***Beyond Classic Designs and Analysis for Health Data.*** *NUS, Singapore, 2022*

# Preparing your computer and R

Before *Session 1* Mon 27th June

Please follow the advice below to prepare for the computer exercises during the course. We are assuming you have (or will have) R (https://cran.r-project.org/) or Rstudio (https://rstudio.com/products/rstudio/download/#download) on your computer. These instructions are simply to set up your work environment so that you are ready to begin: the specific commands needed to do any exercise will be provided with the exercise.

## Organisation of Course Material

Create a folder called "Beyond_Classic" on the desktop of your computer (or elsewhere on the hard disk if you prefer) where you will download the various materials provided for the course. Create three sub-folders of "Beyond_Classic", named "lectures", "exercises" and "reading". For the "exercises" folder, create three "daughter" folders named "data", "docs" and "solutions". Prior to the course, all data sets will be available for you to save in the "data" folder, and prior to each session of the course, the instructions and solutions (including code) will be provided in a Zip package organised in this folder structure.

Once you open R, you should set the working folder to "data", using the menus or a command such as:

```
    setwd("c:/Documents/…./Beyond_Classic/exercises/data")
```

You can see your current working directory by typing `getwd ()`

Some of the exercises will require you to create and save temporary data sets. We advise that you always name such files with a prefix "temp" (e.g. "tempExercise1.rda", "tempFramingham.rdata) so that you can clean up the data folder when the exercise is completed.

## A note on R commands, functions and packages

R comes with a wide range of built-in commands, but there are also many user written functions for performing specific tasks. These functions are stored in packages, and only become available when the package is installed and loaded. Briefly, an R package is installed by the command:
**install.packages**("<package name>") and loaded by the command **library**("<package name>")
and help for a function can be found by calling help("function name") or ?("function name")

The R packages used in *Session 1* are:

- Epi
- EpiDisplay
- epitools
- haven
- vcd

So you should install these as follows:

```
install.packages(c("Epi", “EpiDisplay”, “epitools”, “haven”, "vcd"))
```

Now you are ready for Session 1. A summary of some useful commands and packages used in later sessions are provided in the following pages and may be useful to you during or after the course.

### Installing a package available in Github (https://github.com/)

To install a package from GitHub, you must first install the **devtools** package and then run an install command (for example, to install the design2phase package we use in Session 4):

```
devtools::install_github("Fgraziano/design2phase”)
```

Other Github packages we will use:

DoublingOfCases from https://github.com/nyilin/DoublingOfCase (in Exercise 5.1)
SamplingDesignTools from https://github.com/nyilin/SamplingDesignTools (in Exercise 5.2)

### Homemade functions

An alternative to using R built-in fuctions (and sometimes more convenient) is to write your own R functions.
For example, if you write and run the following code, you will have a function that calculates OR from the 4 numbers in a 2-by-2-table:

```
CALC_OR <- function(a,b,c,d) {
Q1_TAB <- matrix(c(a, b, c, d), nrow = 2, ncol = 2, byrow = T)
ODDS_RATIO <- round((a/b) / (c/d), 3)
ODDS_RATIO
}
```

The OR is computed by typing, for example, `CALC_OR(230, 4885, 106, 4779)`

### Some useful R functions for viewing/manipulating data

`head(data)` returns the first part of a data frame
`colnames(data)` or `rownames(data)` retrieve the row or column names of a matrix or dataframe
`matrix` creates a matrix from the given set of values
`table` creates a contingency table

### Packages used by Session

| | | |
|---|---|---|
| Session 1 | **Ex 1.2:** Epi, epitools, EpiDisplay, haven, readx1, vcd | |
| Session 2 | **Ex 2.1:** car, epiDisplay, haven,MASS | **Ex 2.3:** dplyr, epiDisplay, haven, Matching, survival |
| Session 3 | | **Ex 3.2:** dplyr, haven, Epi, survival |
| Session 4 | **Ex 4.1:** dplyr, haven, Epi, survival, sandwich, summarytools, survey | **Ex 4.2**<br>https://github.com/Fgraziano/design2phase |
| Session 5 | **Ex 5.1:**<br>https://github.com/nyilin/DoublingOfCase | **Ex 5.2**<br>https://github.com/nyilin/SamplingDesignTools |