

Lab6: OLS Regression

QSSBA, Spring 2025

Zhaopeng Qu

Nanjing University

5/13/2025

Section 1

Introduction

Library Packages

```
library(tidyverse)
library(magrittr)
library(haven)
library(ggplot2)
library(AER)
library(stargazer)
```

Reading the Data

```
ca <- read_dta("Data/caschool.dta")
```

Overlook the Data

```
glimpse(ca,width = 90,size='tiny')
ca_small <- ca %>%
  select(c(testscr,str,avginc,meal_pct,calw_pct,el_pct,
           expn_stu,comp_stu))
ca_data <-data.frame(ca_small)
```

```
## Rows: 420
```

```
## Columns: 18
```

```
## $ observat <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
```

```
## $ dist_cod <dbl> 75119, 61499, 61549, 61457, 61523, 62042,
```

```
## $ county <chr> "Alameda", "Butte", "Butte", "Butte", "But
```

```
## $ district <chr> "Sunol Glen Unified", "Manzanita Elementar
```

```
## $ gr_span <chr> "KK-08", "KK-08", "KK-08", "KK-08", "KK-08
```

```
## $ enrl_tot <dbl> 195, 240, 1550, 243, 1335, 137, 195, 888,
```

```
## $ teachers <dbl> 10.90, 11.15, 82.90, 14.00, 71.50, 6.40, 1
```

```
## $ calw_pct <dbl> 0.5102, 15.4167, 55.0323, 36.4754, 33.1086
```

```
## $ meal_pct <dbl> 0.0408, 17.0167, 76.8886, 77.0408, 78.4876
```

OLS Regression

- Simple OLS Regression

$$TestScore = \beta_0 + \beta_1 STR + u$$

```
reg <- lm(testscr ~ str, data = ca_data)
reg
```

```
##
## Call:
## lm(formula = testscr ~ str, data = ca_data)
##
## Coefficients:
## (Intercept)          str
##      698.93         -2.28
```

OLS Regression with S.E

```
##
## Call:
## lm(formula = testscr ~ str, data = ca_data)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-47.727	-14.251	0.483	12.822	48.540

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	698.9330	9.4675	73.825	< 2e-16 ***
str	-2.2798	0.4798	-4.751	2.78e-06 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.58 on 418 degrees of freedom
```

OLS Regression with Robust S.E

```
r_se <- sqrt(diag(vcovHC(reg)))  
r_se
```

```
## (Intercept)          str  
##  10.4605338    0.5243585
```

Multiple OLS Regression

```
attach(ca_data)
spec1 <- lm(testscr ~ str)
spec2 <- lm(testscr ~ str + el_pct)
spec3 <- lm(testscr ~ str + el_pct + meal_pct)
spec4 <- lm(testscr ~ str + el_pct + calw_pct)
spec5 <- lm(testscr ~ str + el_pct + meal_pct + calw_pct)
```

Multiple OLS Regression with Robust S.E.

```
rob_se <- list(  
  sqrt(diag(vcovHC(spec1, type = "HC1"))),  
  sqrt(diag(vcovHC(spec2, type = "HC1"))),  
  sqrt(diag(vcovHC(spec3, type = "HC1"))),  
  sqrt(diag(vcovHC(spec4, type = "HC1"))),  
  sqrt(diag(vcovHC(spec5, type = "HC1")))  
)
```

Regeression Table(1)

```
stargazer(spec1, spec2, spec3, spec4, spec5, type = "latex")
```

Regeression Table(2)

	(1)	(2)
str	-2.280*** (0.480)	-1.101*** (0.380)
el_pct		-0.650*** (0.039)
meal_pct		

Regeression Table(3): Homo S.E

```
stargazer(spec1, spec2, spec3, spec4, spec5,
  type = "latex",no.space=TRUE,header = FALSE,
  column.sep.width = "5pt",digits = 3,
  font.size = "scriptsize",
  dep.var.caption = "Dependent Variable:Test Score",
  dep.var.labels.include = F,df = F,
  omit.stat=c("rsq","ser"),
  notes = "S.E. are shown in the parentheses",
  notes.append = T,
  title = "Test Score and Class Size",
  covariate.labels=c("student-teacher ratio",
                     "english-learners",
                     "free lunch",
                     "low-income program")
)
```

Regeression Table (4): Homo S.E in LaTeX

Table1: Test Score and Class Size

	Dependent Variable: Test Score				
	(1)	(2)	(3)	(4)	(5)
student-teacher ratio	-2.280*** (0.480)	-1.101*** (0.380)	-0.998*** (0.239)	-1.308*** (0.307)	-1.014*** (0.240)
english-learners		-0.650*** (0.039)	-0.122*** (0.032)	-0.488*** (0.033)	-0.130*** (0.034)
free lunch			-0.547*** (0.022)		-0.529*** (0.032)
low-income program				-0.790*** (0.053)	-0.048 (0.061)
Constant	698.933*** (9.467)	686.032*** (7.411)	700.150*** (4.686)	697.999*** (6.024)	700.392*** (4.698)
Observations	420	420	420	420	420
Adjusted R ²	0.049	0.424	0.773	0.626	0.773
F Statistic	22.575***	155.014***	476.306***	234.638***	357.054***

Note:

* p<0.1; ** p<0.05; *** p<0.01
S.E. are shown in the parentheses

Regeression Table(5): Robust S.E

```
stargazer(spec1, spec2, spec3, spec4, spec5,
  type = "latex",no.space=TRUE,header = FALSE,
  column.sep.width = "5pt",digits = 3,
  font.size = "scriptsize",
  dep.var.caption = "Dependent Variable: Test Score",
  dep.var.labels.include = F,df = F,
  omit.stat=c("rsq","ser"),
  se = rob_se # Robust S.E
  notes = "S.E. are shown in the parentheses",
  notes.append = T,
  title = "Test Score and Class Size",
  covariate.labels=c("student-teacher ratio","english-
                    "free lunch","low-income program"
  )
```

Regeression Table(6): Robust S.E in LaTeX

Table1: Test Score and Class Size

	Dependent Variable: Test Score				
	(1)	(2)	(3)	(4)	(5)
student-teacher ratio	-2.280*** (0.519)	-1.101** (0.433)	-0.998*** (0.270)	-1.308*** (0.339)	-1.014*** (0.269)
english-learners		-0.650*** (0.031)	-0.122*** (0.033)	-0.488*** (0.030)	-0.130*** (0.036)
free lunch			-0.547*** (0.024)		-0.529*** (0.038)
low-income program				-0.790*** (0.068)	-0.048 (0.059)
Constant	698.933*** (10.364)	686.032*** (8.728)	700.150*** (5.568)	697.999*** (6.920)	700.392*** (5.537)
Observations	420	420	420	420	420
Adjusted R ²	0.049	0.424	0.773	0.626	0.773
F Statistic	22.575***	155.014***	476.306***	234.638***	357.054***

Note:

* p<0.1; ** p<0.05; *** p<0.01

Robust S.E. are shown in the parentheses

Regeression Table(7): in MS Word

```
stargazer(spec1, spec2, spec3, spec4, spec5, type = "text")
```

Regeression Table(8): in MS Word

```
##
## Table1: Test Score and Class Size
## =====
##                               Dependent Variable: Test
##                               -----
##                               (1)      (2)      (3)      (4)
## -----
## student-teacher ratio -2.280***   -1.101**   -0.998***   -1.3
##                               (0.519)   (0.433)   (0.270)   (0.
## english-learners      -0.650***   -0.122***   -0.4
##                               (0.031)   (0.033)   (0.
## free lunch            -0.547***
##                               (0.024)
## low-income program    -0.7
##                               (0.
## Constant              698.933***  686.032***  700.150***  697.
```

Mutiple OLS Regression

```
model <- lm( testscr~ str + el_pct + expn_stu, data = ca)
summary(model)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + expn_stu, data = ca)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-51.340	-10.111	0.293	10.318	43.181

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	649.577947	15.205719	42.719	< 2e-16 ***
## str	-0.286399	0.480523	-0.596	0.55149
## el_pct	-0.656023	0.039106	-16.776	< 2e-16 ***
## expn_stu	0.000000	0.001410	0.720	0.00042

Joint Hypothesis Testing Using the F-Statistic

The estimated model is

$$\widehat{TestScore} = 649.58 - 0.29 \times str - 0.66 \times english + 0.00387 \times expenditure.$$

(15.21)
(0.48)
(0.04)
(0.00141)

Joint Hypothesis Testing Using the F-Statistic

- H_0 : Both the coefficient on *str* and the coefficient on *expenditure* are zero?

Joint Hypothesis Testing Using the F-Statistic

- H_0 : Both the coefficient on *str* and the coefficient on *expenditure* are zero?
- H_1 : At least one of them is not zero.

Joint Hypothesis Testing Using the F-Statistic

- H_0 : Both the coefficient on *str* and the coefficient on *expenditure* are zero?
- H_1 : At least one of them is not zero.
- The homoskedasticity-only F -Statistic is given by

$$F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})/q}{SSR_{\text{unrestricted}}/(n - k - 1)}$$

Joint Hypothesis Testing Using the F-Statistic

```
linearHypothesis(model, c("str=0", "expn_stu=0"))
```

```
## Linear hypothesis test
```

```
##
```

```
## Hypothesis:
```

```
## str = 0
```

```
## expn_stu = 0
```

```
##
```

```
## Model 1: restricted model
```

```
## Model 2: testscr ~ str + el_pct + expn_stu
```

```
##
```

```
##      Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1      418 89000
```

```
## 2      416 85700   2    3300.3 8.0101 0.000386 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Joint Hypothesis Testing Using the F-Statistic

```
linearHypothesis(model, c("str=0", "expn_stu=0"),
                  white.adjust = "hc1")
```

```
## Linear hypothesis test
##
## Hypothesis:
## str = 0
## expn_stu = 0
##
## Model 1: restricted model
## Model 2: testscr ~ str + el_pct + expn_stu
##
## Note: Coefficient covariance matrix supplied.
##
##      Res.Df Df      F    Pr(>F)
## 1      418
## 2      416  2 5.4337 0.004682 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Joint Hypothesis Testing Using the F-Statistic

overall regression F-statistic

$$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0 \text{ for at least one } j = 1, 2, 3.$$

Joint Hypothesis Testing Using the F-Statistic

```
linearHypothesis(model, c("str=0", "el_pct=0", "expn_stu=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## str = 0
## el_pct = 0
## expn_stu = 0
##
## Model 1: restricted model
## Model 2: testscr ~ str + el_pct + expn_stu
##
##      Res.Df    RSS Df Sum of Sq      F      Pr(>F)
## 1      419 152110
## 2      416   85700   3      66410 107.45 < 2.2e-16 ***
## ---
```

Joint Hypothesis Testing Using the F-Statistic

```
summary(model)$fstatistic # from the model's summary
```

```
##      value      numdf      dendf  
## 107.4547      3.0000 416.0000
```

Mutiple OLS Regression: Poly and Log

- **quadratic regression model:**

$$TestScore_i = \beta_0 + \beta_1 \times income_i + \beta_2 \times income_i^2 + u_i,$$

Mutiple OLS Regression: Poly and Log

- **quadratic regression model:**

$$TestScore_i = \beta_0 + \beta_1 \times income_i + \beta_2 \times income_i^2 + u_i,$$

- **cubic regression model:**

$$TestScore_i = \beta_0 + \beta_1 \times income_i + \beta_2 \times income_i^2 + \beta_3 \times income_i^3 + u_i,$$

Mutiple OLS Regression

```
# fit the quadratic Model
quadratic_model <- lm(testscr ~ avginc + I(avginc^2),
                      data = ca)
# obtain the model summary
coeftest(quadratic_model, vcov. = vcovHC, type = "HC1")
```

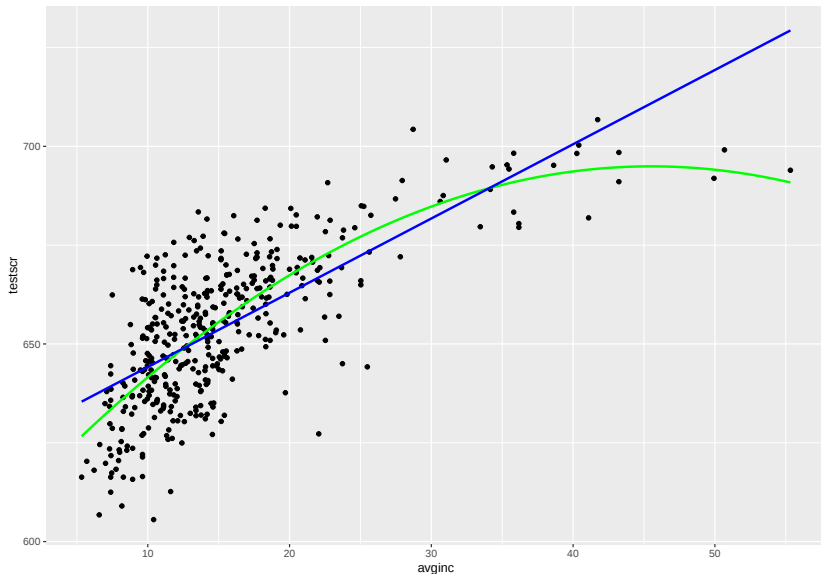
```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 607.3017350   2.9017539 209.2878 < 2.2e-16 ***
## avginc       3.8509947   0.2680941  14.3643 < 2.2e-16 ***
## I(avginc^2)  -0.0423085   0.0047803  -8.8505 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure: Linear and Quadratic Regression

```
xmin <- min(ca$avginc)
xmax <- max(ca$avginc)
predicted <- data.frame(avginc = seq(xmin,xmax,
                                     length.out = 100))
predicted$testscr.hat <- predict(quadratic_model,predicted)

ggplot(data =ca,aes(x=avginc, y=testscr)) +
  geom_point(colour="black")+
  geom_line(data = predicted,aes(x=avginc, y=testscr.hat),
            size = 1,colour="green") +
  stat_smooth(method=lm,se=FALSE,colour="blue")
```

Figure: Linear and Quadratic Regression



Mutiple OLS Regression: Cubic

```
cubic_model <- lm(testscr ~ avginc + I(avginc^2)+I(avginc^3),
```

Joint Hypothesis Testing Using the F-Statistic

- Set up hypothesis matrix for $\beta_2 = 0$ and $\beta_3 = 0$

Joint Hypothesis Testing Using the F-Statistic

- Set up hypothesis matrix for $\beta_2 = 0$ and $\beta_3 = 0$
- thus

$$\begin{aligned}\mathbf{R}\beta &= \mathbf{s} \\ \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} .\end{aligned}$$

Joint Hypothesis Testing Using the F-Statistic

```
R <- rbind(c(0, 0, 1, 0), c(0, 0, 0, 1)) # set up hypothesis matrix  
linearHypothesis(cubic_model, hypothesis.matrix = R,  
                 white.adjust = "hc1")
```

Joint Hypothesis Testing Using the F-Statistic

```
## Linear hypothesis test
##
## Hypothesis:
##  $I(\text{avginc}^2) = 0$ 
##  $I(\text{avginc}^3) = 0$ 
##
## Model 1: restricted model
## Model 2: testscr ~ avginc +  $I(\text{avginc}^2) + I(\text{avginc}^3)$ 
##
## Note: Coefficient covariance matrix supplied.
##
##      Res.Df Df      F      Pr(>F)
## 1      418
## 2      416  2 37.691 9.043e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Logarithms

- three logarithmic regression models:

Case	Population regression function
I.linear-log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$
II.log-linear	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III.log-log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

Case I: X is in Logarithm, Y is not.

```
LinearLog_model <- lm(testscr ~ log(avginc), data = ca)
coeftest(LinearLog_model,
          vcov = vcovHC, type = "HC1") # compute robust summary
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 557.8323      3.8399 145.271 < 2.2e-16 ***
## log(avginc)  36.4197      1.3969  26.071 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Case II: Y is in Logarithm, X is not

```
LogLinear_model <- lm(log(testscr) ~ avginc, data = ca)
coeftest(LogLinear_model,
          vcov = vcovHC, type = "HC1") # compute robust summary
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 6.43936233 0.00289382 2225.210 < 2.2e-16 ***
## avginc      0.00284407 0.00017509   16.244 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Case III: X and Y are in Logarithms

```
LogLog_model <- lm(log(testscr) ~ log(avginc), data = ca)
coeftest(LogLog_model,
          vcov = vcovHC, type = "HC1") # compute robust summary
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 6.3363494   0.0059246 1069.501 < 2.2e-16 ***
## log(avginc) 0.0554190   0.0021446   25.841 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interactions Between Independent Variables

- 1 Interactions between two binary variables.

Interactions Between Independent Variables

- 1 Interactions between two binary variables.
- 2 Interactions between a binary and a continuous variable.

Interactions Between Independent Variables

- 1 Interactions between two binary variables.
- 2 Interactions between a binary and a continuous variable.
- 3 Interactions between two continuous variables.

Interactions Between Two Binary Variables

- Now let

$$HiSTR = \begin{cases} 1, & \text{if } STR \geq 20 \\ 0, & \text{else,} \end{cases}$$

$$HiEL = \begin{cases} 1, & \text{if } PctEL \geq 10 \\ 0, & \text{else.} \end{cases}$$

```
# append HiSTR to CASchools
ca$HiSTR <- as.numeric(ca$str >= 20)
# append HiEL to CASchools
ca$HiEL <- as.numeric(ca$el_pct >= 10)
```

Interactions Between Two Binary Variables

Regression is

$$TestScore = \beta_0 + \beta_1 HiSTR + \beta_2 HiEL + \beta_3 (HiSTR \times HiEL) + u_i$$

```
bi_model <- lm(testscr ~ HiSTR * HiEL, data = ca) # a binary i
coeftest(bi_model, vcov. = vcovHC, type = "HC1") # print a rob
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  664.1433      1.3881  478.4588 < 2.2e-16 ***
## HiSTR        -1.9078      1.9322   -0.9874    0.3240
## HiEL        -18.1629      2.3460   -7.7422  7.502e-14 ***
## HiSTR:HiEL    -3.4943      3.1212   -1.1195    0.2636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interactions Between a Continuous and a Binary Variable

- Answer the question *whether the effect on test scores of decreasing the student-teacher ratio depends on whether there are many or few English learners.*

$$TestScore_i = \beta_0 + \beta_1 str_i + \beta_2 HiEL_i + \beta_2(str_i \times HiEL_i) + u_i.$$

Interactions Between a Continuous and a Binary Variable

```
# estimate the model
bci_model <- lm(testscr ~ str + HiEL + str * HiEL, data = ca)

# print robust summary of coefficients to the console
coeftest(bci_model, vcov. = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  682.24584    11.86781  57.4871  <2e-16 ***
## str          -0.96846     0.58910  -1.6440   0.1009
## HiEL          5.63914    19.51456   0.2890   0.7727
## str:HiEL     -1.27661     0.96692  -1.3203   0.1875
## ---
```

Interactions Between Two Continuous Variables

```
# the interaction between 'PctEL' and 'str'
cci_model <- lm(testscr ~ str + el_pct + str * el_pct,
                 data = ca)
coeftest(cci_model, vcov. = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##              Estimate   Std. Error t value Pr(>|t|)
## (Intercept) 686.3385246   11.7593451  58.3654 < 2e-16 ***
## str         -1.1170183    0.5875135  -1.9013  0.05796 .
## el_pct      -0.6729114    0.3741231  -1.7986  0.07280 .
## str:el_pct   0.0011618    0.0185357   0.0627  0.95005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: The Demand for Economic Journals(Table 8.2)

- The central question is: **how elastic is the demand by libraries for economic journals?**

Example: The Demand for Economic Journals(Table 8.2)

- The central question is: **how elastic is the demand by libraries for economic journals?**
- Dependent Variable: **the number of subscription**

Example: The Demand for Economic Journals(Table 8.2)

- The central question is: **how elastic is the demand by libraries for economic journals?**
- Dependent Variable: **the number of subscription**
- Independent Variable: **the journal's subscription price** in terms of “price per citation”

Example: The Demand for Economic Journals

```
# define and rename variables
data("Journals")
Journals$PricePerCitation <- Journals$price/Journals$citations
Journals$Age <- 2000 - Journals$foundingyear
Journals$Characters <- Journals$charpp * Journals$pages/10^6
Journals$Subscriptions <- Journals$subs
```

- The range of “price per citation” is quite large:

```
# compute summary statistics for price per citation
summary(Journals$PricePerCitation)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.005223	0.464495	1.320513	2.548455	3.440171	24.459459

Example: The Demand for Economic Journals

$$(I) \ln(\text{Subscriptions}_i) = \beta_0 + \beta_1 \ln(\text{PricePerCitation}_i) + u_i$$

$$(II) \ln(\text{Subscriptions}_i) = \beta_0 + \beta_1 \ln(\text{PricePerCitation}_i) + \beta_4 \ln(\text{Age}_i) \\ + \beta_6 \ln(\text{Characters}_i) + u_i$$

$$(III) \ln(\text{Subscriptions}_i) = \beta_0 + \beta_1 \ln(\text{PricePerCitation}_i) \\ + \beta_2 \ln(\text{PricePerCitation}_i)^2 \\ + \beta_3 \ln(\text{PricePerCitation}_i)^3 \\ + \beta_4 \ln(\text{Age}_i) \\ + \beta_5 [\ln(\text{Age}_i) \times \ln(\text{PricePerCitation}_i)] \\ + \beta_6 \ln(\text{Characters}_i) + u_i$$

Example: The Demand for Economic Journals

$$\begin{aligned}(IV) \ln(\text{Subscriptions}_i) = & \beta_0 + \beta_1 \ln(\text{PricePerCitation}_i) + \beta_4 \ln(\text{Age}_i) \\ & + \beta_5 [\ln(\text{Age}_i) \times \ln(\text{PricePerCitation}_i)] \\ & + \beta_6 \ln(\text{Characters}_i) + u_i\end{aligned}$$

Example: The Demand for Economic Journals

```
# Estimate models (I) - (II)
Journals_mod1 <- lm(log(Subscriptions)
                    ~ log(PricePerCitation),
                    data = Journals)

Journals_mod2 <- lm(log(Subscriptions)
                    ~ log(PricePerCitation)
                    + log(Age) + log(Characters),
                    data = Journals)
```

Example: The Demand for Economic Journals

```
# Estimate models (III) - (IV)
Journals_mod3 <- lm(log(Subscriptions) ~
  log(PricePerCitation) + I(log(PricePerCitation)^2)
  + I(log(PricePerCitation)^3) + log(Age)
  + log(Age):log(PricePerCitation) + log(Characters),
  data = Journals)

Journals_mod4 <- lm(log(Subscriptions) ~
  log(PricePerCitation) + log(Age)
  + log(Age):log(PricePerCitation) +
  log(Characters),
  data = Journals)
```

Example: The Demand for Economic Journals

```
# gather robust standard errors in a list
rob_se <- list(sqrt(diag(vcovHC(Journals_mod1, type = "HC1"))),
               sqrt(diag(vcovHC(Journals_mod2, type = "HC1"))),
               sqrt(diag(vcovHC(Journals_mod3, type = "HC1"))),
               sqrt(diag(vcovHC(Journals_mod4, type = "HC1"))))

# generate a Latex table using stargazer
stargazer(Journals_mod1,
          Journals_mod2,
          Journals_mod3,
          Journals_mod4,
          se = rob_se,
          digits = 3,
          column.labels = c("(I)", "(II)", "(III)", "(IV)"))
```

	Dependent Variable: logarithm of subscriptions			
	(1)	(2)	(3)	(4)
$\log(\text{PricePerCitation})$	-0.533*** (0.034)	-0.408*** (0.044)	-0.961*** (0.160)	-0.899*** (0.145)
$I(\log(\text{PricePerCitation})^2)$			0.017 (0.025)	
$I(\log(\text{PricePerCitation})^3)$			0.004 (0.006)	
$\log(\text{Age})$		0.424*** (0.119)	0.373*** (0.118)	0.374*** (0.118)
$\log(\text{Characters})$		0.206** (0.098)	0.235** (0.098)	0.229** (0.096)
$\log(\text{PricePerCitation}) : \log(\text{Age})$			0.156*** (0.052)	0.141*** (0.040)
Constant	4.766*** (0.055)	3.207*** (0.380)	3.408*** (0.374)	3.434*** (0.367)
Observations	180	180	180	180
Adjusted R ²	0.555	0.607	0.622	0.626
Residual Std. Error	0.750	0.705	0.691	0.688

Note:

* p<0.1; ** p<0.05; *** p<0.01
Robust S.E. are shown in the parentheses

Dependent Variable: logarithm of subscriptions						
	(1)	(2)	(3)	(4)	(5)	(6)
str	-0.998*** (0.270)	-0.734*** (0.257)	-0.968 (0.589)	-0.531 (0.342)	-0.772*** (0.256)	83.701*** (28.497)
el_pct	-0.122*** (0.033)	-0.176*** (0.034)				
I(str^2)						-4.381*** (1.441)
I(str^3)						0.075*** (0.024)
meal_pct	-0.547*** (0.024)	-0.398*** (0.033)		-0.411*** (0.029)	-0.417*** (0.028)	-0.418*** (0.029)
log(avginc)		11.569*** (1.819)		12.124*** (1.798)	11.820*** (1.775)	11.800*** (1.778)
HiEL			5.639 (19.515)	5.498 (9.795)	-5.791*** (1.027)	816.075** (327.674)
str:HiEL			-1.277 (0.967)	-0.578 (0.496)		-123.282** (50.213)
I(str^2):HiEL						6.121** (2.542)
I(str^3):HiEL						-0.101** (0.043)
Constant	700.150*** (5.568)	658.552*** (8.642)	682.246*** (11.868)	653.666*** (9.869)	659.356*** (8.421)	122.354 (185.519)
Observations	420	420	420	420	420	420
Adjusted R ²	0.773	0.794	0.305	0.795	0.795	0.799
Residual Std. Error	9.080	8.643	15.880	8.629	8.634	8.547

Note:

*p<0.1; **p<0.05; ***p<0.01
Robust S.E. are shown in the parentheses

Section 2

Resources and References

Some useful online resources

- Hlavac, Marek (2018), stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.

Some useful online resources

- Hlavac, Marek (2018), stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.
- Torres-Reyna, Oscar (2014), Using stargazer to report regression output and descriptive statistics in R

Some useful online resources

- Hlavac, Marek (2018), stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.
- Torres-Reyna, Oscar (2014), Using stargazer to report regression output and descriptive statistics in R
- Zhu, Hao (2019), Create Awesome HTML Table with knitr::kable and kableExtra