HW10 Group 1, Austin Halvorsen

Pedram Jahangiry

Nov 24 2020

# Problems

## Question 1

### (i)

Because this is a log-linear model, we interpret the coefficient on *cigs* as an increase of 10 more cigarettes a day results in a 4.4% lower birthrate.

### (ii)

Holding all other factors constant, our model suggests that white children weigh 5.5% more than a non-white child.

To calculate if this is significant or not, we need to calculate the t-stat, under the hypothesis, and . We get a t value of . This is significant at the 1% level, so we can say that the difference between white and non white babies is statistically significant.

### (iii)

Looking at motheduc, if a mother has an additional year of education, then the estimated birth weight of the is approximately 0.3% higher.

To see if this is significant, we find the t value: This is insignificant at the 1% level so we shouldn’t investigate this relationship.

### (iv)

In order to compute the F-statistic, we would need to reestimate the first equation with the same amount of observations as the second equation.

## Question 2

### (i)

Looking at our t-stat, we get , which means that this is significant at our 1% level and it should remain in our model.

Finding the optimum size, we need to look at the partial derivative. So where:

Since this is measured in hundreds and based on our estimated model, we get 4.41 \* 100 = 441 is the optimal high school size.

### (ii)

Holding *hsize* fixed, the estimated difference in SAT scores between nonblack females and nonblack males is 45 points lower. To see if this is significant, we calculate the t statistic , which shows us that this is significant at the 1% level.

### (iii)

The estimated difference in SAT score between nonblack males and black males is about 170 points.

This is significant at the 1% level.

### (iv)

The estimated score between black females and nonblack females is the difference where our interaction term black and female equal 1 and and black equals 1. This gives us a difference of c

## Question 3

### (i)

Given that , then and Therefore,

We can see here now that our intercept becomes and our slope estimates will change from positive to negative.

### (ii)

The standard error will stay the same, since the spread is not influenced by the sign of the coefficients.

### (iii)

The value will remain unchanged, since we are using the same set of independent variables.

# Computer Exercises

## Question 4

### (i)

df2 <- gpa1  
mrm4 <- lm(colGPA~PC+hsGPA+ACT+mothcoll+fathcoll, df2)  
stargazer(mrm4, type="text")

===============================================  
 Dependent variable:   
 ---------------------------  
 colGPA   
-----------------------------------------------  
PC 0.152\*\*   
 (0.059)   
   
hsGPA 0.450\*\*\*   
 (0.094)   
   
ACT 0.008   
 (0.011)   
   
mothcoll -0.004   
 (0.060)   
   
fathcoll 0.042   
 (0.061)   
   
Constant 1.256\*\*\*   
 (0.335)   
   
-----------------------------------------------  
Observations 141   
R2 0.222   
Adjusted R2 0.193   
Residual Std. Error 0.334 (df = 135)   
F Statistic 7.713\*\*\* (df = 5; 135)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

When adding in *mothcoll* and *fathcoll*, the effect of PC changes from 0.157 to 0.152. The P-value for PC is still less than 0.05, so it is statistically significant.

### (ii)

### (iii)

mrm4a <- lm(colGPA~PC+hsGPA+I(hsGPA^2)+ACT+mothcoll+fathcoll, df2)  
stargazer(mrm4,mrm4a, type='text')

##   
## =================================================================  
## Dependent variable:   
## ---------------------------------------------  
## colGPA   
## (1) (2)   
## -----------------------------------------------------------------  
## PC 0.152\*\* 0.140\*\*   
## (0.059) (0.059)   
##   
## hsGPA 0.450\*\*\* -1.803   
## (0.094) (1.444)   
##   
## I(hsGPA2) 0.337   
## (0.216)   
##   
## ACT 0.008 0.005   
## (0.011) (0.011)   
##   
## mothcoll -0.004 0.003   
## (0.060) (0.060)   
##   
## fathcoll 0.042 0.063   
## (0.061) (0.062)   
##   
## Constant 1.256\*\*\* 5.040\*\*   
## (0.335) (2.443)   
##   
## -----------------------------------------------------------------  
## Observations 141 141   
## R2 0.222 0.236   
## Adjusted R2 0.193 0.202   
## Residual Std. Error 0.334 (df = 135) 0.333 (df = 134)   
## F Statistic 7.713\*\*\* (df = 5; 135) 6.904\*\*\* (df = 6; 134)  
## =================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

summary(mrm4a)

##   
## Call:  
## lm(formula = colGPA ~ PC + hsGPA + I(hsGPA^2) + ACT + mothcoll +   
## fathcoll, data = df2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.78998 -0.24327 -0.00648 0.26179 0.72231   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.040328 2.443038 2.063 0.0410 \*  
## PC 0.140446 0.058858 2.386 0.0184 \*  
## hsGPA -1.802520 1.443552 -1.249 0.2140   
## I(hsGPA^2) 0.337341 0.215711 1.564 0.1202   
## ACT 0.004786 0.010786 0.444 0.6580   
## mothcoll 0.003091 0.060110 0.051 0.9591   
## fathcoll 0.062761 0.062401 1.006 0.3163   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3326 on 134 degrees of freedom  
## Multiple R-squared: 0.2361, Adjusted R-squared: 0.2019   
## F-statistic: 6.904 on 6 and 134 DF, p-value: 2.088e-06

The P-value for 0.115 which is greater than the critical p-value of 0.05, so we probably do not need to add in not needed.

## Question 5

### (i)

df <- wage2  
mrm1 <- lm(lwage~educ+exper+tenure+married+black+south+urban, df)  
pander(summary(mrm1),add.significance.stars = T)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |  |
| **(Intercept)** | 5.395 | 0.1132 | 47.65 | 1.72e-251 | \* \* \* |
| **educ** | 0.06543 | 0.00625 | 10.47 | 2.577e-24 | \* \* \* |
| **exper** | 0.01404 | 0.003185 | 4.409 | 1.161e-05 | \* \* \* |
| **tenure** | 0.01175 | 0.002453 | 4.789 | 1.95e-06 | \* \* \* |
| **married** | 0.1994 | 0.03905 | 5.107 | 3.979e-07 | \* \* \* |
| **black** | -0.1883 | 0.03767 | -5 | 6.839e-07 | \* \* \* |
| **south** | -0.0909 | 0.02625 | -3.463 | 0.0005582 | \* \* \* |
| **urban** | 0.1839 | 0.02696 | 6.822 | 1.618e-11 | \* \* \* |

Fitting linear model: lwage ~ educ + exper + tenure + married + black + south + urban

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Residual Std. Error |  | Adjusted |
| 935 | 0.3655 | 0.2526 | 0.2469 |

Our estimated model is:

Our coefficient on *black*, holding all else constant would indicate that the monthly salary of blacks is 18.83% less than nonblacks. The p-value is very low (.000000068397), which would tell us that this is significant at the 5% level, or that the difference between salaries in black and nonblack is statistically significant.

### (ii)

mrm2 <- lm(lwage~educ+exper+tenure+married+black+south+urban+I(exper^2)+I(tenure^2), df)  
stargazer(mrm1, mrm2, type='text')

===================================================================  
 Dependent variable:   
 -----------------------------------------------  
 lwage   
 (1) (2)   
-------------------------------------------------------------------  
educ 0.065\*\*\* 0.064\*\*\*   
 (0.006) (0.006)   
   
exper 0.014\*\*\* 0.017   
 (0.003) (0.013)   
   
tenure 0.012\*\*\* 0.025\*\*\*   
 (0.002) (0.008)   
   
married 0.199\*\*\* 0.199\*\*\*   
 (0.039) (0.039)   
   
black -0.188\*\*\* -0.191\*\*\*   
 (0.038) (0.038)   
   
south -0.091\*\*\* -0.091\*\*\*   
 (0.026) (0.026)   
   
urban 0.184\*\*\* 0.185\*\*\*   
 (0.027) (0.027)   
   
I(exper2) -0.0001   
 (0.001)   
   
I(tenure2) -0.001\*   
 (0.0005)   
   
Constant 5.395\*\*\* 5.359\*\*\*   
 (0.113) (0.126)   
   
-------------------------------------------------------------------  
Observations 935 935   
R2 0.253 0.255   
Adjusted R2 0.247 0.248   
Residual Std. Error 0.365 (df = 927) 0.365 (df = 925)   
F Statistic 44.747\*\*\* (df = 7; 927) 35.171\*\*\* (df = 9; 925)  
===================================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

summary(mrm2)

Call:  
lm(formula = lwage ~ educ + exper + tenure + married + black +   
 south + urban + I(exper^2) + I(tenure^2), data = df)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.98236 -0.21972 -0.00036 0.24078 1.25127   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 5.3586757 0.1259143 42.558 < 2e-16 \*\*\*  
educ 0.0642761 0.0063115 10.184 < 2e-16 \*\*\*  
exper 0.0172146 0.0126138 1.365 0.172665   
tenure 0.0249291 0.0081297 3.066 0.002229 \*\*   
married 0.1985470 0.0391103 5.077 4.65e-07 \*\*\*  
black -0.1906636 0.0377011 -5.057 5.13e-07 \*\*\*  
south -0.0912153 0.0262356 -3.477 0.000531 \*\*\*  
urban 0.1854241 0.0269585 6.878 1.12e-11 \*\*\*  
I(exper^2) -0.0001138 0.0005319 -0.214 0.830622   
I(tenure^2) -0.0007964 0.0004710 -1.691 0.091188 .   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.3653 on 925 degrees of freedom  
Multiple R-squared: 0.255, Adjusted R-squared: 0.2477   
F-statistic: 35.17 on 9 and 925 DF, p-value: < 2.2e-16

### (iii)

### (iv)

## Question 6

### (i)

df\_dummy <- df1 %>%   
 mutate(ecobuy = case\_when(  
 ecolbs == 0 ~ 0,  
 TRUE ~ 1))  
  
pander(prop.table(table(df\_dummy$ecobuy)))

|  |  |
| --- | --- |
| 0 | 1 |
| 0.3758 | 0.6242 |

From our data, a reported 62.42% of families say that they buy ecolabeled apples.

### (ii)

mrm2 <- lm(ecobuy~ecoprc+regprc+faminc+hhsize+educ+age, df\_dummy)  
pander(summary(mrm2))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **(Intercept)** | 0.4237 | 0.165 | 2.568 | 0.01044 |
| **ecoprc** | -0.8026 | 0.1094 | -7.336 | 6.535e-13 |
| **regprc** | 0.7193 | 0.1316 | 5.464 | 6.63e-08 |
| **faminc** | 0.0005518 | 0.0005295 | 1.042 | 0.2978 |
| **hhsize** | 0.02382 | 0.01253 | 1.902 | 0.05763 |
| **educ** | 0.02478 | 0.008374 | 2.96 | 0.003192 |
| **age** | -0.0005008 | 0.00125 | -0.4007 | 0.6888 |

Fitting linear model: ecobuy ~ ecoprc + regprc + faminc + hhsize + educ + age

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Residual Std. Error |  | Adjusted |
| 660 | 0.4594 | 0.1098 | 0.1016 |

Our estimated model is:

### (iii)

### (iv)

### (v)

### (vi)