

**TUGAS MATA KULIAH
MANAJEMEN DAN ANALISIS DATA DENGAN R
(FINAL EXAM)**

Disusun oleh:

Syndi Nurmawati

NPM 131520220002



**PROGRAM STUDI MAGISTER EPIDEMIOLOGI
FAKULTAS KEDOKTERAN
UNIVERSITAS PADJADJARAN
2023**

Dengan mempergunakan data stroke yang berasal dari link sbb: <http://www.statsci.org/data/oz/stroke.txt> , lakukan analisis dengan aplikasi R sebagai berikut:

Jawab:

Persiapan data (mengakses dataset ke dalam global environment:

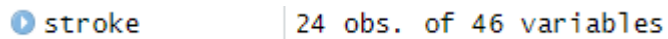
- Mengaktifkan package *data.table*

Kode: `library(data.table)`

- Mengimport data *stroke* dan menjadikannya objek baru

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

Hasil: Terdapat dataset baru *stroke* pada global environment

 `stroke` | 24 obs. of 46 variables

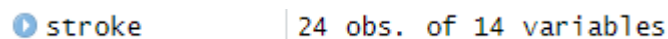
- Mengaktifkan package *dplyr*

Kode: `library(dplyr)`

- Memilih kolom yang akan dipergunakan

Kode: `stroke <- stroke %>% select(c(1:6,39:46))`

Hasil: Dataset *stroke* hanya mengandung data yang akan dipergunakan

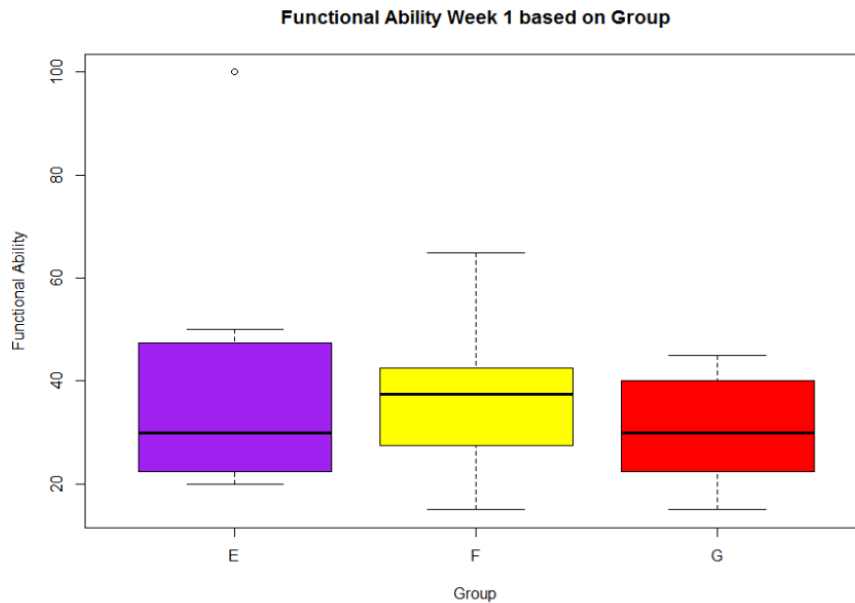
 `stroke` | 24 obs. of 14 variables

1. Menganalisis perbedaan dari Functional Ability pada pekan pertama (Bart1) berdasarkan grup intervensi (Group) dengan visualisasi boxplot. Membuat grafik boxplot dan menginterpretasi dan menyimpulkan hasilnya.

Jawab:

Kode: `boxplot(stroke$Bart1~stroke$Group, xlab = "Group", ylab = "Functional Ability", col = c("purple", "yellow", "red"), main = "Functional Ability Week 1 based on Group")`

Hasil:



Interpretasi:

Dari hasil boxplot terlihat bahwa terdapat perbedaan antara grup yang diberikan intervensi E (experimental program), F (pre-existing program), dan G (non-treatment program). Nilai median antara grup E dan G hampir serupa, namun median grup F didapatkan lebih tinggi dibandingkan 2 grup lainnya. Dari hasil tersebut dapat disimpulkan bahwa pada minggu pertama functional ability dari subjek yang masuk ke dalam grup F memiliki median yang lebih baik dibandingkan kelompok lainnya. Namun bila dilihat dari IQR-nya, subjek pada grup F memiliki rentang yang lebih jauh dibandingkan kedua kelompok lainnya. Hal ini menandakan bahwa subjek yang masuk ke dalam grup F memiliki functional ability yang lebih bervariasi dibandingkan kedua kelompok lainnya. Pada kelompok E terdapat satu titik yang merupakan outlier. Pada data ini harus dilakukan pengecekan apakah data yang di-entry sudah benar atau tidak.

2. Mencek normalitas data dari Functional Ability pekan pertama (Bart1) dengan uji statistik yang sesuai dan menginterpretasikannya.

Jawab:

Uji normalitas pada sampel berjumlah 3 – 5000 sampel dapat dilakukan dengan Shapiro Wilk.

Kode: `shapiro.test(stroke$Bart1)`

Hasil:

```
shapiro-wilk normality test  
  
data:  stroke$Bart1  
W = 0.82449, p-value = 0.0007617
```

Interpretasi:

Hasil uji normalitas menunjukkan hasil $p\text{-value} < 0,05$. Hal ini menandakan bahwa hipotesis null ditolak. Hipotesis null pada uji normalitas adalah bahwa data terdistribusi normal. Berdasarkan hasil ini dapat disimpulkan bahwa data tidak terdistribusi normal.

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

Jawab:

Kode: `stroke$Bart_diff <- stroke$Bart8-stroke$Bart1`

Hasil:

Terbentuk variabel baru dengan nama Bart_diff.

4. Mencek normalitas data dari perubahan Functional Ability (Bart_diff) dari pekan pertama (Bart1) sampai pekan terakhir (Bart8) dengan uji statistik yang sesuai dan menginterpretasikannya.

Jawab:

Kode: `shapiro.test(stroke$Bart_diff)`

Hasil:

```
shapiro-wilk normality test  
  
data:  stroke$Bart_diff  
W = 0.92817, p-value = 0.08875
```

Interpretasi:

Dari hasil uji normalitas didapatkan bahwa $p\text{-value} > 0,05$. Hal ini menandakan bahwa hipotesis null tidak ditolak, dimana hipotesis null untuk uji normalitas adalah bahwa data terdistribusi normal. Oleh karena itu dapat disimpulkan bahwa variabel Bart_diff terdistribusi normal.

5. Mencek kesamaan variance dari perubahan Functional Ability (Bart_diff) antara grup intervensi (Group) dengan uji statistic yang sesuai dan menginterpretasikannya.

Jawab:

Karena variabel Bart_diff terdistribusi normal, maka digunakan Bartlett test untuk menguji kesamaan variance.

Kode: bartlett.test(stroke\$Bart_diff,stroke\$Group)

Hasil

```
Bartlett test of homogeneity of variances  
  
data: stroke$Bart_diff and stroke$Group  
Bartlett's K-squared = 0.39433, df = 2, p-value = 0.8211
```

Interpretasi:

Pada hasil Bartlett test didapatkan hasil p-value > 0,05. Hasil ini menunjukkan bahwa hipotesis null tidak dapat ditolak. Hipotesis null pada Bartlett test adalah varians data sama atau homogen. Oleh karena itu dapat disimpulkan bahwa variabel Bart_diff memiliki varians data yang sama atau homogen.

6. Memplot mean dan 95% Confidence Interval dari nilai perubahan Functional Ability (Bart_diff) berdasarkan grup intervensi (Group) dalam 1 grafik.

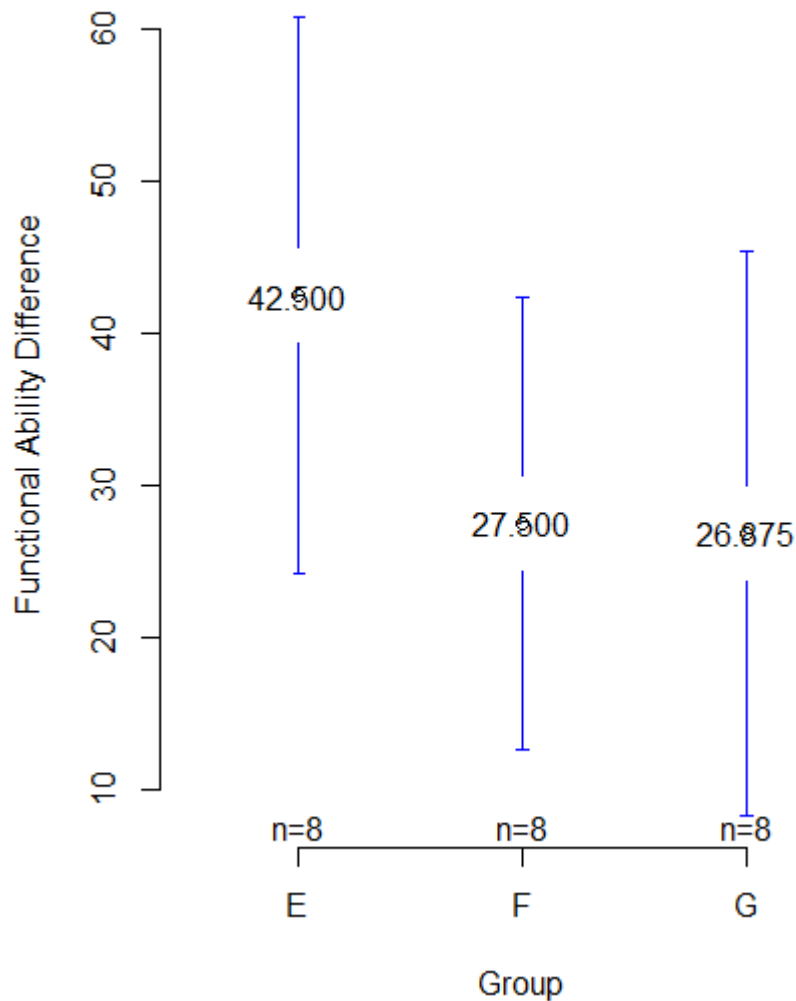
Jawab:

Perlu diaktifkan package gplots.

Kode: plotmeans(Bart_diff~Group, data = stroke, frame = FALSE, mean.labels = TRUE, connect = FALSE, xlab = "Group", ylab = "Functional Ability Difference", main = "Functional Ability Difference based on Group")

Hasil:

Functional Ability Difference based on Group



7. Melakukan uji anova untuk membandingkan rata-rata(mean) nilai perubahan Functional Ability (Bart_diff) antara 3 grup intervensi (Group) dan menginterpretasikannya.

Jawab:

Uji Anova yang dipergunakan adalah one-way ANOVA karena data Bart_diff terdistribusi normal dan memiliki varians yang sama atau homogen.

Kode: `res.aov <- aov(Bart_diff ~ Group, data = stroke)`

Hasil:

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

Interpretasi:

Pada hasil ini p-value > 0,05 oleh karena itu hipotesis null untuk uji ANOVA tidak dapat ditolak. Hipotesis null untuk uji one-way ANOVA adalah bahwa mean antara grup yang dilakukan tes adalah sama. Oleh karena itu dapat disimpulkan bahwa mean antara grup E, F, dan G adalah sama atau tidak terdapat perbedaan yang signifikan.

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

Jawab:

Pada data ini dipakai data dalam bentuk long sehingga dilakukan perubahan menjadi objek baru stroke_long. Variabel Time masih dalam bentuk character sehingga diubah menjadi numerik. Setelah itu dilakukan analisis model linear regresi.

*Kode: model_1 <- lm(Ability~as.numeric(Time) + as.factor(Group) +
as.numeric(Time)*as.factor(Group),data=stroke_long)*

Hasil:

```
Residuals:
    Min       1Q   Median       3Q      Max
-47.812 -13.560  -5.305  13.337  63.854

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      29.82143     5.77401   5.165 6.16e-07 ***
as.numeric(Time)   6.32440     1.14342   5.531 1.07e-07 ***
as.factor(Group)F   3.34821     8.16569   0.410  0.6823
as.factor(Group)G  -0.02232     8.16569  -0.003  0.9978
as.numeric(Time):as.factor(Group)F -1.99405     1.61705  -1.233  0.2191
as.numeric(Time):as.factor(Group)G -2.68601     1.61705  -1.661  0.0984 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.96 on 186 degrees of freedom
Multiple R-squared:  0.2612,    Adjusted R-squared:  0.2413
F-statistic: 13.15 on 5 and 186 DF,  p-value: 5.695e-11
```

- 9. Melakukan ulang Langkah no 8 tanpa variable interaksi di dalam model.**

Jawab:

*Kode: model_2 <- lm(Ability~as.numeric(Time) + as.factor(Group),
data=stroke_long)*

Hasil:

```
Residuals:
    Min       1Q   Median       3Q      Max
-49.332 -13.907  -4.532  15.043  58.394

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    36.8415     3.9712   9.277 < 2e-16 ***
as.numeric(Time)  4.7644     0.6619   7.198 1.42e-11 ***
as.factor(Group)F  -5.6250     3.7147  -1.514  0.13164
as.factor(Group)G -12.1094     3.7147  -3.260  0.00132 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.01 on 188 degrees of freedom
Multiple R-squared:  0.2494,    Adjusted R-squared:  0.2374
F-statistic: 20.82 on 3 and 188 DF,  p-value: 1.084e-11
```

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

Jawab:

Perlu diaktifkan package broom.

Kode: `glance(model_1)`

Hasil:

	r.squared <dbl>	adj.r.squared <dbl>	sigma <dbl>	statistic <dbl>	p.value <dbl>	df <dbl>	logLik <dbl>	AIC <dbl>	BIC <dbl>	deviance <dbl>	df.resi~1 <int>	nobs <int>
1	0.261	0.241	21.0	13.2	5.70e-11	5	-854.	1721.	1744.	81709.	186	192

AIC pada model_1 sebesar 1721

Kode: `glance(model_2)`

Hasil:

	r.squared <dbl>	adj.r.squared <dbl>	sigma <dbl>	statistic <dbl>	p.value <dbl>	df <dbl>	logLik <dbl>	AIC <dbl>	BIC <dbl>	deviance <dbl>	df.resi~1 <int>	nobs <int>
1	0.249	0.237	21.0	20.8	1.08e-11	3	-855.	1720.	1736.	83016.	188	192

AIC pada model_2 sebesar 1720

Pada AIC model pertama (model dengan interaksi) nilai AIC lebih tinggi yaitu 1721. Sedangkan pada AIC model kedua (model tanpa interaksi) memiliki nilai AIC yang sedikit lebih rendah yaitu 1720. Oleh karena itu dapat disimpulkan bahwa model kedua sedikit lebih baik dibandingkan model pertama.

11. Model no 8 dan 9, manakah yang terbaik? Pilih salah satu kemudian interpretasikan hasil dari analisisnya dari model yang dipilih (hubungan antara variable explanatory dengan outcome)

Jawab:

Dari antara kedua model, didapatkan bahwa AIC model_2 sedikit lebih rendah dibandingkan model_1. Berdasarkan hasil ini dapat disimpulkan bahwa bila dibandingkan dengan model_1 maka model_2 merupakan model regresi yang lebih fit terhadap dataset. Pada model_2, didapatkan estimate (β_1) dari Time (x_1) adalah 4,76. Oleh karena itu dapat dikatakan bahwa kenaikan Time sebesar 4,76 kali dapat meningkatkan Functional Ability Score (dalam hal ini Barthell Index Score).

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.

Jawab:

*Kode: `rnDeff<-lme(Ability~as.factor(Group) + as.numeric(Time),`
`data=stroke_long, random=~1|Subject)`*

Hasil:

```

      AIC      BIC    logLik
1467.559 1486.978 -727.7796

Random effects:
Formula: ~1 | Subject
      (Intercept) Residual
StdDev:      20.1068  8.960882

Fixed effects: Ability ~ as.factor(Group) + as.numeric(Time)
              value Std.Error DF   t-value p-value
(Intercept)  36.84152  7.307749 167   5.041432  0.0000
as.factor(Group)F -5.62500 10.177433  21  -0.552693  0.5863
as.factor(Group)G -12.10937 10.177433  21  -1.189826  0.2474
as.numeric(Time)  4.76438  0.282241 167  16.880541  0.0000
Correlation:
              (Intr) a.(G)F a.(G)G
as.factor(Group)F -0.696
as.factor(Group)G -0.696  0.500
as.numeric(Time) -0.174  0.000  0.000

Standardized within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.17717122 -0.62871793  0.02125076  0.65225196  3.00241837

Number of Observations: 192
Number of Groups: 24

```

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**

Jawab:

Kode: gee.exch <- geeglm(Ability~as.factor(Group)+as.numeric(Time),family = gaussian, data=stroke_long,id=as.factor(Subject),wave=as.numeric(Time), corst="exchangeable")

Hasil:

```

Coefficients:
              Estimate Std. err   Wald Pr(>|W|)
(Intercept)      36.8415   8.0038 21.188 4.16e-06 ***
as.factor(Group)F  -5.6250   9.5194   0.349   0.555
as.factor(Group)G -12.1094   9.5198   1.618   0.203
as.numeric(Time)   4.7644   0.6281 57.539 3.31e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable
Estimated Scale Parameters:

              Estimate Std. err
(Intercept)      432.4   98.59
Link = identity

Estimated Correlation Parameters:
              Estimate Std. err
alpha         0.8154 0.04643
Number of clusters: 24 Maximum cluster size: 8

```

- **Auto regressive**

Jawab:

Kode: gee.ar1 <- geeglm(Ability~as.numeric(Time)+as.factor(Group),family = gaussian, data=stroke_long,id=as.factor(Subject),wave=as.numeric(Time), corst="ar1")

Hasil:

```

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

Correlation structure = ar1
Estimated Scale Parameters:

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

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

```

- **Unstructured**

Jawab:

Kode: `gee.un <- geeglm(Ability~as.numeric(Time)+as.factor(Group),family = gaussian, data=stroke_long,id=as.factor(Subject),wave=as.numeric(Time), corst="unstructured")`

Hasil:

```
Call:
geeglm(formula = Ability ~ as.numeric(Time) + as.factor(Group),
       family = gaussian, data = stroke_long, id = as.factor(Subject),
       waves = as.numeric(Time), corstr = "unstructured")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)	
(Intercept)	29.6563	12.0844	6.02	0.014	*
as.numeric(Time)	6.9784	2.9800	5.48	0.019	*
as.factor(Group)F	20.2541	13.9030	2.12	0.145	
as.factor(Group)G	-0.0862	10.8339	0.00	0.994	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	807	317

Link = identity

```

Estimated Correlation Parameters:
      Estimate Std.err
alpha.1:2    0.585  0.2823
alpha.1:3    0.599  0.2293
alpha.1:4    0.616  0.1758
alpha.1:5    0.632  0.1185
alpha.1:6    0.612  0.0844
alpha.1:7    0.629  0.1293
alpha.1:8    0.659  0.2505
alpha.2:3    0.684  0.2280
alpha.2:4    0.710  0.1822
alpha.2:5    0.734  0.1169
alpha.2:6    0.734  0.0689
alpha.2:7    0.766  0.0926
alpha.2:8    0.807  0.2099
alpha.3:4    0.810  0.1521
alpha.3:5    0.832  0.0843
alpha.3:6    0.841  0.0442
alpha.3:7    0.880  0.1026
alpha.3:8    0.928  0.2284
alpha.4:5    0.904  0.0627
alpha.4:6    0.924  0.0563
alpha.4:7    0.993  0.1294
alpha.4:8    1.052  0.2534
alpha.5:6    1.036  0.0871
alpha.5:7    1.105  0.1659
alpha.5:8    1.185  0.2797
alpha.6:7    1.199  0.2093
alpha.6:8    1.293  0.3184
alpha.7:8    1.464  0.3799
Number of clusters: 24 Maximum cluster size: 8

```

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 Unstructure).

Jawab:

- **Exchangeable**

Kode: `exch<-corCompSymm(form = ~ 1 | Subject)`

`gls.exch<-gls(Ability~as.factor(Group)+as.numeric(Time),`

`data=stroke_long, correlation=exch)`

Hasil:

```
AIC  BIC logLik
1468 1487   -728
```

```
Correlation Structure: Compound symmetry
Formula: ~1 | Subject
Parameter estimate(s):
  Rho
0.834
```

Coefficients:

	Value	Std. Error	t-value	p-value
(Intercept)	36.8	7.31	5.04	0.000
as.factor(Group)F	-5.6	10.18	-0.55	0.581
as.factor(Group)G	-12.1	10.18	-1.19	0.236
as.numeric(Time)	4.8	0.28	16.88	0.000

Correlation:

	(Intr)	a. (G)F	a. (G)G
as.factor(Group)F	-0.696		
as.factor(Group)G	-0.696	0.500	
as.numeric(Time)	-0.174	0.000	0.000

Standardized residuals:

Min	Q1	Med	Q3	Max
-2.241	-0.632	-0.206	0.683	2.653

```
Residual standard error: 22
Degrees of freedom: 192 total; 188 residual
```

○ Autoregressive

```
Kode: ar1<-corAR1(form = ~ 1 | Subject)
      gls.ar1<-gls(Ability~as.factor(Group)+as.numeric(Time),
      data=stroke_long, correlation=ar1)
```

Hasil:

```
AIC   BIC logLik
1325 1344   -656
```

```
Correlation Structure: AR(1)
Formula: ~1 | Subject
Parameter estimate(s):
Phi
0.95
```

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	40.0	7.36	5.43	0.000
as.factor(Group)F	-9.8	9.93	-0.98	0.328
as.factor(Group)G	-16.3	9.93	-1.64	0.102
as.numeric(Time)	4.6	0.49	9.41	0.000

Correlation:

	(Intr)	a. (G)F	a. (G)G
as.factor(Group)F	-0.675		
as.factor(Group)G	-0.675	0.500	
as.numeric(Time)	-0.300	0.000	0.000

Standardized residuals:

Min	Q1	Med	Q3	Max
-2.183	-0.615	-0.152	0.725	2.567

Residual standard error: 21.6

Degrees of freedom: 192 total; 188 residual

○ Unstructured

Kode: `un<-corSymm(form = ~ 1 | Subject)`

```
gls.un<-gls(Ability~as.factor(Group)+as.numeric(Time),
data=stroke_long,correlation=un)
```

Hasil:

```

      AIC   BIC logLik
1343 1450   -638

Correlation Structure: General
Formula: ~1 | Subject
Parameter estimate(s):
Correlation:
  1      2      3      4      5      6      7
2 0.943
3 0.894 0.935
4 0.829 0.883 0.954
5 0.733 0.812 0.882 0.906
6 0.615 0.727 0.805 0.844 0.967
7 0.454 0.590 0.672 0.760 0.884 0.943
8 0.299 0.455 0.542 0.644 0.812 0.885 0.966

Coefficients:
              Value Std.Error t-value p-value
(Intercept)    45.7      5.97   7.64 0.0000
as.factor(Group)F -21.9      7.12  -3.07 0.0024
as.factor(Group)G -28.5      7.12  -4.00 0.0001
as.numeric(Time)   4.5      0.71   6.30 0.0000

Correlation:
              (Intr) a. (G)F a. (G)G
as.factor(Group)F -0.595
as.factor(Group)G -0.595 0.500
as.numeric(Time) -0.539 0.000 0.000

Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.1997 -0.5842 0.0218 0.8751 2.5747

Residual standard error: 21.3
Degrees of freedom: 192 total; 188 residual

```

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

Jawab:

Kode: aic = AIC(model_2,gls.exch,gls.ar1,gls.un)

Hasil:

	df	AIC
model_2	5	1720
gls.exch	6	1468
gls.ar1	6	1325
gls.un	33	1343

Hasil AIC dari keempat model ini memperlihatkan bahwa model autoregressive memiliki nilai AIC yang paling rendah. Oleh karena itu dapat disimpulkan bahwa model autoregressive merupakan model regresi yang paling fit terhadap dataset bila dibandingkan dengan model exchangeable, unstructured, dan model regresi linear.