HW12 Patrick Neyland

Patrick Neyland

12/6/2022

## Question 1

Statements 2 and 3 are consequences.

## Question 2

divide everything by income.

## Question 3

False

## Question 4

### Part 1

numerator df = k+1 denominator df = n-k-2

### Part 2

Because there are more explanatory variables on the right-hand side.

### Part 3

F depends on 3 things– R22, n, k

### Part 4

$

Something is wrong. Because is a perfect linear combination of all x variables. This presents perfect colinearity.

## Question 5

### Part 1

### Part 2

model5\_2 <- lm(sleep~ totwrk + educ + age + I(age^2) +yngkid + male, sleep75)  
stargazer(model5\_2, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## sleep   
## -----------------------------------------------  
## totwrk -0.163\*\*\*   
## (0.018)   
##   
## educ -11.713\*\*   
## (5.872)   
##   
## age -8.697   
## (11.329)   
##   
## I(age2) 0.128   
## (0.135)   
##   
## yngkid -0.023   
## (50.276)   
##   
## male 87.755\*\*   
## (34.668)   
##   
## Constant 3,840.852\*\*\*   
## (239.414)   
##   
## -----------------------------------------------  
## Observations 706   
## R2 0.123   
## Adjusted R2 0.115   
## Residual Std. Error 418.027 (df = 699)   
## F Statistic 16.302\*\*\* (df = 6; 699)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

u\_hat <- resid(model5\_2)  
  
reg\_var <- lm(I(u\_hat^2) ~ male, sleep75)  
  
stargazer(reg\_var, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## I(u\_hat2)   
## -----------------------------------------------  
## male -28,849.630   
## (27,296.510)   
##   
## Constant 189,359.200\*\*\*   
## (20,546.360)   
##   
## -----------------------------------------------  
## Observations 706   
## R2 0.002   
## Adjusted R2 0.0002   
## Residual Std. Error 359,414.400 (df = 704)   
## F Statistic 1.117 (df = 1; 704)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part 3

It was higher for women but it was not significant.

## Question 6

### Part 1

model6\_1 <- lm(price ~ lotsize + sqrft + bdrms, hprice1)  
  
coeftest(model6\_1)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.1770e+01 2.9475e+01 -0.7386 0.462208   
## lotsize 2.0677e-03 6.4213e-04 3.2201 0.001823 \*\*   
## sqrft 1.2278e-01 1.3237e-02 9.2751 1.658e-14 \*\*\*  
## bdrms 1.3853e+01 9.0101e+00 1.5374 0.127945   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

coeftest(model6\_1, vcov = vcovHC(model6\_1, type = "HC0"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -21.7703081 36.2843444 -0.6000 0.55013   
## lotsize 0.0020677 0.0012227 1.6912 0.09451 .   
## sqrft 0.1227782 0.0173178 7.0897 3.883e-10 \*\*\*  
## bdrms 13.8525217 8.2836880 1.6723 0.09819 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The biggest difference is in the standard errors for lotsize.

### Part 2

model6\_2 <- lm(log(price) ~ log(lotsize) + log(sqrft) + bdrms, hprice1)  
  
coeftest(model6\_2)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.297042 0.651284 -1.9915 0.04967 \*   
## log(lotsize) 0.167967 0.038281 4.3877 3.307e-05 \*\*\*  
## log(sqrft) 0.700232 0.092865 7.5403 5.006e-11 \*\*\*  
## bdrms 0.036958 0.027531 1.3424 0.18308   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

coeftest(model6\_2, vcov = vcovHC(model6\_2, type = "HC0"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.297042 0.763351 -1.6991 0.09299 .   
## log(lotsize) 0.167967 0.040520 4.1453 8.067e-05 \*\*\*  
## log(sqrft) 0.700232 0.101442 6.9028 9.014e-10 \*\*\*  
## bdrms 0.036958 0.029898 1.2362 0.21984   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The standard errors are not too different.

### Part 3

The log transformation was successful in limiting the impacts of the heteroskedasticity found in the data.

## Question 7

model7\_1 <- lm(voteA ~ prtystrA + democA + log(expendA) + log(expendB), vote1)  
stargazer(model7\_1, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## voteA   
## -----------------------------------------------  
## prtystrA 0.252\*\*\*   
## (0.071)   
##   
## democA 3.793\*\*\*   
## (1.407)   
##   
## log(expendA) 5.779\*\*\*   
## (0.392)   
##   
## log(expendB) -6.238\*\*\*   
## (0.397)   
##   
## Constant 37.661\*\*\*   
## (4.736)   
##   
## -----------------------------------------------  
## Observations 173   
## R2 0.801   
## Adjusted R2 0.796   
## Residual Std. Error 7.573 (df = 168)   
## F Statistic 169.229\*\*\* (df = 4; 168)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

bptest(model7\_1)

##   
## studentized Breusch-Pagan test  
##   
## data: model7\_1  
## BP = 9.0934, df = 4, p-value = 0.05881

u\_hat <- resid(model7\_1)  
#summary(lm( u\_hat^2 ~ lotsize+sqrft+bdrms, data=hprice1))

## Question 8

### Part 1

model8\_1 <- lm(e401k ~ inc + I(inc^2) + age + I(age^2) + male, k401ksubs)  
  
coeftest(model6\_2)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.297042 0.651284 -1.9915 0.04967 \*   
## log(lotsize) 0.167967 0.038281 4.3877 3.307e-05 \*\*\*  
## log(sqrft) 0.700232 0.092865 7.5403 5.006e-11 \*\*\*  
## bdrms 0.036958 0.027531 1.3424 0.18308   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

coeftest(model6\_2, vcov = vcovHC(model6\_2, type = "HC0"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.297042 0.763351 -1.6991 0.09299 .   
## log(lotsize) 0.167967 0.040520 4.1453 8.067e-05 \*\*\*  
## log(sqrft) 0.700232 0.101442 6.9028 9.014e-10 \*\*\*  
## bdrms 0.036958 0.029898 1.2362 0.21984   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Part 2

weights <- k401ksubs$weights  
  
model8\_2 <- lm(e401k ~ inc + I(inc^2) + age + I(age^2) + male, k401ksubs, weights = weights)  
  
summary(model8\_2)

##   
## Call:  
## lm(formula = e401k ~ inc + I(inc^2) + age + I(age^2) + male,   
## data = k401ksubs, weights = weights)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.6970 -0.3719 -0.2149 0.4870 0.9155   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.063e-01 8.110e-02 -6.243 4.48e-10 \*\*\*  
## inc 1.245e-02 5.929e-04 20.993 < 2e-16 \*\*\*  
## I(inc^2) -6.165e-05 4.732e-06 -13.028 < 2e-16 \*\*\*  
## age 2.651e-02 3.922e-03 6.758 1.49e-11 \*\*\*  
## I(age^2) -3.053e-04 4.501e-05 -6.782 1.26e-11 \*\*\*  
## male -3.533e-03 1.208e-02 -0.292 0.77   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4648 on 9269 degrees of freedom  
## Multiple R-squared: 0.09428, Adjusted R-squared: 0.09379   
## F-statistic: 193 on 5 and 9269 DF, p-value: < 2.2e-16

No important differences.