HW7 Patrick Neyland

Patrick Neyland

2022-10-18

## Question 1

1. and (iii) can cause the usual OLS t statistic to be invalid.

## Question 2

### Part i

### Part ii

I expect both and to be positive.

### Part iii

10% increase in pop will result in .66% increase in rent, not 6.6% as stated in the problem. Because it is a log-log equation, there is no scaling of the coefficients.

### Part iv

.0056/.0017

[1] 3.294118

qt(.99, 60)

[1] 2.390119

Because we reject the null hypothesis at the 1% significance level.

Or we can compute the p-value.

pt(3.294, 60, lower.tail = FALSE)

[1] 0.0008300033

Easily reject the null because the p-value(0.00083) is less than 0.01()

## Question 3

### Part i

### Part ii

### Part iii

## Question 4

### Part i

### Part ii

q = 2

((209444.99-165644.51)/2)/(165644.51/86)

[1] 11.37026

1 - pf(11.37026, 2, 86)

[1] 4.155306e-05

### Part iii

((.829-.820)/3)/(1-(.829/83))

[1] 0.003030266

### Part iv

Homoscedasticity assumption is violated. You cannot count on the f-statistic to fit the F distribution—you cannot rely on the F test.

## Question 5

### Part i

df <- filter(k401ksubs, fsize == 1)  
length(df$fsize)

[1] 2017

The data set contains 2017 single person households.

### Part ii

model5\_2 <- lm(nettfa ~ inc + age, data = df)  
stargazer(model5\_2, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 nettfa   
-----------------------------------------------  
inc 0.799\*\*\*   
 (0.060)   
   
age 0.843\*\*\*   
 (0.092)   
   
Constant -43.040\*\*\*   
 (4.080)   
   
-----------------------------------------------  
Observations 2,017   
R2 0.119   
Adjusted R2 0.118   
Residual Std. Error 44.683 (df = 2014)   
F Statistic 136.465\*\*\* (df = 2; 2014)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

No, there are no surprises in the the slope estimates.

### Part iii

I think it shows that a lot of people start out with a lot of debt early in their careers—student loans, mortgage, etc.

### Part iv

linearHypothesis(model5\_2, "age=1")

Linear hypothesis test  
  
Hypothesis:  
age = 1  
  
Model 1: restricted model  
Model 2: nettfa ~ inc + age  
  
 Res.Df RSS Df Sum of Sq F Pr(>F)   
1 2015 4026886   
2 2014 4021048 1 5837.7 2.9239 0.08743 .  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The p-value is 0.08743. At the 1% signigicance level, we do not reject the null hypothesis ### Part v

model5\_5 <- lm(nettfa ~ inc, data = df)  
stargazer(model5\_5, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 nettfa   
-----------------------------------------------  
inc 0.821\*\*\*   
 (0.061)   
   
Constant -10.571\*\*\*   
 (2.061)   
   
-----------------------------------------------  
Observations 2,017   
R2 0.083   
Adjusted R2 0.082   
Residual Std. Error 45.592 (df = 2015)   
F Statistic 181.599\*\*\* (df = 1; 2015)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The estimated coefficient on is not much different from the estimate in part ii.

## Question 6

### Part i

model6\_1 <- lm(log(psoda) ~ prpblck + log(income) + prppov, data = discrim)  
stargazer(model6\_1, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.073\*\*   
 (0.031)   
   
log(income) 0.137\*\*\*   
 (0.027)   
   
prppov 0.380\*\*\*   
 (0.133)   
   
Constant -1.463\*\*\*   
 (0.294)   
   
-----------------------------------------------  
Observations 401   
R2 0.087   
Adjusted R2 0.080   
Residual Std. Error 0.081 (df = 397)   
F Statistic 12.604\*\*\* (df = 3; 397)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Yes, is statistically different from zero at a 5% level in a two-sided test. However, it is not significant at the 1% level.

### Part ii

df1 <- na.omit(discrim)  
cor(df1$lincome, df1$prppov)

[1] -0.8468178

cor(log(df1$income), df1$prppov)

[1] -0.8468178

summary(model6\_1)

Call:  
lm(formula = log(psoda) ~ prpblck + log(income) + prppov, data = discrim)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.32218 -0.04648 0.00651 0.04272 0.35622   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -1.46333 0.29371 -4.982 9.4e-07 \*\*\*  
prpblck 0.07281 0.03068 2.373 0.0181 \*   
log(income) 0.13696 0.02676 5.119 4.8e-07 \*\*\*  
prppov 0.38036 0.13279 2.864 0.0044 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.08137 on 397 degrees of freedom  
 (9 observations deleted due to missingness)  
Multiple R-squared: 0.08696, Adjusted R-squared: 0.08006   
F-statistic: 12.6 on 3 and 397 DF, p-value: 6.917e-08

The correlation between and is -0.847 Yes, each of the variables is statistically significant.

| Variable | p-value |  |  |  |
| --- | --- | --- | --- | --- |
| prpblck | 0.018 |  |  |  |
| log(income) | 0.000 |  |  |  |
| prppov | 0.004 |  |  |  |

#df <- mutate()  
model6\_3 <- lm(log(psoda) ~ prpblck + I(prpblck+prppov) + log(income), data = discrim)  
stargazer(model6\_1, model6\_3, type = "text")

===========================================================  
 Dependent variable:   
 ----------------------------  
 log(psoda)   
 (1) (2)   
-----------------------------------------------------------  
prpblck 0.073\*\* -0.308\*\*   
 (0.031) (0.152)   
   
I(prpblck + prppov) 0.380\*\*\*   
 (0.133)   
   
log(income) 0.137\*\*\* 0.137\*\*\*   
 (0.027) (0.027)   
   
prppov 0.380\*\*\*   
 (0.133)   
   
Constant -1.463\*\*\* -1.463\*\*\*   
 (0.294) (0.294)   
   
-----------------------------------------------------------  
Observations 401 401   
R2 0.087 0.087   
Adjusted R2 0.080 0.080   
Residual Std. Error (df = 397) 0.081 0.081   
F Statistic (df = 3; 397) 12.604\*\*\* 12.604\*\*\*   
===========================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

linearHypothesis(model6\_1, c("prpblck = 0", "prppov = 0"))

Linear hypothesis test  
  
Hypothesis:  
prpblck = 0  
prppov = 0  
  
Model 1: restricted model  
Model 2: log(psoda) ~ prpblck + log(income) + prppov  
  
 Res.Df RSS Df Sum of Sq F Pr(>F)   
1 399 2.8331   
2 397 2.6284 2 0.20464 15.454 3.441e-07 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Part iv

model6\_4 <- lm(log(psoda) ~ prpblck + log(income) +   
 prppov + log(hseval), data = discrim)  
stargazer(model6\_4, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 log(psoda)   
-----------------------------------------------  
prpblck 0.098\*\*\*   
 (0.029)   
   
log(income) -0.053   
 (0.038)   
   
prppov 0.052   
 (0.134)   
   
log(hseval) 0.121\*\*\*   
 (0.018)   
   
Constant -0.842\*\*\*   
 (0.292)   
   
-----------------------------------------------  
Observations 401   
R2 0.184   
Adjusted R2 0.176   
Residual Std. Error 0.077 (df = 396)   
F Statistic 22.313\*\*\* (df = 4; 396)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

summary(model6\_4)

Call:  
lm(formula = log(psoda) ~ prpblck + log(income) + prppov + log(hseval),   
 data = discrim)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.30652 -0.04380 0.00701 0.04332 0.35272   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -0.84151 0.29243 -2.878 0.004224 \*\*   
prpblck 0.09755 0.02926 3.334 0.000937 \*\*\*  
log(income) -0.05299 0.03753 -1.412 0.158706   
prppov 0.05212 0.13450 0.388 0.698571   
log(hseval) 0.12131 0.01768 6.860 2.67e-11 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.07702 on 396 degrees of freedom  
 (9 observations deleted due to missingness)  
Multiple R-squared: 0.1839, Adjusted R-squared: 0.1757   
F-statistic: 22.31 on 4 and 396 DF, p-value: < 2.2e-16

Ceteris paribus, as the increases by 1%, increases by 0.121% on average. The two-sided p-value for is 2.67e-11.

### Part v

linearHypothesis(model6\_4, c("prpblck-prppov=0"))

Linear hypothesis test  
  
Hypothesis:  
prpblck - prppov = 0  
  
Model 1: restricted model  
Model 2: log(psoda) ~ prpblck + log(income) + prppov + log(hseval)  
  
 Res.Df RSS Df Sum of Sq F Pr(>F)  
1 397 2.3498   
2 396 2.3493 1 0.00052445 0.0884 0.7664