



PELATIHAN SOFTWARE R UNTUK ANALISIS STATISTIK¹

Herlina Hanum¹, Yuli Andriani¹, Dwi Setyawan²

¹FMIPA Universitas Sriwijaya

linhanum@gmail.com

yulinabila@yahoo.com

²Fakultas Pertanian Universitas Sriwijaya

dsetyawan@unsri.ac.id

ABSTRAK

R is a very useful software in statistical analysis. Its open-source nature makes it easy for users to download and use it freely. Its relatively small size allows R to be installed and used on a laptop because it doesn't use much laptop memory. Users are also facilitated to learn statistical analysis using R because there are many learning resources available which are also easily accessible. But the existence and ease of this has not been widely known by users of statistical analysis. To further introduce R to the public using statistical analysis, R training was conducted through the community service program (PPM) of Sriwijaya University. In this training, the introduction of R along with its potential use and development was carried out in statistical analysis. Participants were guided while practicing. The material studied included downloading and installing R, adding packages, calling data, naming and determining variable values, as well as some simple statistical analyzes. Statistical analysis studied descriptive, regression, and ANOVA. During training, participants can already download, install, and use R in accordance with the training material. For further use according to the needs of the participants, they are given some reference materials contained in a flash disk. In addition participants were also shown how to search for the required syntax via the internet.

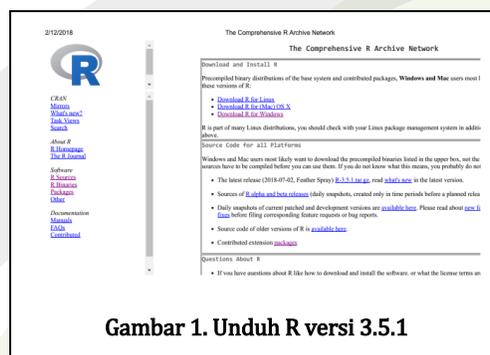
Keywords : *R software, open source, statistical analysis, PPM*

¹ Dibiayai oleh PNBPDana Universitas Sriwijaya Sesuai dengan Surat Perjanjian Penugasan Tenaga Pelaksana Pengabdian kepada Masyarakat Aplikasi IPTEK dan Pengembangan Seni Budaya Lokal Universitas Sriwijaya Nomor :0022.174/UN9/SB3.LP2M.PM/2018, tanggal 16 Agustus 2018

I. PENDAHULUAN

Analisis data yang menggunakan metode statistika memerlukan perhitungan yang banyak. Adanya software komputer yang khusus digunakan untuk analisis data statistika sangat membantu mempermudah analisis tersebut. Software yang banyak digunakan untuk analisis statistika umumnya adalah software yang berbayar dengan harga cukup mahal. Sebagai contoh yaitu SAS, SPSS, dan Minitab. Pada umumnya, untuk menghindari pembayaran, pengguna

memilih untuk menggunakan software bajakan atau software lawas yang sudah dibebas-bayarkan oleh produsennya. Akhir-akhir ini ada beberapa software gratis (open source) yang dikembangkan untuk analisis statistic dan matematik yaitu software R. Menurut Verzani (2018) software R memiliki beberapa kelebihan. Diantara kelebihan tersebut adalah R open source dan dapat dijalankan pada UNIX, Windows, dan Macintosh, R memiliki system bantuan yang sangat baik, R memiliki kapabilitas grafik yang sangat baik, R adalah bahasa yang sangat kuat dengan banyak fungsi-fungsi statistic yang tersedia, serta mudah dipelajari. Fungsi-fungsi dalam R mudah digunakan dan dikembangkan sesuai dengan perkembangan statistika dan kebutuhan pengguna. Salah satu pengembangan yang dilakukan adalah dalam Hanum dkk (2015) yang menggunakan fungsi gamma, gamma tak lengkap, dan invers gamma tak lengkap untuk membangun fungsi-fungsi untuk sebaran gamma-pareto. Contoh lain adalah paket visreg untuk visualisasi regresi yang dikembangkan oleh Breheny & Burchett (2017).



Software R dapat diunduh secara bebas dari website [http:// Cran R](http://cran.r-project.org/). Cara mengunduh R, struktur data dalam R dan pengenalan beberapa fungsi dalam R antara lain dapat dilihat dalam Torfs & Brauer (2014). Selain gratis, software R juga bervolume kecil karena pengguna hanya mengambil bagian paket program (package) yang diperlukan saja. Dengan menggunakan R, analisis akan andal karena Paket program dalam R selalu diperbarui. Cran R menyediakan petunjuk yang lengkap cara penggunaan setiap paket yang dibuat.

Namun software R ini belum banyak dikenal dan digunakan oleh pengguna analisis statistik. Hal ini kemungkinan besar karena ketidaktahuan tentang adanya software R atau karena ketidaktahuan cara menggunakannya.

Program PPM ini bertujuan untuk meningkatkan pengetahuan dan kemampuan dosen tentang analisis statistika menggunakan software. Diakhir program diharapkan peserta sudah menguasai analisis statistika menggunakan R.

Untuk mencapai tujuan tersebut dalam program PPM ini akan dilakukan pelatihan pengenalan R dan penggunaannya. Khalayak sasaran akan dibimbing secara teori dan praktek tentang penggunaan R dalam analisis statistika. Pembimbingan secara teori tentang R dapat membangkitkan ketertarikan untuk mempelajari R.. Praktek pengunduhan, penginstalan, input data, dan analisis statistic menggunakan R, akan sangat membantu dalam penguasaan R oleh khalayak sasaran.



II. METODE PELAKSANAAN

Dalam pengabdian ini digunakan metode ceramah untuk menjelaskan tentang R. Ceramah dan diskusi digunakan dalam review metode statistika. Selanjutnya peserta akan praktek cara mengunduh dan menginstal R. Setelah itu dipraktekkan juga cara mengunduh dan menginstal serta memanggil paket yang diperlukan. Sebagai materi inti adalah ceramah, diskusi, disertai praktek analisis data menggunakan R.

Pelaksanaan PPM

Kegiatan PPM dilaksanakan di Laboratorium Komputasi Matematika FMIPA Universitas Sriwijaya pada tanggal 23 November 2018. Kegiatan ini diikuti oleh 15 peserta. Mereka adalah pengguna analisis statistika. Dalam kegiatan ini peserta dianjurkan untuk membawa laptop dengan tujuan agar hasil pelatihan dapat langsung dimanfaatkan secara pribadi. Dari hasil pelatihan ini peserta sudah berhasil mengunduh dan menginstal program R di laptop masing-masing. Program R yang diinstal adalah R base versi terbaru yaitu 3.5.1. Peserta juga dikenalkan dengan versi R studio dan diberikan alamat web tempat mengunduhnya.

Peserta dikenalkan beberapa package (paket) program yang sudah ada dalam R base. Paket-paket tersebut cukup untuk analisis data dan penyajian gambar yang standar. Paket-paket tersebut sudah dapat dipanggil jika diperlukan. Untuk memanggil paket program dapat digunakan `library()`. Salah satu paket yang sudah dipanggil adalah paket datasets dengan `library(datasets)`. Peserta sudah dapat memanggil data yang ada di dalam paket datasets tersebut.

Pemanggilan paket dapat juga dilakukan dengan mengklik tulisan package yang ada di bagian atas program R. Setelah diklik akan muncul semua paket yang sudah terinstal di R yang digunakan. Kemudian klik nama paket yang diinginkan. Mereka juga dapat melihat kegunaan dan syntax dalam setiap paket dengan cara mengunduh file pdf untuk paket yang diperlukan. Untuk paket yang belum ada dalam R yang diunduh, dapat diunduh dari web cran R package dan diinstal ke dalam R yang digunakan. Paket yang diperlukan dapat diketahui dari hasil pencarian di google tentang program komputasi yang diperlukan. Misalnya paket `xlsx` digunakan untuk mengimpor data dalam format excel ke dalam R. Paket `xlsx` dapat diunduh untuk membuka web <https://cran.r/packages/xlsx>

Paket yang sudah diunduh selanjutnya diinstal ke R dengan cara :
klik package => install package from computer=> cari paket di computer => klik nama paket
Paket `xlsx` baru dapat digunakan jika semua paket yang mendasarinya sudah diinstal ke R. Misalnya paket `xlsx` memerlukan antara lain paket `rJava`. Jika tidak ada `rJava` maka akan keluar pesan : *Error: package 'rJava' required by 'xlsx' could not be found.* Agar paket `xlsx` dapat digunakan, kita harus menginstal semua paket yang diperlukan oleh `xlsx`.

Dalam pelatihan ini peserta juga belajar memanggil data yang mau dianalisis. Data yang ada dalam paket dataset R sudah dapat mereka panggil dengan menuliskan nama data yang dituju setelah paket dataset dipanggil. Data yang berhasil dipanggil adalah `cars` dan `iris`. Data yang ada di laptop peserta juga sudah dapat mereka panggil ke R menggunakan syntax `read.table`.

```
mydata <- read.table("c:/mydata.csv", header=TRUE,  
  sep=";", row.names="id")  
sebagai contoh dibuka file banjir.csv dan diberi nama suka2  
suka2<-read.table("D:banjir.csv",sep=";")  
kemudian kolom 2 dari data suka2 diberi nama hujan dan kolom 3 diberi nama air  
> hujan<-suka2[1:39,2] > air <-suka2[1:39,3]
```



III. HASIL DAN PEMBAHASAN

Deskripsi data

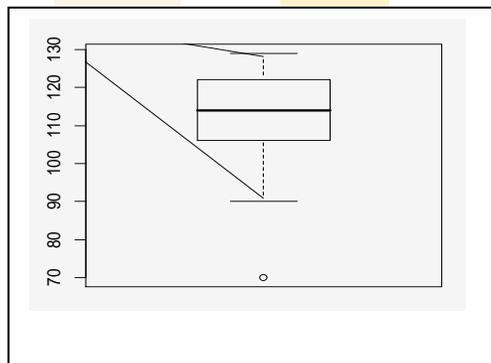


Gambar 2. Unduh paket program dalam R

Dalam analisis statistika sangat penting untuk mengetahui kondisi data antara lain ukuran pemusatan, ukuran penyebaran, dan sebaran data. Untuk mengetahui hal hal tersebut sangat mudah jika menggunakan R. Syntax yang digunakan sebagian besar sama dengan nama ukuran tersebut. Berikut analisis deskriptif yang digunakan untuk peubah air.

```
mean(air)      # nilai tengah atau mean
[1] 112.5641
> var(air)# varian
[1] 151.0945
> IQR(air) # interquartile range
[1] 16
> quantile(air, 0.50) # quantile 50%
50%
114
```

Untuk melihat sebaran suatu peubah dapat digunakan boxplot maupun histogram. Syntax `boxplot(air)` menghasilkan diagram boxplot seperti di bawah ini.





Regresi linier

Untuk analisis statistik, peserta sudah dapat melakukan analisis regresi menggunakan perintah **lm()**. Misalnya untuk menganalisis regresi peubah hujan terhadap peubah air digunakan syntax

```
fit<-lm(air~hujan)
```

dengan **fit** adalah nama untuk hasil analisis regresi peubah hujan terhadap air.

Untuk melihat hasil analisis regresi telah dilakukan dengan cara menuliskan nama hasil analisis yaitu **fit** atau menggunakan syntax **summary(fit)**. Penulisan nama **fit** hanya menghasilkan nilai penduga parameter regresi. Sementara penggunaan **summary(fit)** memberikan hasil analisis secara lengkap sebagai berikut :

```
Residuals:
  Min   1Q Median   3Q   Max
-41.840 -7.741  1.507  9.420 16.958

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.116e+02  2.293e+00  48.688  <2e-16 ***
hujan       1.363e-04  1.713e-04   0.795   0.431
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.35 on 37 degrees of freedom
Multiple R-squared:  0.01681, Adjusted R-squared: -0.00976
F-statistic: 0.6327 on 1 and 37 DF, p-value: 0.4314
```

Pemeriksaan galat model sudah dapat mereka lakukan menggunakan syntax **plot(fit)**. Analisis statistik lainnya yang dipelajari dalam pelatihan ini adalah analisis variansi (ANOVA).

Agar peserta dapat melanjutkan pembelajaran tentang analisis data statistic dengan R, peserta dibekali dengan beberapa bahan bacaan tentang syntax program R untuk dapat dipelajari dan dicoba sendiri sesuai keperluan masing-masing. Bahan bacaan tersebut berupa buku-buku maupun artikel mengenai pemrograman menggunakan R. Semua bahan bacaan tersebut dimuat dalam flashdisk.

Untuk analisis yang belum ada dalam bahan bacaan yang diberikan, peserta juga belajar cara mendapatkan syntax untuk analisis yang diperlukan melalui pencarian di google. Dengan menuliskan analisis yang diinginkan di pencarian google, akan didapatkan beberapa blog yang menyediakan syntax tersebut. Blog yang dimaksud antara lain R archive Dengan cara ini peserta dapat mengembangkan pengetahuan dan keterampilannya terhadap analisis data statistik menggunakan R.

IV. KESIMPULAN DAN SARAN

Kesimpulan

Analisis statistik menggunakan R sangat mudah dipelajari oleh peserta. Hal ini dapat dilihat dari peserta yang dapat dengan mudah melakukan langkah-langkah yang diajarkan. Peserta juga



dipermudah dengan adanya bahan bacaan yang diberikan dalam pelatihan ini. Selain itu peserta tidak merasa kesulitan untuk melanjutkan pembelajaran analisis statistic menggunakan R karena dapat dipelajari dengan mudah melalui beberapa blog di internet yang membahas penggunaan R.

Saran

Program R tidak hanya dapat digunakan untuk analisis statistik. Tetapi juga dapat digunakan untuk bidang matematika yang lain. Misalnya bidang optimasi dan aktuaria. Dalam program R juga tersedia syntax-syntax yang dapat digunakan untuk materi pembelajaran di sekolah. Untuk itu disarankan agar dilakukan pelatihan R untuk masalah-masalah tersebut.

DAFTAR PUSTAKA

- [1] Breheny, P & Burchett.W, 2017, Visualization of Regression Models Using visreg, The R Journal, <https://journal.r-project.org/archive/2017/RJ-2017-046/index.html>
- [2] Hanum, H., Wigena, A. H., Djuraidah, A., & Mangku, I. W. (2016). Modeling Gamma-Pareto distributed data using GLM Gamma. Global Journal of Pure and Applied Mathematics, 12(4)
- [3] <https://cran.r-project.org>, bbmle package, Tools for General Maximum Likelihood Estimation terakhir diupdate 30 Desember 2017.
- [4] Torfs.P & Brauer.C, 2014, A (very) short introduction to R, dilihat tanggal 19 Februari 2018 di <http://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro>.
- [5] Verzani, J, 2018, simpleR Using R for Introductory Statistics, dilihat tanggal 20 Maret 2018 di <https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>.