3.R

Gebruiker

Tue Sep 18 16:11:52 2018

library(TeachingDemos)  
txtStart("Homework 3.txt")

## Output being copied to text file,  
## use txtStop to end

# Homework Assignment 3  
# Name: Born? Djavdan   
# ANR: 749817  
  
# Exercise 1  
load("~/Study/Semester 3/Econometrics for ECO/R Files/R Data/WAGE2.RData")  
attach(data)  
  
# a) In this dataset;   
# The number of observations is 935.   
# The unit of observations is people.  
# The number of variables is 17.   
  
# b)  
mean(wage)

## [1] 957.9455

mean(IQ)

## [1] 101.2824

sd(IQ)

## [1] 15.05264

# The sample average of salary = $957.95 per month.  
# The sample average of IQ = 101.2824 IQ points.  
# The sample standard deviation of IQ = 15.05264 IQ points.  
  
# c)   
regIQ <- lm(wage~IQ)  
regIQ

##   
## Call:  
## lm(formula = wage ~ IQ)  
##   
## Coefficients:  
## (Intercept) IQ   
## 116.992 8.303

summary(regIQ)

##   
## Call:  
## lm(formula = wage ~ IQ)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -898.7 -256.5 -47.3 201.1 2072.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 116.9916 85.6415 1.366 0.172   
## IQ 8.3031 0.8364 9.927 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 384.8 on 933 degrees of freedom  
## Multiple R-squared: 0.09554, Adjusted R-squared: 0.09457   
## F-statistic: 98.55 on 1 and 933 DF, p-value: < 2.2e-16

length(regIQ$fitted.values)

## [1] 935

regIQ$coefficients

## (Intercept) IQ   
## 116.991565 8.303064

# The equation is wage = 116.99 + 8.30 \* IQ.   
# The n = 935.  
# The R^2 = 0.09554  
8.30 \* 15

## [1] 124.5

# When all else equal, if IQ points increase by 15 units, monthly wage increases by $124,50 on average.   
# No, IQ does not explain most of the variable, because the R-squared is low.  
# Meaning that the omitted variables are important in this case when it comes to explaining monthly wage.   
  
# d)   
regIQl <- lm(lwage~IQ)  
regIQl

##   
## Call:  
## lm(formula = lwage ~ IQ)  
##   
## Coefficients:  
## (Intercept) IQ   
## 5.886994 0.008807

summary(regIQl)

##   
## Call:  
## lm(formula = lwage ~ IQ)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.09324 -0.25547 0.02261 0.27544 1.21487   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.8869944 0.0890206 66.13 <2e-16 \*\*\*  
## IQ 0.0088072 0.0008694 10.13 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3999 on 933 degrees of freedom  
## Multiple R-squared: 0.09909, Adjusted R-squared: 0.09813   
## F-statistic: 102.6 on 1 and 933 DF, p-value: < 2.2e-16

length(regIQl$fitted.values)

## [1] 935

regIQl$coefficients

## (Intercept) IQ   
## 5.886994354 0.008807156

# lwage = 5.887 + 0.0088 \* IQ  
# n = 935; R^2 = 0.09909  
0.0088 \* 15

## [1] 0.132

# When all else equal, if IQ points increase by 15 units, the approximate percentage increase in predicted wage is 13.2%.  
  
# e)  
LIQ <- log(IQ)  
lm(lwage~LIQ)

##   
## Call:  
## lm(formula = lwage ~ LIQ)  
##   
## Coefficients:  
## (Intercept) LIQ   
## 2.942 0.833

# The elasticity of wage with respect to IQ is explained by: ln(wage)=2.942+0.833\*ln(IQ)  
  
# f)  
0.833 \* 15

## [1] 12.495

# When all else equal, if IQ points increase by 15%, the approximate percentage increase in predicted wage is 12.495%  
# Both answers are very close, because they both represent the percentage change.   
  
detach(data)  
# Exercise 2  
load("~/Study/Semester 3/Econometrics for ECO/R Files/R Data/MEAP93.RData")  
attach(data)  
# a)   
# The more one spends more per student in dollar, the less extra value the students seem to get out of it when it comes to math test pass rate.   
# Therefor I conclude that letting the effect be explained by a diminishing effect seems more appropriate.   
  
# b)  
summary(math10)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.90 16.62 23.40 24.11 30.05 66.70

sd(math10)

## [1] 10.49361

30.05 - 16.62

## [1] 13.43

# Math pass rate:  
# The mean = 24.11%; The range = 1.9% to 66.7%; The standard deviation = 10.49%.   
# In case you mean 'interquartile range' by 'range', IQR = |k3 - k1| = 30.05% - 16.62% = 13.43%.   
summary(expend)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3332 3821 4145 4377 4659 7419

sd(expend)

## [1] 775.7897

4659 - 3821

## [1] 838

# Spending:   
# The mean = $4377; The range = $3332 to $7419; The standard deviation = $775.79.   
# In case you mean 'interquartile range' by 'range', IQR = |k3 - k1| = 4659 - 3821 = 838.   
  
# c)   
# This model the change in percentage of the math test pass rate is equal to the change in percentage of the expendatures per student in dollars \* Beta1.   
# When the Expendatures per student in dollars increases 10%, the math test pass rate increases by 10% \* Beta1.   
# Thus percentage point change math10 = Beta1 / 10 \* 10%.   
  
# d)  
regMath <- lm(math10~log(expend))  
regMath

##   
## Call:  
## lm(formula = math10 ~ log(expend))  
##   
## Coefficients:  
## (Intercept) log(expend)   
## -69.34 11.16

summary(regMath)

##   
## Call:  
## lm(formula = math10 ~ log(expend))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.343 -7.100 -0.914 6.148 39.093   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -69.341 26.530 -2.614 0.009290 \*\*   
## log(expend) 11.164 3.169 3.523 0.000475 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.35 on 406 degrees of freedom  
## Multiple R-squared: 0.02966, Adjusted R-squared: 0.02727   
## F-statistic: 12.41 on 1 and 406 DF, p-value: 0.0004752

length(regMath$fitted.values)

## [1] 408

regMath$coefficients

## (Intercept) log(expend)   
## -69.34116 11.16440

# math10 = -69.34116 + 11.16440 \* ln(expend)  
# n = 408; R^2 = 0.02966  
# No spending per student does not explain well the variation in math pass rate, because R-squared is not large.   
# Meaning that the omitted variables are important in this case when it comes to explaining the pass rate for the math test.   
  
# e)   
11.16440/10

## [1] 1.11644

# If spending increases by 10%, the estimated percentage point increase in math10 is (11.16440/10=) 1.12% increase.   
# Thus the estimated spending effect is 1.12%.   
  
# f)  
# No, the estimate does not capture the ceteris paribus effect of expenditure on math pass rate.   
# There are more variables beside expenditure per student that have an effect on the math pass rate, which are not controlled for in the model.   
# For example the condition in which the children grow up.   
# Thus it is not ceteris paribus, and the estimate B1 does not capture the ceteris paribus effect of expenditure on math pass rate.   
  
# g) This not much of a worry in this data set, because a value given in percentage (the math test pass rate) can never be more than 100%.   
  
detach(data)  
txtStop()