

Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Disusun oleh:

Syndi Nurmawati NPM 131520220002



Program Studi Magister Epidemiologi
Fakultas Kedokteran
Universitas Padjadjaran
2023

1. Mengakses dataset ke dalam *global environment*

Sumber dataset: <http://www.statsci.org/data/oz/stroke.txt>

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

Hasil:

Terdapat dataset stroke dalam *global environment*. Dataset terdiri dari 24 observasi dari 46 variabel

▶ stroke | 24 obs. of 46 variables

2. Membuat visualisasi grafik garis dari perkembangan nilai kemampuan motorik (*functional ability score*) dari setiap subjek menggunakan variabel Bart

Data berbentuk wide sehingga harus diubah menjadi bentuk long.

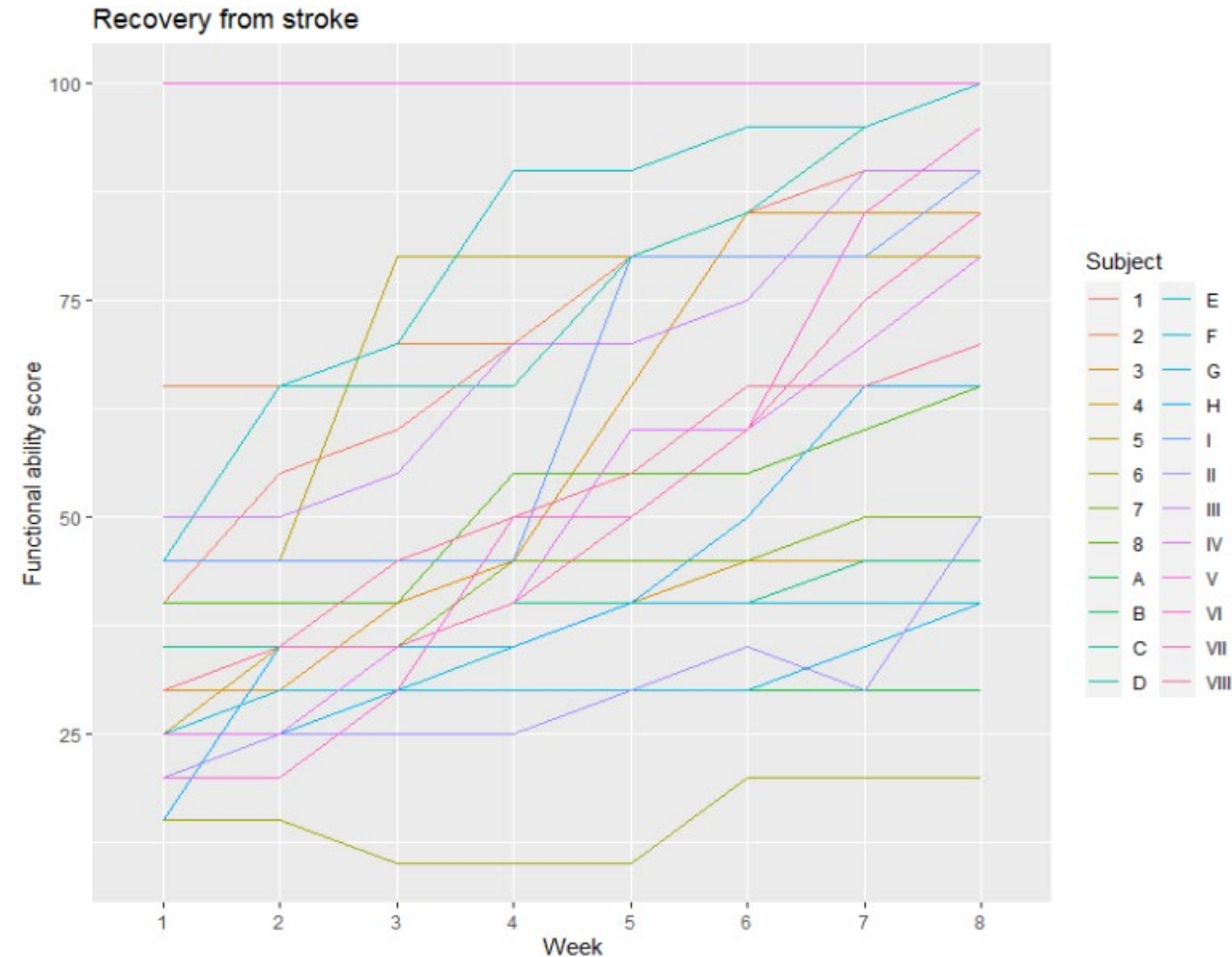
```
Kode: stroke_long = stroke %>% select(c(1:6,39:46)) %>%  
pivot_longer(cols=Bart1:Bart8, names_to = "Time",  
names_prefix = "Bart", values_to = "Ability")
```

Hasil:

```
stroke_long | 192 obs. of 8 variables
```

Setelah itu dibuat grafik garis seperti di sebelah kanan.

```
Kode: ggplot(data = stroke_long, aes(x = Time, y = Ability,  
group = Subject, color = Subject)) + geom_line()
```



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

- Dibuat mean untuk masing-masing grup berdasarkan time

Kode: `stroke_mean <- stroke_long %>% group_by (Group, Time) %>% summarise(mean_val = mean(Ability))`

Hasil: terbentuk objek baru yaitu `stroke_mean` yang terdiri dari 24 observasi dari 3 variabel dengan perincian seperti tampilan di sebelah kanan.

▶ `stroke_mean` | 24 obs. of 3 variables

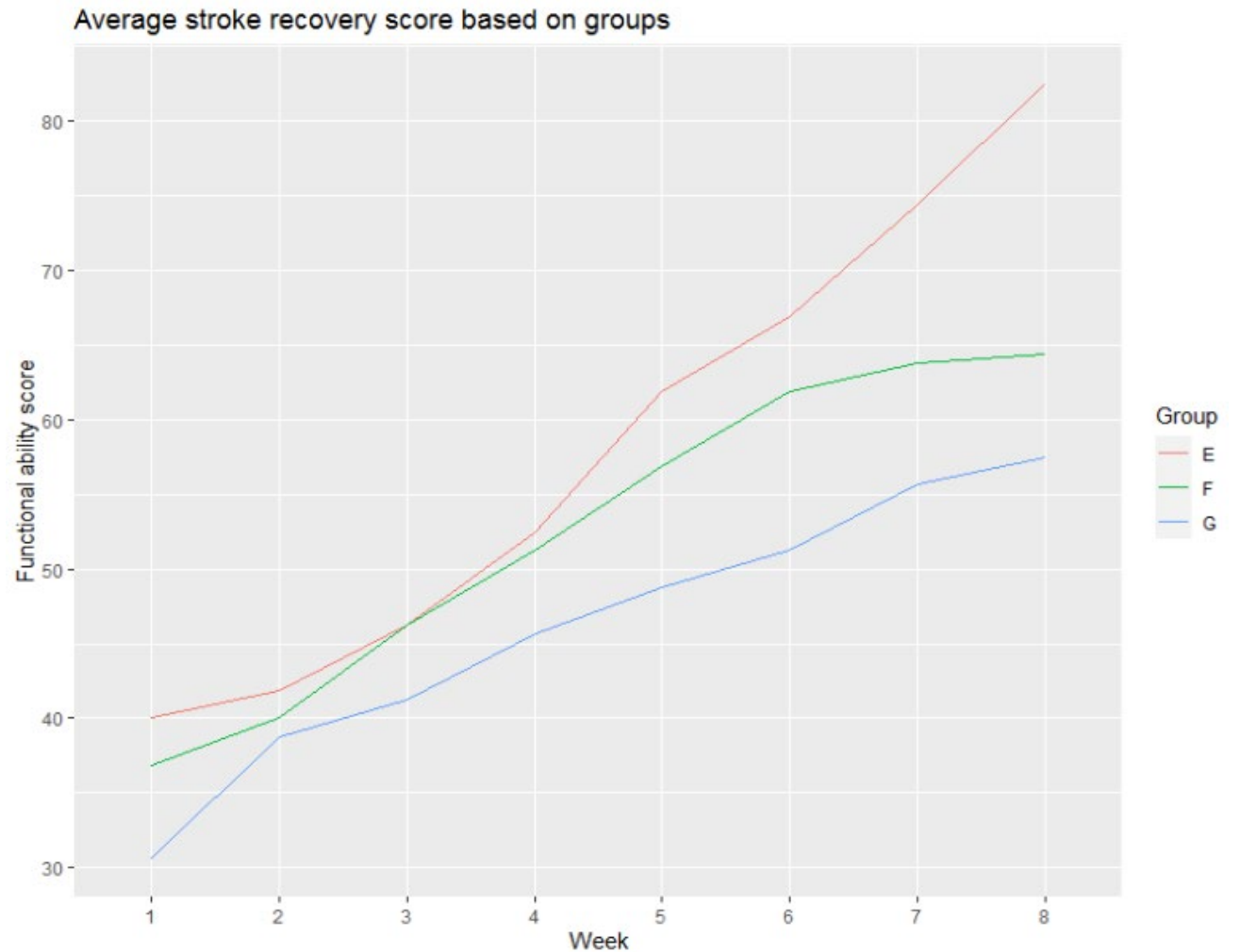
	Group	Time	mean_val
1	E	1	40.000
2	E	2	41.875
3	E	3	46.250
4	E	4	52.500
5	E	5	61.875
6	E	6	66.875
7	E	7	74.375
8	E	8	82.500
9	F	1	36.875
10	F	2	40.000
11	F	3	46.250
12	F	4	51.250
13	F	5	56.875
14	F	6	61.875
15	F	7	63.750
16	F	8	64.375
17	G	1	30.625
18	G	2	38.750
19	G	3	41.250
20	G	4	45.625
21	G	5	48.750
22	G	6	51.250
23	G	7	55.625
24	G	8	57.500

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

- Setelah dibuat mean, baru dibuat grafik per grup

Kode: `ggplot(data = stroke_mean, aes(x = Time, y = mean_val, group = Group, color = Group)) +
geom_line() + labs(title = "Average stroke recovery
score based on groups",
x = "Week", y =
"Functional ability score")`

Hasil seperti lampiran di sebelah kanan.

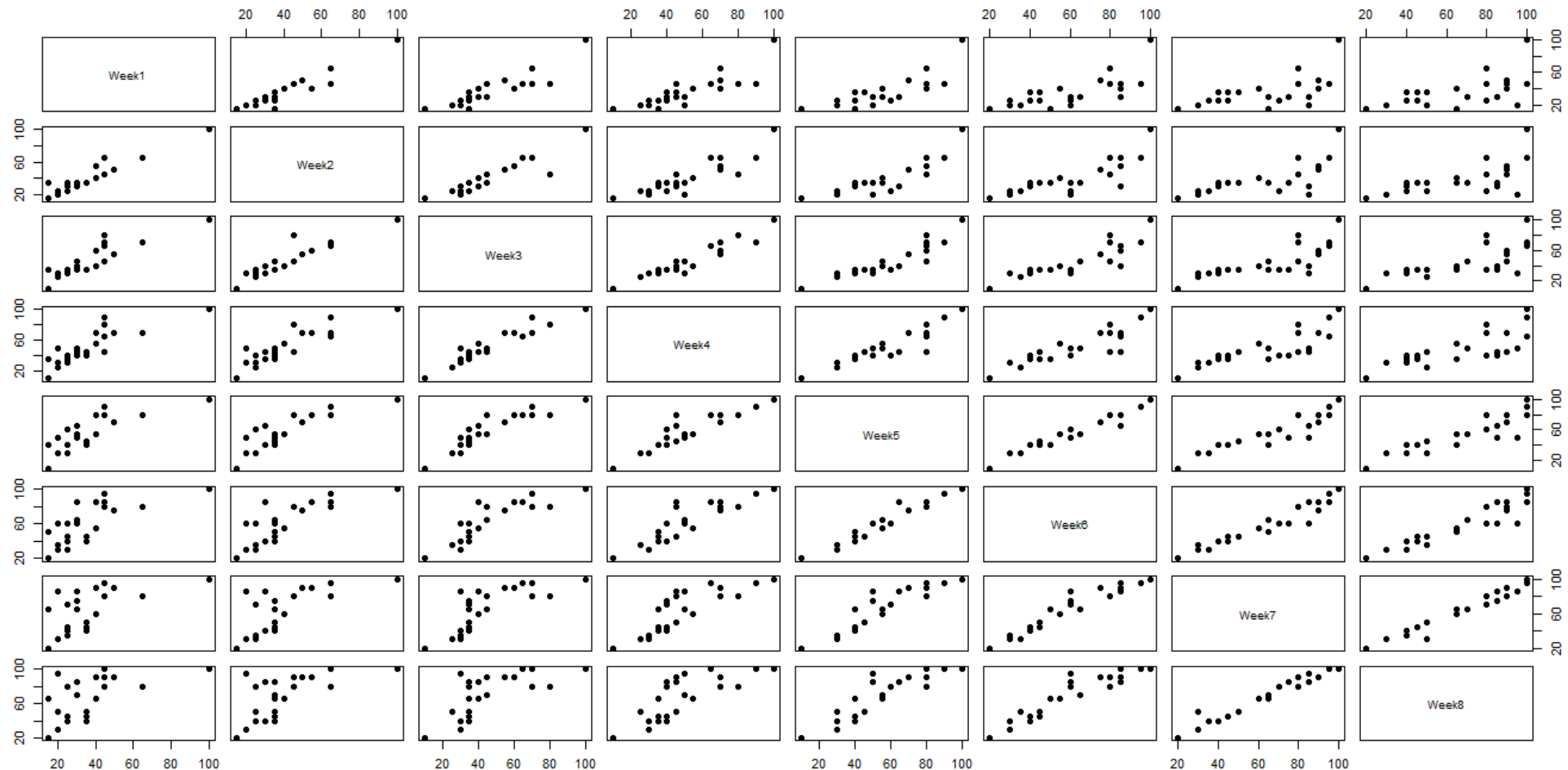


4. Membuat *matrix scatter plot* dari nilai fungsi motorik antar waktu/pekan

- Matrix scatterplot dibuat dengan mempergunakan data wide (data awal)
- Dilakukan perubahan variabel dari integer menjadi numerik pada variabel Bart1 – Bart8
- Variabel Bart1 – Bart8 di-rename menjadi Week1 – Week8 untuk memudahkan pembacaan
- Setelah itu dibuat matrix scatterplot

Kode: `pairs(stroke_wide[,Week1:Week8],pch = 19)`

Hasil terlihat di bawah.



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

- Correlation matrix dibuat dari data wide

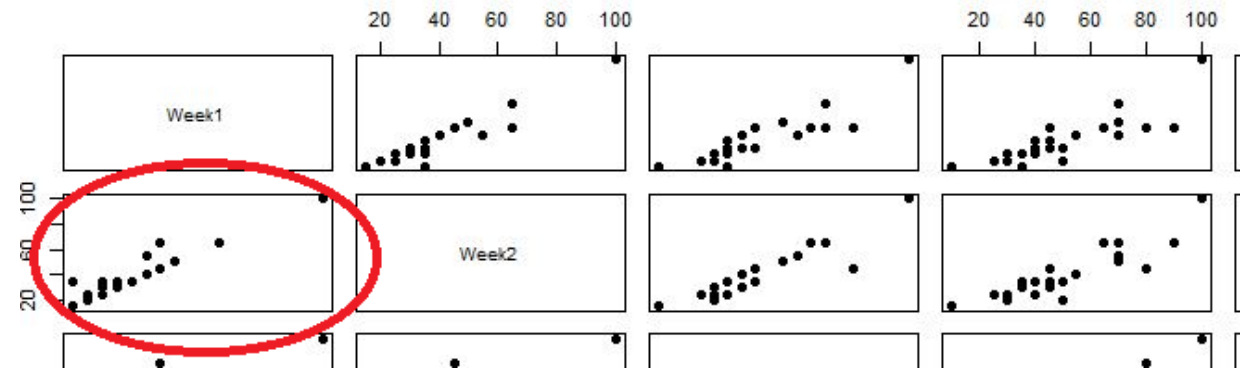
Kode: `res <- cor(stroke_3)`

Hasil tampak seperti tabel di sebelah kanan

	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8
Week1	1.0000000	0.9280364	0.8820153	0.8306544	0.7936646	0.7125630	0.6163514	0.5544246
Week2	0.9280364	1.0000000	0.9225559	0.8774061	0.8466833	0.7895900	0.7041487	0.6425992
Week3	0.8820153	0.9225559	1.0000000	0.9530914	0.9092148	0.8542616	0.7667288	0.7007907
Week4	0.8306544	0.8774061	0.9530914	1.0000000	0.9215159	0.8786341	0.8313352	0.7716004
Week5	0.7936646	0.8466833	0.9092148	0.9215159	1.0000000	0.9734304	0.9149511	0.8819552
Week6	0.7125630	0.7895900	0.8542616	0.8786341	0.9734304	1.0000000	0.9569344	0.9266933
Week7	0.6163514	0.7041487	0.7667288	0.8313352	0.9149511	0.9569344	1.0000000	0.9776126
Week8	0.5544246	0.6425992	0.7007907	0.7716004	0.8819552	0.9266933	0.9776126	1.0000000

6. Interpretasi dan simpulkan *matrix scatter plot* dan koefisien korelasi tersebut

Scatterplot matrix dan *correlation matrix* menampilkan hubungan antara beberapa variabel, dalam hal ini antara Week1 hingga Week8. Pada *scatterplot matrix* kotak pertama baris kedua (lingkaran merah) menunjukkan hubungan antara Week1 (x) dan Week2 (y) dimana diduga terdapat hubungan karena plot membentuk garis lurus. Hubungan tersebut dikonfirmasi dengan hasil perhitungan koefisien korelasi pada *correlation matrix* kotak paling kiri kedua dari atas dimana didapatkan koefisien korelasi 0,93 (lingkaran biru).

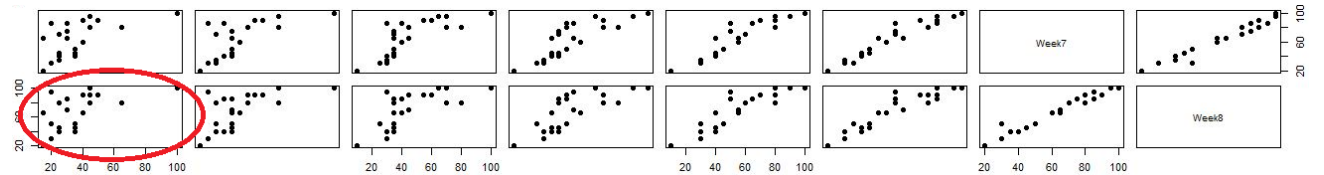


	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8
Week1	1.0000000	0.9280364	0.8820153	0.8306544	0.7936646	0.7125630	0.6163514	0.5544246
Week2	0.9280364	1.0000000	0.9225559	0.8774061	0.8466833	0.7895900	0.7041487	0.6425992
Week3	0.8820153	0.9225559	1.0000000	0.9530914	0.9092148	0.8542616	0.7667288	0.7007907
Week4	0.8306544	0.8774061	0.9530914	1.0000000	0.9215159	0.8786341	0.8313352	0.7716004
Week5	0.7936646	0.8466833	0.9092148	0.9215159	1.0000000	0.9734304	0.9149511	0.8819552
Week6	0.7125630	0.7895900	0.8542616	0.8786341	0.9734304	1.0000000	0.9569344	0.9266933
Week7	0.6163514	0.7041487	0.7667288	0.8313352	0.9149511	0.9569344	1.0000000	0.9776126
Week8	0.5544246	0.6425992	0.7007907	0.7716004	0.8819552	0.9266933	0.9776126	1.0000000

6. Interpretasi dan simpulkan *matrix scatter plot* dan koefisien korelasi tersebut

Hasil ini berbeda dengan kotak pojok kiri bawah pada *scatterplot matrix* yang menunjukkan hubungan antara Week1 (x) dan Week8 (y) dimana plot tidak membentuk garis lurus (lingkaran merah). Hasil *correlation matrix* menunjukkan bahwa didapatkan koefisien korelasi sebesar 0,55 (lingkaran biru), lebih kecil dibandingkan koefisien korelasi antara Week1 dan Week2.

Dari hasil ini terlihat bahwa semakin jauh jarak antar minggu, maka semakin kecil koefisien korelasinya.



	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8
Week1	1.0000000	0.9280364	0.8820153	0.8306544	0.7936646	0.7125630	0.6163514	0.5544246
Week2	0.9280364	1.0000000	0.9225559	0.8774061	0.8466833	0.7895900	0.7041487	0.6425992
Week3	0.8820153	0.9225559	1.0000000	0.9530914	0.9092148	0.8542616	0.7667288	0.7007907
Week4	0.8306544	0.8774061	0.9530914	1.0000000	0.9215159	0.8786341	0.8313352	0.7716004
Week5	0.7936646	0.8466833	0.9092148	0.9215159	1.0000000	0.9734304	0.9149511	0.8819552
Week6	0.7125630	0.7895900	0.8542616	0.8786341	0.9734304	1.0000000	0.9569344	0.9266933
Week7	0.6163514	0.7041487	0.7667288	0.8313352	0.9149511	0.9569344	1.0000000	0.9776126
Week8	0.5544246	0.6425992	0.7007907	0.7716004	0.8819552	0.9266933	0.9776126	1.0000000

7. Menghitung *intercept* dan *slope*, beserta *Standard Error*-nya masing-masing dari hubungan fungsi motorik dengan waktu/pekan setiap subjek serta mempresentasikan hasilnya dalam bentuk tabel

- Dilakukan transformasi menjadi variabel numerik
- Dilakukan perhitungan intercept, slope, dan standard error

Kode: `fm1 <- lmList(Ability ~ Time1 | Subject, data = stroke_long)`

Hasil untuk Intercept ada di sebelah kiri dan hasil untuk slope ada di sebelah kanan.

, , (Intercept)

	Estimate	Std. Error	t value	Pr(> t)
1	38.5714286	4.03722	9.5539575	4.691036e-17
2	61.9642857	4.03722	15.3482558	4.139627e-32
3	14.4642857	4.03722	3.5827341	4.643985e-04
4	26.0714286	4.03722	6.4577676	1.526780e-09
5	48.7500000	4.03722	12.0751407	1.254066e-23
6	10.1785714	4.03722	2.5211832	1.278521e-02
7	31.2500000	4.03722	7.7404748	1.595320e-12
8	34.1071429	4.03722	8.4481754	2.941363e-14
A	21.0714286	4.03722	5.2192916	6.152007e-07
B	34.1071429	4.03722	8.4481754	2.941363e-14
C	32.1428571	4.03722	7.9616313	4.641098e-13
D	42.3214286	4.03722	10.4828145	1.852355e-19
E	48.5714286	4.03722	12.0309095	1.638129e-23
F	24.8214286	4.03722	6.1481486	7.307372e-09
G	22.3214286	4.03722	5.5289106	1.471632e-07
H	13.0357143	4.03722	3.2288838	1.539380e-03
I	30.0000000	4.03722	7.4308558	8.783934e-12
II	15.5357143	4.03722	3.8481218	1.783285e-04
III	39.8214286	4.03722	9.8635765	7.485004e-18
IV	11.6071429	4.03722	2.8750335	4.653108e-03
V	100.0000000	NaN	NaN	NaN
VI	0.8928571	4.03722	0.2211564	8.252839e-01
VII	15.3571429	4.03722	3.8038905	2.098796e-04
VIII	25.3571429	4.03722	6.2808424	3.753902e-09

, , Time1

	Estimate	Std. Error	t value	Pr(> t)
1	7.2619048	0.7994887	9.083186	7.462018e-16
2	2.6190476	0.7994887	3.275903	1.319663e-03
3	9.7023810	0.7994887	12.135732	8.697407e-24
4	2.6785714	0.7994887	3.350356	1.030670e-03
5	5.0000000	0.7994887	6.253997	4.298044e-09
6	1.0714286	0.7994887	1.340142	1.823099e-01
7	2.5000000	0.7994887	3.126999	2.137175e-03
8	3.8095238	0.7994887	4.764950	4.566548e-06
A	1.4285714	0.7994887	1.786856	7.606449e-02
B	0.8928571	0.7994887	1.116785	2.659461e-01
C	1.6071429	0.7994887	2.010213	4.627610e-02
D	7.2619048	0.7994887	9.083186	7.462018e-16
E	7.2619048	0.7994887	9.083186	7.462018e-16
F	2.2619048	0.7994887	2.829189	5.333069e-03
G	1.8452381	0.7994887	2.308023	2.242213e-02
H	6.5476190	0.7994887	8.189758	1.281897e-13
I	7.5000000	0.7994887	9.380996	1.301105e-16
II	3.2142857	0.7994887	4.020427	9.334977e-05
III	6.4285714	0.7994887	8.040853	2.973049e-13
IV	8.3928571	0.7994887	10.497781	1.693303e-19
V	0.0000000	NaN	NaN	NaN
VI	11.1904762	0.7994887	13.997041	1.209852e-28
VII	7.9761905	0.7994887	9.976614	3.819924e-18
VIII	5.8928571	0.7994887	7.370782	1.219004e-11

- Pada tabel di samping terlihat intercept dan slope untuk masing-masing subjek. Namun standard error untuk setiap subjek, baik standard error intercept maupun standard error slope. Saya sudah mencoba untuk mencari alasan mengapa hal ini terjadi dan mengubah rumus. Saya juga sudah mencoba dengan melakukan transformasi data dan *centering*. Dari seluruh hal yang dapat saya kerjakan, saya masih tetap belum dapat menemukan alasan mengapa *standard error* tersebut sama untuk semua subjek. Setelah saya lakukan pencarian via internet, ada yang mengatakan bahwa *standard error* yang sama merupakan tanda bahwa data *perfect fit* dengan model. Namun masih ada kemungkinan bahwa saya melakukan kesalahan pada saat pemrosesan data.
- Untuk subjek V, slope 0 karena tidak terdapat perubahan *functional ability score* pada setiap kali kontrol.

Subject	Intercept	Standard Error	Slope	Standard Error
1	38.57	4.04	7.26	0.799
2	61.96	4.04	2.62	0.799
3	14.46	4.04	9.70	0.799
4	26.07	4.04	2.68	0.799
5	48.75	4.04	5.00	0.799
6	10.18	4.04	1.07	0.799
7	31.25	4.04	2.50	0.799
8	34.11	4.04	3.81	0.799
A	21.07	4.04	1.43	0.799
B	34.11	4.04	0.89	0.799
C	32.14	4.04	1.61	0.799
D	42.32	4.04	7.26	0.799
E	48.57	4.04	7.26	0.799
F	24.82	4.04	2.26	0.799
G	22.32	4.04	1.84	0.799
H	13.03	4.04	6.55	0.799
I	30.00	4.04	7.50	0.799
II	15.53	4.04	3.21	0.799
III	39.82	4.04	6.43	0.799
IV	11.61	4.04	8.39	0.799
V	100.0	NaN	0	NaN
VI	0.89	4.04	11.19	0.799
VII	15.36	4.04	7.98	0.799
VIII	25.36	4.04	5.89	0.799