Data Exercise 1

Monica Komer

February 4, 2019

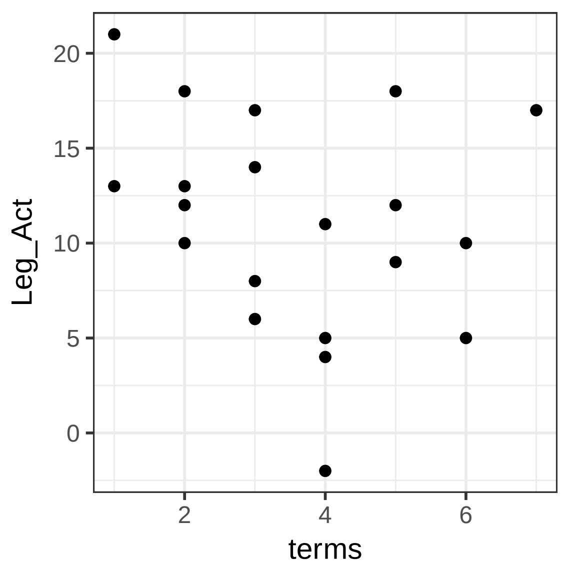
```
net install PS813_EX1, from(https://faculty.polisci.wisc.edu/weimer/)
```

```
PS813_EX1 0720
```

```
save "EX1.dta"
```

1. A plot of Legislative Activity by Terms in Office

```
d <- readstata13::read.dta13(here("data/EX1.dta"))
ggplot(d, aes(y = Leg_Act, x = terms)) +
  geom_point()
```




```
corXY <- cor(d$Leg_Act, d$terms)
corXY
```

```
## [1] -0.242168
```

The correlation between Legislative Activity and Terms in Office is -0.242168.

2. Estimating linear regression

```
model <- lm(d$Leg_Act ~ d$terms)
summary(model)
```

```
##
## Call:
## lm(formula = d$Leg_Act ~ d$terms)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7245  -3.6779   0.0547   3.2117   8.7172
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.9799     3.0453   4.591 0.000227 ***
## d$terms      -0.8139     0.7686  -1.059 0.303626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.689 on 18 degrees of freedom
## Multiple R-squared:  0.05865,    Adjusted R-squared:  0.006348
## F-statistic: 1.121 on 1 and 18 DF,  p-value: 0.3036
```

```r
alpha <- model$coefficients[1]
beta <- model$coefficients[2]
```

Regression coefficients: α = 13.979927 and β = -0.8138686

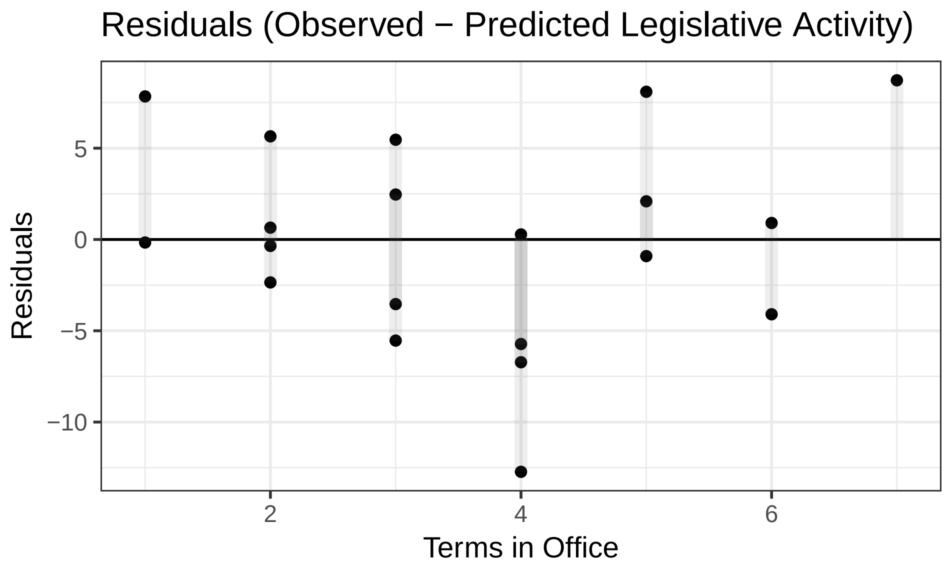
3. Computing residuals

```r
p_Leg_Act <- predict(model)

resid <- d$Leg_Act - p_Leg_Act
```

4. Plot of Residuals

```r
ggplot(d, aes(y = resid, x = terms)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_col(alpha = .1, width = .1, position = "dodge") +
  labs(title = "Residuals (Observed - Predicted Legislative Activity)",
       x = "Terms in Office",
       y = "Residuals")
```



5. $Cor(Y, \hat{Y})$

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

$Cor(Y, \hat{Y})$ = 0.242168

6. $Cor(Y, \hat{Y})^2$ vs. R^2.

```
r2 <- summary(model)$r.squared
```

R^2 = 0.0586454

7. Hypothesis test

The null hypothesis, $H_0 : \beta = 0$, and the alternative hypothesis, $H_a : \beta \neq 0$, are tested in the linear regression above. The p-value associated with the `term` coefficient in the linear regression above is larger than 0.05. This indicates that the number of terms in office is not statistically significant. For this reason, there is not enough evidence to accept the alternative hypothesis and I fail to reject the null hypothesis that the number of terms in office has no statistically significant relationship to legislative activity.

8. Discussion

The results above indicate that α = 13.979927 and β = -0.8138686. The intercept is 13.979927, meaning that legislative acitivity is, on average, 13.979927 when the number of terms in office is 0.The slope is -0.8138686, which indicates that each additional term in office is associated with a change of -0.8138686 on the index of legislative activity among state assembly members. While this seems to imply a negative relationship between the number of terms in office and legislative activity, the results from the linear regression do *not* provide sufficient evidence to accept the altnernative hypothesis that the number of terms impacts legislative activity. The p-value on the `term` coefficient is 0.3036, which is larger than 0.05. I therefore, fail to reject the null hypothesis.