

# Descriptive Statistics by R

Ling-Chieh Kung

September 22, 2014

## 1 The environment

1. Changing work directory: `setwd()` (with `/` or `\\`), `getwd()`.
2. Changing GUI setting.
3. Opening new or old script files.

## 2 Basic operations

1. Making comments: `#`.
2. Direct calculations.
3. Creating new variables: `<-`.
4. Listing and removing objects: `ls()`, `rm()`, `rm(list = ls())`.
5. Data types:
  - (a) Boolean values (`TRUE` and `FALSE`).
  - (b) Single numbers (scalars).
  - (c) Characters (within `"` and `"`).
6. Arithmetic:
  - (a) Common operators: `+`, `-`, `*`, `/`, `%%`, `^`, `()`.
  - (b) Common functions: `sqrt(x)`, `factorial(x)`, `abs(x)`, `round(x)`, `floor(x)`, `ceiling(x)`, `max(x, y)`, `min(x, y)`.
  - (c) Some mathematics functions: `exp(x)`, `pi`, `log(x, e)`, `sin(x)` (and her friends).
7. Logic operators:
  - (a) `==`, `<`, `>`, `<=`, `>=`, `!=`.

- (b) |, &, !.
- (c) any(), all().

8. Searching in the manual: ? and ??.

### 3 Vectors

1. Generating a sequence of numbers:

- (a) 1:20, 5:-5.
- (b) seq(from, to, by), seq(from, to, length), seq(from, to, along).
- (c) rep(x, times), rep(x, each).

2. Concatenation with c().

3. Subscripting (indexing):

- (a) The indices start from 1.
- (b) x[2], x[-2], x[c(2, 4)], x[c(-2, -4)].
- (c) which(x >= 1).

4. Recycling.

5. Logic operators on vectors.

### 4 Visualizing data with data graphs

1. Histograms and frequency distributions:

- (a) hist(x, breaks, right, main, xlab, ylim) and box().
- (b) freqMG <- hist(x, breaks, right, plot = FALSE)
- (c) freqMG\$breaks, freqMG\$mids, and freqMG\$counts.

2. Frequency polygons:

- (a) plot(x, y, type).
- (b) plot(x, y, type, main, xlab, ylab, xlim, ylim, lwd).
- (c) Use freqMG\$mids, freqMG\$counts and range(freqMG\$breaks).

3. Pie charts:

- (a) pie(x, labels).
- (b) pie(x, labels, clockwise, main).
- (c) paste().

4. Bar charts:

(a) `barplot(x, names.arg)`.

(b) `barplot(x, names.arg, xlab, ylim, main)`.

5. Scatter plots:

(a) `plot(x, y)`.

(b) `plot(x, y, main, xlab, ylab, xlim, ylim)`.

## 5 Describing data with statistics

1. Central tendency:

(a) `table(x)`.

(b) `median(x)`, `sum(x)`, `length(x)`, `mean(x)`.

(c) `quantile(x)`, `quantile(x, p)`.

2. Variability:

(a) `range(x)`, `max(x)`, `min(x)`.

(b) Calculating MAD by combining `sum()`, `abs()`, and `length()`.

(c) For sample data: `var(x)`, `sd(x)`.

(d) For population data: Do adjustments by ourselves.

(e) Coefficient of variation.

3. Correlation:

(a) For sample data: `cov(x, y)`, `cor(x, y)`.

(b) For population data: Do adjustments by ourselves.

4. `summary(x)`.