

#### Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in F Conduct Linear Regression

R Example

Exercise

# Class 1: Basics in R R section for EPH 705 Statistical Methods in Epidemiology II

### Min Lu

Division of Biostatistics University of Miami

Spring 2017



## Overview

Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

# Object:

Install R and R studio Import data file in R Conduct Linear Regression

- 2 R Example
- Exercise





# Why R?

Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

### R is Smart and Free!







### Ross Ihaka Robert Gentleman

**Bill Venables** 

イロト イポト イヨト イヨト 三日

Check out examples at https://minlu.shinyapps.io/BasicR\_example/ https://minlu.shinyapps.io/Plot\_function/



# Install R and R studio

#### Class 1: Basics in R

Min Lu

### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

• Download latest version of R at **Windows**:

https://cran.r-project.org/bin/windows/base/
Mac:

https://cran.r-project.org/bin/macosx/

• Download R studio at www.rstudio.com/products/rstudio/download/ Choose the latest version for your Windows or Mac.



# Import data file in R studio

#### Class 1: Basics in R

#### Import data file in R

0 0 0	RStudio			
💽 📲 😴 🖌 🔒 🚔 🕞 🕜 Go to file/function			惠 Project: (	(None) 👻
inference.utilities.1.4.R × () icm treatmentEffect.1.5.R* × >	Environment History			
○ ○ □ Source Q Ž• □ → > → Source •	Contract Import Dat	aset - ) 🖉 💿		List •
26	Global Environment •	٩		
27 - <i>***********************************</i>	Data			
28 ##	0 dta	2231 obs. of 41 variables		
29 ## global parameters	1 dtg.test	1000 obs. of 4 variables		
31	label	chr [1:1000, 1:20] "1" "1" "1" "1" "1" "1	" "1" "1" .	
32	max0.5	num [1:1000, 1:20] 1 1 1 1 1 1 1 1 1 1		
33 ## plotting parameters	MBST2005dich	4972 obs. of 10 variables		
34 cut.dominate <95	MBST2005poly	4972 obs. of 10 variables		
35 cex.dominate <- c(.1, .5, 1)	O result	10 obs. of 12 variables		_
36	- Teaure	che [1,1000, 1,20] 117 117 117 117 117 117		
37 ## range over which treatment effect is defined		nm [1:1000, 1:20] 1 1 1 1 1 1 1 1 1 1		
38 tme.ronge <- c(0, 10)	Valuer	Num [1.1000, 1.20] 1 1 1 1 1 1 1 1 1 1		
49	vutues	50		_
41 ## number of clusters in subarroup analysis	D and and a sta	50		
42 nclus <- 10	cex.dominate	hum [1:5] 0.1 0.5 1		
43	cuc.dointridice	ets [1:6] "slaash suiles" "slaash ssias"	le" "elhes	-
44	exptitist	chr [1:6] miberich.smilley miberich.spiru	ts milberi	cn
34:1 [] (Untitled)   R Script		- [assoc] adverse ]		-
Console ×/ Ø	Files Plots Package	is Help Viewer		
the deal on date of	🔍 Install 🕥 Update	G		
Diseksnasa laadad fear / RData]	Name	Description	Version	
[workspace toaded from ~/.kbata]	System Library			
> cut.dominate <95	abind	Combine Multidimensional Arrays	1.4-5	8
<pre>&gt; cex.dominate &lt;- c(.1, .5, 1) &gt; source("causal.inference.utilities.1.4.R")</pre>	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1	0
Error in file(filename, "r", encoding = encoding) :	assertthat	Easy pre and post assertions.	0.1	0
cannot open the connection In addition: Warning message:	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.0.4	0
In file(filename, "r", encoding = encoding) :	base64enc	Tools for base64 encoding	0.1-3	8
cannot open file 'causal.inference.utilities.1.4.R': No su	bdsmatrix	Routines for Block Diagonal Symmetric matrices	1.3-2	8
ch file or directory	BH	Boost C++ Header Files	1.60.0-2	0
> tme.range	BiocGenerics	S4 generic functions for Bioconductor	0.16.1	0
<pre>&gt; cex.dominate</pre>	BiocInstaller	Install/Update Bioconductor, CRAN, and github	1.20.3	0
[1] 0.1 0.5 1.0	bitops	Bitwise Operations	1.0-6	0
> Cut.dominate	hoot	Rootstran Functions (Originally by Angelo Canty	1 3-18	0
>		for S)	210 10	~
	caTools	Tools: moving window statistics, GIF, Base64,	1.17.1	63

5/13



# Linear Regression

#### Class 1: Basics in R

Min Lu

### Object:

Install R and R studio Import data file in

Conduct Linear Regression

R Example

Exercise

We are going to run linear regression in R and save publishable result.

### Linear Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

where p = number of independent or predictor variables.



Class 1: Basics in R

Min Lu

Object:

Install R and R studio Import data file in F Conduct Linear Regression

R Example

Exercise

### Import data and run linear regression

```
# Set working directory first
# Read in the data
BostonHousing <- read.csv("BostonHousing.csv")[, -1]
# Display the data.
head(BostonHousing[, 1:4])</pre>
```

##		town	tract	lon	lat
##	1	Nahant	2011	-70.9550	42.2550
##	2	Swampscott	2021	-70.9500	42.2875
##	3	Swampscott	2022	-70.9360	42.2830
##	4	Marblehead	2031	-70.9280	42.2930
##	5	Marblehead	2032	-70.9220	42.2980
##	6	Marblehead	2033	-70.9165	42.3040

# Run linear regression
object <- lm(cmedv ~ rm + crim + zn + tax, data = BostonHousing)
# Display result using summary(object)</pre>



Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

### Make publishable result

```
# Result looks messy
mytable <- summary(object)$coefficients
mytable</pre>
```

##		Estimate	Std. Error	t value	Pr(> t )
##	(Intercept)	-20.74664918	2.788507802	-7.440054	4.407615e-13
##	rm	7.60427384	0.407762762	18.648770	2.459906e-59
##	crim	-0.13648598	0.038040220	-3.587939	3.660388e-04
##	zn	0.03567246	0.012355481	2.887177	4.055118e-03
##	tax	-0.01084358	0.002022447	-5.361613	1.261320e-07



Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

```
R Example
```

Exercise

### Make publishable result

```
# Give pretty variable names
rownames(mytable) <- c("Intercept", "Room number", "Crime rate",
    "Residential land", "Property tax")
# Give pretty decimal
mytable[, 1:3] <- round(mytable[, 1:3], 2)
mytable[, 4] <- round(mytable[, 4], 3)
mytable</pre>
```

##		Estimate	Std.	Error	t	value	Pr(> t )
##	Intercept	-20.75		2.79		-7.44	0.000
##	Room number	7.60		0.41		18.65	0.000
##	Crime rate	-0.14		0.04		-3.59	0.000
##	Residential land	0.04		0.01		2.89	0.004
##	Property tax	-0.01		0.00		-5.36	0.000



Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

### Store publishable result

Save as a table write.csv(mytable, "BostonHousingResult.csv")

Save in Latex

```
# have to intall package: install.packages("xtable")
library(xtable)
latex_table <- xtable(mytable)</pre>
```



# In class exercise

Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

Data:Housing data (BostonHousing.csv) is 506 census tracts of Boston from the 1970 census. The dataframe contains the original data by Harrison and Rubinfeld (1979), with 506 observations on 14 variables, medv being the target variable:

variable name	discreption
crim	per capita crime rate by town
zn	proportion of residential land zoned for lots over 25,000 sq.ft
indus	proportion of non-retail business acres per town
chas	Charles River dummy variable $(= 1 \text{ if tract bounds river; } 0 \text{ otherwise})$
nox	nitric oxides concentration (parts per 10 million)
rm	average number of rooms per dwelling
age	proportion of owner-occupied units built prior to 1940
dis	weighted distances to five Boston employment centres
rad	index of accessibility to radial highways
tax	full-value property-tax rate per USD 10,000
ptratio	pupil-teacher ratio by town
b	$1000(B-0.63)^2$ where B is the proportion of blacks by town
lstat	percentage of lower status of the population
town	name of town
tract	census tract
lon	longitude of census tract
lat	latitude of census tract
cmedv	corrected median value of owner-occupied homes in USD 1000's

Task:Choose variables "rad", "dis", and "b" to predict "crim" in a linear regression model and check the result.



### Take home exercise

#### Class 1: Basics in R

Min Lu

#### Object:

Install R and R studio Import data file in R Conduct Linear Regression

R Example

Exercise

Make a summary table of a linear model of your own. Choose your own outcome variable with 3-4 predictors. Save the result in an excel (.csv) file.



## Class over

Class 1: Basics in R

Min Lu

Object:

Install R and R studio Import data file in Conduct Linear Regression

R Example

Exercise

