| **Nama:**  **Adrian Halim**  **NIM: 064002200043** | C:\Users\RPL-SI 02\Pictures\288px-Trisakti_Logo.svg.png | **MODUL 2**  **Nama Dosen:**  **Dedy Sugiarto** |
| --- | --- | --- |
| **Hari/Tanggal:**  **Rabu, 12 JULI 2023** | **Praktikum Statistika** | **Nama Asisten Labratorium:**   1. **Elen Fadilla Estri 064002000008** 2. **Rukhy Zaifa Aduhalim 064002000041** |

**Tipe Data, Filter Data & Koneksi ke Database**

1. **Teori Singkat**

R (juga dikenal sebagai GNU S) adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik. R dibuat oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru, dan kini dikembangkan oleh R Development Core Team, di mana Chambers merupakan anggotanya. R dinamakan sebagian setelah nama dua pembuatnya (Robert Gentleman dan Ross Ihaka), dan sebagian sebagian dari permainan nama dari S.

Bahasa R kini menjadi standar de facto di antara statistikawan untuk pengembangan perangkat lunak statistika, serta digunakan secara luas untuk pengembangan perangkat lunak statistika dan analisis data. R merupakan bagian dari proyek GNU. Kode sumbernya tersedia secara bebas di bawah Lisensi Publik Umum GNU, dan versi biner prekompilasinya tersedia untuk berbagai sistem operasi. R menggunakan antarmuka baris perintah, meski beberapa antarmuka pengguna grafik juga tersedia.

R menyediakan berbagai teknik statistika (permodelan linier dan nonlinier, uji statistik klasik, analisis deret waktu, klasifikasi, klasterisasi, dan sebagainya) serta grafik. R, sebagaimana S, dirancang sebagai bahasa komputer sebenarnya, dan mengizinkan penggunanya untuk menambah fungsi tambahan dengan mendefinisikan fungsi baru. Kekuatan besar dari R yang lain adalah fasilitas grafiknya, yang menghasilkan grafik dengan kualitas publikasi yang dapat memuat simbol matematika. R memiliki format dokumentasi seperti LaTeX, yang digunakan untuk menyediakan dokumentasi yang lengkap, baik secara daring (dalam berbagai format) maupun secara cetakan.

RStudio merupakan integrated development environment (IDE) khusus bagi bahasa pemrograman R. Software ini menyediakan R console, code editor dengan syntax highlighting, code completion dan direct execution, environment, history, connections, dan fitur-fitur tambahan lainnya seperti file manager, packages manager, help, plot viewer, hingga project versioning menggunakan git. RStudio sebenarnya memiliki dua versi, yaitu open source (gratis) dan commercial edition (berbayar). RStudio juga tidak hanya terbatas dalam bentuk aplikasi dekstop, melainkan terdapat versi RStudio Server, yaitu RStudio yang dapat diakses melalui browser yang terhubung dengan suatu jaringan komputer. Untuk saat ini, versi RStudio yang akan dijelaskan hanyalah RStudio open source berbasis dekstop saja.

1. **Alat dan Bahan**

Hardware : Laptop/PC

Software : R Studio

1. **Elemen Kompetensi**

Terdapat beberapa tipe data di R antara lain vektor, matriks dan data frame. Cantumkan setiap output yang dihasilkan dari console Rstudio, ke kolom yang sudah disediakan.

* 1. Latihan pertama – Vektor

1. Tuliskan Perintah berikut ini di window R Script:

| a <- c(1,2,-5,0.3,6,-2,4) # numeric vector  b <- c("one","two","three") # character vector  c <- c(TRUE,TRUE,TRUE,FALSE,TRUE) #logical vector  print(a)  print(b)  print(c) |
| --- |

Output:

|  |
| --- |

* 1. Latihan Kedua – Matriks

1. Seluruh kolom dalam sebuah matriks harus memiliki tipe yang sama (numerik semua, karakter semua, dll) dan memiliki panjang yang sama.

\*gunakan nama variable dengan nama anda masing-masing

| #MATRIKS  cells <- c(3,15,-27,38)  r\_nama <- c("R1", "R2")  c\_nama <- c("C1", "C2")  nama\_matrix <- matrix(cells, nrow=2, ncol=2, byrow=TRUE,  dimnames=list(r\_nama, c\_nama))  print(nama\_matrix) |
| --- |

Output:

|  |
| --- |

* 1. Latihan Ketiga – Data Frame

1. Mengubah data input menjadi data frame

\*gunakan nama variable dengan nama anda masing-masing

| nama1 <- c(1,2,3,4)  nama2 <- c("red", "white", "red", NA)  nama3 <- c(TRUE,TRUE,TRUE,FALSE)  dataku <- data.frame(nama1, nama2, nama3)  names(dataku) <- c("ID","Color","Passed") # variable names  print(dataku) |
| --- |

Output:

|  |
| --- |

2. Selanjutnya ketikkan perintah dibawah ini

| data\_nama <- data.frame(id = letters[1:10], x = 1:10, y = 11:20)  print (data\_nama) |
| --- |

Output:

|  |
| --- |

Beberapa fungsi penting terkait data frame :

head() - shows first 6 rows

tail() - shows last 6 rows

dim() - returns the dimensions of data frame (i.e. number of rows and number of columns)

nrow() - number of rows

ncol() - number of columns

str() - structure of data frame - name, type and preview of data in each column

names() - shows the names attribute for a data frame, which gives the column names.

* 1. Latihan Keempat – Koneksi ke Database

1. Buat sebuah nama database terlebih dahulu dengan nama houseprices di phpmyadmin, Lalu klik menu import

|  |
| --- |

2. Pilih file yang ingin di import ke database (untuk file nama\_excel.csv), Ceklis the first line of the file contains the table column name untuk membuat baris pertama pada file excel tersebut menjadi nama atributnya atau nama kolomnya pada database.

|  |
| --- |

3. Klik go, Rename nama tabel sesuai dengan nama anda

|  |
| --- |

4. Kembali ke R Studio, lalu instal dahulu library yang dibutuhkan pada RStudio. Jika belum tersedia, maka lakukan instruksi:

install.packages(“RMySQL”)

|  |
| --- |

5. Lalu jalankan perintah dibawah ini

| library(RMySQL)  con = dbConnect(MySQL(), user = 'root', password = '', dbname = 'houseprices', host = 'localhost')  dbListTables(con)  myQuery <- "select \* from nama;"  df <- dbGetQuery(con, myQuery) |
| --- |

Output:

|  |
| --- |

* 1. Latihan Kelima – Teknik Filter Data

1. Jalankan perintah dibawah ini:

\*Perintah ini akan menampilkan 86 baris data hasil filter.

| library(dplyr)  df<-filter(df,Brick=="No")  print(df) |
| --- |

Output:

|  |
| --- |

2. Jalankan perintah dibawah ini:

\*Perintah ini akan menampilkan 105 baris data hasil filter.

| library(dplyr)  df<-filter(df,Brick=="No"|Neighborhood=="East")  print(df) |
| --- |

Output:

|  |
| --- |

* 1. Latihan Keenam – Tugas

Buat sebuah database serta tabel di dalamnya (bisa gunakan data teman dipraktikum ke 1). Lakukan koneksi R ke database serta berikan beberapa filter data sesuai yang anda inginkan. Tampilkan data tersebut

1. Buatlah terlebih dahulu Database baru dengan format PS2[NamaAnda]

|  |
| --- |

2. Import file berformat csv yang telah anda buat sebelumnya di Praktikum 1, dimana file tersebut berisi 6 kolom dan 20 baris data.

|  |
| --- |

3. Koneksikan R ke Database tersebut sebagaimana yang telah anda lakukan pada Elemen Kompetensi 1 dimodul kedua ini dengan menyesuaikan kembali nama Database baru yang sudah dibuat.

|  |
| --- |

4. Lakukan filter data terhadap Kolom Gender, untuk melihat berapa baris data Pria/Wanita (Pilih salah 1).

|  |
| --- |

5. Lampirkan Screenshot

✅Hasil import file kedalam Database

✅Berhasilnya koneksi R ke Database

✅Kode serta hasil filter di RStudio

|  |
| --- |

1. **File Praktikum**

Github Repository:

|  |
| --- |

1. **Soal Latihan**

Soal:

1. Ada berapa tipe data yang data di deklrasikan oleh Rstudio?
2. Bagaimana cara memanggil sebuah tabel di dalam database ke Rstudio?

Jawaban:  
1. RStudio sebenarnya adalah lingkungan pengembangan terintegrasi (IDE) untuk R, yang digunakan untuk analisis data dan pemrograman statistik. RStudio sendiri tidak memiliki tipe data sendiri, karena tipe data yang digunakan dalam RStudio berasal dari bahasa pemrograman R.

R, sebagai bahasa pemrograman, memiliki beberapa tipe data dasar yang umum digunakan. Berikut adalah beberapa tipe data yang sering digunakan dalam R:

Angka (numeric): Tipe data ini digunakan untuk menyimpan bilangan desimal atau angka dengan titik koma.

Integer: Tipe data ini digunakan untuk menyimpan bilangan bulat.

Karakter (character): Tipe data ini digunakan untuk menyimpan teks atau karakter.

Logika (logical): Tipe data ini digunakan untuk menyimpan nilai kebenaran (TRUE atau FALSE).

Faktor (factor): Tipe data ini digunakan untuk menyimpan variabel kategorikal atau faktor.

Vektor (vector): Tipe data ini digunakan untuk menyimpan kumpulan elemen dengan tipe data yang sama.

Matriks (matrix): Tipe data ini digunakan untuk menyimpan kumpulan elemen dalam bentuk matriks dua dimensi.

Data frame: Tipe data ini digunakan untuk menyimpan data dalam bentuk tabel dua dimensi, dengan kolom-kolom yang mungkin berbeda tipe data.

2. Untuk memanggil sebuah tabel di dalam database ke RStudio, Anda perlu terlebih dahulu menghubungkan RStudio ke database yang sesuai. Anda dapat menggunakan package R yang disebut DBI (Database Interface) untuk melakukan koneksi ke database. Berikut adalah langkah-langkah umumnya:

Instal dan muat paket DBI di RStudio menggunakan perintah install.packages("DBI") dan library(DBI).

Gunakan fungsi dbConnect() untuk membuat koneksi ke database. Anda perlu memberikan informasi seperti tipe database, host, nama pengguna, kata sandi, dan nama database.

Contoh: con <- dbConnect(RSQLite::SQLite(), dbname = "nama\_database.sqlite") (contoh untuk SQLite).

Setelah terhubung, Anda dapat menggunakan fungsi-fungsi seperti dbListTables() untuk melihat daftar tabel yang tersedia di database.

Gunakan fungsi dbReadTable() untuk membaca isi tabel ke dalam RStudio.

Contoh: data <- dbReadTable(con, "nama\_tabel").

Pastikan Anda mengganti "nama\_database.sqlite" dengan nama database yang sesuai dan "nama\_tabel" dengan nama tabel yang ingin Anda baca. Selain itu, perhatikan bahwa proses ini akan bervariasi tergantung pada jenis database yang Anda gunakan.

1. **Kesimpulan**
   1. Dalam pengerjaan praktikum Statistika

Dalam pengerjaan statistika, tipe data memainkan peran penting. Tipe data yang digunakan dalam analisis statistika mempengaruhi jenis metode analisis yang dapat dilakukan dan bagaimana data tersebut diinterpretasikan. Berikut adalah beberapa tipe data yang umum digunakan dalam statistika:

Data Numerik: Tipe data numerik terdiri dari data kuantitatif yang dapat diukur. Contoh tipe data numerik meliputi data berkelanjutan seperti tinggi, berat badan, suhu, dan data diskrit seperti jumlah pengunjung, jumlah kejadian, dan sebagainya.

Data Kategorikal: Tipe data kategorikal menggambarkan karakteristik atau kategori yang tidak dapat diukur secara kuantitatif. Contoh tipe data kategorikal termasuk jenis kelamin, status perkawinan, jenis pekerjaan, dan sebagainya. Data kategorikal dapat dikelompokkan menjadi dua jenis: nominal dan ordinal. Data nominal tidak memiliki urutan atau tingkatan tertentu (misalnya, warna favorit), sedangkan data ordinal memiliki urutan atau tingkatan (misalnya, tingkat kepuasan: rendah, sedang, tinggi).

Data Waktu: Tipe data waktu digunakan untuk merekam informasi tentang waktu atau tanggal. Contoh tipe data waktu termasuk tanggal lahir, waktu transaksi, atau data deret waktu. Data waktu juga dapat dianalisis menggunakan metode statistika khusus seperti analisis deret waktu.

Data Spasial: Tipe data spasial digunakan untuk merepresentasikan data yang berkaitan dengan lokasi atau wilayah geografis. Contoh tipe data spasial termasuk koordinat geografis, peta wilayah, atau data geospasial. Analisis statistika spasial melibatkan pemodelan dan analisis data geografis.

Pemahaman tipe data yang digunakan dalam pengerjaan praktikum statistika sangat penting untuk memilih dan menerapkan metode analisis yang tepat. Selain itu, pemahaman karakteristik dan sifat data tersebut juga penting untuk memastikan interpretasi yang benar terhadap hasil analisis.

* 1. Kita juga dapat mengetahui

Dalam kesimpulannya, terdapat beberapa tipe data dasar yang digunakan dalam bahasa pemrograman R dan diakses melalui RStudio. Tipe data dasar tersebut meliputi angka (numeric), integer, karakter (character), logika (logical), faktor (factor), vektor (vector), matriks (matrix), dan data frame.

Untuk memanggil sebuah tabel dari database ke RStudio, Anda perlu melakukan koneksi ke database terlebih dahulu dengan menggunakan paket R yang sesuai dengan jenis database yang digunakan. Setelah terhubung, Anda dapat menggunakan fungsi-fungsi khusus seperti dbListTables() untuk melihat daftar tabel yang tersedia di database dan dbReadTable() untuk membaca isi tabel ke dalam RStudio sebagai data frame.

Penting untuk mengikuti langkah-langkah yang relevan untuk koneksi ke database dan menggunakan fungsi yang tepat sesuai dengan jenis database yang digunakan. Pastikan juga untuk menginstal paket-paket yang diperlukan dan mengacu pada dokumentasi atau panduan penggunaan yang relevan.

1. **Cek List (✔)**

| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| --- | --- | --- | --- |
| **Selesai** | **Tidak Selesai** |
| **1.** | Latihan Pertama | ✅ |  |
| **2.** | Latihan Kedua | ✅ |  |
| **3.** | Latihan Ketiga | ✅ |  |
| **4.** | Latihan Keempat | ✅ |  |
| **5.** | Latihan Kelima | ✅ |  |
| **6.** | Latihan Keenam | ✅ |  |

1. **Formulir Umpan Balik**

| **No** | **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| --- | --- | --- | --- |
| **1.** | Latihan Pertama | 3 Menit | Menarik |
| **2.** | Latihan Kedua | 5 Menit | Menarik |
| **3.** | Latihan Ketiga | 4 Menit | Menarik |
| **4.** | Latihan Keempat | 10 Menit | Menarik |
| **5.** | Latihan Kelima | 10 Menit | Menarik |
| **6.** | Latihan Keenam | 13 Menit | Menarik |

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang