

Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Visualisasi Grafik Garis, Matrix Scatter Plot dan Koefisien Korelasi

Dibuat untuk memenuhi Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Oleh:

Nurul Hidayah Chairunnisa

NPM: 131520220001



Program Studi Magister Epidemiologi

Fakultas Kedokteran

Universitas Padjadjaran

2023

1. Menggunakan data stroke pada link yang diberikan dan penjelasan datanya, dilakukan akses dan download dataset ke dalam global environment RStudio.

Jawab:

-) Mengaktifkan *package* yang dibutuhkan

```
library(data.table)
```

```
library(tidyverse)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(lme4)
```

-) Melakukan *import* file stroke dari link yang diberikan

```
stroke <- fread("http://www.statsci.org/data/oz/stroke.txt")
```

2. Membuat visualisasi grafik garis dari perkembangan nilai kemampuan motorik (functional ability score) dari setiap subyek menggunakan variabel Bart dengan tampilan sesuai contoh.

Jawab:

-) Untuk membuat visualisasi grafik garis seperti yang dicontohkan, dilakukan perubahan bentuk data dari *wide data* menjadi *long data*.

```
stroke_long = stroke %>% select(c(1:6,39:46)) %>%
```

```
  pivot_longer(cols=Bart1:Bart8, names_to = "Time", names_prefix = "Bart",
```

```
    values_to = "Ability")
```

-) Membuat *dataframe* baru pada *global environment*

```
stroke_graph <- stroke_long %>%
```

```
  mutate(label = if_else(Time == max(Time),
```

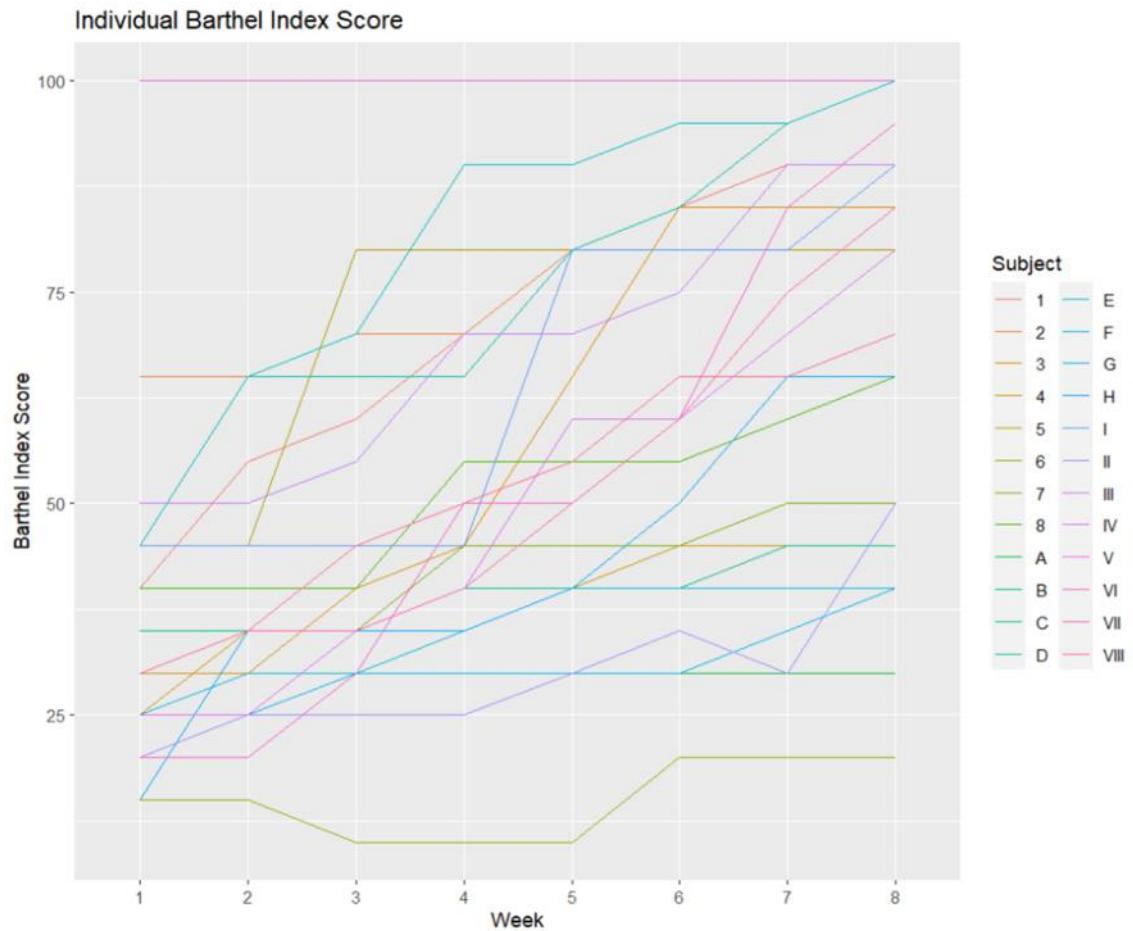
```
    as.character(Subject), NA_character_)) %>%
```

```
  ggplot(aes(x = Time, y = Ability, group = Subject, colour = Subject)) + geom_line()
```

-) Memberikan label pada grafik yang sudah dibuat

```
stroke_graph + labs(title = "Individual Barthel Index Score", x="Week", y="Barthel  
Index Score")
```

Hasil grafik yang didapat:



3. Membuat grafik nilai rata-rata perkembangan fungsi motorik secara total dan masing-masing yang divisualisasikan pada 1 grafik sesuai contoh.

Jawab:

) Melakukan duplikasi *dataframe*

```
stroke_new <- stroke_long
```

) Melakukan penghitungan rerata nilai perkembangan fungsi motorik

```
stroke_new <- stroke_long %>%
```

```
  group_by(Group, Time) %>%
```

```
  mutate(BSI_mean = mean(Ability)) %>%
```

```
  as.data.frame()
```

) Membuat *dataframe* baru pada *global environment*

```
stroke_group <- stroke_new %>%
```

```
  mutate(label = if_else(Time == max(Time), as.character(Group),
```

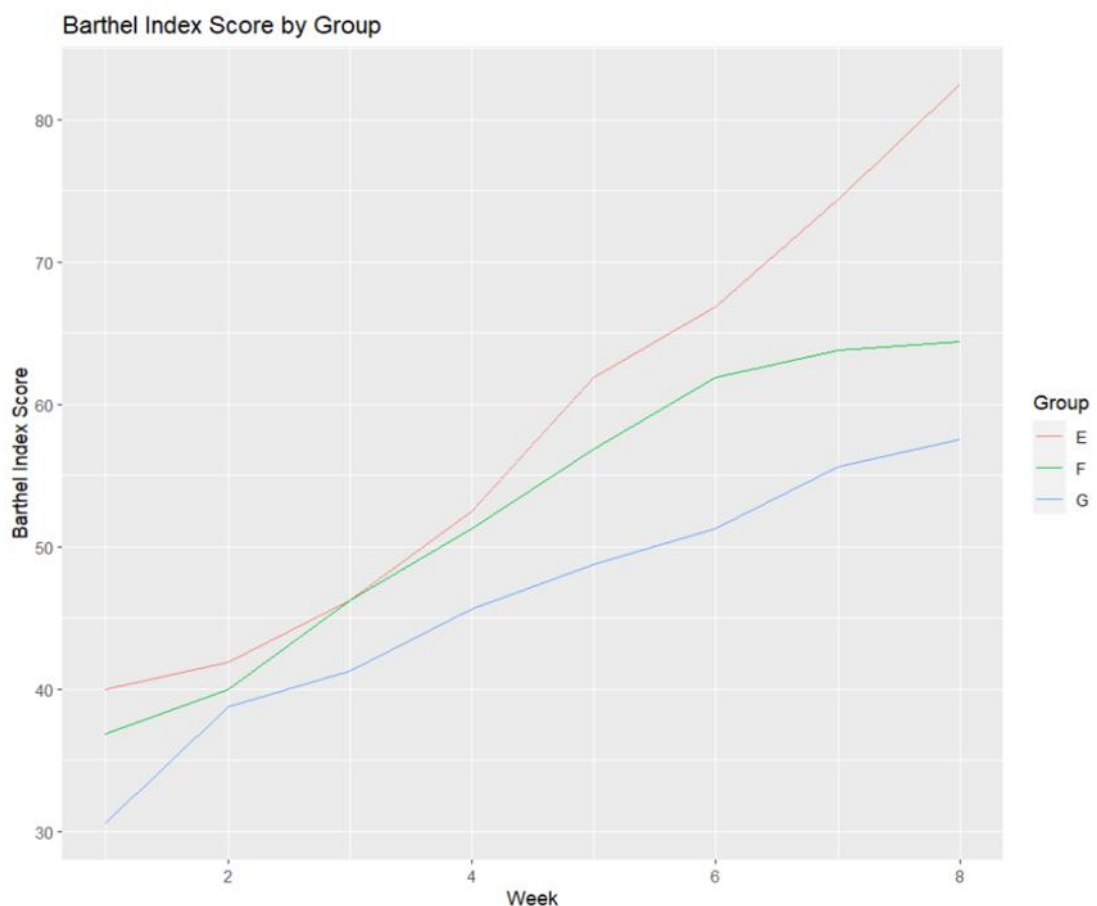
```
    NA_character_)) %>%
```

```
  ggplot(aes(x = Time, y = BSImean, group = Group, colour = Group)) + geom_line()
```

) Memberikan label pada grafik yang sudah dibuat

```
stroke_group + labs(title = "Barthel Index Score by Group", x="Week", y="Barthel  
Index Score")
```

Hasil grafik yang didapat:



4. Membuat Matrix Scatter plot dari nilai fungsi motorik antar waktu/pekan sebagaimana contoh visualisasi.

Jawab:

) Untuk dapat membuat Matrix Scatter plot seperti pada cobtoh, dilakukan transformasi data dari *long data* menjadi *wide data*.

```
stroke_wide = stroke_long %>%
  pivot_wider(names_from = "Time",
    names_prefix = "Week",
    values_from = "Ability")
```

) Membuat Matrix Scatter plot
`pairs(stroke_wide[, 7:14], pch=19)`

Hasil yang didapat:



5. Menghitung dan membuat tabel silang koefisien korelasi nilai fungsi motorik antar waktu/pekan sesuai contoh.

Jawab:

```
cor(stroke_wide$Week1, stroke_wide$Week2)
cor(stroke_wide$Week1, stroke_wide$Week3)
```

cor(stroke_wide\$Week1, stroke_wide\$Week4)
cor(stroke_wide\$Week1, stroke_wide\$Week5)
cor(stroke_wide\$Week1, stroke_wide\$Week6)
cor(stroke_wide\$Week1, stroke_wide\$Week7)
cor(stroke_wide\$Week1, stroke_wide\$Week8)

cor(stroke_wide\$Week2, stroke_wide\$Week3)
cor(stroke_wide\$Week2, stroke_wide\$Week4)
cor(stroke_wide\$Week2, stroke_wide\$Week5)
cor(stroke_wide\$Week2, stroke_wide\$Week6)
cor(stroke_wide\$Week2, stroke_wide\$Week7)
cor(stroke_wide\$Week2, stroke_wide\$Week8)

cor(stroke_wide\$Week3, stroke_wide\$Week4)
cor(stroke_wide\$Week3, stroke_wide\$Week5)
cor(stroke_wide\$Week3, stroke_wide\$Week6)
cor(stroke_wide\$Week3, stroke_wide\$Week7)
cor(stroke_wide\$Week3, stroke_wide\$Week8)

cor(stroke_wide\$Week4, stroke_wide\$Week5)
cor(stroke_wide\$Week4, stroke_wide\$Week6)
cor(stroke_wide\$Week4, stroke_wide\$Week7)
cor(stroke_wide\$Week4, stroke_wide\$Week8)

cor(stroke_wide\$Week5, stroke_wide\$Week6)
cor(stroke_wide\$Week5, stroke_wide\$Week7)
cor(stroke_wide\$Week5, stroke_wide\$Week8)

cor(stroke_wide\$Week6, stroke_wide\$Week7)
cor(stroke_wide\$Week6, stroke_wide\$Week8)

cor(stroke_wide\$Week7, stroke_wide\$Week8)

Hasil yang didapatkan:

		Week						
		1	2	3	4	5	6	8
W e e k	2	0.93						
	3	0.88	0.92					
	4	0.93	0.88	0.95				
	5	0.79	0.85	0.91	0.92			
	6	0.71	0.79	0.85	0.88	0.97		
	7	0.62	0.70	0.77	0.83	0.92	0.96	
	8	0.55	0.64	0.70	0.77	0.88	0.93	0.98

6. Interpretasi dan simpulan dari Matrix Scatter plot dan koefisien korelasi di atas.

Pada matrix scatter plot, gambaran sebaran yang berbentuk garis miring ke kanan atas menunjukkan hubungan linear positif. Semakin miring garisnya, maka hubungannya semakin kuat. Jika dilihat dari hasil di atas, terlihat hubungan antar pekan tergambar sebagai garis miring ke kanan atas, sehingga hubungannya linear positif. Sebaran yang terjadi antara pekan yang berbeda 1 pekan (misal Week 1 pada sumbu y dengan Week 2 pada sumbu x atau Week 3 pada sumbu y dengan Week 4 pada sumbu x) memiliki kemiringan yang lebih curam jika dibandingkan dengan sebaran antara pekan yang berbeda jauh (misal Week 1 pada sumbu y dan Week 5 pada sumbu x). Maka dapat dikatakan bahwa semakin dekat waktu/pekan pengukuran yang terjadi maka hubungannya semakin kuat, begitu juga sebaliknya.

Pada tabel koefisien korelasi, masing-masing angka menunjukkan berapa dekat hubungan variabel yang berbeda. Variabel dapat dilihat pada baris dan kolom, dan angka pada *cell* menunjukkan hubungan antara keduanya. Angka yang positif menunjukkan adanya hubungan yang positif dan sebaliknya. Semakin dekat angkat tersebut dengan angka 1 atau -1 maka semakin kuat hubungannya, dan angka 0 menunjukkan tidak terdapat hubungan. Pada hasil di atas, terlihat bahwa hubungan antar pekan seluruhnya berada pada angka yang positif sehingga disimpulkan terdapat hubungan yang positif. Selain itu dapat pula dilihat bahwa hubungan antara pekan yang berbeda 1 pekan nilai koefisiennya lebih mendekati angka 1 jika dibandingkan dengan pekan yang berbeda jauh. Sehingga dapat disimpulkan bahwa semakin jauh perbedaan waktu pengukuran, maka hubungannya lebih lemah dibandingkan dengan waktu pengukuran yang lebih dekat.

7. Menghitung *intercept* dan *slope*, beserta *standard error* masing-masing, dari hubungan fungsi morotik dengan waktu/pekan setiap subyek, serta mempresentasikan hasilnya dalam bentuk tabel sesuai contoh.

Jawab:

) Menghitung *intercept* dan *slope* dan *standard error* masing-masing
`summary(fit1 <- (lmList(Ability ~ Time | Subject, data = stroke_new)))`

Hasil yang didapatkan:

Subject	Intercept	(std. error)	Slope	(std. error)
1	38.57	4.037	7.26	0.799
2	61.96	4.037	2.62	0.799
3	14.46	4.037	9.70	0.799
4	26.07	4.037	2.68	0.799
5	48.75	4.037	5.00	0.799
6	10.17	4.037	1.07	0.799
7	31.25	4.037	2.50	0.799
8	34.11	4.037	3.81	0.799
9	21.07	4.037	1.43	0.799
10	34.11	4.037	0.89	0.799
11	32.14	4.037	1.61	0.799
12	42.32	4.037	7.26	0.799
13	48.57	4.037	7.26	0.799
14	24.82	4.037	2.26	0.799
15	22.32	4.037	1.85	0.799
16	13.04	4.037	6.55	0.799
17	30.00	4.037	7.50	0.799
18	15.54	4.037	3.21	0.799
19	39.82	4.037	6.43	0.799
20	11.61	4.037	8.39	0.799
21	100	NaN	0	NaN
22	0.89	4.037	11.19	0.799
23	15.36	4.037	7.98	0.799
24	25.36	4.037	5.89	0.799

Hasil tersebut didapatkan setelah memastikan bahwa baik Ability maupun Time merupakan data numerik, dan sudah dicoba menggunakan beberapa formula lainnya seperti `lm()` dan `lmer()`, namun hasil yang didapatkan masih sama. Apabila menambahkan logaritma pada Ability maka hasil intercept dan slope berbeda besaran angkanya, namun standard error yang didapatkan tetap sama pada semua subjek.

Setelah mencari penjelasan mengenai nilai standard error yang persis sama dikatakan bahwa hal tersebut dapat terjadi apabila model yang di digunakan memiliki error variance

yang homogen atau desain yang digunakan sangat balance (misal jumlah observasi yang sama). Selain itu terdapat pula penjelasan bahwa hal tersebut terjadi karena adanya overfitting yang dapat dilihat dari residual standard error sebesar 10^{-12} , namun pada hasil ini, residual standard errornya adalah 5.1812, sehingga penjelasan tersebut mungkin tidak dapat diaplikasikan pada kasus ini.

```
Residual standard error: 5.181279 on 144 degrees of fr  
eedom
```