residuals

First, I create a separate variable for the predicted values using the previous model and the predict() function:

```
PS1Data$p_Leg_Act <- predict(fit)
```

Second, I calculate the residuals as the difference beteen the observed andpredicted values of the dependent variable:

```
PS1Data$Residuals <- PS1Data$Leg_Act - PS1Data$p_Leg_Act
```

Subsequently, I plot the residuals against the independent variable:

```
plot(PS1Data$terms, PS1Data$Residuals,
     xlab = "Number of terms in office",
     ylab= "Residuals",
     cex.lab = 0.8, cex.axis= 0.8, pch=20)
```
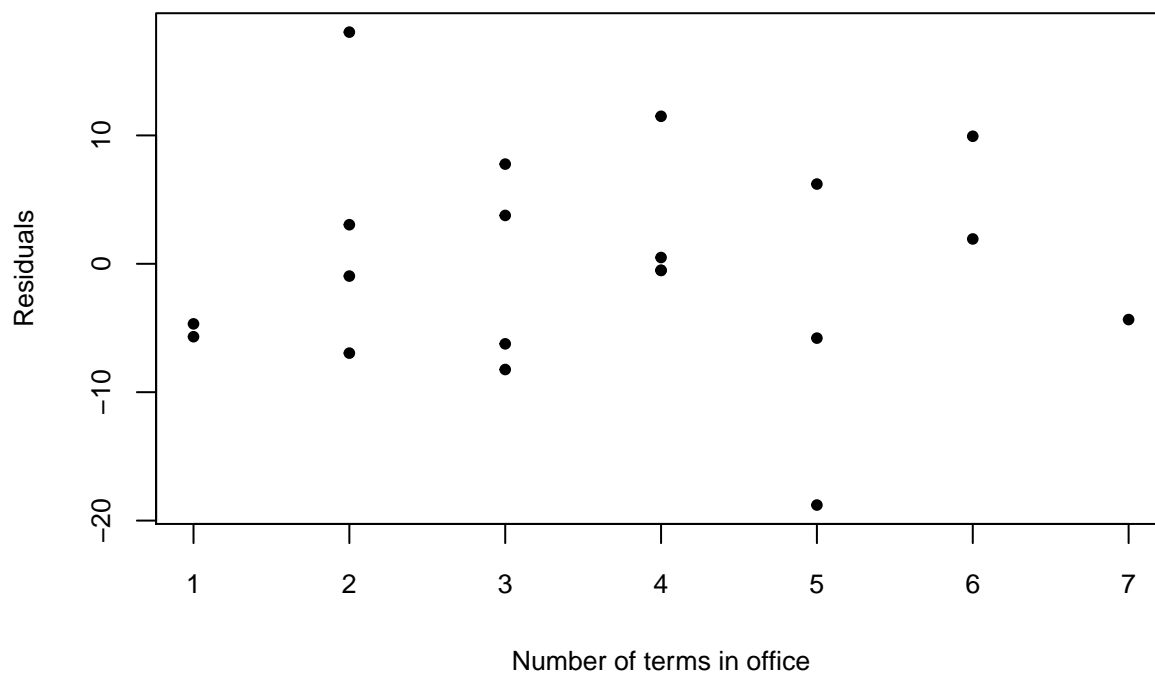

In order to get the correlation of the observed and predicted values of the dependent variable, I use the cor() function:

```
correlation <- cor(PS1Data$Leg_Act, PS1Data$p_Leg_Act)
correlation
```

```
## [1] 0.3336006
```

I now square the correlation coefficient compare to the reported R^2:

```
correlation*correlation
```

```
## [1] 0.1112894
```