



UNIVERSITAS
INDONESIA

Veritas, Probitas, Iustitia

Analisis Data Menggunakan R

.....

Sigit Wibowo


📌 sswibowo.github.io

15 & 16 September 2021

Cara Mengunduh & Instalasi di Windows

1. Unduh R dari laman <https://cran.r-project.org>, lalu *install* R (ikuti wizard window)
2. Buat folder di `c:\RStudio\download`.
3. Buka <https://www.rstudio.com/products/rstudio/download/>, dan klik “Download RStudio”
 - Di laman berikutnya, klik “Desktop”
 - Di laman berikutnya, klik “Download RStudio Desktop”
 - Di laman berikutnya, pilih sistem operasi Anda dan klik tautan unduh
 - Unduh ke folder `c:\RStudio\download`
 - Pada Windows, pada *bar* unduhan di bagian bawah jendela akan terbuka.

Cara Mengunduh & Instalasi di Windows (lanjutan)

- Klik panah di sebelah “Save”, pilih “Save as”, dan pilih `c:\RStudio\download` sebagai folder tujuan
- 4. Buka `c:\RStudio\download` dan klik dua kali pada icon “RStudio-(latest version)”,
 - Ini akan memulai panduan instalasi (*installation wizard*), lalu klik “next”.
- 5. Pastikan “Destination Location”, klik “Browse” dan pilih `c:\RStudio`.
- 6. Klik “next”. Klik “Install”, dan “Finish”.
- 🍏 Untuk pengguna Mac, sebelum loading `RStudio-[1.x.x].dmg`, masuk ke Security & Privacy lalu *unlock* (klik icon , masukkan *username* dan *password*) agar bisa loading .dmg.

Setting pada RStudio

1. Klik ikon Windows di sudut kiri bawah layar Anda, lalu “All programs”, gulir ke bawah ke RStudio dan klik icon RStudio
2. Buka RStudio (tahap ini mungkin memakan waktu beberapa saat pertama kali).
 - Buka “tools”, “global options”.
 - Klik pada “appearance” dan ubah ukuran font dan Editor Theme, jika perlu
3. Secara *default*, RStudio akan membuat dokumen pdf menggunakan perangkat lunak pdf “RStudio Viewer”
 - Jika Anda ingin menggunakan *pdf viewer* lainnya (biasanya Adobe Acrobat), buka “Tools,” “Global Options,” “Sweave,” dan pilih “system viewer.”

Instalasi Paket

1. Klik pada “Tools ” lalu “Install packages”
2. Masukkan nama paket di bidang putih (di bawah “Packages (separate multiple with space or comma)”), dan klik “Install”
3. Anda juga dapat mengklik tab “tools”, “check for updates”.
4. Untuk pelatihan ini, pastikan paket ini sudah terpasang sebelum sesi dimulai:
 - tidyverse, readr, ggplot2, ggExtra, ggrepel, cowplot, aplpack, moments, brms

Topik

1. Sesi 1


1.1 Dasar-dasar R

1.2 Data Wrangling

1.3 Visualisasi Data

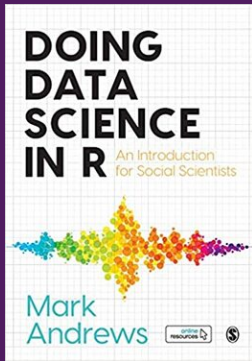
2. Sesi 2

Pustaka

 Andrews, M. (2021). *Doing Data Science in R: An Introduction for Social Scientists*, SAGE

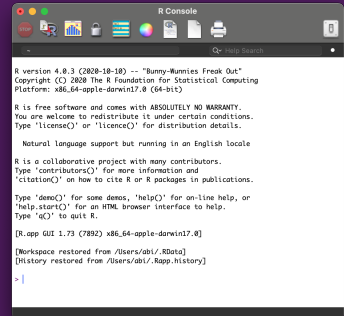
➤ www.mjandrews.org/book/ddsr/

 github.com/mark-andrews/ddsr



R

- R adalah pengembangan dari bahasa S oleh John Chamber dkk di AT&T Bell Labs pada akhir 1970-an dan awal 1980-an
- R sering disebut juga sebagai "GNU S" adalah piranti *open source* dengan lisensi sesuai dengan GNU *general public licence* (GPL 2)
- Robert Gentleman dan Ross Ihaka (University of Auckland) menginisiasi pengembangan R



```
R Console
Qt - Help Search

R version 4.0.3 (2020-10-10) -- "Bunny-Munnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

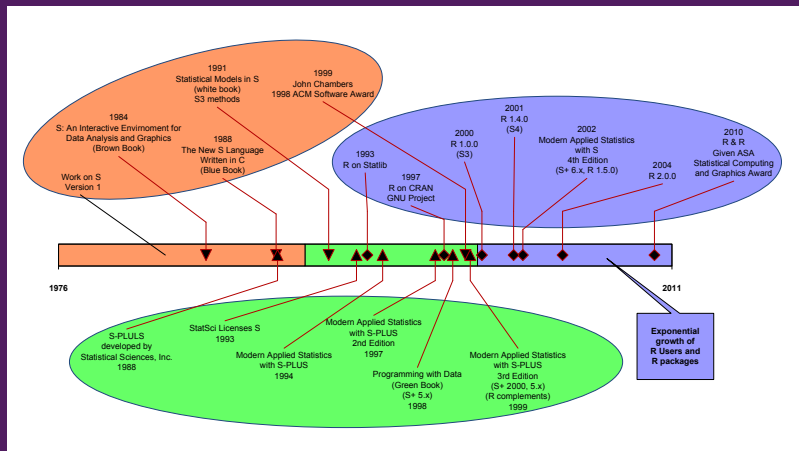
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.73 (7892) x86_64-apple-darwin17.0]

[Workspace restored from /Users/abi/.RData]
[History restored from /Users/abi/.Rapp.history]

>|
```


Timeline



Sumber: Yollin (2011).

Lebih lanjut: github.com/revodavid/20-years-of-R, youtu.be/iq_bIXElx-U

Mengapa Menggunakan R?

- R gratis dan tersedia di platform apa pun (Mac, PC, Linux) dan juga, melalui Rstudio, di *web browser*.
 - Ini berarti kita memiliki akses ke R kapan pun dan di mana pun ketika kita membutuhkannya
- R menghasilkan gambar berkualitas tinggi untuk publikasi
 - Melalui paket `\pkg{lattice}`, berbagai plot berguna mudah diproduksi
- R membantu pengolahan data dengan cara yang efisien dan efektif untuk analisis statistik.

Mengapa Menggunakan R?

- R menghasilkan penelitian yang dapat direproduksi
 - R memberikan catatan yang tepat tentang bagaimana analisis dilakukan
 - Perintah bisa diedit, jalankan kembali, berkomentar, dibagikan, dll.
- R selalu *up-to-date*
 - Banyak metode analisis baru muncul pertama kali di R

Mengapa Menggunakan R?

- Ada banyak paket yang tersedia yang mengotomatisasi tugas-tugas tertentu.
 - CRAN (Comprehensive R Archive Network) repository berisi lebih dari 3000 paket yang menyediakan berbagai kemampuan tambahan
- R dapat dikombinasikan dengan alat lain.
 - R dapat digunakan dalam bahasa pemrograman (seperti Python) atau dalam lingkungan *script* untuk mengotomatisasi analisis data
 - Lihat Anaconda Navigator untuk *deep learning* dan *big data*

Mengapa Menggunakan R?

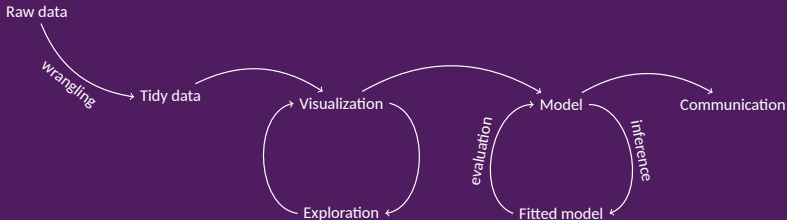
- R sangat banyak digunakan di penelitian akademis, dan mendapatkan pangsa pasar di industri
- Meskipun R sangat komprehensif, contoh hal-hal berguna dapat dipelajari dengan cukup mudah.

- Lebih lanjut

www.dummies.com/programming/r/the-benefits-of-using-r/
atau www.msperlin.com

Dasar-dasar

- Tukey (1962): *data analysis*
- Cleveland (2001): *data science*
 - ...“*computational and statistical methods applied to all the aspects of data analysis.*”



Gambar: Alur kerja *data science*. Sumber: Andrews (2021)

Langkah-langkah dalam R

0. R console
1. R sebagai kalkulator
2. Variables & assignment
3. Vectors
4. Data frame
5. Struktur data lain
6. Functions
7. Scripts
8. Instalasi dan loading packages
9. Membuka dan membaca data
10. Working directory, RStudio projects, & clean workspaces

The image shows the RStudio interface with four main panels. The top-left panel is the Editor, showing an R script file named 'chapter03.R'. The top-right panel is the Environment, which is currently empty. The bottom-left panel is the Console, displaying the R version and copyright information. The bottom-right panel is the Files panel, showing a directory tree with various chapter folders.

1. R console

```
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]
```

2. Editor

```
# manipulasi data #
11 glimpse(blp_df)
12 select(blp_df, spell:prev.rt)
13 select(blp_df, 2:5) # columns 2 to 5
14 select(blp_df, starts_with('p'))
15 select(blp_df, ends_with('t'))
16 select(blp_df, contains('rt'))
17 select(blp_df, matches('^rt|rt$'))
18 select(blp_df, -participant)
19
20
21 # ganti nama variabel
22 select(blp_df, subject=participant, reaction_time=rt)
23 rename(blp_df, subject=participant, reaction_time=rt)
24 rename_all(blp_df, ~str_replace_all(., '\\.', '_'))
25
26 ## Memilih observasi dengan 'slice' dan 'filter'
27 slice(blp_df, c(10, 20, 50, 100, 500))
28 slice(blp_df, 10:100)
```

3. Workspace & history

Environment is empty

4. Files, plots, help

Files | Plots | Packages | Help | Viewer

- New Folder
- Delete
- Rename
- More

Name	Size	Modified
..		
chapter_02_introduction		
chapter_03_wrangling		
chapter_04_visualization		
chapter_05_exploratory		
chapter_07_reproducible		
chapter_08_inference		
chapter_09_linear		
chapter_10_logistic		
chapter_11_generalized		
chapter_12_multilevel		
chapter_13_nregression		
chapter_14_sem		
chapter_15_hpc		
chapter_16_shiny		

R sebagai kalkulator

```
> 3 + 3
[1] 6
> 3 * 3
[1] 9
> 3 * 3
[1] 9
> 3 / 3
[1] 1
> 3 ^ 3
[1] 27
> 12 * 5 - 9 / 5 ^ 2
[1] 59.64
> 2 / (3 * 4)
[1] 0.1666667
```

Assignment

- operator: <-
- fungsi: assign
- tanda sama dengan: =
 - harus digunakan untuk menyatakan argumen dalam suatu fungsi

Variable assignment

```
> y <- 7
> y
[1] 7
> assign("e",2.7813)
> e
[1] 2.7813
> s = sqrt(49)
> s
[1] 7
> r <- rnorm(n=3)
> r
[1] -0.126625  0.339415 -0.093429
```

Struktur Data

- Struktur data menjelaskan bagaimana data disusun
- Dalam R, umumnya data disusun sebagai:
 - vectors
 - matrix
 - array
 - dataframe
 - timeseries
 - list

Orientasi Obyek

- Semua dalam R adalah object
 - Gunakan fungsi `ls` and `objects` untuk melihat semua obyek yang ada di *current workspace*

Listing objects

```
> x <- c(3.1416,2.7183)
> m <- matrix(rnorm(9),nrow=3)
> tab <- data.frame(store=c("downtown","eastside","airport"),
sales=c(32,17,24))
> cities <- c("Cengkareng","Tanjung Priok","Halim")
> ls()
[1] "cities" "e"      "m"      "r"      "s"      "tab"    "x"      "y"
```

Jenis Data

- Seluruh R objects memiliki type atau storage mode
- Fungsi `typeof` memperlihatkan jenis suatu obyek
- Types atau jenis:
 - character: "a", "abc"
 - numeric: 2, 5.5
 - integer: 2L
 - logical: TRUE, FALSE
 - complex: 1+4i

Object type

```
> x
[1] 3.1416 2.7183
> typeof(x)
[1] "double"
> cities
[1] "Cengkareng" "Tanjung Priok" "Halim"
> typeof(cities)
[1] "character"
```

Kelas Obyek

- Seluruh R objects memiliki class
- Fungsi class menampilkan kelas suatu obyek
- R memiliki banyak kelas, basic class:
 - numeric
 - character
 - data.frame
 - matrix

Object type

```
> m
      [,1]      [,2]      [,3]
[1,]  0.7607100 -0.6803638  0.7672464
[2,] -0.7387365  0.9563487  1.0087268
[3,] -0.8776376 -0.7847578 -1.0441825
> class(m)
[1] "matrix"
```

Vector

- R adalah bahasa matriks/vektor
 - vektor dapat dibuat dengan menggunakan `c`, fungsi menggabungkan
 - Ketika nilai tunggal dapat digunakan, sebuah vektor bisa digunakan dan R akan menjalankan operasi vektor

```
> constants <- c(3.1416,2.7183,1.4142,1.6180)
> names(constants) <- c("pi","euler","sqrt2","golden")
> constants
      pi  euler sqrt2 golden
3.1416 2.7183 1.4142 1.6180
> constants^2
      pi  euler sqrt2 golden
9.869651 7.389155 1.999962 2.617924
```

Vector

```
> ## Vektor x
> x <- c(10, 5, 3, 6)
> x
[1] 10  5  3  6
> ## Vektor y
> y <- c(x, 0.55, x, x)
> y
[1] 10.00  5.00  3.00  6.00  0.55 10.00  5.00  3.00  6.00
[10] 10.00  5.00  3.00  6.00
> round(y, 3)
[1] 10.00  5.00  3.00  6.00  0.55 10.00  5.00  3.00  6.00
[10] 10.00  5.00  3.00  6.00
```


Data Frame

- Data frame dibuat saat kita membaca dalam konten file data, seperti file '.csv' atau '.xlsx'
- Sebagai pengenalan, kita menggunakan perintah fungsi 'data.frame()'

```
> contoh_df <- data.frame(name = c('unyil',  
'ucrit', 'usro'), age = c(11, 13, 10))  
> contoh_df  
   name age  
1 unyil  11  
2 ucrit  13  
3 usro   10  
> class(contoh_df)  
[1] "data.frame"
```

Struktur data lain

- *Matrices*
- *Lists*
- *Arrays*

Matrices

```
> ## Matriks
> x <- 1:8
> ## konversi vektor menjadi matriks
> dim(x) <- c(2, 4)
> x
      [,1] [,2] [,3] [,4]
[1,]    1    3    5    7
[2,]    2    4    6    8
> ## Cara lain
> x <- matrix(1:8, 2, 4)
> x
      [,1] [,2] [,3] [,4]
[1,]    1    3    5    7
[2,]    2    4    6    8
```

Functions

- Fungsi umum berperilaku sesuai dengan kelas argumentasinya, seperti:
 - plot
 - print
 - summary

```
> methods(plot)[1:15]  
[1] "plot.acf"           "plot.data.frame"    "plot.decomposed.ts"  
[4] "plot.default"       "plot.dendrogram"    "plot.density"  
[7] "plot.ecdf"          "plot.factor"        "plot.formula"  
[10] "plot.function"      "plot.hclust"        "plot.histogram"  
[13] "plot.HoltWinters"    "plot.isoreg"        "plot.lm"
```

Contoh *Functions*

- Dengan menggunakan vektor

```
> primes <- c(2, 3, 5, 7, 11, 13, 17, 19, 23, 29)
> sum(primes)
[1] 129
> mean(primes)
[1] 12.9
> sd(primes)
[1] 9.024042
> var(primes)
[1] 81.43333
```

- Pilihan argumen `trim=0.1`

```
> median(primes, trim=0.1)
[1] 12
```

R packages

- Semua fungsi R disimpan dalam paket (packages)
- Distribusi R standar mencakup paket *core* dan paket *recommended*:
 - *Core R packages*: base, utils, stats, methods, graphics, grDevices, datasets
 - *Core R packages*: boot, rpart, foreign, MASS, cluster, Matrix, tidyverse, dplyr
 - Paket tambahan dapat diunduh melalui R GUI atau melalui fungsi `install.packages`
- Ketika R awalnya loading, hanya paket inti R yang dimuat secara *default*
 - Paket tambahan dimuat melalui perintah `library`
 - Paket dataset dibuat dapat diakses melalui perintah `dataset`

Topik

1. Sesi 1

1.1 Dasar-dasar R

1.2 Data Wrangling

1.3 Visualisasi Data

2. Sesi 2

Data Wrangle

- Definisi
 - “proses mengambil data dalam bentuk aslinya yang tidak terstruktur, berantakan, atau rumit dan mengubahnya menjadi format yang bersih dan rapi yang memungkinkan eksplorasi data, visualisasi, dan akhirnya pemodelan dan analisis statistik dapat berjalan secara efisien dan relatif mudah.”
- Istilah lain
 - “*data cleaning, data munging, data pre-processing, data preparation*”

Membaca Data dalam Data Frame

- Contoh perintah yang umumnya digunakan:
 - `read_csv`: untuk file dengan nilai pada setiap baris dipisahkan dengan koma
 - `read_tsv`: untuk file yang nilainya dipisahkan oleh tab
 - `read_delim`: untuk file yang nilainya dipisahkan oleh pembatas arbitrer seperti '|', ':', ';', dll.
 - » `read_csv` dan `read_tsv` adalah kasus khusus dari perintah `'read_delim'` yang lebih umum.
 - `read_table`: untuk file yang nilainya dipisahkan oleh satu atau lebih, dan kemungkinan spasi putih yang tidak konsisten.

Data Wrangling menggunakan dplyr

- “Kata kerja” seperti:
 - select, rename, slice, filter, mutate, arrange, group_by, summarize
- Menggabungkan (*merge & join*) sekumpulan data menggunakan perintah:
 - inner_join, left_join, right_join, full_join
- Perintah lain:
 - pivot_longer, pivot_wider.

Operator %>%

- Operator '%>%' di R dikenal sebagai “pipe”.
- Hal ini memungkinkan kita untuk membuat urutan fungsi, kadang-kadang dikenal sebagai “pipe”, yang menghindari penggunaan berulang fungsi yang bersifat *nested* atau struktur data sementara.
- Hasilnya biasanya kode yang sangat bersih, mudah dibaca, dan tidak berantakan.
- Di RStudio, pintasan keyboard Ctrl+Shift+M mengetik '%>%'.

Menggabungkan Data Frame

- `binds`: operasi sederhana yang secara vertikal menumpuk data frame yang berbagi variabel umum, atau menumpuk secara horizontal data frame yang memiliki jumlah pengamatan yang sama.
- `joins`: operasi umum dalam database relasional menggunakan SQL dan memungkinkan kita untuk menggabungkan tabel terpisah sesuai dengan kata kunci yang sama
- `set operations`: fungsi `'intersect'`, `'union'`, dll., memungkinkan kita untuk menggabungkan data frame “yang memiliki variabel identik” menggunakan operasi yang ditetapkan.

Topik

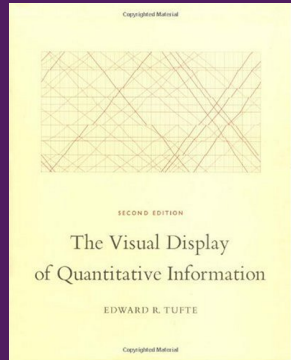
1. Sesi 1

- 1.1 Dasar-dasar R
- 1.2 Data Wrangling
- 1.3 Visualisasi Data

2. Sesi 2

Tufte (1983, 2001)

- *Above all else show the data*
- *Avoid distorting what the data have to say*
- *Present many numbers in a small space*
- *Encourage the eye to compare different pieces of data*
- *Reveal the data at several levels of detail, from a broad overview to the fine structure*

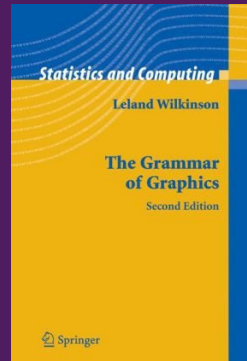


Visualisasi Data

- Histogram, density plots, bar plots
 - digunakan untuk menampilkan distribusi nilai variabel kontinu dan diskrit.
- Boxplots
 - digunakan untuk menampilkan distribusi nilai variabel kontinu
 - lebih terkait erat dengan deskripsi statistik yang *robust* dan diperlakukan sebagai kelas tersendiri
- Scatterplots
 - digunakan untuk menampilkan data bivariat, atau hubungan antara dua variabel
 - biasanya digunakan dalam kasus dengan dua variabel kontinu, tetapi juga dapat digunakan, meskipun mungkin dengan modifikasi tambahan, ketika satu variabel diskrit

Plotting menggunakan ggplot

- *"grammar of graphics"*
 - sekumpulan sistem aturan untuk memetakan variabel dalam kumpulan data ke properti plot (misalnya bentuk, ukuran, warna, posisi)
- `library(ggplot2)`



Histogram, density plots, barplots

- Histogram dan density plots
 - Salah satu cara paling sederhana dan umumnya berguna untuk memvisualisasikan distribusi nilai-nilai variabel individu dan bersifat kontinu.
- Barplots
 - Digunakan untuk nilai-nilai variabel diskrit

Tukey boxplots

- Plot kotak, juga dikenal sebagai plot *box-and-whisker*, digunakan untuk menampilkan distribusi nilai dari suatu variabel.
 - "Kotak" memanjang dari persentil 25 ke 75
 - "Garis" atau "pita" di dalam kotak adalah nilai median, yang juga merupakan persentil 50
 - "*Whisker*" meluas ke titik terjauh di atas persentil 75, atau di bawah persentil 25, yang berada dalam 1,5 kali rentang antarkuartil (rentang dari 25 ke persentil 75)
 - Setiap titik di luar 1,5 kali rentang antarkuartil di atas persentil 75 atau di bawah persentil 25 diwakili oleh sebuah titik dan diklasifikasikan sebagai "pencilan"

Scatterplots

- Menampilkan pasangan nilai sebagai titik dalam ruang dua dimensi.
- Fitur:
 - Menambahkan distribusi marginal
 - *Smoothing function (loess=locally estimated scatterplot smoothing)*
 - Bubbleplots

Facet Plots

- Memungkinkan kita menghasilkan beberapa subplot terkait, yang mana setiap subplot menampilkan beberapa subset data

Tips

</> Rstudio cheat sheet: www.rstudio.com/resources/cheatsheets/

</> ggplot2 cheat sheet:
github.com/rstudio/cheatsheets/raw/master/data-visualization.pdf

 Tufte in R: motioninsocial.com/tufte/

Topik

1. Sesi 1

2. Sesi 2

2.1 Analisis Eksplorasi Data

2.2 Model Statistik dan Inferensi Statistik

Eksplorasi Data: Sebuah Pengantar

- Tukey (1977)
 - *“It is important to understand what you can do before you learn to measure how well you seem to have done it.”*

Eksplorasi Data: Sebuah Pengantar

- Tukey (1977)
 - *"It is important to understand what you can do before you learn to measure how well you seem to have done it."*
- Measures
 - *central tendency* (lokasi)
 - *dispersion*
 - *skewness* (bentuk atau *shape*)
 - *kurtosis*
- Ekplorasi grafis untuk distribusi univariat

Topik

1. Sesi 1

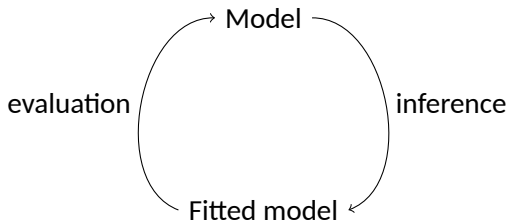
2. Sesi 2

2.1 Analisis Eksplorasi Data

2.2 Model Statistik dan Inferensi Statistik

Pengantar

- *Statistical modelling & statistical inference*



- Contoh menggunakan data `housing.csv`

Inferensi Statistik

- Inferensi statistik adalah inferensi dari nilai-nilai variabel yang tidak diketahui dalam model statistik.
- Ada dua pendekatan utama untuk inferensi statistik:
 - pendekatan berbasis “klasik”, “frequentist”, atau “teori sampel”
 - pendekatan Bayesian

Contoh



The screenshot shows the top of The Guardian website. The header includes a 'Sign in' link, a 'Subscribe' button with a right arrow, and the 'The Guardian' logo with the tagline 'News website of the year'. Below the header is a navigation bar with links for 'News', 'Opinion', 'Sport', 'Culture', and 'Lifestyle', along with a hamburger menu icon. A secondary navigation bar lists various global regions: 'World', 'Europe', 'US', 'Americas', 'Asia', 'Australia', 'Middle East', 'Africa', 'Inequality', and 'Global development'. The main content area features a 'World news' section with the headline 'Heads, Belgium wins - and wins'. Below the headline is the byline 'Charlotte Denny and Sarah Dennis' and the date 'Fri 4 Jan 2002 02:00 GMT'. There are social media sharing icons for Facebook, Twitter, and Email. The article text begins with a memo to all teams playing Belgium in the World Cup, followed by a paragraph about mathematicians' observations on eurozone coins, and then a paragraph about King Albert's appearance on Belgian coins.

Sign in

Subscribe →

The Guardian
News website of the year

News Opinion Sport Culture Lifestyle

World Europe US Americas Asia Australia Middle East Africa Inequality Global development

World news
Heads, Belgium wins - and wins

Charlotte Denny and Sarah Dennis
Fri 4 Jan 2002 02:00 GMT

f t e

Memo to all teams playing Belgium in the World Cup this year: don't let them use their own coins for the toss.

Mathematicians say the coins issued in the eurozone's administrative heartland are more likely to land heads up than down.

While the notes which began circulating in the 12 members of the eurozone on January 1 are all the same, the coins show national symbols on one side and a map of Europe on the other.

King Albert, who appears on Belgian coins, appears to be a bit of a lightweight, according to Polish mathematicians Tomasz Gliszczynski and Wacław Zawadowski. The two professors and their students at the Podlaska Academy in Siedlce spun a Belgian one euro coin 250 times, and found it landed heads up 140 times. The cent coins proved even more likely to land heads up.

- Populasi adalah distribusi probabilitas/himpunan nilai yang mungkin dari jumlah sisi kepala koin dalam sampel percobaan $n = 250$, yang merupakan distribusi probabilitas lebih dari $0, 1, \dots n$.
- m adalah realisasi dari variabel acak Y yang nilainya mungkin adalah $0 \dots n$ dan distribusi probabilitasnya adalah $Pop(Y)$.
- Salah satu pilihan yang layak untuk model ini adalah distribusi binomial: $Y \sim \text{Binomial}(\theta, n = 250)$.

Pendekatan Klasik: MLE

- Jika Y adalah variabel acak yang terdistribusi sebagai distribusi binomial dengan parameter θ dan ukuran sampel n , yang dapat kita nyatakan secara ringkas sebagai $Y \sim \text{Binomial}(\theta, n)$, maka probabilitas bahwa Y mengambil nilai m adalah sebagai berikut:

$$P(Y = m|\theta, n) = \text{Binomial}(Y = m|\theta, n) = \binom{n}{m} \theta^m (1-\theta)^{n-m}.$$

- Fungsi likelihood:

$$L(\theta|m, n) = \binom{n}{m} \theta^m (1-\theta)^{n-m},$$

dengan θ diasumsikan mengambil nilai dalam interval $\Theta = [0, 1]$ dan nilai n dan m diasumsikan tetap

Pendekatan Bayesian*

- Pendekatan Bayesian bergantung pada hasil matematika abad ke-18 yang dikenal sebagai “aturan Bayes” atau “teorema Bayes” untuk menghitung distribusi probabilitas atas nilai-nilai yang mungkin dari variabel yang tidak diketahui.
- Distribusi probabilitas ini, yang dikenal sebagai “distribusi posterior”, memberikan kita informasi mengenai probabilitas bahwa variabel yang tidak diketahui mengambil nilai tertentu, bergantung pada asumsi model.



UNIVERSITAS
INDONESIA

Veritas, Probitas, Iustitia
