

Lab 4

Alexander Hernandez

09/06/22

1) Babyboomer data

Time of birth (24-HOUR), Sex (1-GIRL,2-BOY), Birth Weight (GRAMS),
Number of minutes after midnight

a) Import the 'babyboom.dat.txt'

```
babyboom_data = read.table('http://jse.amstat.org/datasets/babyboom.dat.txt')
```

b) How many observations are recorded?

```
nrow(babyboom_data)
```

```
## [1] 44
```

c) Print first 5 observations

```
head(babyboom_data, 5)
```

```
##      V1 V2   V3  V4
## 1     5  1 3837   5
## 2  104  1 3334  64
## 3  118  2 3554  78
## 4  155  2 3838 115
## 5  257  2 3625 177
```

d) Print last 5 observations

```
tail(babyboom_data, 5)
```

```
##      V1 V2   V3  V4
## 40 2104  2 2121 1264
## 41 2123  2 3150 1283
## 42 2217  1 3866 1337
## 43 2327  1 3542 1407
## 44 2355  1 3278 1435
```

2) Car Crashes with harm (people or property) and at least 1 vehicle towed

a) Import the data in R using appropriate R code

```
file_car_crashes = file.choose()
car_crashes = read.csv(file_car_crashes)
```

b) Print the first 5 observations

```
head(car_crashes, 5)
```

```
##      X  dvcat  weight  dead  airbag  seatbelt  frontal  sex  age0Focc  yearacc  yearVeh
## 1 1  25-39  25.069  alive   none   belted      1    f      26     1997     1990
## 2 2  24-Oct  25.069  alive  airbag   belted      1    f      72     1997     1995
## 3 3  24-Oct  32.379  alive   none    none      1    f      69     1997     1988
## 4 4  25-39  495.444  alive  airbag   belted      1    f      53     1997     1995
## 5 5  25-39  25.069  alive   none   belted      1    f      32     1997     1988
##      abcat  occRole  deploy  injSeverity  caseid
## 1  unavail  driver      0              3 2:03:01
## 2  deploy   driver      1              1 2:03:02
## 3  unavail  driver      0              4 2:05:01
## 4  deploy   driver      1              1 2:10:01
## 5  unavail  driver      0              3 2:11:01
```

3) consider a data set where the columns are separated by \$

Data in problem-3-data.csv

```
file_problem3 = file.choose()
test_data = read.table(file_problem3, sep="$", header=T)
test_data
```

```
##      Col1 Col2 Col3
## 1      1     2     3
## 2      4     5     6
## 3      7     8     9
## 4      a     b     c
```

4) Baby Weight

```
library(readstata13)
file_bweight = file.choose()
baby_weight = read.dta13(file_bweight)
```

a) Identify dimensions of data

```
dim(baby_weight)
```

```
## [1] 4642  23
```

b) Extract the variables included in the datasets

```
colnames(baby_weight)
```

```
## [1] "bweight"      "mmarried"     "mhispanic"    "fhispanic"    "foreign"
## [6] "alcohol"      "deadkids"     "mage"         "medu"         "fage"
## [11] "fedu"         "nprenatal"    "months1b"     "order"        "msmoke"
## [16] "mbsmoke"      "mrace"        "frace"        "prenatal"     "birthmonth"
## [21] "lbweight"     "fbaby"        "prenatal1"
```

5) Biochemist publications - unable to complete

Data: <http://www.stata-press.com/data/lf2/couart2.dta>

Neither the 'readstat13' or 'foreign' packages can read this file

Claims the 20kb file is 8.1 GB

a) Import the data in R - Showing how you would do this without errors

```
# webdata_publications = read.dta("https://www.stata-press.com/data/lf2/couart2.dta")
```

Unable to use direct read from url due to error: the 'wininet' method is deprecated for http:// and https:// URLs No alternative argument in read.dta or read.dta13 method ## b) List the variables included in the data

```
# colnames(webdata_publications)
```

c) State the dimension of the data

```
# dim(webdata_publications)
```