Exercise\_chap\_04.R

2022-09-28

Some Exercises from chapter o4

rm(list=ls())  
suppressPackageStartupMessages(library(mosaic))

browseURL(“<http://www.principlesofeconometrics.com/poe5/data/def/newbroiler.def>”)

load(url("http://www.principlesofeconometrics.com/poe5/data/rdata/newbroiler.rdata"))  
head(newbroiler)

## year q y p pb pcorn pf qprod lexpts popgro  
## 1 1950 14.3 7863 2.88382 1.29461 2.48133 NA 1826.616 -0.1718388 NA  
## 2 1951 15.1 7953 2.80385 1.40385 2.77308 NA 1975.677 -0.1648426 1.7161  
## 3 1952 15.3 8071 2.75849 1.36604 2.69057 NA 1981.376 -0.1922588 1.7286  
## 4 1953 15.2 8319 2.67041 1.06742 2.34831 NA 2052.745 -0.1668136 1.6635  
## 5 1954 15.8 8276 2.39405 1.01859 2.35688 NA 2154.065 -0.1749048 1.7711  
## 6 1955 14.7 8675 2.50000 1.01119 2.09328 NA 2055.664 -0.1671042 1.7760

Units of the variables:

p - price (no unit), just price index

q - per capita consumption (pound) Model 1: lin-lin model

m1 <- lm(q ~ p, data = newbroiler)  
  
coef(m1)

## (Intercept) p   
## 57.95564 -18.40283

coef(m1)[2]

## p   
## -18.40283

more general output

summary(m1)

##   
## Call:  
## lm(formula = q ~ p, data = newbroiler)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.182 -5.563 -2.941 5.981 13.402   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.956 2.575 22.50 < 2e-16 \*\*\*  
## p -18.403 1.661 -11.08 4.58e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.513 on 50 degrees of freedom  
## Multiple R-squared: 0.7105, Adjusted R-squared: 0.7047   
## F-statistic: 122.7 on 1 and 50 DF, p-value: 4.584e-15

Interprate the parameters:

Intercept:

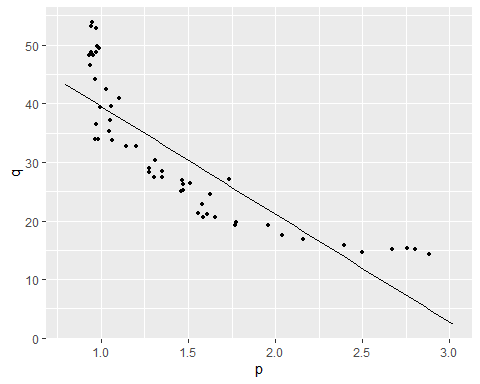
57.956: The expected level of per capita consumption when price is zero.

Slope:

-18.403: A 1 unit increase in price leads to (approximately ) 18.40 pound decrease in the level of expected/average/ per capita consumption q. Interprate R^2:

71.05% of variation in per capita consumption is explained by the variation in price. : the remaining 29% is the unexplained part

plotModel(m1)



confidence interval of the parameter

confint(m1)

## 2.5 % 97.5 %  
## (Intercept) 52.78315 63.12813  
## p -21.73935 -15.06631

Prediction: Predict the E(q|p) at a given value of p, p= 0.98

E(q|p = 0.98)

f <- makeFun(m1)  
f(0.98)

## 1   
## 39.92087

f(p = 0.98)

## 1   
## 39.92087

f(p = 0.98, interval = "confidence", level = 0.95)

## fit lwr upr  
## 1 39.92087 37.51904 42.32269

Model 2: lin-log

m2 <- lm(q ~ log(p), data = newbroiler)  
summary(m2)

##   
## Call:  
## lm(formula = q ~ log(p), data = newbroiler)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.286 -4.186 -1.747 3.818 11.006   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 41.2111 0.9898 41.63 <2e-16 \*\*\*  
## log(p) -31.9078 2.1584 -14.78 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.223 on 50 degrees of freedom  
## Multiple R-squared: 0.8138, Adjusted R-squared: 0.8101   
## F-statistic: 218.5 on 1 and 50 DF, p-value: < 2.2e-16

Interprate the parameters:

coef(m2)[2]/100

## log(p)   
## -0.319078

-0.319078 : a 1% increase in price, leads to (approximately) 0.319 pound decrease in the level of per capita consumption q

Interprate R^2:

81.38% of variation in per capita consumption is explained by the variation in log price.

confidence interval of the parameter

confint(m2)

## 2.5 % 97.5 %  
## (Intercept) 39.22304 43.19923  
## log(p) -36.24300 -27.57260

Prediction: Predict the E(q|p) at a given value of p, p= 0.98

E(q|p = 0.98) Manually:

coef(m2)[1]-(coef(m2)[2]\*log(0.98))

## (Intercept)   
## 40.56651

more simply

f <- makeFun(m2)  
f(0.98)

## 1   
## 41.85576

f(p = 0.98)

## 1   
## 41.85576

f(p = 0.98, interval = "confidence", level = 0.95)

## fit lwr upr  
## 1 41.85576 39.80697 43.90455

Model 3: log- linear model

m3 <- lm(log(q) ~ p, data = newbroiler)  
summary(m3)

##   
## Call:  
## lm(formula = log(q) ~ p, data = newbroiler)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.25251 -0.12407 -0.05208 0.14091 0.28644   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.33079 0.06323 68.50 <2e-16 \*\*\*  
## p -0.66422 0.04078 -16.29 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1599 on 50 degrees of freedom  
## Multiple R-squared: 0.8414, Adjusted R-squared: 0.8382   
## F-statistic: 265.2 on 1 and 50 DF, p-value: < 2.2e-16

Interpretation the parameters:

100\*coef(m3)[2]

## p   
## -66.42237

a 1 unit increase in price leads to (approximately) a 66.4 % decrease in q. Prediction in the log -linear model:

E(q|p) at a given value of p, p= 0.98 (mean of p)

E(q|p = 0.98) Manually

coef(m3)[1]+coef(m3)[2]\*0.98

## (Intercept)   
## 3.679846

exp(coef(m3)[1]+coef(m3)[2]\*0.98)

## (Intercept)   
## 39.64029

yn <- function(x){exp(coef(m3)[1]+coef(m3)[2]\*x)} # not corrected   
yn(0.98) # not corrected

## (Intercept)   
## 39.64029

Alternatively

y <- makeFun(m3)   
y(0.98) # predict q for the uncorrected depe.var.

## 1   
## 39.64029

However, there is one problem here. what is that? Since the dep. var. is in log form, we have log transformed error /residual terms. That means e = log(q) - b1 -p We have to adjust that. How? Look at your book on page (175 -178

s2 <- deviance(m3)/m3$df.residual # The variance in the model, sigma squared   
  
yc <- function(x){exp(coef(m3)[1]+coef(m3)[2]\*x+s2/2)} # corrected   
yc(0.98) # corrected

## (Intercept)   
## 40.15032

Alternatively, using the mosaic::makeFun()

y(0.98)\*exp(s2/2) # must adjust

## 1   
## 40.15032

Remember: the mosaic::makeFun() function predict the dependent variable in levels but w/o se correction

# we have to adjsut

we have estimated three different models. Now let us compare the three models in terms of R^2 values, i.e., which model has the highest R^2 value. The best model is the one with the highest R^2 value.  
R^2 from model 1

summary(m1)$r.squared

## [1] 0.7105309

R^2 from model 2

summary(m2)$r.squared

## [1] 0.8138129

R^2 from model 3

summary(m3)$r.squared

## [1] 0.8413902

But this is explained variation in the log dep. variable. That means we have the explained variation in log form of the dep.variable Change to explained variation in dep. variable in levels. How? A generalized R^2 measure (page 176): comparing two models, y = and log(y)=

cor(newbroiler$q, yc(newbroiler$p))

## [1] 0.9047122

cor(newbroiler$q, yc(newbroiler$p))^2

## [1] 0.8185042

Alternatively

cor(newbroiler$q, y(newbroiler$p)\*exp(s2/2))^2

## [1] 0.8185042