Assignment Template

Your Name Here

2024-07-31

This is an **Quarto document**, to be used as a template for your assignment submissions. It uses **R Markdown**, which is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. You can put code, code output, math, and text in a single document, all of which are then compiled into a single pdf document.

We will only need the very basics for your assignments – don't worry about making beautiful documents at this stage. For more details on making Quarto documents, see https://quarto.org/docs/get-started/hello/rstudio.html.

Structure

At the top of the `.qmd` file you will see some code in-between two `---`. This is known as "yaml" code. Don't worry about what "yaml" stands for, change only the title, author and date.

R code can be included in **code chunks**, as in the example below, which loads the "wooldridge" package, which contains all of the datasets from the Wooldridge (2012) textbook. Remember that all packages must first be installed in R system before it can be loaded into an R session.

````
```{r warning=FALSE}
library(wooldridge)
ls("package:wooldridge")
```
````

When you click the **Render** button a document will be generated that includes both content as well as the output of any **embedded R code chunks** within the document. The "warning=FALSE" is a "chunk option". In this example we suppress all warning messages. When the document is compiled, the below is obtained.

```r
library(wooldridge)
ls("package:wooldridge")
```

```
  [1] "admnrev"       "affairs"       "airfare"       "alcohol"
  [5] "apple"         "approval"      "athlet1"       "athlet2"
  [9] "attend"        "audit"         "barium"        "beauty"
 [13] "benefits"      "beveridge"     "big9salary"    "bwght"
 [17] "bwght2"        "campus"        "card"          "catholic"
 [21] "cement"        "census2000"    "ceosal1"       "ceosal2"
 [25] "charity"       "consump"       "corn"          "countymurders"
 [29] "cps78_85"      "cps91"         "crime1"        "crime2"
 [33] "crime3"        "crime4"        "discrim"       "driving"
 [37] "earns"         "econmath"      "elem94_95"     "engin"
 [41] "expendshares"  "ezanders"      "ezunem"        "fair"
 [45] "fertil1"       "fertil2"       "fertil3"       "fish"
 [49] "fringe"        "gpa1"          "gpa2"          "gpa3"
 [53] "happiness"     "hprice1"       "hprice2"       "hprice3"
 [57] "hseinv"        "htv"           "infmrt"        "injury"
 [61] "intdef"        "intqrt"        "inven"         "jtrain"
 [65] "jtrain2"       "jtrain3"       "jtrain98"      "k401k"
 [69] "k401ksubs"     "kielmc"        "labsup"        "lawsch85"
 [73] "loanapp"       "lowbrth"       "mathpnl"       "meap00_01"
 [77] "meap01"        "meap93"        "meapsingle"    "minwage"
 [81] "mlb1"          "mroz"          "murder"        "nbasal"
 [85] "ncaa_rpi"      "nyse"          "okun"          "openness"
 [89] "pension"       "phillips"      "pntsprd"       "prison"
 [93] "prminwge"      "rdchem"        "rdtelec"       "recid"
 [97] "rental"        "return"        "saving"        "school93_98"
[101] "sleep75"       "slp75_81"      "smoke"         "traffic1"
[105] "traffic2"      "twoyear"       "volat"         "vote1"
[109] "vote2"         "voucher"       "wage1"         "wage2"
[113] "wagepan"       "wageprc"       "wine"
```

Study the `.qmd` file and compare with what you observe in the pdf after clicking "Render".

When you write your assignment answer, use the format shown below. See the "Cheatsheet for R Markdown" https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf for tips on getting various effects, such as bolding a word. We'll come to typesetting math later in this document.

Question 1 (Wooldridge Chapter 2 Exercise C2) The data set in `ceosal2` contains information of chief executive officers for U.S. corporations. The variable *salary* is annual compensation in thousands of dollars, and *ceoten* is prior number of years as company CEO.

(i) Find the average salary and the average tenure in the sample.

(ii) How many CEOs are in their first year as CEO (i.e., *ceoten*=0)? What is the longest tenure as company CEO?

(iii) Estimate the simple regression model

$$\log(salary) = \beta_0 + \beta_1 ceoten + u\,,$$

and report your results in the usual form. What is the (approximate) predicted percentage increase in salary given one more year as a CEO?

Answers

Remark: Prior to answering the question, you should explore the data set. Parts (i) and (ii) of the question itself is also encouraging you to get to know your data set better.

```
df <- ceosal2    # Not necessary, but I like to do this
str(df)          # Gives a nice overview of the data
```

```
'data.frame':   177 obs. of  15 variables:
 $ salary  : int  1161 600 379 651 497 1067 945 1261 503 1094 ...
 $ age     : int  49 43 51 55 44 64 59 63 47 64 ...
 $ college : int  1 1 1 1 1 1 1 1 1 1 ...
 $ grad    : int  1 1 1 0 1 1 0 1 1 1 ...
 $ comten  : int  9 10 9 22 8 7 35 32 4 39 ...
 $ ceoten  : int  2 10 3 22 6 7 10 8 4 5 ...
 $ sales   : num  6200 283 169 1100 351 19000 536 4800 610 2900 ...
 $ profits : int  966 48 40 -54 28 614 24 191 7 230 ...
 $ mktval  : num  23200 1100 1100 1000 387 3900 623 2100 454 3900 ...
 $ lsalary : num  7.06 6.4 5.94 6.48 6.21 ...
 $ lsales  : num  8.73 5.65 5.13 7 5.86 ...
 $ lmktval : num  10.05 7 7 6.91 5.96 ...
 $ comtensq: int  81 100 81 484 64 49 1225 1024 16 1521 ...
 $ ceotensq: int  4 100 9 484 36 49 100 64 16 25 ...
 $ profmarg: num  15.58 16.96 23.67 -4.91 7.98 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

(i)

```
cat("Average CEO salary in the sample:", mean(df$salary),
    "thousand dollars per year.\n")
cat("Average CEO tenure in the sample:", mean(df$ceoten), "years.\n")
```

```
Average CEO salary in the sample: 865.8644 thousand dollars per year.
Average CEO tenure in the sample: 7.954802 years.
```

(ii)

```
cat("No. of CEOs in their first year: ",sum(df$ceoten==0) ,".\n", sep="")
cat("Longest tenure in sample:", max(df$ceoten), "years.\n")
```

```
No. of CEOs in their first year: 5.
Longest tenure in sample: 37 years.
```

(iii) Regression of log(*salary*) on *ceoten* (with intercept)

```
mdl <- lm(log(salary)~ceoten, data=df)
mdl_sum <- summary(mdl)
coef(mdl_sum)
cat("n = ", nobs(mdl), ", R-sqr = ", round(mdl_sum$r.squared,3), "\n", sep="")
```

```
               Estimate  Std. Error   t value      Pr(>|t|)
(Intercept) 6.505497982 0.067991070 95.681654 4.843250e-153
ceoten      0.009723632 0.006364478  1.527797  1.283683e-01
n = 177, R-sqr = 0.013
```

Predicted percentage increase in salary given one more year as a CEO is approximately 0.97 percent.

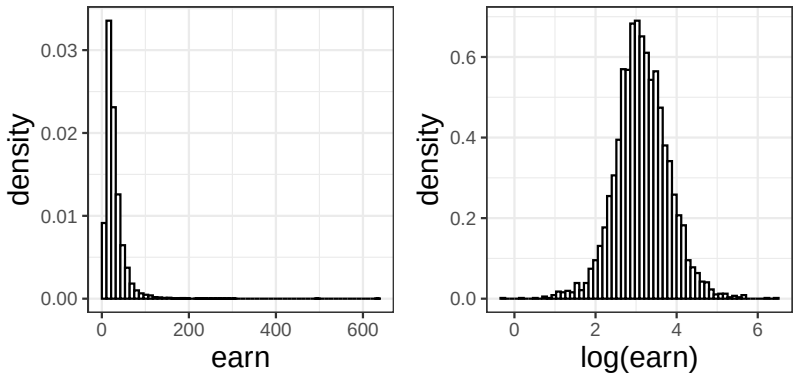
As an exercise you should download one of the class data sets into the data subfolder of your working director, and try to download it. E.g.,

```
library(tidyverse)
library(patchwork)
dat <- read_csv("data\\earnings2019.csv", show_col_types=FALSE)
head(dat,3)
```

```
# A tibble: 3 x 11
    age height  educ feduc meduc tenure  wexp race   male  earn totalwork
  <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <chr> <dbl> <dbl>     <dbl>
1    59     67    12     3     3      5    30 White     0 36.3       1652
2    43     63    10     4     3      7    13 White     1  6.46      1548
3    28     74    12     2     3      6     9 White     1 13.1       2460
```

The following code generates a figure.

```
plot_theme <- theme_bw() + theme(aspect.ratio = 1, axis.title = element_text(size=14))
p1 <- ggplot(dat, aes(x = earn)) +
      geom_histogram(aes(y = after_stat(density)), fill = "white", color="black",
                     bins=60, boundary = 0) +
      scale_x_continuous("earn") + plot_theme
p2 <- ggplot(dat, aes(x = log(earn))) +
      geom_histogram(aes(y = after_stat(density)), fill = "white", color="black",
                     bins=60, boundary = 0) +
      scale_x_continuous("log(earn)") + xlab("ln earn") + plot_theme
p1 | p2
```



Including Math

Math is typeset using LaTeX. Study the code in the original `.qmd` document to see how I typeset Questions 2.33 and 2.35 from Exercise 2.3.3 of the Notes, shown below. To put math "inline", enclose the LaTeX code in-between `$...$` (no space after the first `$`, no space before the last `$`) To display math (show it in its own line), put it in its own paragraph between double dollar signs.

Exercise 2.33 Let $\{X_i\}_{i=1}^n$ be a sample of n observations of a random variable X with population mean $E(X) = \mu$. Show that the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$ minimizes the function

$$f(\hat{\mu}) = \sum_{i=1}^{n}(X_i - \hat{\mu})^2.$$

Exercise 2.35 Show that

$$\begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}.$$

Explain why this shows that the matrix

$$\begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix}$$

is positive definite. Find the stationary point of the function

$$f(x, y) = 5x^2 + 10xy + 10y^2.$$

Is this point a maximum point, a minimum point, or neither?

Here is another example:

$$\begin{aligned}\widetilde{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2.\end{aligned}$$

Resources

- To learn Latex for typesetting math, go to https://www.overleaf.com/ register for a free account, and follow the tutorial at https://www.overleaf.com/learn. Nothing to install on your computer. Write code on the left of the screen and compile to see the output on the right.

References

Wooldridge, Jeffrey M., (2012). Introductory econometrics : a modern approach. Mason, Ohio. South-Western Cengage Learning.