

Final Examination Manajemen dan Analisis Data dengan R

Sri Hartati

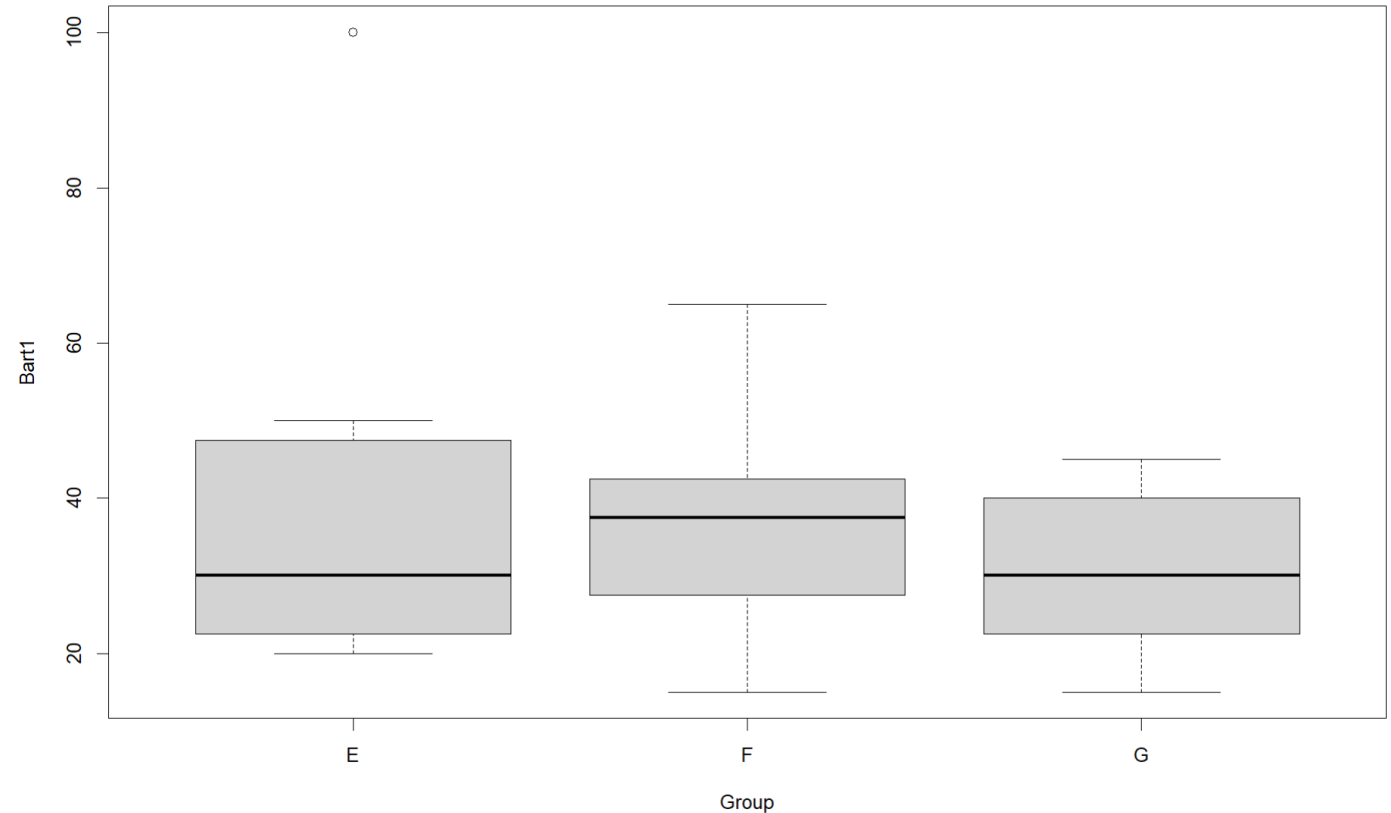
131520220006

Program Studi Magister Epidemiologi
Universitas Padjadjaran
2023

1. Boxplot dari functional ability pada pekan pertama (Bart1) berdasarkan grup intervensi

Interpretasi:

- Subjek di Grup F mempunyai nilai median skor Bart1 dan nilai maksimum skor Bart1 yang lebih tinggi dibandingkan dengan dua grup yang lainnya.
- Grup E memiliki *interquartile range* yang lebih lebar, dengan pola positively skewed (nilai median lebih dekat ke Q1).
- Grup F dan G memiliki nilai minimum yang sama, yaitu 15.
- Secara umum, nilai Bart1 di ketiga grup memiliki pola distribusi yang berbeda (dilihat dari nilai median dan distribusi sebarannya (*spread*)).



2. Memeriksa normalitas data dari functional ability pekan pertama (Bart1) dengan uji statistik

Hipotesis untuk menguji normalitas:

- H_0 : Data terdistribusi normal.
- Karena jumlah sampel yang kecil maka digunakan fungsi shapiro.test untuk menguji normalitas.

```
> shapiro.test(stroke$Bart1)

      Shapiro-Wilk normality test

data:  stroke$Bart1
W = 0.82449, p-value = 0.0007617
```

Interpretasi:

- Uji normalitas menunjukkan nilai p-value < 0.05 berarti terdapat cukup bukti untuk menolak null hypothesis yang menyatakan data terdistribusi normal.
- Maka data pada variabel *Bart1* tidak terdistribusi secara normal.

3. Menghitung perubahan functional ability dari pekan pertama (Bart1) sampai pekan terakhir (Bart8) dan membuatnya menjadi variabel baru (Bart_diff)

```
> #NOMOR 3
> #Membuat variabel Bart_diff
> stroke$Bart_diff <- stroke$Bart8 - stroke$Bart1
> str(stroke$Bart_diff)
int [1:24] 45 30 40 55 0 75 55 40 50 15 ...
```

4. Memeriksa normalitas dari variabel Bart_diff dengan uji statistik yang sesuai dan interpretasinya

Hipotesis untuk menguji normalitas:

- H_0 : Data terdistribusi normal.
- Karena jumlah sampel yang kecil maka digunakan fungsi shapiro.test untuk menguji normalitas.

```
> shapiro.test(stroke$Bart_diff)

      Shapiro-Wilk normality test

data:  stroke$Bart_diff
W = 0.92817, p-value = 0.08875
```

Interpretasi:

- Uji normalitas menunjukkan nilai p-value >0.05 berarti tidak terdapat cukup bukti untuk menolak *null hypothesis* yang menyatakan data terdistribusi normal.
- Maka data pada variabel Bart_diff terdistribusi normal.

5. Memeriksa kesamaan varians dari variabel Bart_diff di antara grup intervensi dengan uji statistik yang sesuai dan interpretasinya

Hipotesis untuk menguji kesamaan varians:

- H_0 : Tidak terdapat perbedaan varians di antara populasi
- H_A : Terdapat perbedaan varians di antara populasi (setidaknya dua grup memiliki perbedaan)
- Bartlett's test untuk menguji kesamaan varians pada data yang terdistribusi normal

```
> bartlett.test(Bart_diff ~ Group, data=stroke)

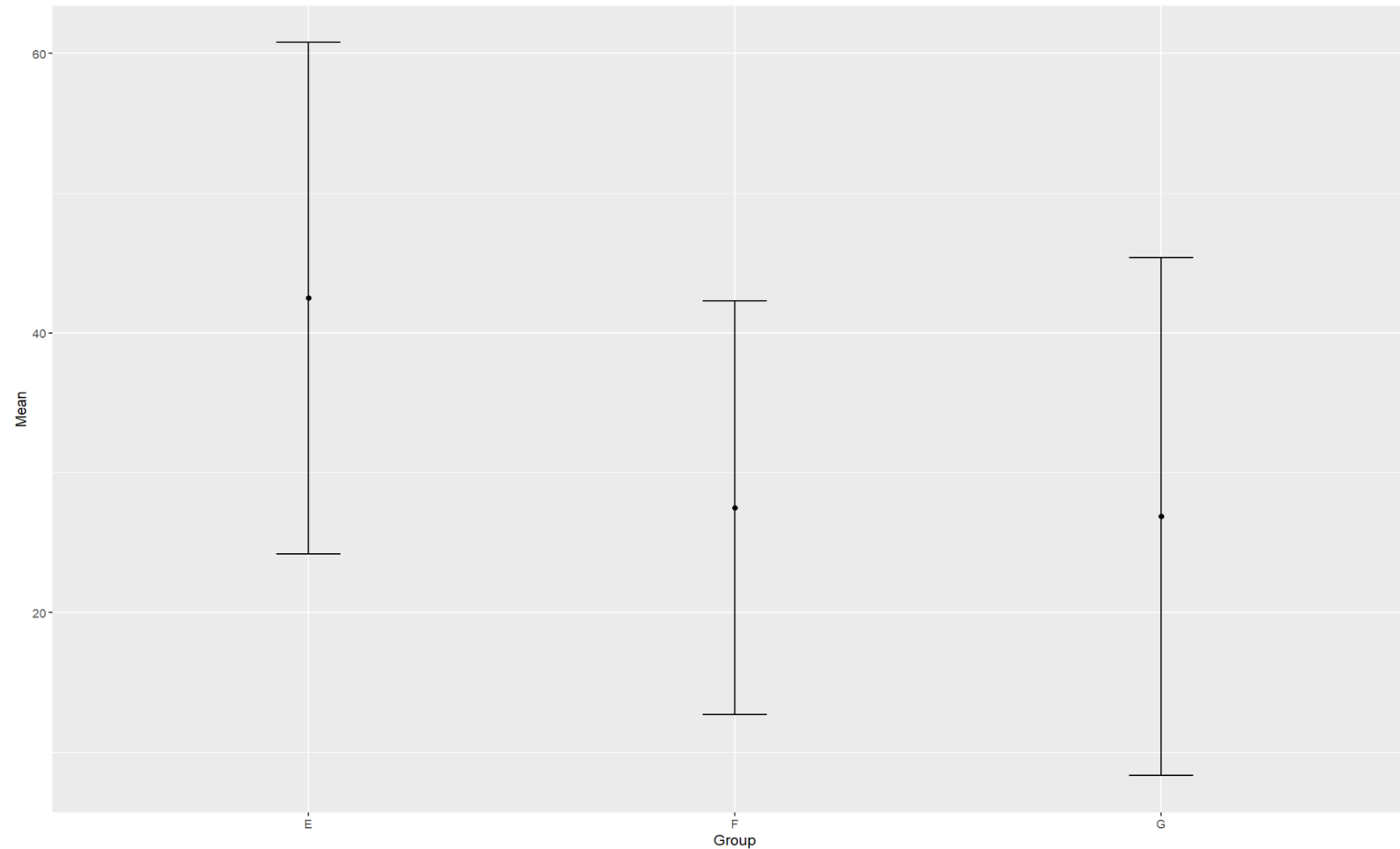
Bartlett test of homogeneity of variances

data:  Bart_diff by Group
Bartlett's K-squared = 0.39433, df = 2, p-value = 0.8211
```

Interpretasi:

- Nilai p-value dari tes ini adalah **0.8211**, menunjukkan bahwa tidak terdapat cukup bukti untuk menolak null hypothesis, maka variabel Bart_diff memiliki kesamaan varians.

6. Membuat plot mean dan 95% CI dari variabel Bart_diff berdasarkan grup intervensi dalam 1 grafik.



Plot dari mean variable Bart_diff berdasarkan grup intervensi, dengan nilai 95% confidence interval nya

7. Melakukan uji anova untuk membandingkan rata-rata (mean) dari variabel Bart_diff di antara tiga grup intervensi dan menginterpretasikannya.

Hipotesis untuk *Repeated Measures ANOVA*:

- H_0 : tidak terdapat perbedaan nilai rata-rata (mean) dari tiga grup yang diuji:
 $H_0: \mu_1 = \mu_2 = \mu_3$; (μ = nilai rata-rata grup)
- H_A : terdapat perbedaan nilai rata-rata (mean) dari tiga grup yang diuji (setidaknya di antara dua grup)

```
> #Melakukan uji anova
> res.aov <- aov(Bart_diff ~ Group, data = stroke)
> summary(res.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	2	1252	626.0	1.461	0.255
Residuals	21	8997	428.4		

Interpretasi:

- Nilai p-value > 0.05 menunjukkan bahwa tidak terdapat cukup bukti untuk menolak *null hypothesis*, sehingga dapat disimpulkan bahwa tidak terdapat perbedaan di antara nilai rata-rata Bart-diff di antara tiga grup yang diuji.

8. Melakukan analisis model linear regresi dengan Functional Ability (Bartlet) sebagai outcome(y) dan explanatory variables meliputi: waktu (time), grup intervensi (group), dan interaksi waktu dan grup intervensi.

```
> mod_lm1 <- lm(ability ~ time + Group + time*Group, data = stroke_long)
> summary(mod_lm1)

Call:
lm(formula = ability ~ time + Group + time * Group, data = stroke_long)

Residuals:
    Min       1Q   Median       3Q      Max
-46.875 -12.500  -5.625   14.531   60.000

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    40.000     7.768   5.150 7.24e-07 ***
time2           1.875     10.985   0.171 0.864675
time3           6.250     10.985   0.569 0.570145
time4          12.500     10.985   1.138 0.256777
time5          21.875     10.985   1.991 0.048063 *
time6          26.875     10.985   2.447 0.015455 *
time7          34.375     10.985   3.129 0.002067 **
time8          42.500     10.985   3.869 0.000156 ***
GroupF          -3.125     10.985  -0.284 0.776394
GroupG          -9.375     10.985  -0.853 0.394633
time2:GroupF     1.250     15.535   0.080 0.935965
time3:GroupF     3.125     15.535   0.201 0.840819
time4:GroupF     1.875     15.535   0.121 0.904078
time5:GroupF    -1.875     15.535  -0.121 0.904078
time6:GroupF    -1.875     15.535  -0.121 0.904078
time7:GroupF    -7.500     15.535  -0.483 0.629883
time8:GroupF   -15.000     15.535  -0.966 0.335656
time2:GroupG     6.250     15.535   0.402 0.687965
time3:GroupG     4.375     15.535   0.282 0.778582
time4:GroupG     2.500     15.535   0.161 0.872346
time5:GroupG    -3.750     15.535  -0.241 0.809549
time6:GroupG    -6.250     15.535  -0.402 0.687965
time7:GroupG    -9.375     15.535  -0.603 0.547010
time8:GroupG   -15.625     15.535  -1.006 0.315966
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.97 on 168 degrees of freedom
Multiple R-squared:  0.2668,    Adjusted R-squared:  0.1664
F-statistic: 2.658 on 23 and 168 DF,  p-value: 0.0001788
```

9. Melakukan analisis model linear regresi dengan Functional Ability (Bartlet) sebagai outcome(y) dan explanatory variables meliputi: waktu (time), grup intervensi (group).

```
> mod_lm2 <- lm(ability ~ time + Group, data = stroke_long)
> summary(mod_lm2)

Call:
lm(formula = ability ~ time + Group, data = stroke_long)

Residuals:
    Min       1Q   Median       3Q      Max
-48.411 -13.854  -4.635   15.130   58.255

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    41.745     4.872   8.568 4.40e-15 ***
time2           4.375     6.163   0.710 0.478667
time3           8.750     6.163   1.420 0.157369
time4          13.958     6.163   2.265 0.024694 *
time5          20.000     6.163   3.245 0.001396 **
time6          24.167     6.163   3.921 0.000125 ***
time7          28.750     6.163   4.665 5.96e-06 ***
time8          32.292     6.163   5.240 4.42e-07 ***
GroupF         -5.625     3.774  -1.491 0.137822
GroupG        -12.109     3.774  -3.209 0.001576 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.35 on 182 degrees of freedom
Multiple R-squared:  0.25,    Adjusted R-squared:  0.2129
F-statistic: 6.741 on 9 and 182 DF,  p-value: 2.522e-08
```

10. Menghitung AIC model no 8 dan 9, serta menginterpretasikan perbandingan nilai AIC nya.

- Hasil perhitungan AIC dari kedua model:

```
> #menghitung AIC dari setiap model
> aictab(cand.set = models, modnames = mod.names)
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcwt	Cum.Wt	LL
model2_tanpa interaksi	11	1733.48	0.00	1	1	-855.01
model1_interaksi	25	1763.50	30.02	0	1	-852.84

Interpretasi:

- Model dengan nilai AIC yang lebih kecil menunjukkan model yang lebih fit, yaitu model 2 tanpa penambahan interaksi di dalam model.

11. Manakah yang terbaik di antara model 8 dan 9? Pilih salah satu kemudian interpretasikan hasil dari analisisnya dari model yang dipilih (hubungan antara variable explanatory dengan outcome).

- Model dengan nilai AIC yang lebih kecil menunjukkan model yang lebih fit, yaitu model 2 tanpa penambahan interaksi di dalam model.

Interpretasi:

- Intercept** = Nilai mean skor Barthel pada subjek di Group E di minggu pertama adalah 41.745.*
- Perbedaan nilai mean skor Barthel pada minggu kedua dan minggu pertama adalah 4.375.
- Perbedaan nilai mean skor Barthel pada minggu ketiga dan minggu pertama adalah 8.750.
- Perbedaan nilai mean skor Barthel pada minggu keempat dan minggu pertama adalah 13.958.*
- Perbedaan nilai mean skor Barthel pada minggu kelima dan minggu pertama adalah 20.00.*
- Perbedaan nilai mean skor Barthel pada minggu keenam dan minggu pertama adalah 24.167.*
- Perbedaan nilai mean skor Barthel pada minggu ketujuh dan minggu pertama adalah 28.750.*
- Perbedaan nilai mean skor Barthel pada minggu kedelapan dan minggu pertama adalah 32.292.*
- Group F** = Subjek pada Grup F memiliki nilai mean skor Barthel 5.625 lebih rendah dibandingkan dengan subjek pada Grup E.
- Group G** = Subjek pada Grup G memiliki nilai mean skor Barthel 12.109 lebih rendah dibandingkan dengan subjek pada Grup E.*

Model:

Skor Barthel = $41.745 + 4.375(\text{time2}) + 8.750(\text{time3}) + 13.958(\text{time4}) + 20(\text{time5}) + 24.167(\text{time6}) + 28.750(\text{time7}) + 32.292(\text{time8}) - 5.625(\text{groupF}) - 12.109(\text{groupG})$

* Nilai estimate dengan signifikan p-value

```
> mod_lm2 <- lm(ability ~ time + Group, data = stroke_long)
> summary(mod_lm2)

Call:
lm(formula = ability ~ time + Group, data = stroke_long)

Residuals:
    Min       1Q   Median       3Q      Max
-48.411 -13.854  -4.635   15.130   58.255

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   41.745      4.872    8.568 4.40e-15 ***
time2          4.375      6.163     0.710 0.478667
time3          8.750      6.163     1.420 0.157369
time4         13.958      6.163     2.265 0.024694 *
time5         20.000      6.163     3.245 0.001396 **
time6         24.167      6.163     3.921 0.000125 ***
time7         28.750      6.163     4.665 5.96e-06 ***
time8         32.292      6.163     5.240 4.42e-07 ***
GroupF        -5.625      3.774    -1.491 0.137822
GroupG       -12.109      3.774    -3.209 0.001576 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.35 on 182 degrees of freedom
Multiple R-squared:  0.25,    Adjusted R-squared:  0.2129
F-statistic: 6.741 on 9 and 182 DF,  p-value: 2.522e-08
```

12. Melakukan analisis mixed model (random intercept) menggunakan package nlme. Functional Ability (Bartlet) sebagai outcome(y) dan explanatory variables meliputi: waktu(time/week), grup intervensi (group), dan Random intercept.

```
> mm1 <- lmer(ability ~ time + Group + (1 | Subject) , data = stroke_long)
> summary(mm1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: ability ~ time + Group + (1 | Subject)
Data: stroke_long

REML criterion at convergence: 1430.5

Scaled residuals:
    Min       1Q   Median       3Q      Max
-2.11630 -0.65706  0.01248  0.63196  3.05608

Random effects:
Groups   Name             Variance Std.Dev.
Subject (Intercept)  403.96     20.099
Residual                82.86      9.103
Number of obs: 192, groups: Subject, 24

Fixed effects:
              Estimate Std. Error    df t value Pr(>|t|)
(Intercept)   41.745      7.403  23.511   5.639 8.97e-06 ***
time2          4.375      2.628 161.000   1.665  0.09788 .
time3          8.750      2.628 161.000   3.330  0.00108 **
time4         13.958      2.628 161.000   5.312 3.56e-07 ***
time5         20.000      2.628 161.000   7.611 2.16e-12 ***
time6         24.167      2.628 161.000   9.197 < 2e-16 ***
time7         28.750      2.628 161.000  10.941 < 2e-16 ***
time8         32.292      2.628 161.000  12.289 < 2e-16 ***
GroupF        -5.625     10.177  21.000  -0.553  0.58631
GroupG       -12.109     10.177  21.000  -1.190  0.24739
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

13. Melakukan ulang analisis dengan functional ability (Bartlett) sebagai outcome (y) dan explanatory variables meliputi: waktu (time/week), grup intervensi (group) dengan General Estimating Equation (GEE) dengan correlation structure: exchangeable, auto regressive, unstructured.

```
library(geepack)
#exchangeable
gee.exc<-geeglm(ability ~ as.factor(Group) + as.numeric(time), family=gaussian,
               data = stroke_long, id = as.factor(Subject), wave = as.numeric(time), corst="exchangeable")
summary(gee.exc)

#unstructured
gee.un<-geeglm(ability ~ as.factor(Group) + as.numeric(time), family=gaussian,
               data = stroke_long, id = as.factor(Subject), wave = as.numeric(time), corst="unstructured")
summary(gee.un)

#auto regressive
gee.ar1<-geeglm(ability ~ as.factor(Group) + as.numeric(time), family=gaussian,
                data=stroke_long, id=as.factor(Subject), wave=as.numeric(time), corst="ar1")
summary(gee.ar1)

library(nlme)
rndeff<-lme(ability~as.factor(Group) + as.numeric(time), data=stroke_long,
            random=~1|Subject)
summary(rndeff)
```

14. Mengingat GEE tidak dapat mengeluarkan AIC, dengan menggunakan statement **gls**, menghitung AIC dari model GLS dengan ketiga struktur korelasi di atas (Exchangeable, Auto regressive, dan Unstructured).

```
exch<-corCompSymm(form = ~ 1 | Subject)
gls.exch<-gls(ability~as.factor(Group)+as.numeric(time), data=stroke_long,
             correlation=exch)

ar1<-corAR1(form = ~ 1 | Subject)
gls.ar1<-gls(ability~as.factor(Group)+as.numeric(time), data=stroke_long,
             correlation=ar1)

un<-corSymm(form = ~ 1 | Subject)
gls.un<-gls(ability~as.factor(Group)+as.numeric(time), data=stroke_long,
            correlation=un)
```

15. Membuat tabel untuk Membandingkan AIC dari model dengan korelasi struktur Exchangeable, Auto regressive, dan Unstructured, dengan AIC linear regresi model (Model dari instruksi no 9). Interpretasikan dan simpulkan.

	df	AIC
gls.exch	6	1468
gls.ar1	6	1325
gls.un	33	1343
mod_lm2	11	1732

- Model dengan nilai AIC yang lebih kecil menunjukkan model yang lebih fit, yaitu model GEE dengan korelasi struktur auto regressive.

```
> gee.ar1<-geeglm(ability ~ as.factor(Group) + as.numeric(time), family=gaussian,
+               data=stroke_long, id=as.factor(Subject), wave=as.numeric(time), corst="ar1")
> summary(gee.ar1)

Call:
geeglm(formula = ability ~ as.factor(Group) + as.numeric(time),
       family = gaussian, data = stroke_long, id = as.factor(Subject),
       waves = as.numeric(time), corstr = "ar1")

Coefficients:
              Estimate Std.terr   Wald Pr(>|W|)
(Intercept)    39.757    7.690 26.73  2.3e-07 ***
as.factor(Group)F  -9.420    8.863  1.13   0.288
as.factor(Group)G -15.963    8.895  3.22   0.073 .
as.numeric(time)   4.617    0.603 58.63  1.9e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = ar1
Estimated Scale Parameters:

              Estimate Std.terr
(Intercept)    436    98.6
Link = identity

Estimated Correlation Parameters:
              Estimate Std.terr
alpha    0.926    0.0202
Number of clusters:  24  Maximum cluster size: 8
```