

BAB I

Pengenalan R dan R-Studio

Capaian Pembelajaran

- Mahasiswa mampu mengenal perangkat lunak statistika open source R dan R-Studio
- Mahasiswa mengetahui istilah-istilah dasar dalam komputasi statistika
- Mahasiswa dapat mengoperasikan perangkat lunak statistika open source R dan R-Studio menggunakan dasar-dasar algoritma komputasi untuk memformulasikan penyelesaian masalah
- Mahasiswa mampu menjelaskan konsep teoritis dan prinsip-prinsip pokok statistika dengan benar dan baik

1.1 Perangkat Lunak R dan R-Studio

Seiring perkembangan kemampuan komputasi dan perangkat lunak, analisis data eksplorasi telah berkembang jauh melampaui ruang lingkup aslinya. Penggerak utama dari disiplin ini mengikuti pesatnya perkembangan teknologi baru, akses kepada data yang lebih besar, dan penggunaan analisis kuantitatif yang lebih besar dalam berbagai disiplin ilmu.

Bahasa R merupakan bahasa pemrograman dan perangkat lunak untuk analisis statistik (juga dikenal sebagai GNU S). R adalah suatu sistem untuk analisis data yang termasuk kelompok *software* statistik *open source* yang tidak memerlukan lisensi atau gratis, yang dikenal dengan *freeware*. Sampai saat ini, R sangat populer di kalangan pengguna statistika di Indonesia untuk keperluan analisis data. R mirip dengan bahasa S, yang diperluas dan dikembangkan pada 1980-an di AT&T Bell Labs (sekarang Alcatel-Lucent) (J Horton and Kleinman, 2015).

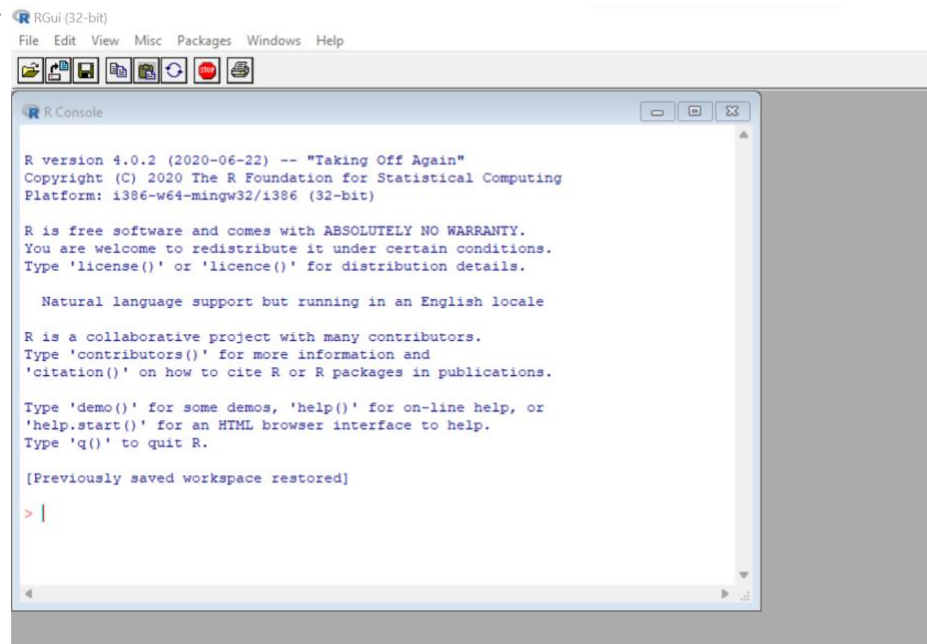
R adalah program komputer statistik yang tersedia melalui Internet di bawah Lisensi Publik Umum (GPL). Artinya, itu dilengkapi dengan lisensi yang memungkinkan Anda menggunakannya secara bebas, mendistribusikannya, atau bahkan menjualnya, selama penerima memiliki hak yang sama dan kode sumber tersedia secara bebas (Dalgaard, 2008).

Kelebihan dari R :

1. Efektif dalam pengelolaan data dan fasilitas penyimpanan. Ukuran file yang disimpan jauh lebih kecil dibanding *software* lainnya.
2. Lengkap dalam operator perhitungan *array*,

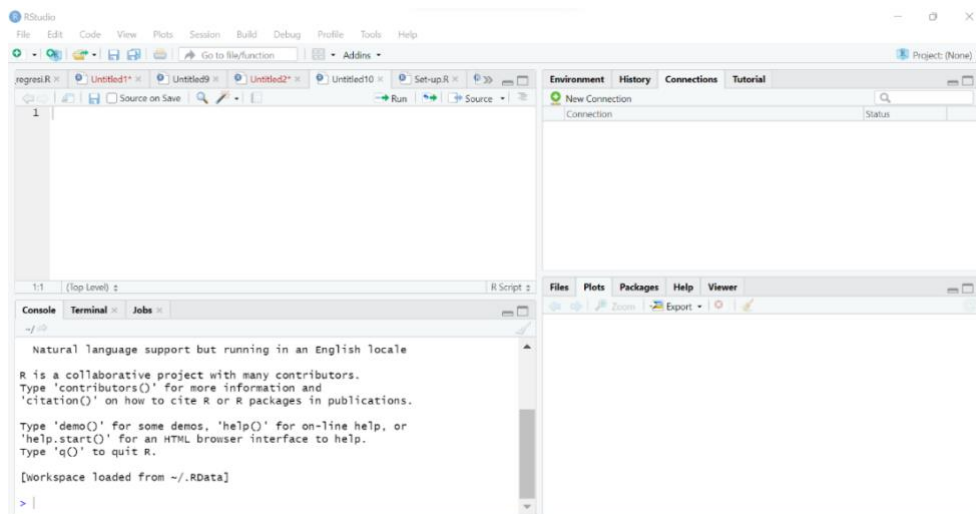
3. Lengkap dan terdiri dari koleksi *tools* statistik yang terintegrasi untuk analisis data, diantaranya, mulai statistik deskriptif, fungsi probabilitas, berbagai macam uji statistik, hingga *time series*.
4. Tampilan grafik yang menarik dan fleksibel ataupun *costumized*
5. Dapat dikembangkan sesuai keperluan dan kebutuhan dan sifatnya yang terbuka, setiap orang dapat menambahkan fitur-fitur tambahan dalam bentuk paket ke dalam software R
6. R bersifat *multiplatform*, yakni dapat diinstall dan digunakan baik pada system operasi Windows, UNIX/LINUX maupun pada Macintosh. Untuk dua system operasi disebutkan terakhir diperlukan sedikit penyesuaian.

Sampai saat ini software R dikembangkan oleh semua penggunanya yang terhimpun dalam naungan R *core team* yang merupakan pekerja keras dan sukarelawan (*voulentir*). R merupakan sebuah lingkungan interaktif untuk komputasi secara statistik dan grafik-grafik, R dapat di *download* pada <https://cran.r-project.org/>. Gambar 1 menunjukkan tampilan jendela pada *software R*.



Gambar 1. Tampilan jendela pada *software R*.

Dalam perkembangannya, R-Studio memungkinkan pengguna menjalankan R dengan lebih mudah, produk ini tersedia pada <http://www.rstudio.com/>. R-Studio hanya bisa diinstall jika pengguna sudah memiliki *software R* didalamnya. R-Studio memiliki empat jendela dalam satu *frame*, hal ini memudahkan pengguna dalam melihat tampilan output yang dihasilkan.



Gambar 2. Tampilan jendela pada *software* R-Studio

Dengan data yang besar, akan sangat sulit untuk mengentry ulang data yang dimiliki jika ingin menggunakan program yang berbeda. *Copy* data secara manual juga memakan banyak waktu dan energi, selain itu kesalahan sangat mungkin terjadi. Untuk memudahkan mengolah data, ada beberapa cara *import* data dari Excel ke R, salah satunya adalah dengan mengubah format excel (*.xls, *.xlsx) ke dalam format .csv (misalnya `read.csv`).

R-Studio menyediakan fitur "*project*" yang sangat berguna yang memungkinkan pengguna beralih dengan cepat antar proyek. Setiap proyek mungkin memiliki direktori kerja, ruang kerja, dan kumpulan file di komponen `source`. Nama proyek saat ini terdaftar di paling kanan dari *main application toolbar* pada combobox yang memungkinkan seseorang untuk beralih antara membuka proyek, membuka proyek yang sudah ada, atau membuat proyek baru. Proyek baru hanya membutuhkan nama dan direktori kerja. Fitur ini sangat sesuai untuk RStudio, karena ketika dijalankan sebagai aplikasi web, ada kebutuhan untuk membuat *serialize* dan pulihkan sesi karena sifat koneksi web. Berpindah antar proyek semudah memilih proyek yang terbuka (Verzani, 2011).

Kemampuan analisis R pada dasarnya didefinisikan dalam fungsi-fungsi yang dikemas dalam bentuk paket atau pustaka. Beberapa pustaka telah menjadi bagian dari paket R yang secara otomatis diinstal ketika kita menginstal R. Sebagian besar pustaka harus diinstal secara khusus sesuai kebutuhan dengan menuliskan `library()`. Beberapa pustaka juga telah diintegrasikan dalam pustaka RGUI R-Commander sehingga dapat diakses melalui menu, sedangkan sebagian besar sisanya harus diakses melalui CLI. Di samping itu, pengguna R masih mungkin mendefinisikan sendiri fungsi-fungsi R, baik dengan menggabungkan definisi yang telah ada, maupun mendefinisikan fungsi yang sama sekali baru.

R bekerja secara mendasar dengan model tanya-jawab: memasukkan baris dengan perintah dan tekan Enter. Kemudian programnya melakukan sesuatu, mencetak hasilnya jika relevan, dan meminta lebih banyak masukan. Ketika R siap untuk input maka prompt akan dicetak, ">" (Dalgaard, 2008).

1.2. Perintah Dasar dalam Pemograman R

Perintah paling sederhana pada R adalah dengan memanfaatkan R sebagai kalkulator, contoh perintah yang dapat dilakukan adalah

```
> 50+60  
[1] 110
```

Dengan perintah tersebut, mesin dapat membaca kalkulator sederhana, bahwa 50+60 hasilnya adalah 110. [1] di depan hasil adalah bagian dari cara R mencetak angka dan vektor. Untuk hasil yang hanya memiliki satu angka, tidak terlalu terlihat kegunaannya, tetapi menjadi begitu berarti ketika hasilnya vector yang panjang. Angka dalam tanda kurung adalah indeks dari angka pertama itu garis. Pertimbangkan kasus menghasilkan 15 angka acak dari normal distribusi:

```
> rnorm(10)  
[1] -0.7888575 -0.1341562 -0.4567463 -3.0869651  0.6399005  
[6]  0.8964738 -0.1187127 -0.6530804  0.1392005 -0.4033104
```

Nilai 0.8964738 adalah hasil elemen ke[6] dari bentuk `rnorm(10)`.

Selain sebagai kalkulator, bahasa R juga dapat memberikan penamaan (symbol) pada variabel seperti contoh berikut.

```
> data = c(10,27,24,13,22,17,19,23,20,49)
```

Output yang diberikan pada penamaan data adalah membaca data dalam bentuk kolom dengan perintah `c()`. Jika kita memanggil data maka akan keluar data yang kita ketik sebelumnya.

```
> data  
[1] 10 27 24 13 22 17 19 23 20 49
```

[1] artinya data setelah tanda itu adalah data nomor pertama (baris pertama). Jika data yang kita input salah, misalnya data nomor 4 seharusnya 12 dan bukan 13, maka

```
> data[4]=12  
> data  
[1] 10 27 24 12 22 17 19 23 20 49
```

Jika data yang kita ketik kurang, misalnya seharusnya ada (15) diantara data ketiga dan keempat, maka

```
> data=c(data[1:3],15,data[4:10])
> data
[1] 10 27 24 15 12 22 17 19 23 20 49
```

Beberapa perintah dalam R beserta output yang dihasilkan mengikuti entri pada data

```
> mean(data)
[1] 22.4

> median(data)
[1] 21

> sum(data)
[1] 224

> data^2
[1] 100 729 576 169 484 289 361 529 400 2401

> data^2-mean(data)
[1] 77.6 706.6 553.6 146.6 461.6 266.6 338.6
506.6
[9] 377.6 2378.6

> var(data)
[1] 113.3778

> sd(data)
[1] 10.6479

> length(data)
[1] 10

> sum(data^2-mean(data))/length(data)-1
[1] 580.4
> sum(data^2-mean(data))/(length(data)-1)
[1] 646

> rank(data)
[1] 1 9 8 2 6 3 4 7 5 10

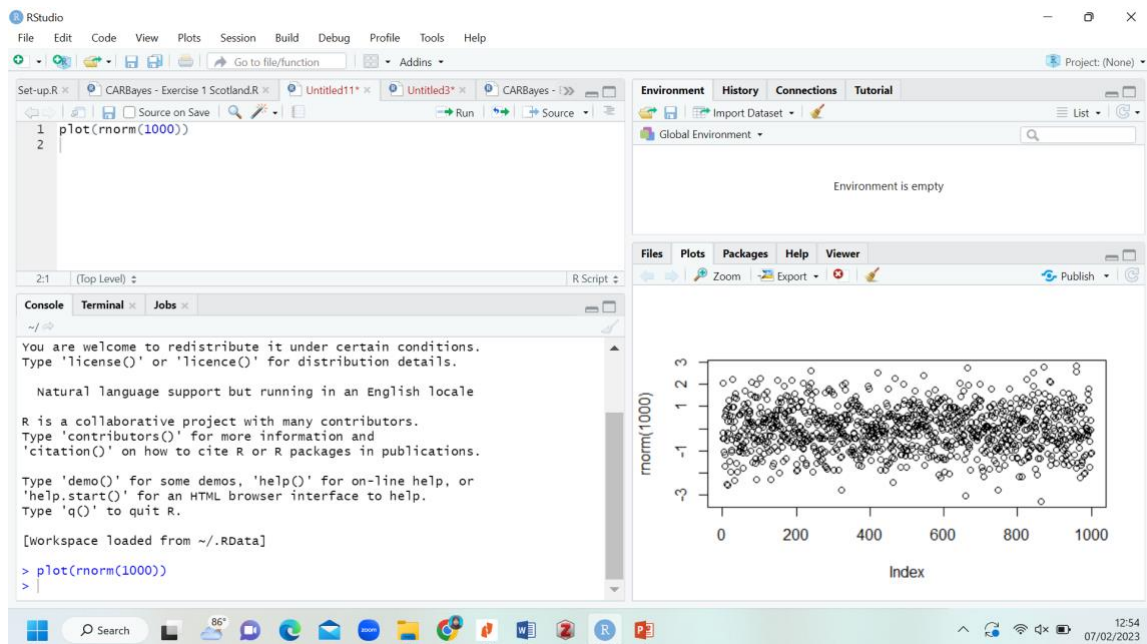
> data<25
[1] TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE
FALSE

> sum(data<25)
[1] 8
```

Jika ada 1000 pengamatan dari distribusi normal $N(0,1)$ maka dapat disimulasikan sebagai berikut

```
>rnorm(1000)
>plot(rnorm(1000))
```

Perintah ini menarik 1000 angka secara acak dari distribusi normal (`rnorm` = random normal) dan memplotnya dalam grafik. Hasil dapat dilihat pada Gambar 3.



Gambar 3. Output R untuk distribusi normal dari 1000 angka acak

Dua karakter “<-” harus dibaca sebagai satu simbol: panah yang menunjuk ke variabel yang diberi nilai, ini dikenal sebagai operator penugasan. Spasi di sekitar operator umumnya diabaikan oleh R. R memiliki *case sensitive*, `b` dan `B` adalah berbeda. Sehingga pengetikan yang salah akan menampilkan pesan *error*. Sebagai contoh:

```
> b <- 5-3
> b
[1] 2
> B
> Error: object "B" not found
```

Tanda # dapat digunakan untuk memberi keterangan pada pengetikan program di R.

```
> 5*3 # comments like this are ignored in R
[1] 15
```

1.3 Latihan

Ketikkan perintah berikut pada layar R-Studio anda dan perhatikan output yang dihasilkan

```
>5*5
>sqrt(100) #sqrt adalah sintaks untuk akar
>log(100)
>log10(100)

>x<-3*5 #x adalah tanda untuk simbol variabel
>x
>x+10
>latihan<-sqrt(x+10)
>latihan
>y<-1:10
>y
>y>8
>y^2
>log(y)
>z<-1:9
>z
>z<-1:3
>z
>100*y+z #notice recycling
>rnorm(15) #rnorm adalah random normal
>plot(rnorm(1000))
```

Tugas

Berikut adalah data berat dan tinggi badan dari 6 sampel :

Berat (kg) : 60, 72, 57, 90, 95, 72

Tinggi (m) : 1.75, 1.80, 1.65, 1.90, 1.74, 1.91

Gunakan R-Studio untuk menentukan :

- Indeks massa tubuh dengan rumus : $\text{imt} <- \text{weight} / \text{height}^2$
- Rataan data (mean)
- Simpangan baku
- Plot yang sesuai untuk data tinggi dan berat badan

REFERENSI{Citation}

Bruce, P., Bruce, A. and Gedeck, P. (2020) *Practical Statistics for Data Scientists*. United States of America: O'Reilly Media, Inc.,.

Dalgaard, P. (2008) *Introductory Statistics with R*. second. Denmark: Springer.

J Horton, N. and Kleinman, K. (2015) *Using R and RStudio for Data anagement, Statistical Analysis ,and Graphics*. second. Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Verzani, J. (2011) *Getting started with Rstudio*. United States of America: O'Reilly Media, Inc.