CPE - Med Stat: Assignment 01

Your name here

Contents

Before you start the tutorial, please check out this web site and watch the video in it.

In this tutorial, you will learn how to:

- Practice using RMarkdown to create neat and reproducible reports
- Load libraries into the R environment so you can use them in your code chunks
- Load a dataset and inspect it in the environment panel
- Create a `table 1` of your dataset

Practice using RMarkdown

Change the YAML using the documentation for HTML features to include

- floating table of contents (toc)
- paged table printing
- a flatly theme (see section 3.1.4 in the documentation)

Use inline formatting to create

- Hyperlinks. For example, you can find documentation to the use of HTML inline formatting here and here.
- Text written in **Bold font**
- Third level section headers
- Unordered list items
- Text marked as `inline code` using `backticks`
- The use of blockquotes

Coding in R

The chunk below is our setup chunk. This is where we load our libraries (in this case, we are loading `tidyverse` and `table1`). We can also dataset our dataset from the file which you can find in a folder called "data".

Running the chunk above will introduce a dataset into your environment.

The dataset is a random sample from the National Health and Nutrition Examination Survey (NHANES), a survey designed to "assess the health and nutritional status of adults and children in the United States."

The variables in this dataset include:

- `sbp`: the systolic blood pressure
- `dbp`: the diastolic blood pressure

- `age` age in years
- `sex` sex assigned at birth
- `income` an indicator for the income class
- `smoker` an indicator of participant's smoking status

Use chunk options to prevent the code from showing in your final report.

Display the contents of the dataset by typing `nhanes` in the code chunk below

```
## # A tibble: 1,000 x 6
##      sbp   dbp   age sex    income                smoker
##    <dbl> <dbl> <dbl> <fct>  <fct>                 <fct>
##  1   123    77    59 Male   $55,000+              Never
##  2   126    71    55 Female <NA>                  Past
##  3   179    64    64 Female < $25,000             Current
##  4   138    78    26 Female $55,000+              Never
##  5   127    73    54 Female $55,000+              Past
##  6   130    77    61 Male   $55,000+              Never
##  7   145    96    23 Male   $25,000 to < $55,000 Past
##  8   129    63    55 Male   $25,000 to < $55,000 Current
##  9    95    66    23 Female < $25,000             Current
## 10   154    46    80 Female <NA>                  Never
## # i 990 more rows
```

Generate a summary of the dataset (use the function `summary`)

```
##       sbp              dbp              age            sex
##  Min.   : 83.0   Min.   :  0.00   Min.   :20.00   Female:518
##  1st Qu.:111.0   1st Qu.: 65.00   1st Qu.:32.00   Male  :482
##  Median :121.0   Median : 73.00   Median :47.00
##  Mean   :123.5   Mean   : 73.05   Mean   :47.72
##  3rd Qu.:134.0   3rd Qu.: 80.00   3rd Qu.:61.00
##  Max.   :234.0   Max.   :111.00   Max.   :80.00
##  NA's   :42      NA's   :42
##                   income        smoker
##  < $25,000           :156   Never  :567
##  $25,000 to < $55,000:254   Past   :266
##  $55,000+            :480   Current:167
##  NA's                :110
##
##
##
```

Use inline code to print out the total number of variables and the number of observed units in our dataset.

Answer: The number of variables in our data-set is 6 and the number of observed units is 1000

Create `Table 1`

In most published articles, there is a "Table 1" containing descriptive statistics for the sample. This may include, for example, the mean and standard deviation for continuous variables, the frequency and proportion for categorical variables, and perhaps also the number of missing values.

The brute force method of creating such a table would be to compute each statistic for each variable of interest and then copy and paste the results into a table. But an easier way is to use the functions from the `table1` package. Please use the documentation provided in this link to create the table below.

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Never	Past	Current	Overall
	(N=567)	(N=266)	(N=167)	(N=1000)
Systolic blood pressure (mm/Hg)				
Mean (SD)	122 (17.5)	127 (18.0)	122 (17.0)	124 (17.7)
Median [Min, Max]	120 [89.0, 234]	126 [90.0, 203]	120 [83.0, 179]	121 [83.0, 234]
Missing	30 (5.3%)	4 (1.5%)	8 (4.8%)	42 (4.2%)
Diastolic blood pressure (mm/Hg)				
Mean (SD)	73.2 (10.5)	72.2 (14.2)	73.8 (11.4)	73.1 (11.7)
Median [Min, Max]	73.0 [46.0, 108]	73.0 [0, 111]	74.0 [39.0, 107]	73.0 [0, 111]
Missing	30 (5.3%)	4 (1.5%)	8 (4.8%)	42 (4.2%)
Age (years)				
Mean (SD)	46.3 (17.5)	53.1 (17.4)	43.9 (15.1)	47.7 (17.4)
Median [Min, Max]	46.0 [20.0, 80.0]	55.5 [21.0, 80.0]	41.0 [21.0, 80.0]	47.0 [20.0, 80.0]
Sex				
Female	352 (62.1%)	91 (34.2%)	75 (44.9%)	518 (51.8%)
Male	215 (37.9%)	175 (65.8%)	92 (55.1%)	482 (48.2%)
income				
< $25,000	73 (12.9%)	36 (13.5%)	47 (28.1%)	156 (15.6%)
$25,000 to < $55,000	128 (22.6%)	77 (28.9%)	49 (29.3%)	254 (25.4%)
$55,000+	303 (53.4%)	127 (47.7%)	50 (29.9%)	480 (48.0%)
Missing	63 (11.1%)	26 (9.8%)	21 (12.6%)	110 (11.0%)

An extension

Extend the work by adding your own creative spin on what you've learned in this lab. For instance, you can create a new table1 or count the number of distinct values of the income variable using the `count` function. You may add a new variable using the `mutate` function and display it in a table. Just be creative!