Chapter 5: Variables & Manipulation

Exercises

Brian Fogarty

15 May 2018

Contents

EXERCISE I

Using an abbreviated version of the 2015 UK Millennium Cohort survey dataset (`mcs.dta`), perform the following exercises. Note: the survey was carried out in 2015 to 14 year-old pupil in the UK. The dataset consists of nearly 12,000 observations and 52 variables. You need to use the `haven` package to read-in the data.

Perform recoding and labeling on the variables below. Note: for all the variables, you need to first convert them to factors using the `as.factor()` function.

1. `mths` - rename as `maths` and label the values as `1='1.  Strongly Disagree';2='2.  Disagree';3='3. Agree';4='4.  Strongly Agree'`. This variable includes pupils' responses to whether they were good at mathematics.

2. `scien` - rename as `science` and label the values as `1='1. Strongly Disagree';2='2. Disagree';3='3.  Agree';4='4.  Strongly Agree'`. This variable includes pupils' responses to whether they were good at science.

3. `sex` - rename as `gender` and label the values as `0='0.  Female';1='1.  Male'`.

4. `best` - rename as `bestsch` and label the values as `1='1.  Never'; 2='2.  Sometimes';3='3.  Most Times';4='4.  Always'`. This variable includes pupils' responses to how often they do their best at school.

5. `games` - rename as `vidgames` and label the values as `1='1.  Never';2='2.  Less Half Hr';3='3.  Half Hr to Hr';4='4.  1-2 Hrs'; 5='5.  2-3 Hrs';6='6.  3-5 Hrs';7='7.  5-7 Hrs'; 8='8.  More 7 Hrs'`. This variable includes pupils' responses for many hours per weekday do they play video games.

6. `sibl_fl` - rename as `siblings` and recode to create a dummy variable which equals 0 if the pupil has no siblings and equals 1 if the pupil has 1 or more siblings - `0='0.  No Siblings';1:10='1.  Siblings'`.

ANSWERS FOR EXERCISE I

Read-in 2015 Millennium Cohort Study data.

```
setwd("c:/QSSD/Exercises/Chapter 5 - Exercises")
getwd()
```

```
[1] "c:/QSSD/Exercises/Chapter 5 - Exercises"
```

```
library(haven)
mcs <- read_dta("mcs.dta")
```

Question 1.1

```
library(car)
```

```
Warning: package 'car' was built under R version 3.4.3
```

```
mcs$mths <- as.factor(mcs$mths)
table(mcs$mths)
```

```
   1    2    3    4
 598 1827 5958 3118
```

```
mcs$maths <- recode(mcs$mths, "1='1. Strongly Disagree';2='2. Disagree';
                 3='3. Agree';4='4. Strongly Agree'")
table(mcs$maths)
```

```
1. Strongly Disagree          2. Disagree             3. Agree
                 598                 1827                 5958
    4. Strongly Agree
                3118
```

Question 1.2

```
mcs$scien <- as.factor(mcs$scien)
table(mcs$scien)
```

```
   1    2    3    4
 500 1993 6166 2834
```

```r
mcs$science <- recode(mcs$scien, "1='1. Strongly Disagree';2='2. Disagree';
                      3='3. Agree';4='4. Strongly Agree'")
table(mcs$science)
```

```
1. Strongly Disagree          2. Disagree           3. Agree
                 500                 1993               6166
    4. Strongly Agree
                2834
```

Question 1.3

```r
mcs$sex <- as.factor(mcs$sex)
table(mcs$sex)
```

```
   0    1
5926 5946
```

```r
mcs$gender <- recode(mcs$sex, "0='0. Female';1='1. Male'")
table(mcs$gender)
```

```
0. Female    1. Male
     5926       5946
```

Question 1.4

```r
mcs$best <- as.factor(mcs$best)
table(mcs$best)
```

```
   1    2    3    4
  35 1058 6469 3937
```

```r
mcs$bestsch <- recode(mcs$best, "1='1. Never';2='2. Sometimes';
                      3='3. Most Times';4='4. Always'")
table(mcs$bestsch)
```

```
    1. Never  2. Sometimes 3. Most Times    4. Always
          35          1058          6469         3937
```

Question 1.5

```r
mcs$games <- as.factor(mcs$games)
table(mcs$games)
```

```
   1    2    3    4    5    6    7    8
2160 1565 1343 1751 1583 1478  770  862
```

```
mcs$vidgames <- recode(mcs$games, "1='1. Never';2='2. Less Half Hr';3='3. Half Hr to Hr';
                       4='4. 1-2 Hrs'; 5='5. 2-3 Hrs';6='6. 3-5 Hrs';7='7. 5-7 Hrs';
                       8='8. More 7 Hrs'")
table(mcs$vidgames)
```

```
       1. Never  2. Less Half Hr 3. Half Hr to Hr       4. 1-2 Hrs
          2160             1565             1343             1751
     5. 2-3 Hrs        6. 3-5 Hrs       7. 5-7 Hrs     8. More 7 Hrs
          1583             1478              770              862
```

Question 1.6

```
table(mcs$sibl_fl)
```

```
   0    1    2    3    4    5    6    7    8    9   10
1678 5203 2990 1322  417  178   50   21   10    2    1
```
```
mcs$siblings <- recode(mcs$sibl_fl, "0='0. No Siblings';1:10='1. Siblings'")
table(mcs$siblings)
```

```
0. No Siblings    1. Siblings
         1678          10194
```

EXERCISE II

Provide the level of measurement for the variables you recoded in Exercise I.

ANSWERS FOR EXERCISE II

1. `maths` - ordinal

2. `science` - ordinal

3. `gender` - nominal

4. `bestsch` - ordinal

5. `vidgames` - ordinal/high ordinal

6. `siblings` - nominal

EXERCISE III

Using the dataset of the 2011 Scottish Census with measures of deprivation and commuting for all Scottish postcodes (`depdata.csv`), perform the following exercises.

1. What is the level of measurement of the variable `pcnt_unemployed`?

2. Collapse `pcnt_unemployed` into 4 categories. The categories should be 0-5, 6-10, 11-15, and 16-20. What is the level of measurement of this collapsed variable?

3. Label the collapsed categories as `"low"` for 0-5, `"med low"` for 6-10, `"med high"` for 11-15, `"high"` for 16-20.

ANSWERS FOR EXERCISE III

Question 3.1

The variable is ratio-level.

Question 3.2

```
depdata <- read.csv("depdata.csv")

depdata$pcnt_unemployed2 <- cut(depdata$unemployed,
                                breaks=c(0,5,10,15,20))
table(depdata$pcnt_unemployed2)
```

```
 (0,5]  (5,10] (10,15] (15,20]
    43      33      25       9
```

This variable is now ordinal-level.

Question 3.3

```
depdata$pcnt_unemployed3 <- cut(depdata$unemployed,
                                breaks=c(0,5,10,15,20),
                                labels=c("low","med low","med high","high"))

table(depdata$pcnt_unemployed3)
```

```
    low  med low med high     high
     43       33       25        9
```