

Exercise #7 – Working with data in R

1. Open your R script and add a new section for this exercise
2. Check that the Rosmap (“RM”) dataset is loaded in R. If not, please load the dataset in R (see Exercise #5)
3. Without attaching the RM dataset (use \$ to call variables), use R to answer the following questions:
 - a. What is the classification of the age of death (age_death) variable?
 - b. What is the classification of the male sex (msex) variable?
 - c. What is the mean global pathology score (gpath)?
 - d. What is the mean global cognitive score at baseline (global_bl)?
4. We are now going to demonstrate potential issues with the attach() function using two datasets from different studies. Please refer to the code output on the next page to answer the following questions:
 - a. What did we name the two datasets in R? (*Step 1*)
 - b. Which variable names appear in both the HIV and RM dataset? (*Step 2*)
 - c. What does it mean to attach both the HIV and RM datasets in R? (*Step 3*)
 - d. Are the results from code chunk (A) from the HIV or RM dataset?
 - e. Are the results from code chunk (B) from the HIV or RM dataset?
 - f. Are the results from code chunk (C) from the HIV or RM dataset?
 - g. Are the results from code chunk (D) from the HIV or RM dataset?
 - h. Of the above four code chunks, which were the hardest to determine? Why?
 - i. Name one potential issue with using the attach() function for data analysis in R.

Note: In our course, we will not attach data because we believe it can lead to confusion or errors as demonstrated in part 4. Moving forward, we will only use the \$ to call variables.

DATA SETUP

You do *not* need to run this code on your own. However, you should use these three steps for reference to understand how the data was loaded and attached.

Step 1. Import two datasets into R

```
RM <- read.csv("RM_xsect.csv")
HIV <- read.csv("HIV_xsect.csv")
```

Step 2. Variable names from each dataset

```
names(RM)
```

```
## [1] "ranid"      "age_death"  "educ"      "msex"
## [5] "time_in_study" "gpath"     "global_b1" "global_lv"
## [9] "pathoAD"     "apoe4d"    "ceradsc"   "braaksc"
## [13] "cad"        "cad_year"
```

```
names(HIV)
```

```
## [1] "i"      "ranid"  "eversmk" "msex"   "bmi"    "wt_kg"  "ht_m"
## [8] "age"    "HIVpos" "grs64"   "grs59"  "grs5"   "center" "age_dx"
## [15] "RS2"    "RS1"
```

Step 3. Attaching both datasets

```
attach(RM)
attach(HIV)
```

CODE CHUNKS FOR EXERCISE 7, PART 4

Code chunk (A)

```
mean(msex)
```

```
## [1] 0.2052786
```

Code chunk (B)

```
mean(RM$time_in_study)
```

```
## [1] 7.254506
```

Code chunk (C)

```
mean(grs64)
```

```
## [1] 9.915212
```

Code chunk (D)

```
mean(HIV$eversmk)
```

```
## [1] 0.2644184
```