

SPPH 604 001 Lab Exercise: Data accessing (CCHS)

Contents

Question 1: [60% grade]	1
1(a) Importing dataset	1
1(b) Subsetting according to eligibility	2
1(c) Identify variables	2
1(d) Retaining necessary variables	2
1(e) Creating analytic dataset	3
1(f) Number of columns and variable names	3
Question 2: Table 1 [20% grade]	3
Question 3: [20% grade]	4
3(a) Subset	4
3(b) Recode	4
3(c) Regression	4
3(d) Reporting odds ratio	5
Knit your file	5

We will use the following article:

Palis, Marchand & Oviedo-Joeke. (2020). The relationship between sense of community belonging and self-rated mental health among Canadians with mental or substance use disorders. *Journal of Mental Health*, 29(2): 168-175. DOI: [10.1080/09638237.2018.1437602](https://doi.org/10.1080/09638237.2018.1437602) (available in the “Library Online Course Reserves”: [open link](#)).

- Download the [CCHS MH topical index](#)
- Download the [CCHS MH Data Dictionary](#)

Question 1: [60% grade]

1(a) Importing dataset

```
# Importing dataset  
load("Data/accessing/cchsMH.RData")
```

1(b) Subsetting according to eligibility

Subset the dataset according to the eligibility criteria / restriction specified in the paper

- Identify the variable needed for eligibility criteria

Hint

- Read the first paragraph of **Analytic sample** (page 2) for the eligibility criteria
- Eligibility criteria was determined based on only **one** variable. Only work with 'YES' category.

```
# your code here
```

1(c) Identify variables

- Identify the outcome variable
- Identify the explanatory variable
- Identify the potential confounders
- Identify the survey weight variable

- **Hint**

1. Read

- first and second paragraphs of **Study variables** for the outcome, explanatory and confounding variables
- third paragraph of the **Statistical analyses** for the survey weights variable.

2. There were **five** potential confounders.

3. Potentially useful functions for this exercise:

- `%in%`
- `levels`
- `recode`
- `subset`
- `as.factor`
- `relevel`
- or dplyr ways: `filter`, `select`

```
# your code here
```

1(d) Retaining necessary variables

In the dataset, retain only the variables associated with outcome measure, explanatory variable, potential confounders and survey weight. There should be eight variables (one outcome, one exposure, five confounders, and one survey weight).

```
# your code here
```

1(e) Creating analytic dataset

Outcome variable has a category ‘NOT STATED’, but for our analysis, we will omit anyone associated with this category. Similarly, for explanatory variable, we have categories such as DON’T KNOW, REFUSAL and NOT STATED. We will omit anyone with these categories.

- Assign missing values for categories such as DON’T KNOW, REFUSAL and NOT STATED.
- Recode the variables as shown in Table 1 in the article. You can use any function/package of your choice. Here is an example (but feel free to use other functions. In R there are many other ways to do this same task.

```
## your code here
# levels(your.data.frame$your.age.variable) <-
#   list("15 to 24 years" = c("15 TO 19 YEARS", "20 TO 24 YEARS"),
#       "25 to 34 years" = c("25 TO 29 YEARS", "30 TO 34 YEARS"),
#       "35 to 44 years" = c("35 TO 39 YEARS", "40 TO 44 YEARS"),
#       "45 to 54 years" = c("45 TO 49 YEARS", "50 TO 54 YEARS"),
#       "55 to 64 years" = c("55 TO 59 YEARS", "60 TO 64 YEARS"),
#       "65 years or older" = c("65 TO 69 YEARS", "70 TO 74 YEARS",
#                               "75 TO 79 YEARS", "80 YEARS OR MORE"))
```

1(f) Number of columns and variable names

Report the number of columns in your analytic dataset, and the variable names.

```
# your code here
```

Question 2: Table 1 [20% grade]

Reproduce Table 1 presented in the article (or see below). Omit the ‘Main source of income’ variable. The table you produce should report numbers as follows, with all columns as shown in the table:

	Total n(%)	Poor or Fair n(%)	Good n(%)	Very good or excellent n(%)
Study sample	2628 (100)	1002 (38.1)	885 (33.7)	741 (28.2)
Community belonging				
- Very weak	480 (18.3)	282 (28.1)	118 (13.3)a	80 (10.8)a
- Somewhat weak	857 (32.6)	358 (35.7)	309 (34.9)	190 (25.6)
- Somewhat strong	1005 (38.2)	288 (28.7)	362 (40.9)	355 (47.9)
- Very strong	286 (10.9)	74 (7.4)a	96 (10.8)a	116 (15.7)a
Sex				
- Females	1407 (53.5)	616 (61.5)	487 (55.0)	304 (41.0)
- Males	1221 (46.5)	386 (38.5)	398 (45.0)	437 (59.0)
Age group				
- 15 to 24 years	740 (28.2)	191 (19.1)	264 (29.8)	285 (38.5)
- 25 to 34 years	475 (18.1)	141 (14.1)	167 (18.9)	167 (22.5)
- 35 to 44 years	393 (15.0)	185 (18.5)	119 (13.4)a	89 (12.0)a
- 45 to 54 years	438 (16.6)	220 (22.0)	139 (15.7)	79 (10.7)a
- 55 to 64 years	379 (14.4)	198 (19.7)	113 (12.8)a	68 (9.2)a

- 65 years or older	203 (7.7)	67 (6.6)a	83 (8.4)a	53 (7.1)b
Race/Ethnicity				
- Non-white	458 (17.4)	184 (18.4)	140 (15.8)	134 (18.1)
- White	2170 (82.6)	818 (81.6)	745 (84.2)	607 (81.9)
Main source of income				
- Employment	1054 (40.1)	289 (28.8)	386 (43.6)	379 (51.1)
Income ^d				
- Worker's	160 (6.1)	91 (9.1)a	44 (5.0)b	25 (3.4)c
Compensation ^e				
- Senior Benefits ^f	134 (5.1)	57 (5.7)a	42 (4.7)b	35 (4.7)
- Other ^g	184 (7.0)	82 (8.2)a	60 (6.8)a	42 (5.7)b
- Not applicable ^h	851 (32.4)	402 (40.1)	263 (29.7)	186 (25.1)
- Not Stated ⁱ	245 (9.3)	81 (8.1)a	90 (10.2)a	74 (10.0)

^a Coefficient of variation between 16.6 and 25.0%. ^b Coefficient of variation between 25.1 and 33.3%. ^c Coefficient of variation > 33.3%. ^d Employment Income: Wages/salaries or self-employment. ^e Worker's compensation: Employment insurance or worker's compensation or social assistance/welfare. ^f Senior Benefits: Benefits from Canada or Quebec Pension Plan or job related retirement pensions, superannuation and annuities or RRSP/RRIF of Old Age Security and Guaranteed Income Supplement. ^g Other: Dividends/interest or child tax benefit or child support or alimony or other or no income. ^h Not applicable: Respondents who live in a household with only one person. The income variable "main source of personal income" is applicable only to those that live in a household of more than one person. ⁱ Not Stated: Question was not answered (don't know, refusal, not stated).

```
# your code here
require(tableone)
```

Question 3: [20% grade]

3(a) Subset

Subset the dataset excluding 'Very good or excellent' responses from the self-rated mental health variable

```
# your code here
```

3(b) Recode

Recode self-rated mental health variable and make it a binary variable: 'Good' vs. 'Poor' (simplifying category labels only). Convert that variable to a factor variable with 'Poor' being the reference level.

```
# your code here
```

3(c) Regression

Run a logistic regression model for finding the relationship between community belonging (Reference: Very weak) and self-rated mental health (Reference: Poor) among respondents with mental or substance use disorders. Adjust the model for three confounders: sex, age, and race/ethnicity. Do not need to report summary of the model.

```
# your code here
```

3(d) Reporting odds ratio

Report the odds ratios and associated confidence intervals. Publish or jtools package could be useful to report the odds ratios with confidence intervals.

```
# your code here
```

Knit your file

Please knit your file once you finished and submit the knitted PDF or doc file. Please also fill-up the following table:

Group name: ** xyz **

Student initial	% contribution
Student 1 initial	x%
Student 2 initial	x%
Student 3 initial	x%