First steps with R
Practical: first steps with R
Jacques van Helden
2017-01-08

Contents

Scope ................................................. 1
**R** is a calculator .................................. 2
Assign a value to a variable ........................... 2
Computing with variables ............................... 2
Variables need to be updated ........................... 2
Updating variable contents ............................. 3
Vectors of values ...................................... 3
Series ................................................. 3
Computing with vectors ................................. 3
Scatter plot ........................................... 4
Line plot ............................................. 4
Variables can also contain strings ....................... 5
String concatenation ................................... 5
Carl’s preferred distribution ........................... 6
Plotting the Poisson distribution ....................... 6
Getting help for **R** functions ......................... 6
Exercise: improve Poisson density plot ............... 7
Improve the plot: type = histogram .................... 7
Improve the plot: Add a title .......................... 7
Improve the plot: define axis labels ................... 8
Improve the plot: add a legend ......................... 9
Poisson: a family of curves ........................... 9
Solution: a family of Poisson curves ................... 9
Solution: a family of Poisson curves .................. 10
Solution: a family of Poisson curves ................. 11
Solution: a family of Poisson curves ................. 12
Solution: a family of Poisson curves ................. 13
Solution: a family of Poisson curves ................ 14
Solution: a family of Poisson curves ................. 15
Solution: a family of Poisson curves ................ 16
Solution: a family of Poisson curves ................. 17
Solution: a family of Poisson curves ................. 18
Solution: a family of Poisson curves ................. 19
Solution: a family of Poisson curves ................. 20
Solution: a family of Poisson curves ................. 21
Solution: a family of Poisson curves ................. 22
Before finishing – keep track of your session .......... 23

Scope

In this session we will explore basic manipulations of variables.

- Assigning a value to a variable
• Basic operations on numbers

**R is a calculator**

**Convention:**
- Dark boxes: commands to type in RStudio **Console** (bottom-left panel).
- White boxes: the result you should obtain.

**Example:** compute a simple addition.

```r
2 + 5
```

[1] 7

**Assign a value to a variable**

In **R** `<-` means “create a variable and assign its value.”

**Example:**
- create a variable named `a`,
- assign the value 2 to this variable,
- `print` the result.

```r
a <- 2
print(a)
```

[1] 2

**Computing with variables**

**Example:**
- create a variable named `b` having value 5,
- compute `a + b` and store the result in a variable named `c`,
- `print` the result.

```r
b <- 5
c <- a + b
print(c)
```

[1] 7

**Variables need to be updated**

**Example:**
- change the value of `a` to 3,
- `print` the value of `c`
- Is this the correct result for `c = a + b`? Why?

```r
a <- 3  # Change the value of a
print(c)  # Print the value of c
```

[1] 7
## Check whether c equals a + b

c == a + b

[1] FALSE

Note: == is used to test whether two variables have the same content.

### Updating variable contents

**Example:**

- change the value of a to 27,
- recompute and print the value of c

```r
a <- 27 ## Change the value of a
c <- a + b
print(c) ## Print the value of c
```

[1] 32

```r
c == a + b

[1] TRUE
```

### Vectors of values

The simplest data structure in R is a vector. In the previous example, the variable a was actually a vector with a single value.

**Example:** create a variable named `three.numbers`, and initialize it with a vector with values 27, 12 and 3000.

**Tips:** - variable names can comprize several parts, separated by dots. - the function `c()` combines several values into a vector

```r
three.numbers <- c(27,12,3000)
print(three.numbers)
```

[1] 27 12 3000

### Series

The simple way to create a series of numbers. The column operator permits to generate all integer values between two limits.

```r
x <- 0:14
print(x)
```

[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

### Computing with vectors

R handles vectors in a very convenient way. An operation on a vector applies to all its elements.
```r
x <- 1:10 # Define a series from 1 to 10
dprint(x)

[1]  1  2  3  4  5  6  7  8  9 10

y <- x^2 # Compute the square of each number
dprint(y)

[1]  1  4  9 16 25 36 49 64 81 100

Scatter plot

x <- -10:10
y <- x^2
plot(x, y)

Line plot

x <- -10:10
y <- x^2
plot(x, y, type="l")
```
Variables can also contain strings

# The # symbol allows to insert comments in R code

# Define a vector named "whoami", and
# containing two names
whoami <- c("Denis", "Siméon")
print(whoami) # Comment at the end of a line

[1] "Denis" "Siméon"

String concatenation

# Define a vector named "names", and
# containing two names
whoami <- c("Denis", "Siméon")

# Paste the values of a vector of string
print(paste(sep=" ", whoami[1], whoami[2]))

[1] "Denis Siméon"
Carl’s preferred distribution

The function `dpois()` computes the Poisson density, i.e. the probability to observe exactly \( x \) successes in a series of independent trials with equal probability.

The Poisson distribution is defined by a single parameter: the expected number of successes \( \lambda \) (read “lambda”).

\[
P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}
\]

```r
x <- 0:14  # Define the X values from 0 to 14
y <- dpois(x, lambda = 2.5)  # Poisson density
print(y)  # Check the result
```

Plotting the Poisson distribution

```r
x <- 0:14  # Define the X values from 0 to 14
y <- dpois(x, lambda = 2.5)  # Poisson density
plot(x,y)  # Check the result
```

This first plot is not very nice. Let us get some help to improve it.

Getting help for R functions

Need help? Type `help()`.

```r
help(plot)
```

A question? Type `?`
Result: R displays the help message for the function `dpois()`.

Exercise: improve Poisson density plot

1. Do not (yet) look the next slide.
2. Read the help page for the `dpois()` function.
3. Draw a plot that provides a didactic illustration of the Poisson density.

Improve the plot: type = histogram

```r
x <- 0:14
lambda <- 2.54
y <- dpois(x, lambda)
plot(x, y, type="h")
```

Improve the plot: Add a title

```r
plot(x, y, type="h", lwd=5, col="blue",
     main="Poisson density")
```
Improve the plot: define axis labels

```r
plot(x, y, type="h", lwd=5, col="blue",
     main="Poisson density",
     xlab="x = number of successes",
     ylab="P(X=x)")
```

**Poisson density**

![Poisson density plot](image)

**Improved plot with axis labels**

```r
plot(x, y, type="h", lwd=5, col="blue",
     main="Poisson density",
     xlab="x = number of successes",
     ylab="P(X=x)")
```

**Poisson density**

![Poisson density plot](image)

**Improved plot with axis labels**
Improve the plot: add a legend

```r
plot(x, y, type="h", lwd=5, col="blue",
     main="Poisson density",
     xlab="x = number of successes",
     ylab="P(X=x)")
legend("topright", paste("lambda =", lambda))
```

**Poisson density**

![Poisson density plot](image)

**Poisson: a family of curves**

**Exercice:** explore the properties of the Poisson density function, by changing the range of $x$ values, and the $\lambda$ parameter.

**Solution: a family of Poisson curves**

$\lambda = 0.01$

```r
lambda <- 0.01
x <- 0:20
plot(x, dpois(x, lambda=lambda), type="h",
     col="darkblue", lwd=5, xlab="X", ylab="dpois(x)")
legend("topright", paste("lambda =", lambda))
```
Solution: a family of Poisson curves

\( \lambda = 0.1 \)

```r
lambda <- 0.1
dx <- 0:20
plot(dx, dpois(x, lambda=lambda), type="h",
     col="darkblue", lwd=5, xlab="X", ylab="dpois(x)"
)
legend("topright", paste("lambda=", lambda))
```
Solution: a family of Poisson curves

\[ \text{lambda} = 1 \]

```r
lambda <- 1
x <- 0:20
plot(x, dpois(x, lambda=lambda), type="h",
     col="darkblue", lwd=5, xlab="X", ylab="dpois(x)")
legend("topright", paste("lambda=", lambda))
```
Solution: a family of Poisson curves

\( \lambda = 2 \)

```r
lambda <- 2
x <- 0:20
plot(x, dpois(x, lambda=lambda), type="h",
    col="darkblue", lwd=5, xlab="X", ylab="dpois(x)"
legend("topright", paste("lambda=", lambda))
```
Solution: a family of Poisson curves

\[
\lambda = 5
\]

\[
\text{lambda} \leftarrow 5 \\
\text{x} \leftarrow 0:20
\]

\[
\text{plot(x, dpois(x,lambda=lambda), type="h",} \\
\hspace{1cm} \text{col="darkblue", lwd=5, xlab="x",ylab="dpois(x)"})
\]

\[
\text{legend("topright", paste("lambda=",lambda))}
\]
Solution: a family of Poisson curves

\( \lambda = 5 \)

\[
\lambda <- 5 \\
x <- 0:40 \\
plot(x, dpois(x,lambda=lambda), type="h", 
    col="darkblue", lwd=5, xlab="X", ylab="dpois(x)") \\
legend("topright", paste("lambda=",lambda))
\]
Solution: a family of Poisson curves

\[ \lambda = 10 \]

\[
\begin{align*}
\text{lambda} &\leftarrow 10 \\
\text{x} &\leftarrow 0: (4\times \text{lambda}) \\
\text{plot(x, dpois(x, lambda=\text{lambda}), type="h",}
\text{ col="darkblue", lwd=5, xlab="X", ylab="dpois(x)"})
\text{legend("topright", paste("lambda=", lambda))}
\end{align*}
\]
Solution: a family of Poisson curves

\(\lambda = 20\)

```r
lambda <- 20
x <- 0:(4*lambda)
plot(x, dpois(x, lambda=lambda), type="h",
    col="darkblue", lwd=2, xlab="X", ylab="dpois(x)"
legend("topright", paste("lambda=", lambda))
```
Solution: a family of Poisson curves

\[ \lambda = 30 \]

```r
lambda <- 30
x <- 0:(2*lambda)
pplot(x, dpois(x,lambda=lambda), type="h",
     col="darkblue", lwd=2, xlab="X",ylab="dpois(x)"
legend("topright", paste("lambda=",lambda))
```
Solution: a family of Poisson curves

\[
\lambda = 50
\]

\[
\lambda \leftarrow 50
\]

\[
x \leftarrow 0:(2\lambda)
\]

\[
\text{plot}(x, \text{dpois}(x,\lambda=\lambda), \text{type}="h", \text{col}="\text{darkblue}", \text{lwd}=2, \text{xlab}="X", \text{ylab}="\text{dpois}(x)"
\]

\[
\text{legend}(\text{"topright", paste("\lambda=",\lambda))}
\]
Solution: a family of Poisson curves

\( \lambda = 100 \)

```r
lambda <- 100
x <- 0:(2*lambda)
plot(x, dpois(x,lambda=lambda), type="h",
    col="darkblue", lwd=2, xlab="X", ylab="dpois(x)"
legend("topright", paste("lambda=",lambda))
```
Solution: a family of Poisson curves

\( \lambda = 1000 \)

```r
lambda <- 1000
x <- 0:(2*lambda)
plot(x, dpois(x,lambda=lambda), type="h",
     col="darkblue", lwd=2, xlab="X",ylab="dpois(x)"
legend("topright", paste("lambda=",lambda))
```
Solution: a family of Poisson curves

\( \lambda = 10000 \)

\[
\begin{align*}
\lambda & \leftarrow 10000 \\
x & \leftarrow 0:(2*\lambda) \\
plot(x, \text{dpois}(x, \lambda=\lambda), \text{type="h"}, \\
\quad \text{col="darkblue", lwd=2, xlab="X", ylab="dpois(x)"}) \\
\text{legend("topright", paste("lambda=", \lambda))}
\end{align*}
\]
Solution: a family of Poisson curves

\( \lambda = 100000 \)

\[
\text{plot}(95000:105000, \text{dpois}(95000:105000, \lambda=100000), \text{type}="h", \text{col}=\"darkblue\", \text{xlab}=\"X\", \text{ylab}=\"dpois(x)\")
\]
Before finishing – keep track of your session

Tractability is an important issue in sciences. The R function `sessionInfo()` summarizes information about the versions of R, the operating system, and all the libraries used during a session.

```
sessionInfo()
```

R version 3.3.2 (2016-10-31)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: macOS Sierra 10.12.2

locale:

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] knitr_1.15.1

loaded via a namespace (and not attached):
[1] backports_1.0.4 magrittr_1.5 rprojroot_1.1 tools_3.3.2
[5] htmltools_0.3.5 yaml_2.1.14 Rcpp_0.12.8 stringi_1.1.2
[9] rmarkdown_1.3 stringr_1.1.0 digest_0.6.10 evaluate_0.10