ARE212_PS3

Student: XX

2024-02-15

Install packages

```r
### Define path/working directory and date
path<-'C:/Users/52554/Documents/GitHub/are_212/pset_3/code/'
path_data <- 'C:/Users/52554/Documents/GitHub/are_212/pset_3/data/'
knitr::opts_chunk$set(setwd = path)
date <- Sys.Date()
print(date)
```

```
## [1] "2024-02-15"
```

```r
## Function to install packages and call libraries
install <- function(packages){
  new.packages <- packages[!(packages %in% installed.packages()[, "Package"])]
  if (length(new.packages))
    install.packages(new.packages, dependencies = TRUE)
  sapply(packages, require, character.only = TRUE)
}
required.packages <- c("readr", "haven", "dplyr", "estimatr", "devtools",
                       "rdrobust", "rdd", "ggplot2", "tidyverse", "pacman", "psych",
                       "stargazer", "tinytex")
install(required.packages)
```

```
## Loading required package: readr

## Loading required package: haven

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Loading required package: estimatr

## Warning: package 'estimatr' was built under R version 4.3.2

## Loading required package: devtools

## Loading required package: usethis

##
## Attaching package: 'devtools'

## The following object is masked _by_ '.GlobalEnv':
##
##     install

## Loading required package: rdrobust

## Loading required package: rdd

## Loading required package: sandwich

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: AER

## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## Loading required package: survival

## Loading required package: Formula
```

```
## Loading required package: ggplot2

## Loading required package: tidyverse

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x car::recode()   masks dplyr::recode()
## x purrr::some()   masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
## Loading required package: pacman

## Warning: package 'pacman' was built under R version 4.3.2

## Loading required package: psych
##
## Attaching package: 'psych'
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
##
## The following object is masked from 'package:car':
##
##     logit
##
## Loading required package: stargazer
##
## Please cite as:
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
##
## Loading required package: tinytex

##     readr     haven     dplyr  estimatr  devtools  rdrobust       rdd   ggplot2
##      TRUE      TRUE      TRUE      TRUE      TRUE      TRUE      TRUE      TRUE
## tidyverse    pacman     psych stargazer   tinytex
##      TRUE      TRUE      TRUE      TRUE      TRUE
```

```
p_load(dplyr, haven, readr, knitr, psych, ggplot2,stats4, stargazer, lmSupport, magrittr, qwraps2, Jmis
```

```
## Installing package into 'C:/Users/52554/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## Warning: package 'lmSupport' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
## Warning: unable to access index for repository http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib
##   no fue posible abrir la URL 'http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/4.3/PACKAGES'

## Warning in p_install(package, character.only = TRUE, ...):

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'lmSupport'

## Warning in p_load(dplyr, haven, readr, knitr, psych, ggplot2, stats4, stargazer, : Failed to install/
## lmSupport
```

Problem 1. Please check for missing values (as in section 3)

```r
data_or <- read_dta(paste0(path_data,"pset3_2024.dta"))
ls(data_or)
```

```
##  [1] "brand"       "co"          "country"     "country1"    "country2"
##  [6] "country3"    "country4"    "country5"    "domestic"    "firm"
## [11] "fuel"        "height"      "horsepower"  "loc"         "luxury"
## [16] "ngdp"        "ngdpe"       "pop"         "pr"          "price"
## [21] "princ"       "qu"          "segment"     "type"        "weight"
## [26] "width"       "year"        "yearsquared"
```

```r
anyNA(data_or)
```

```
## [1] FALSE
```

```r
data <- data_or %>% drop_na()
data <- data %>%
  mutate(lprice=log(price)) %>%
  mutate(lqu=log(qu))
## We notice there is no any NA in our database
```

Problem 2. Get the summary statistics for price: sample mean, standard deviation, minimum and maximum. Construct a 99% confidence interval for the sample average of price

```r
summary_maker <-
  list("Price" =
         list("min" = ~ min(data$price),
              "max" = ~ max(data$price),
              "mean (sd)" = ~ qwraps2::mean_sd(data$price)))
whole <- summary_table(data, summary_maker)
whole
```

	data (N = 57)
Price	
min	8.13041400909424
max	48.0754890441895
mean (sd)	20.77 ± 8.54

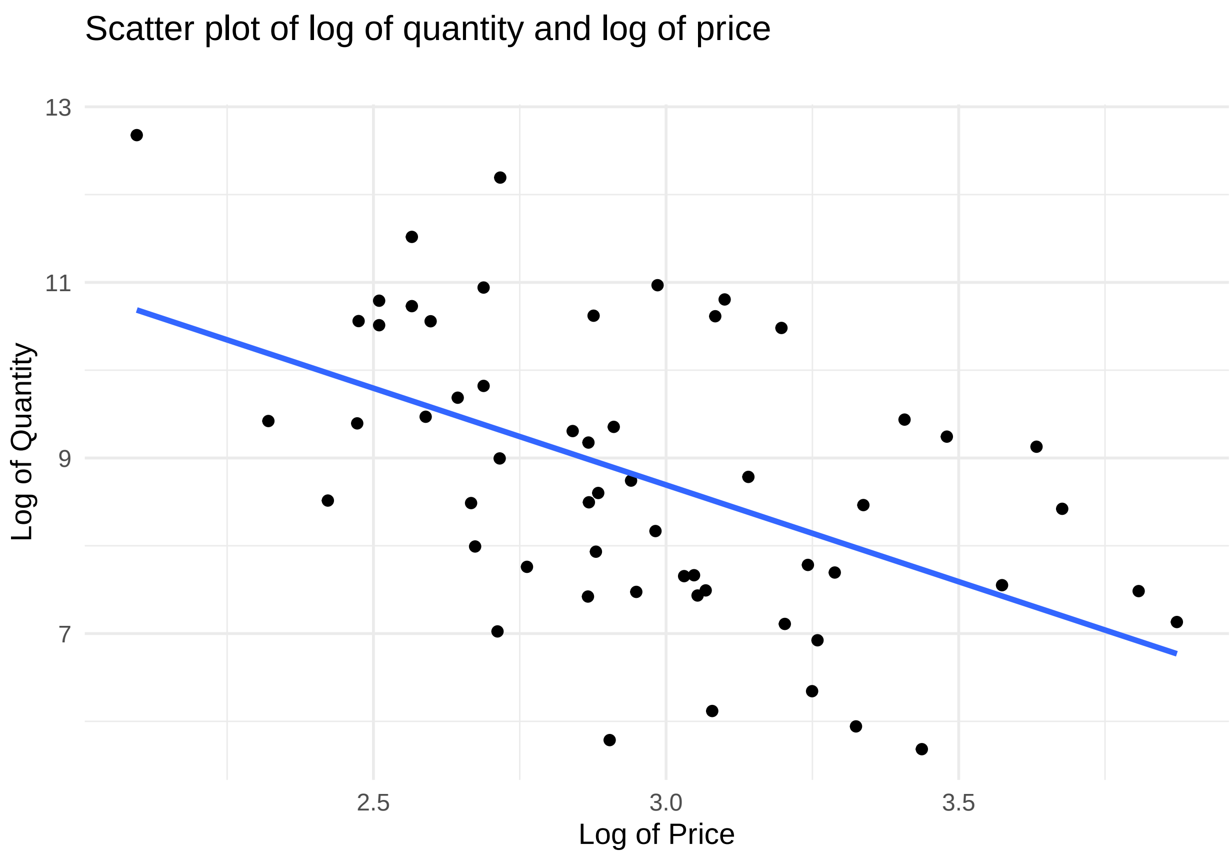
```r
df <- nrow(data)-1 ## degree freedoms
pbar<-mean(data$price)
diff <- (data$price - pbar)^2

#v_pbar<-pbar*(1-pbar)/nrow(data$price)
v_pbar<-sum(diff)/(nrow(data)-1)
se_pbar<-sqrt(v_pbar)
se_mean <- se_pbar/sqrt(nrow(data))
cn<-qt(0.995, df,lower.tail=TRUE)
bottomCI<-pbar-cn * se_mean
topCI<-pbar+cn * se_mean
print(c(bottomCI, topCI))
```

[1] 17.75517 23.79069

Problem 3. Create two new variables log of price and log of quantity, lprice and lqu. Create the scatter plot of the two variables lqu and lprice. What is the estimated OLS linear model slope associated with this scatter plot? Estimate a regression to answer this.

```r
ggplot(data, aes( x=lprice, y=lqu )) +
  geom_point() + geom_smooth(formula = y ~ x, se = FALSE, method = "lm") +
  theme_minimal() +
  labs(y="Log of Quantity", x="Log of Price", subtitle="", title="Scatter plot of log of quantity and l
```




```r
Y<-data$lqu
X<-cbind(1, data$lprice)

b <- (solve(t(X)%*%X))%*%(t(X)%*%Y)
print(b)
```

```
     [,1]
```

[1,] 15.303260 [2,] -2.203468

Problem 4. Regress lqu on fuel, luxury, domestic, and a constant, create the residuals elqu Regress lprice on fuel, luxury, domestic, and a constant, create the residuals elprice Scatter plot the residuals elqu on vertical axis and elprice on horizontal axis. What is the estimated OLS slope associated with this scatter plot? Estimate a regression (no constant) to answer this and explain what theorem underlies the fact that this slope is the marginal effect of lprice on lqu in a regression that also features fuel,luxury, domestic, and a constant.

```r
##Reg 1
Y<-data$lqu
X<-cbind(1, data$fuel, data$luxury, data$domestic)
```

```r
b <- (solve(t(X)%*%X))%*%(t(X)%*%Y)
print(b)
```

```
       [,1]
```

[1,] 10.9239821 [2,] -0.3049930 [3,] -0.2616135 [4,] 2.4008093

```r
P <- (X %*% solve(t(X)%*%X) %*% t(X))
M <- diag(nrow(P))-P
e <- M %*% Y
elqu <- M %*% Y
# SSR <- t(e) %*% e
# SST <- t(Y)%*%Y
# SSE <- t(b) %*% t(X)  %*% X  %*% b
# R2<-1-(SSR/SST)
# reg <- lm(Y~X-1,)
# summary(reg)


##Reg 2
Y<-data$lprice
X<-cbind(1, data$fuel, data$luxury, data$domestic)

b <- (solve(t(X)%*%X))%*%(t(X)%*%Y)
print(b)
```

```
        [,1]
```

[1,] 1.82649309 [2,] 0.12496107 [3,] 0.50002347 [4,] -0.02752284

```r
P <- (X %*% solve(t(X)%*%X) %*% t(X))
M <- diag(nrow(P))-P
e <- M %*% Y
elprice <- M %*% Y
reg <- lm(data$lprice ~ data$fuel + data$luxury+ data$domestic)
summary(reg)
```

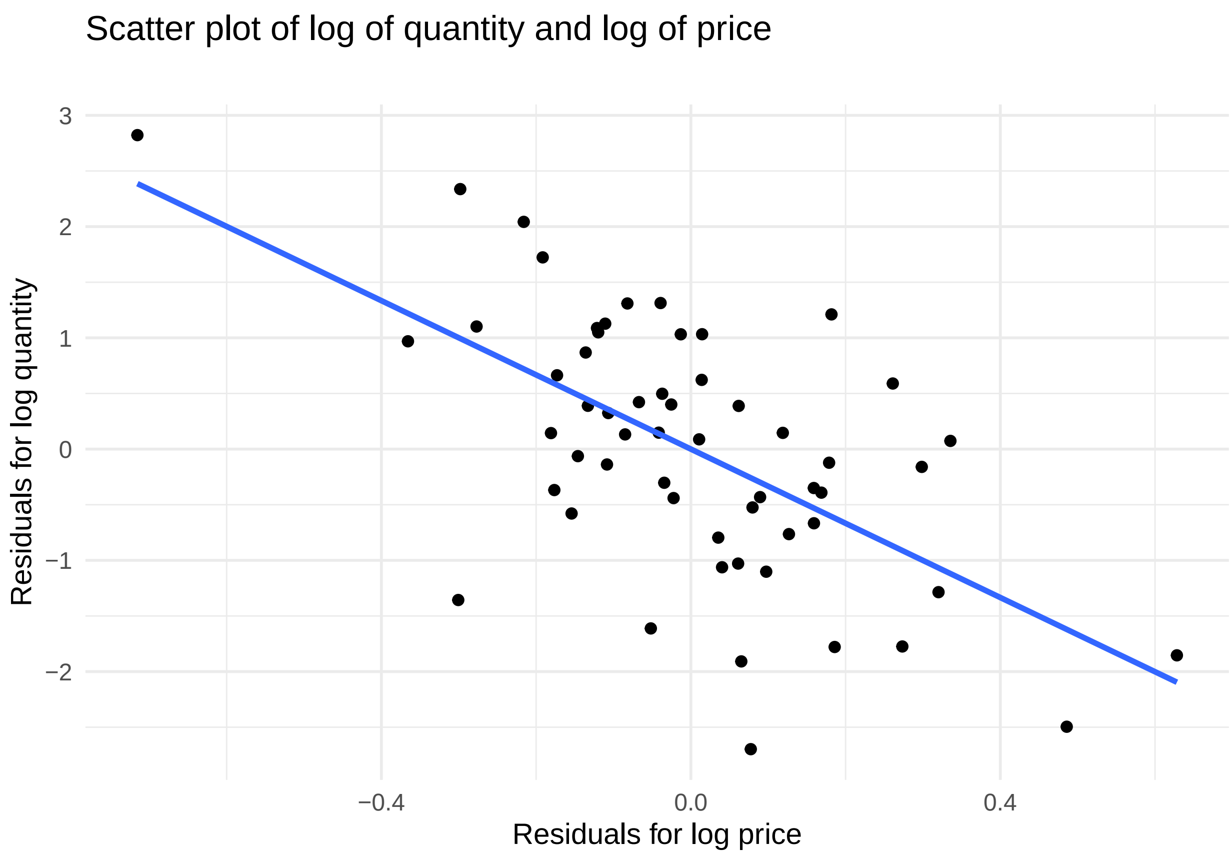
Call: lm(formula = data$lprice\ data$fuel + data$luxury + data$domestic)

Residuals: Min 1Q Median 3Q Max -0.71534 -0.12130 -0.02231 0.11886 0.62825

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.82649 0.13386 13.644 < 0.0000000000000002 $\textbf{\textit{data}}fuel$0.124960.015048.3090.000000000036 $*$
$*$ $*$ *data* ***luxury* 0.50002 0.13775 3.630 0.000639** data$domestic -0.02752 0.06875 -0.400 0.690535
— Signif. codes: 0 '***' **0.001** *'' 0.01* " 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2208 on 53 degrees of freedom Multiple R-squared: 0.6796, Adjusted R-squared: 0.6614 F-statistic: 37.47 on 3 and 53 DF, p-value: 0.0000000000003902

```r
ggplot(data, aes( x=elprice, y=elqu )) +
  geom_point() + geom_smooth(formula = y ~ x, se = FALSE, method = "lm") +
  theme_minimal() +
  labs(y="Residuals for log quantity", x="Residuals for log price", subtitle="", title="Scatter plot of
```




```r
#Reg 3
reg <- lm(elqu~elprice-1,)
summary(reg)
```

Call: lm(formula = elqu ~ elprice - 1)

Residuals: Min 1Q Median 3Q Max -2.4388 -0.5148 0.0865 0.6506 1.8171

Coefficients: Estimate Std. Error t value Pr(>|t|)
elprice -3.336 0.571 -5.841 0.000000275 *** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9179 on 56 degrees of freedom Multiple R-squared: 0.3786, Adjusted R-squared: 0.3675 F-statistic: 34.12 on 1 and 56 DF, p-value: 0.0000002749

Problem 5. Why is the slope estimate in 3 not equal to the one in 4? Theoretically speaking, when would they be equal?

```r
#Omiited Variables, ortogonality
```

Problem 6. Please interpret the OLS slope point estimate size, sign of the slope lprice estimate in 4. What is the pvalue for the estimated lprice coefficient? Use the stat tables for this.

```
reg <- lm(elqu~elprice-1,)
summary(reg)
```

Call: lm(formula = elqu ~ elprice - 1)

Residuals: Min 1Q Median 3Q Max -2.4388 -0.5148 0.0865 0.6506 1.8171

Coefficients: Estimate Std. Error t value Pr(>|t|)
elprice -3.336 0.571 -5.841 0.000000275 *** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9179 on 56 degrees of freedom Multiple R-squared: 0.3786, Adjusted R-squared: 0.3675 F-statistic: 34.12 on 1 and 56 DF, p-value: 0.0000002749

```
##The coef is -3.336 , this means while the t value es |5.841|, significant at 1% level.
```

Problem 7. Can you reject that the marginal effect of lprice on lqu is -4 conditional on all else equal (fuel, luxury, domestic, and a constant)? Do five steps in Hypothesis Testing at the 5% significance level against a twosided alternative. Get critical values from the relevant stats table.

```
reg <- lm(elqu~elprice-1,)
summary(reg)
```

Call: lm(formula = elqu ~ elprice - 1)

Residuals: Min 1Q Median 3Q Max -2.4388 -0.5148 0.0865 0.6506 1.8171

Coefficients: Estimate Std. Error t value Pr(>|t|)
elprice -3.336 0.571 -5.841 0.000000275 *** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9179 on 56 degrees of freedom Multiple R-squared: 0.3786, Adjusted R-squared: 0.3675 F-statistic: 34.12 on 1 and 56 DF, p-value: 0.0000002749

```
#Solve with Matrices
Y<-elqu
X<-elprice
b <- (solve(t(X)%*%X))%*%(t(X)%*%Y)
print(b)
```

```
      [,1]
```

[1,] -3.335674

```
P <- (X %*% solve(t(X)%*%X) %*% t(X))
M <- diag(nrow(P))-P
e <- M %*% Y  #e<-Y-X%*%b
df<-nrow(Y)-1
```

```r
s2<-as.numeric(t(e)%*%e)/df

vb<-s2*solve(t(X)%*%X) ##variance
vb
```

```
     [,1]
```

[1,] 0.3260789

```r
seb<-sqrt(diag(vb))
##The coef is -3.336 , this means while the t value es |5.841|, significant at 1% level.

seb_elqu<- sqrt(vb[1,1])
lprice_test <- -4
tn <- ( b[1,1] - lprice_test )/seb_elqu
tn
```

[1] 1.163376

```r
# Tn is low, so we fail to reject that b=-4
```

Problem 8. Estimate the sample data correlation of all these variables with each other: lqu, lprice, fuel,, weight, luxury, domestic. Suppose the population model is given by

$$\text{lqu}_i = \beta_0 + \beta_1 * \text{lprice}_i + \beta_2 * \text{domestic}_i + \beta_3 * \text{fuel}_i + \beta_4 * \text{luxury}_i + \epsilon_i *$$

and we estimate the model

$$\text{lqu}_i = \alpha_0 + \alpha_1 * \text{lprice}_i + \alpha_3 * \text{fuel}_i + \alpha_4 * \text{luxury}_i + \epsilon_i *$$

```r
data1<-data %>%
  select(lqu, lprice, fuel, weight, luxury, domestic)
(corr <- cor(data1))
```

```
          lqu     lprice       fuel      weight      luxury   domestic
```

lqu 1.0000000 -0.5121756 -0.3030351 -0.44153421 -0.12611097 0.57723120 lprice -0.5121756 1.0000000 0.7739287 0.91246510 0.50805990 0.08792450 fuel -0.3030351 0.7739287 1.0000000 0.82759388 0.30973689 0.15311909 weight -0.4415342 0.9124651 0.8275939 1.00000000 0.50441087 0.04385066 luxury -0.1261110 0.5080599 0.3097369 0.50441087 1.00000000 0.04803253 domestic 0.5772312 0.0879245 0.1531191 0.04385066 0.04803253 1.00000000

```r
#The correlation between fuel and domestic is positive, 0.3097369. Supposing that domestic is relevant
```

Problem 9. If I told you that research shows that advertising expenditures by car model are positively correlated with lprice and that when including advertising in addition to all factors in (8.b), the estimated weight coefficient does not change at all. What does this imply about the sample correlation between advertising and weight of cars in the sample?

##This mean both variables are no correlated, in the corr matrix the corr would be zero.

Problem 10. Suppose that research showed that the log of advertising is, on average, 5 times the log of price. Construct that advertising variable based on this fact and include it in a regression in addition to lprice and the other covariates in 8.b. Explain what happened.

Answer: Multicollinearity (Perfect combination of another variable). The the model does not calculate a coeff for advertising.

```
data <- data %>%
  mutate(advertising=5*log(price))

reg <- lm(data$lqu ~ data$lprice + data$fuel + data$luxury)
summary(reg)
```

Call: lm(formula = data$lqu ~ data$lprice + data$fuel + data$luxury)

Residuals: Min 1Q Median 3Q Max -2.9173 -1.1039 -0.2641 0.8686 2.8187

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.3902 1.7853 9.741 0.000000000000209 *** data$lprice -3.5886 0.8617 -4.164 0.000115 *** data$fuel
0.2167 0.1420 1.526 0.132968
data$luxury 1.5357 0.9668 1.588 0.118127
— Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.387 on 53 degrees of freedom Multiple R-squared: 0.3166, Adjusted R-squared: 0.2779 F-statistic: 8.184 on 3 and 53 DF, p-value: 0.0001431

```
reg <- lm(data$lqu ~ data$lprice + data$fuel + data$luxury + data$advertising)
summary(reg)
```

Call: lm(formula = data$lqu ~ data$lprice + data$fuel + data$luxury + data$advertising)

Residuals: Min 1Q Median 3Q Max -2.9173 -1.1039 -0.2641 0.8686 2.8187

Coefficients: (1 not defined because of singularities) Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.3902 1.7853 9.741 0.000000000000209 *** data$lprice -3.5886 0.8617 -4.164 0.000115 *** data$fuel
0.2167 0.1420 1.526 0.132968
data$luxury 1.5357 0.9668 1.588 0.118127 data$advertising NA NA NA NA
— Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.387 on 53 degrees of freedom Multiple R-squared: 0.3166, Adjusted R-squared: 0.2779 F-statistic: 8.184 on 3 and 53 DF, p-value: 0.0001431

Problem 11. Please estimate a specification that allows you to test the following. Research shows that luxury goods have a different price elasticity than nonluxury goods. The null hypothesis is that the marginal effect in lprice on log qu does not differ by luxury classification of the car. Write out the regression model that allows you to estimate and perform a hypothesis test for this null.

$$\text{lqu}_i = \beta_0 + \beta_1 * \text{lprice}_i + \beta_2 * \text{domestic}_i + \beta_3 * \text{fuel}_i + \beta_4 * \text{luxury}_i + \beta_4 * \text{lprice*luxury}_i + \epsilon_i *$$

```
data <- data %>%
  mutate(price_luxury=lprice*luxury)

reg <- lm(lqu ~ lprice+fuel+luxury+price_luxury, data)
summary(reg)
```

Call: lm(formula = lqu ~ lprice + fuel + luxury + price_luxury, data = data)

Residuals: Min 1Q Median 3Q Max -2.8728 -0.9882 -0.2149 0.8798 2.8211

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.5070 1.7756 9.860 0.000000000000169 ***lprice -3.8136 0.8730 -4.368 0.000059960224506***
fuel 0.2782 0.1487 1.871 0.0669 .
luxury 42.6068 31.3393 1.360 0.1798
price_luxury -10.8868 8.3033 -1.311 0.1956
— Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.378 on 52 degrees of freedom Multiple R-squared: 0.3385, Adjusted R-squared: 0.2876 F-statistic: 6.651 on 4 and 52 DF, p-value: 0.0002117

```
# reg <- lm(data$lqu ~ data$lprice+data$domestic+data$fuel+data$luxury)
# summary(reg)
```

##Answer, the interaction coef is not significant at all, wth a coef of -10.88 and t=|1.311|, so we fail to reject that the interaction plays an important role on modelign lqu.

Problem 12. Regress lqu on a constant, fuel, lprice, luxury, domestic, weight. (eq 12) Test the joint hypothesis that. beta_domestic= 1.5; beta_fuel=60*beta_-weight at the 1 percent significance level.

```
#a). Fit based test
#we estimate the OLS estimator of y on X
Y<-data$lqu
X<-cbind(1, 60*data$fuel, data$lprice, data$luxury,  -1.5*data$domestic, data$weight)
b<-solve(t(X)%*%X)%*%t(X)%*%Y
b
```

```
         [,1]
```

[1,] 17.616710641 [2,] 0.001030686 [3,] -3.835083206 [4,] 1.278030470 [5,] -1.570670070 [6,] 0.001414092

```
yr<-data$lqu+60*data$fuel+data$lprice+data$luxury-1.5*data$domestic+data$weight
i<-c(rep(1,nrow(data)))
M0<-diag(nrow(data))-i%*%solve(t(i)%*%i)%*%t(i)
er<-M0%*%yr
ssrr<-as.numeric(t(er)%*%er)

e<-data$lqu-X%*%b
ssr<-as.numeric(t(e)%*%e)
df <- length(yr) - 1
s2 <- as.numeric(t(e)%*%e)/df
F_fitBased<-(ssrr-ssr)/(2*(s2))
F_fitBased
```

[1] 3674911

```r
#a). Wald test
Y<-data$lqu
X<-cbind(1, data$fuel, data$lprice, data$luxury,  data$domestic, data$weight)

b <- (solve(t(X)%*%X))%*%(t(X)%*%Y)
print(b)
```

```
          [,1]
```

[1,] 17.616710641 [2,] 0.061841144 [3,] -3.835083206 [4,] 1.278030470 [5,] 2.356005105 [6,] 0.001414092

```r
P <- (X %*% solve(t(X)%*%X) %*% t(X))
M <- diag(nrow(P))-P
e <- M %*% Y  #e<-Y-X%*%b
df<-length(Y)-6
s2<-as.numeric(t(e)%*%e)/df
vb<-s2*solve(t(X)%*%X) ##variance

#s2 from above estimate of sigma squared
# const - fuel - lprice - luxury - domestic - weight
R_1=c(0,   1 ,     0 ,      0 ,     0,        -60)
R_2=c(0,   0 ,     0 ,      0 ,     1,         0)
R<-t(cbind(R_1,R_2))
q<-c(0,1.5)

VRbq<-s2* R %*% solve(t(X) %*% X) %*% t(R)
Fw<-t(R %*% b-q) %*% solve(VRbq) %*% (R %*% b-q)
#divide by 2 , J=2
(Fw<-Fw * 0.5)
```

```
      [,1]
```

[1,] 4.131466

```r
## They are no the same
```

Problem 13. Without running any additional regressions and starting from the baseline regression in reg question 8.a. A) Will omitting fuel create an OVB problem the OLS estimator of lprice? B)Compute the variance inflated factor (VIF) for the variable height to be potentially also included into the (reg of question 8.a) model to explain the variation in lqu. Feel free to use the lm canned function to get what you need for the VIFj, for all j. C)Will including this variable height with the others in (model in question 8.a) result in multicollinearity problems?

$$\text{lqu}_i = \beta_0 + \beta_1 * \text{lprice}_i + \beta_2 * \text{domestic}_i + \beta_3 * \text{fuel}_i + \beta_4 * \text{luxury}_i + \epsilon_i *$$

```
#a)
##I would say It could create a problem since fuel and lprice have a strong correlation of 0.7739287.
corr
```

lqu lprice fuel weight luxury domestic

lqu 1.0000000 -0.5121756 -0.3030351 -0.44153421 -0.12611097 0.57723120 lprice -0.5121756 1.0000000 0.7739287 0.91246510 0.50805990 0.08792450 fuel -0.3030351 0.7739287 1.0000000 0.82759388 0.30973689 0.15311909 weight -0.4415342 0.9124651 0.8275939 1.00000000 0.50441087 0.04385066 luxury -0.1261110 0.5080599 0.3097369 0.50441087 1.00000000 0.04803253 domestic 0.5772312 0.0879245 0.1531191 0.04385066 0.04803253 1.00000000

```
#b)
reg <- lm(lqu ~ lprice + domestic + fuel + luxury, data)
summary(reg)
```

Call: lm(formula = lqu ~ lprice + domestic + fuel + luxury, data = data)

Residuals: Min 1Q Median 3Q Max -2.4388 -0.5148 0.0865 0.6506 1.8171

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.01657 1.22679 13.871 < 0.0000000000000002 ***lprice -3.33567 0.59259 -5.629 0.000000737197*** domestic 2.30900 0.29706 7.773 0.000000000291 ** *fuel 0.11184 0.09845 1.136 0.261*
luxury 1.40630 0.66403 2.118 0.039
— Signif. codes: 0 '*' ***0.001*** '' *0.01* " 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9526 on 52 degrees of freedom Multiple R-squared: 0.6839, Adjusted R-squared: 0.6596 F-statistic: 28.12 on 4 and 52 DF, p-value: 0.000000000001861

```
(vif1 <- 1/(1-summary(reg)$r.squared))
```

[1] 3.163388

```
reg <- lm(lqu ~ lprice + domestic + fuel + luxury + height, data)
summary(reg)
```

Call: lm(formula = lqu ~ lprice + domestic + fuel + luxury + height, data = data)

Residuals: Min 1Q Median 3Q Max -2.36422 -0.52199 0.02135 0.57366 1.82002

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.88452 3.09250 4.166 0.00012 ***lprice -3.35883 0.58658 -5.726 0.000000549045*** domestic 2.40205 0.30083 7.985 0.000000000154 ** *fuel 0.09854 0.09785 1.007 0.31862*
luxury 1.36592 0.65764 2.077 0.04286
height 0.03037 0.02091 1.453 0.15242
— Signif. codes: 0 '*' ***0.001*** '' *0.01* " 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9426 on 51 degrees of freedom Multiple R-squared: 0.6964, Adjusted R-squared: 0.6667 F-statistic: 23.4 on 5 and 51 DF, p-value: 0.000000000003885

```
(vif2 <- 1/(1-summary(reg)$r.squared))
```

[1] 3.294295

```
##We notice adding height increased r.squared from 3.16 to 3.29, this indicates a possible problem of m
```

Problem 14.Suppose a car salesman told you that the conditional variance in the unobserved determinants of the log quantity (lqu) for luxury cars is three times the variance for nonluxury cars.

##A) Which assumption no longer holds when we derive the statistical properties of the OLS estimators for the linear model in Reg 8a. B) Let the variance of the disturbance of log quantity for luxury=1 be 3 times the variance for luxury=0. In R create a matrix Omega, its inverse, and the positive definite matrix C such that the inverse of Omega = C C' as derived in lecture. C) ## Homoscedasticity is not not held under this scenario.

```
table(data$luxury)
```

0 1 54 3

```
nonlux <- 54
lux <- 3

topm <- diag(nonlux)
toplux <- matrix(data=0, nrow=nonlux, ncol=lux)
top <- cbind(topm, toplux)
toplux <- matrix(data=0, ncol=nonlux, nrow=lux)
topm <- diag(lux)*3
bottom <- cbind(toplux,topm)
A <- rbind(top,bottom)

A.eig <- eigen(solve(A))
D<-diag((A.eig$values))
V<-A.eig$vectors
C <- A.eig$vectors %*% diag(sqrt(A.eig$values)) %*% solve(A.eig$vectors)
```