Preliminaries - How to Use R

Deepayan Sarkar

Indian Statistical Institute, Delhi

What is R

R is a free software environment for data analysis and graphics
What does that mean? How should you use R?

R is like a calculator

```
> 34 * 23
[1] 782
> 27 / 7
[1] 3.857143
> exp(2)
[1] 7.389056
> 2^10
[1] 1024
```

R has standard mathematical functions

```
> sqrt(5 * 125)

[1] 25

> log(120)

[1] 4.787492

> factorial(10)

[1] 3628800

> log(factorial(10))

[1] 15.10441
```

R has standard mathematical functions

```
> choose(15, 5)

[1] 3003

> factorial(15) / (factorial(10) * factorial(5))

[1] 3003
```

R has standard mathematical functions

```
> choose(15, 5)

[1] 3003

> factorial(15) / (factorial(10) * factorial(5))

[1] 3003

> choose(1500, 2)

[1] 1124250

> factorial(1500) / (factorial(1498) * factorial(2))

[1] NaN
```

R supports variables

```
> x <- 2
> y <- 10
> x^y

[1] 1024

> y^x

[1] 100

> factorial(y)

[1] 3628800

> log(factorial(y), base = x)

[1] 21.79106
```

R can compute on vectors

```
> N <- 15
> x <- seq(0, N)
> N

[1] 15

> x

 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15

> choose(N, x)

 [1]    1   15  105  455 1365 3003 5005 6435 6435 5005 3003
[12] 1365  455  105   15    1
```

R has functions for statistical calculations

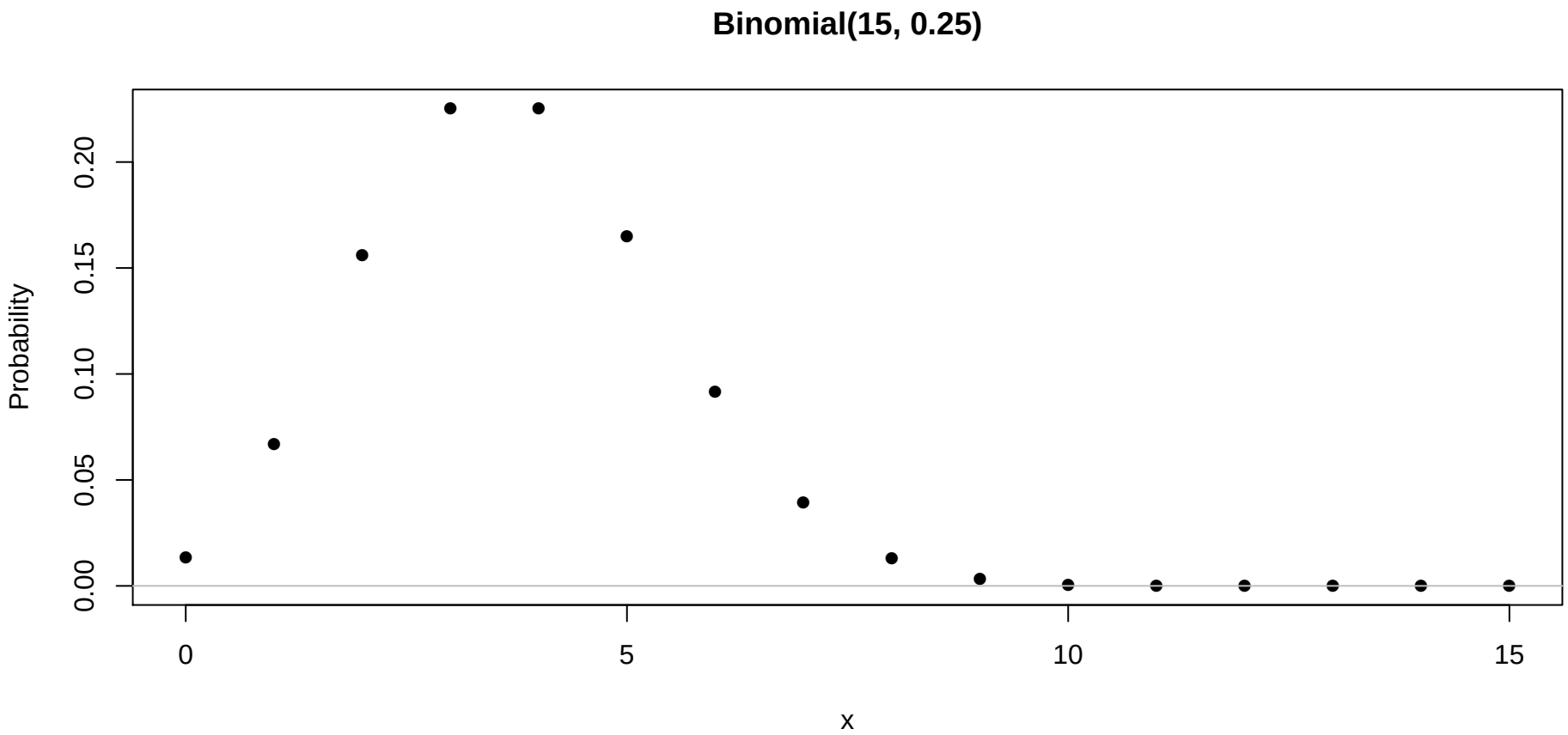
```
> p <- 0.25
> choose(N, x) * p^x * (1-p)^(N-x)

 [1] 1.336346e-02 6.681731e-02 1.559070e-01 2.251991e-01
 [5] 2.251991e-01 1.651460e-01 9.174777e-02 3.932047e-02
 [9] 1.310682e-02 3.398065e-03 6.796131e-04 1.029717e-04
[13] 1.144130e-05 8.800998e-07 4.190952e-08 9.313226e-10

> dbinom(x, size = N, prob = p)

 [1] 1.336346e-02 6.681731e-02 1.559070e-01 2.251991e-01
 [5] 2.251991e-01 1.651460e-01 9.174777e-02 3.932047e-02
 [9] 1.310682e-02 3.398065e-03 6.796131e-04 1.029717e-04
[13] 1.144130e-05 8.800998e-07 4.190952e-08 9.313226e-10
```

R has functions that work on vectors

```
> p.x <- dbinom(x, size = N, prob = p)
> sum(x * p.x) / sum(p.x)

[1] 3.75

> N * p

[1] 3.75
```

R can draw graphs

```
> plot(x, p.x, ylab = "Probability", pch = 16)
> title(main = sprintf("Binomial(%g, %g)", N, p))
> abline(h = 0, col = "grey")
```



R can simulate random variables

```
> cards <- as.vector(outer(c("H", "D", "C", "S"), 1:13, paste))
> cards

 [1] "H 1"  "D 1"  "C 1"  "S 1"  "H 2"  "D 2"  "C 2"  "S 2"
 [9] "H 3"  "D 3"  "C 3"  "S 3"  "H 4"  "D 4"  "C 4"  "S 4"
[17] "H 5"  "D 5"  "C 5"  "S 5"  "H 6"  "D 6"  "C 6"  "S 6"
[25] "H 7"  "D 7"  "C 7"  "S 7"  "H 8"  "D 8"  "C 8"  "S 8"
[33] "H 9"  "D 9"  "C 9"  "S 9"  "H 10" "D 10" "C 10" "S 10"
[41] "H 11" "D 11" "C 11" "S 11" "H 12" "D 12" "C 12" "S 12"
[49] "H 13" "D 13" "C 13" "S 13"
```

R can simulate random variables

```
> cards <- as.vector(outer(c("H", "D", "C", "S"), 1:13, paste))
> cards

 [1] "H 1"  "D 1"  "C 1"  "S 1"  "H 2"  "D 2"  "C 2"  "S 2"
 [9] "H 3"  "D 3"  "C 3"  "S 3"  "H 4"  "D 4"  "C 4"  "S 4"
[17] "H 5"  "D 5"  "C 5"  "S 5"  "H 6"  "D 6"  "C 6"  "S 6"
[25] "H 7"  "D 7"  "C 7"  "S 7"  "H 8"  "D 8"  "C 8"  "S 8"
[33] "H 9"  "D 9"  "C 9"  "S 9"  "H 10" "D 10" "C 10" "S 10"
[41] "H 11" "D 11" "C 11" "S 11" "H 12" "D 12" "C 12" "S 12"
[49] "H 13" "D 13" "C 13" "S 13"

> sample(cards, 10)

 [1] "H 3"  "D 12" "C 5"  "C 4"  "C 13" "C 10" "H 12" "D 3"
 [9] "H 13" "H 8"

> sample(cards, 10)

 [1] "S 10" "D 11" "C 8"  "S 13" "D 8"  "S 2"  "C 9"  "C 13"
 [9] "C 6"  "D 12"
```

R can simulate random variables

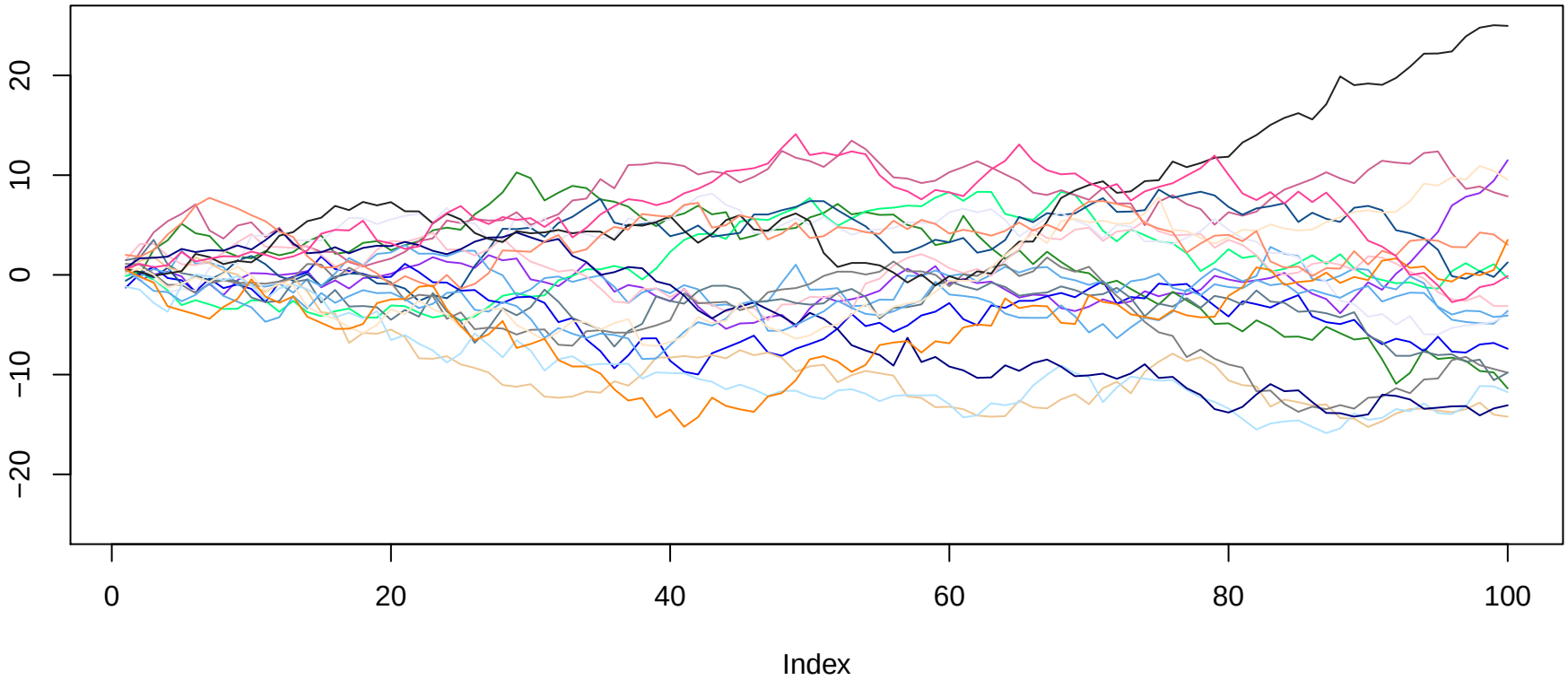
```
> z <- rnorm(50, mean = 0, sd = 1)
> z

 [1] -0.43734856 -1.59634253  1.19745713  0.91645465
 [5]  1.44872899  0.28898227 -1.02667319 -0.97724505
 [9]  0.63212570 -0.69665043 -0.91740888  0.68692774
[13] -0.75772614 -0.01072082  0.33976521  0.50415652
[17]  1.11073892 -0.26828238 -0.13919286 -1.16920920
[21]  0.31562780  1.84191405  0.30534616  1.97173971
[25] -1.04247380 -0.07432403 -1.52985575  0.36464759
[29]  2.35655486 -0.19895015 -0.85968426 -3.49660781
[33]  0.51259987  0.07533628  0.52598052 -1.39535647
[37] -0.89986166 -0.56584650  0.81804095 -1.10243954
[41] -0.67729072 -1.15418113 -0.16885468  0.76021102
[45]  0.94273852  0.42434328  1.28477143  1.05343133
[49] -0.19333798 -0.38565279
```

R is in fact a full programming language

- Variables
- Functions
- Control flow structures
  - For loops, while loops
  - If-then-else (branching)

Example: random walk

```
> plot(1:100, type = "n", ylim = c(-25, 25), ylab = "")
> for (i in 1:20) {
+     z <- rnorm(100, mean = 0, sd = 1)
+     lines(cumsum(z), col = sample(colors(), 1))
+ }
```



R gives access to an extensive toolset

- Most standard data analysis methods are already implemented
- Can be extended by writing add-on packages
- Thousands of add-on packages are available

Drawbacks

- Learning R needs some effort
- Not point-and-click software
- Command-line interface

Good practices

- Use a good interface (R Studio is the most popular one)
- Save your code in a script file (makes it easier to reproduce later)