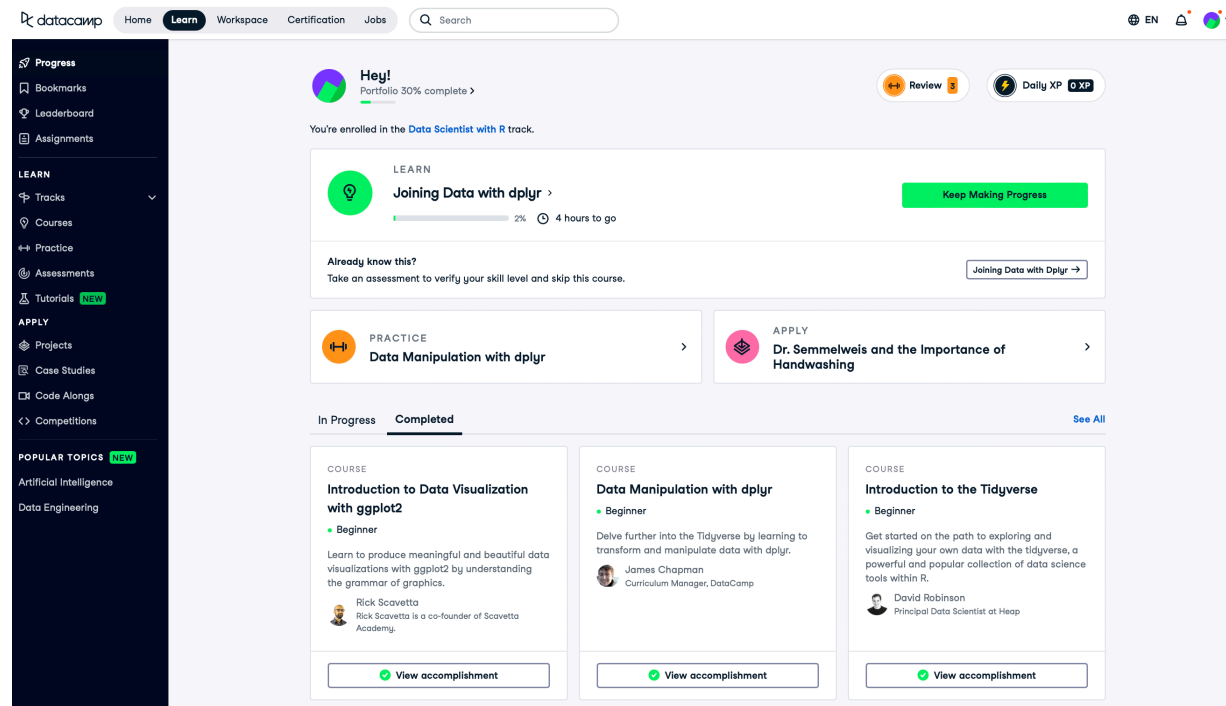


## 1 Programming Setup

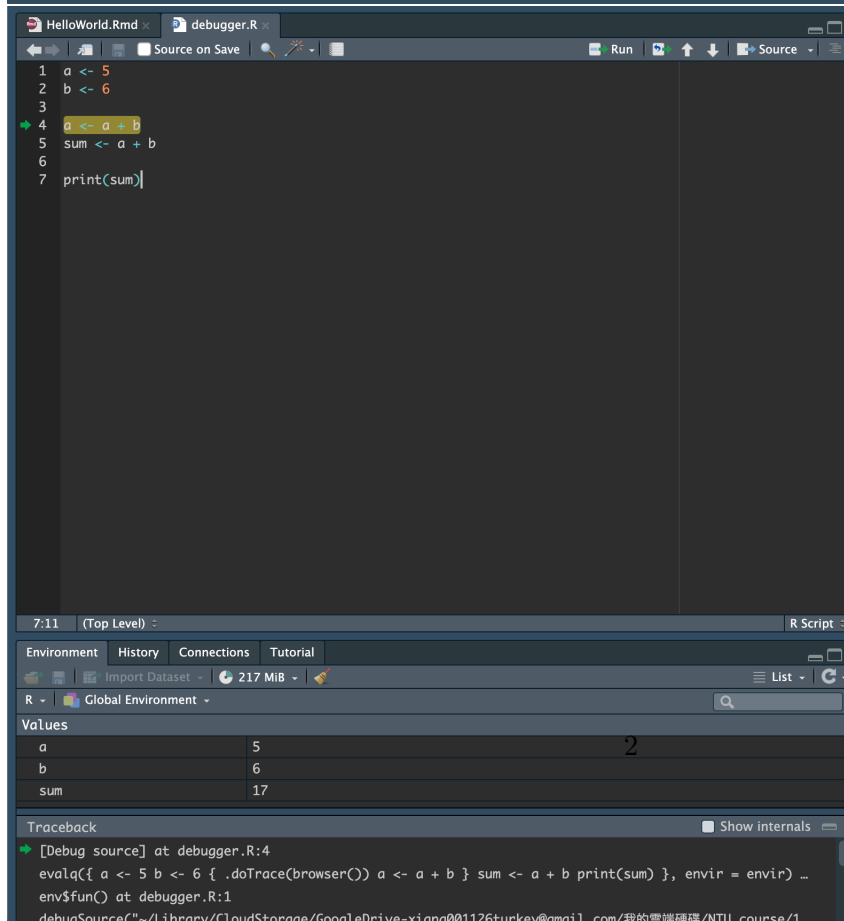
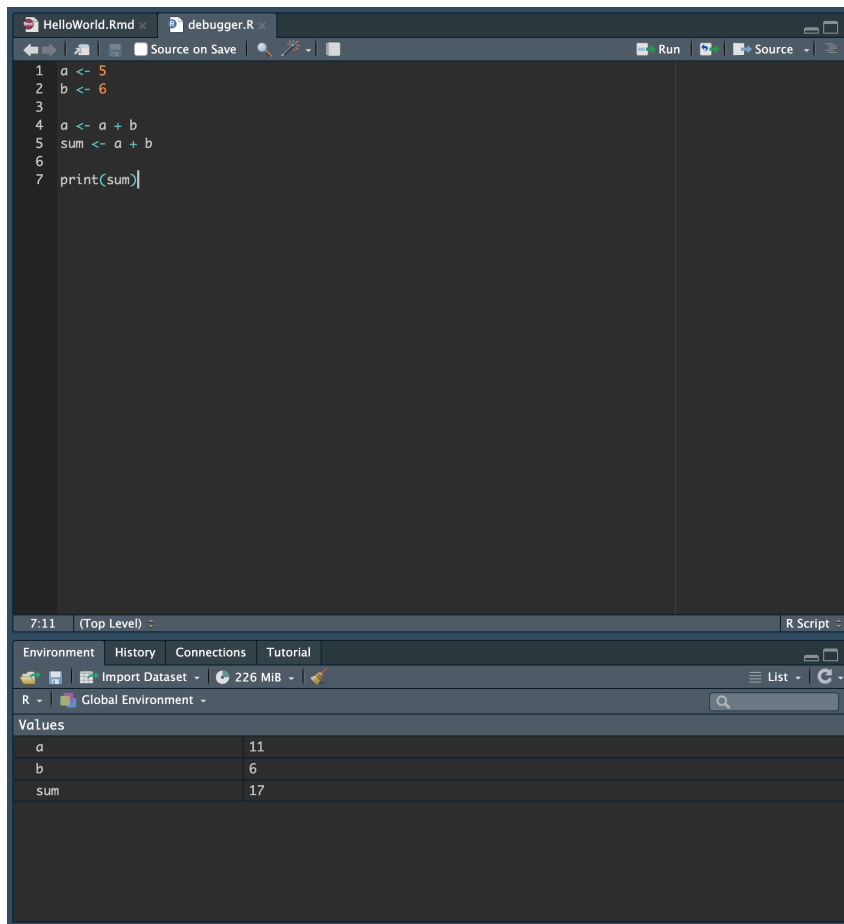
### 1.1 Setup Datacamp



### 1.2 R

For installation process and HelloWorld example, please refer to the file located at `"/homework/homework1/R_code"`

## 1.3 Debugger



## 1.4 Setup Github

This screenshot shows the main page of a GitHub repository named 'ECON-5211-Labro-Economics-I' by user 'xiaaaang1126'. The repository is public and has 1 branch (main) and 0 tags. The file list shows a 'new folder' with 2 commits, and subfolders 'homework', 'lecturenotes', '.DS\_Store', and '.gitattributes'. A large 'Add a README' button is prominent in the center. On the right, there are sections for 'About' (no description), 'Releases' (no releases), and 'Packages' (no packages).

**Repository: ECON-5211-Labro-Economics-I** (Public)

1 Branch, 0 Tags

Go to file

Add file > Code >

**new folder** bc5750b · yesterday 2 Commits

File/Folder	Commit Message	Commit Date
homework	new folder	yesterday
lecturenotes	Initial commit	yesterday
.DS_Store	Initial commit	yesterday
.gitattributes	Initial commit	yesterday

**README**

**Add a README**

Help people interested in this repository understand your project by adding a README.

**Add a README**

**About**

No description, website, or topics provided.

Activity

0 stars

1 watching

0 forks

**Releases**

No releases published

[Create a new release](#)

**Packages**

No packages published

[Publish your first package](#)

This screenshot shows the 'R\_code' folder within the 'homework1' directory. The file list includes '.DS\_Store', 'HelloWorld.Rmd', 'HelloWorld.pdf', and 'debugger.R'. The commit history for this folder shows four commits, all with the message 'doing HW1' and dated 'now'.

**Files**

main

Go to file

homework

homework1

**R\_code**

- .DS\_Store
- HelloWorld.Rmd
- HelloWorld.pdf
- debugger.R

pic

question

- .DS\_Store
- .DS\_Store
- Labor\_Economics\_HW2.pdf

lecturenotes

- .DS\_Store
- .gitattributes

**ECON-5211-Labro-Economics-I / homework / homework1 / R\_code**

Add file > ...

xiaaaang1126 doing HW1 fa907ee · now History

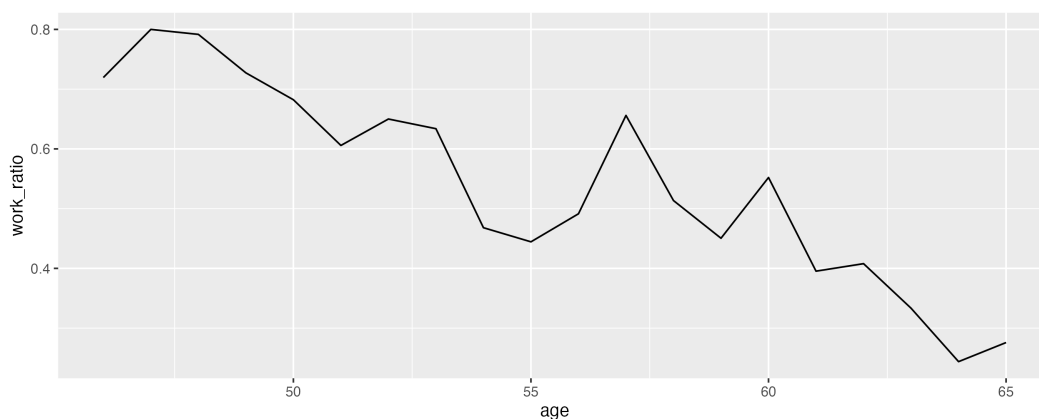
Name	Last commit message	Last commit date
..		
.DS_Store	doing HW1	now
HelloWorld.Rmd	doing HW1	now
HelloWorld.pdf	doing HW1	now
debugger.R	doing HW1	now

## 2 Sign up NBER working paper series

1. The title of the second paper listed on the NBER weekly working paper series is "*Why Survey-Based Subjective Expectations are Meaningful and Important*", working paper number 32199
2. I have downloaded a paper interesting me most at the path `"/homework/homework1/path"` with title "*Why Survey-Based Subjective Expectations are Meaningful and Important*"

## 3 Sign up SRDA

1. The year of the PSFD dataset I downloaded is 2000
2. The working rate against age in the dataset on 2000 year is



## 4 Roy Model

### 4.1 Review

1. We first consider  $\mathbb{E}[w_0|\text{Migrate}]$ , which can be expressed as

$$\begin{aligned}\mathbb{E}[w_0|\text{Migrate}] &= \mu_0 + \mathbb{E}[\epsilon_0|w_1 > w_0 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\mu_1 + \epsilon_1 > \mu_0 + \epsilon_0 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\nu > \mu_0 - \mu_1 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\frac{\nu}{\sigma_\nu} > z] \\ &= \mu_0 + \sigma_0 \mathbb{E}[\frac{\epsilon_0}{\sigma_0}|\frac{\nu}{\sigma_\nu} > z]\end{aligned}$$

in order to compute the conditional expectation of  $\epsilon_0$ , we consider the following steps

- (a) For  $\epsilon_0$  and  $\nu$ , we can express their linear relationship with regression coefficient formula<sup>1</sup>

---

<sup>1</sup>Please refer to <https://www.youtube.com/watch?v=TRcfZ2awa-o> for the proof of the formula

$$\mathbb{E}[\epsilon_0|\nu] = \frac{\sigma_{0,\nu}}{\sigma_\nu^2} \nu$$

(b) Use the formula after changing terms

$$\mathbb{E}\left[\frac{\epsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu}\right] = \frac{1}{\sigma_0} \underbrace{\frac{\sigma_{0,\nu}}{\sigma_\nu^2} \frac{\nu}{\sigma_\nu}}_{\substack{\text{by above} \\ \text{formula}}} \frac{1}{\sigma_\nu^{-2}} \frac{1}{\sigma_\nu} = \frac{\sigma_{0,\nu}}{\sigma_0 \sigma_\nu} \frac{\nu}{\sigma_\nu} = \rho_{0,\nu} \frac{\nu}{\sigma_\nu} \quad (1)$$

(c) From the preceding expectation formula, we can simplify the conditional expectation

$$\begin{aligned} \mathbb{E}[w_0|\text{Migrate}] &= \mu_0 + \sigma_0 \mathbb{E}\left[\frac{\epsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu} > z\right] \\ &\stackrel{(1)}{=} \mu_0 + \rho_{0,\nu} \sigma_0 \mathbb{E}\left(\frac{\nu}{\sigma_\nu} \middle| \frac{\nu}{\sigma_\nu} > z\right) \\ &= \mu_0 + \rho_{0,\nu} \sigma_0 \frac{\phi(z)}{1 - \Phi(z)} \end{aligned}$$

where  $\Phi(z)$  and  $\phi(z)$  are respectively defined as the pdf and cdf of  $z$ .

Besides, note that the coefficient of Inverse Mill Ratio can be rewritten as

$$\rho_{0,\nu} \sigma_0 = \underbrace{\frac{\sigma_{0,\nu}}{\sigma_\nu} = \frac{\sigma_{0,1} - \sigma_0^2}{\sigma_\nu}}_{\substack{\sigma_{0,\nu} = \text{Cov}(\epsilon_0, \epsilon_1 - \epsilon_0) \\ = \sigma_{0,1} - \sigma_0^2}} = \frac{\sigma_0 \sigma_1}{\sigma_\nu} \left( \frac{\sigma_{0,1} - \sigma_0^2}{\sigma_0 \sigma_1} \right) = \frac{\sigma_0 \sigma_1}{\sigma_\nu} \left( \rho - \frac{\sigma_0}{\sigma_1} \right)$$

Follow the same procedure, we can derive  $\mathbb{E}[w_1|\text{Migrate}]$  as well. Therefore, the conditional expectation can be expressed in a quite symmetric way

$$\begin{cases} \mathbb{E}[w_0|\text{Migrate}] = \mu_0 + \frac{\sigma_0 \sigma_1}{\sigma_\nu} \left( \rho - \frac{\sigma_0}{\sigma_1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ \mathbb{E}[w_1|\text{Migrate}] = \mu_1 + \frac{\sigma_0 \sigma_1}{\sigma_\nu} \left( \frac{\sigma_1}{\sigma_0} - \rho \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \end{cases}$$

2. If  $Q_0 > 0$  and  $Q_1 < 0$ , then we have

$$\begin{cases} \rho - \frac{\sigma_0}{\sigma_1} > 0 \\ \frac{\sigma_1}{\sigma_0} - \rho < 0 \end{cases} \Leftrightarrow \rho = \frac{\sigma_{0,1}}{\sigma_0 \sigma_1} > \max\left\{\frac{\sigma_0}{\sigma_1}, \frac{\sigma_1}{\sigma_0}\right\} \Leftrightarrow \sigma_{0,1} > \max\{\sigma_0^2, \sigma_1^2\}$$

which is impossible, hence such situation doesn't occur given people choose to migrate. In economics sense, this contradiction suggest that people would not leave for the lower proportion of  $w_1$  income distribution when they are in the upper proportion of  $w_0$  income distribution

## 4.2 Simulation

1. First, let's pick some values for the parameters  $\{\mu_0, \mu_1, \sigma_0, \sigma_1, \sigma_{0,1}, C\}$

```
1 # 1. Pick my favorite value for set of parameters
2 mu_0 <- 100
3 mu_1 <- 100
4 sigma_0 <- 3
5 sigma_1 <- 6
6 sigma_01 <- 5
7 cost <- 50
8 n <- 1000000
9
```

2. After simulation, store the result of  $\{\epsilon_0, \epsilon_1\}$  in `data.table`

```
1 # 2. Simulate the epsilon_0 and epsilon_1
2 set.seed(123)
3 reps <- 1000000
4 par.est.DT <- data.table(matrix(0, reps, 6))
5 epsilon_0 <- rnorm(n, mean = 0, sd = sigma_0**2)
6 epsilon_1 <- rnorm(n, mean = 0, sd = sigma_1**2)
7
```

3. and create the columns for  $w_0$  and  $w_1$

```
1 # 3. Create the columns for w0 and w1
2 w_0 <- mu_0 + epsilon_0
3 w_1 <- mu_1 + epsilon_1
4 par.est.DT[, 1] <- epsilon_0
5 par.est.DT[, 2] <- epsilon_1
6 par.est.DT[, 3] <- w_0
7 par.est.DT[, 4] <- w_1
8
```

4. also, generate a column  $I$  that takes binary value

```
1 # 4. Generate the column I that take binary value.
2 par.est.DT[, 5] <- ifelse(par.est.DT[, 2] > par.est.DT[, 1], 1, 0)
3
```

5. Then compute  $\mathbb{E}[w_0|I], \mathbb{E}[w_1|I], Q_0, Q_1$  from data

```
1 # 5. Calculate E[w0|I], E[w1|I], Q0, Q1 from data
2 E_epsilon0_D1 <- par.est.DT[V5 == 1, mean(V1)]
3 E_epsilon1_D1 <- par.est.DT[V5 == 1, mean(V2)]
4 E_w0_D1 <- par.est.DT[V5 == 1, mean(V3)]
5 E_w1_D1 <- par.est.DT[V5 == 1, mean(V4)]
6
```

6. Use equation (1) and (2) to compute conditional expectation of  $w_0$  and  $w_1$

```
1 # 6. Calculate RHS of equation (1) and (2)
2 nu <- epsilon_1 - epsilon_0
3 par.est.DT[, 6] <- nu
4 sigma_nu <- par.est.DT[, sd(V5)]
```

```

5 Z <- (mu_0 - mu_1 + cost)/sigma_nu
6 mill <- function(x) {
7   pnorm(x, lower.tail=FALSE, log.p=TRUE) - dnorm(x, log=TRUE)
8 }
9 RHS_1 <- mu_0 + (sigma_0*sigma_1/sigma_nu) * (sigma_01/(sigma_0*sigma_1) - sigma_
0/sigma_1) * mill(Z)
10 RHS_2 <- mu_1 + (sigma_0*sigma_1/sigma_nu) * (sigma_0/sigma_1 - sigma_01/(sigma_0
11 *sigma_1)) * mill(Z)
12 # comparison
13 E_w0_D1 # 98.24212
14 E_w1_D1 # 127.8435
15 RHS_1   # 136.8422
16 RHS_2   # 63.15782
17

```

7. The column  $\mathbb{I}$  and  $\mathbb{E}[w_1|\mathbb{I} = 1]$  are observable in the real world, but the other variable  $\{\epsilon_0, \epsilon_1, Q_0, Q_1, \phi(Z), \Phi(Z)\}$  and  $\mathbb{E}[w_0|\mathbb{I} = 1]$  are unobservable

## 5 Roy Model is Everywhere

1. An example that immediately comes to my mind is getting master (or ph.D) degree or not. The time spent on studying a degree or the risk of not finishing it are costly, nonetheless, the degree itself is also desirable when seeking a job. It might need argue that  $Q_0 < 0$  or  $Q_0 > 0$ , but I think most people will agree on  $Q_1 > 0$ .
2. Consider the Roy model

$$\begin{cases} w_0 = \mu_0 + \epsilon_0 \\ w_1 = \mu_1 + \epsilon_1 \end{cases}$$

where studying a higher degree represents 1, while not studying represents 0. It's assumed that  $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ , where  $i \in \{0, 1\}$ . Note that the pain or time studying takes are captured by constant cost  $C$  and one chooses studying higher degree only if  $\mathbb{E}[w_1] - C > \mathbb{E}[w_0]$ .