

Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Final Exam

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

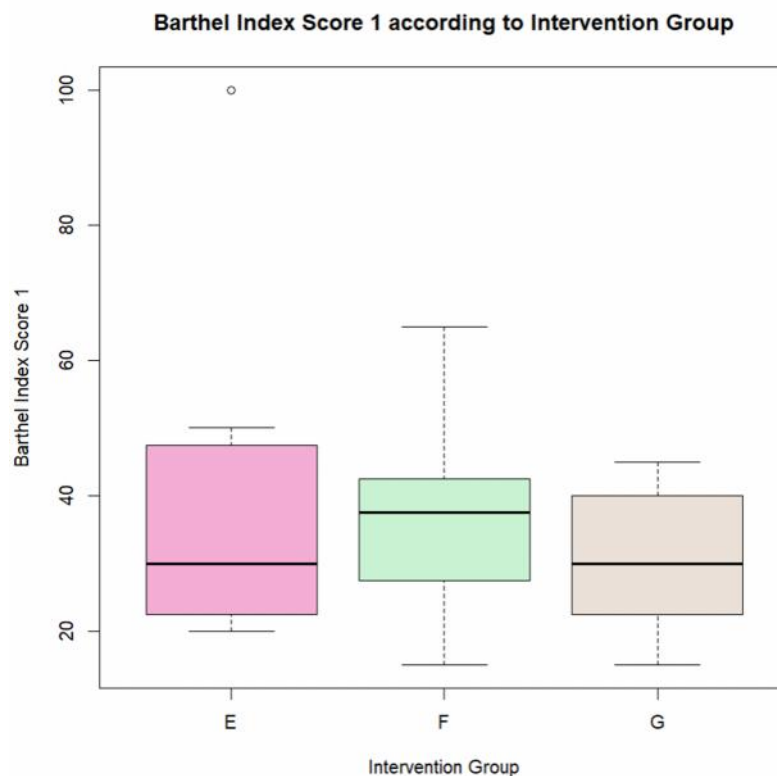
2023

1. 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:

```
boxplot(stroke$Bart1~stroke$Group,xlab = "Intervention Group", ylab = "Barthel Index Score 1", col = c("#F4AED6", "#C9F3D3", "#EBE1D9"), main = "Barthel Index Score 1 according to Intervention Group")
```

Hasil:



Pada boxplot terlihat bahwa nilai median pada Barthel Index Score (BIS) di minggu pertama memiliki nilai lebih tinggi pada Grup F yaitu grup yang mendapatkan lengan program yang sudah diterapkan sebelumnya. Nilai minimum BIS pada grup E (grup yang mendapatkan intervensi program) lebih tinggi dibandingkan dengan grup F dan grup G (grup yang tidak mendapatkan intervensi). Nilai maksimum BIS tertinggi adalah pada grup F. Pada grup E terdapat outlier yang memiliki nilai BIS 100 pada minggu pertama, dan tidak didapatkan outlier pada grup lainnya. Kesimpulannya, grup dengan lengan program yang sudah ada sebelumnya memiliki nilai median tertinggi pada BIS minggu pertama, namun nilai secara keseluruhan lebih bervariasi karena range minimum dan maksimum yang lebih besar dibanding dengan kedua grup lainnya.

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

Jawab:

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

Hasil:

```
Shapiro-Wilk normality test

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

Uji normalitas data Shapiro-Wilk sebaiknya digunakan pada data yang memiliki ukuran sampel yang kecil (<50 sampel). Pada uji Shapiro-Wilk, hipotesis null-nya adalah “distribusi sampel adalah normal”. Pada uji normalitas yang telah dilakukan, didapatkan bahwa hasil p-value <0,05 yang berarti hipotesis null ditolak, sehingga disimpulkan bahwa distribusi sampel data tersebut tidak normal.

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

Jawab:

```
stroke_wide <- stroke_wide %>%
  mutate(Bart_diff=Week8-Week1) %>%
  as.data.frame()
```

Hasil:

Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8	Bart_diff
45	45	45	45	80	80	80	90	45
20	25	25	25	30	35	30	50	30
50	50	55	70	70	75	90	90	40
25	25	35	40	60	60	70	80	55
100	100	100	100	100	100	100	100	0
20	20	30	50	50	60	85	95	75
30	35	35	40	50	60	75	85	55
30	35	45	50	55	65	65	70	40
40	55	60	70	80	85	90	90	50
65	65	70	70	80	80	80	80	15
30	30	40	45	65	85	85	85	55
25	35	35	35	40	45	45	45	20

Variabel baru (Bart_diff) telah berhasil ditambahkan pada dataset stroke_wide.

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:

```
shapiro.test(stroke_wide$Bart_diff)
```

Hasil:

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

Setelah dilakukan perhitungan perubahan Functional Ability (Bart_diff) dan kembali melakukan pengecekan dengan uji normalitas Shapiro-Wilk, didapatkan bahwa p-value >0,05. Sehingga hipotesis null diterima, yang berarti distribusi sampel pada data tersebut normal.

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

Jawab:

```
bartlett.test(Bart_diff ~ Group, stroke_wide)
```

Hasil:

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

Bartlett test digunakan untuk melihat apakah terdapat perbedaan pada varians atau homogenitas varians. Pada Bartlett test, hipotesis null-nya adalah “variens dari seluruh populasi adalah sama”. Dari hasil tes yang sudah dilakukan, didapatkan bahwa nilai p-value >0,05 sehingga hipotesis null diterima. Hal ini berarti varians yang ada pada data tersebut adalah 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:

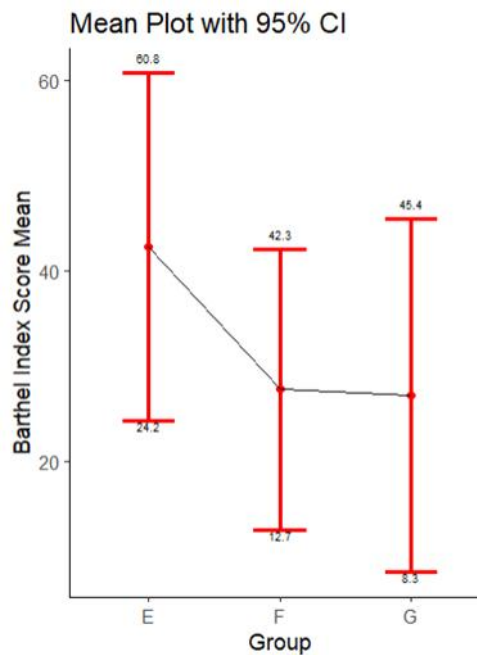
- a. Menghitung mean, lower CI dan upper CI

```
dt <- stroke_wide %>%  
  dplyr::group_by(Group) %>%  
  dplyr::summarise(  
    mean = mean(Bart_diff),  
    lci = t.test(Bart_diff, conf.level = 0.95)$conf.int[1],  
    uci = t.test(Bart_diff, conf.level = 0.95)$conf.int[2])
```

	Group	mean	lci	uci
1	E	42.500	24.21098	60.78902
2	F	27.500	12.67893	42.32107
3	G	26.875	8.32342	45.42658

- b. Membuat plot

```
pl1 <- ggplot(data=dt)  
pl1 <- pl1 + geom_line(aes(x = Group, y = mean), group = 1)  
pl1 <- pl1 + geom_point(aes(x=Group, y=mean), color = "red")  
pl1 <- pl1 + geom_errorbar(aes(x=Group, ymin=lci, ymax=uci), width = 0.4,  
  color = "red", size = 1)  
pl1 <- pl1 + geom_text(aes(x=Group, y=lci, label = round(lci,1)), size=2, vjust=1)  
pl1 <- pl1 + geom_text(aes(x=Group, y=uci, label = round(uci,1)), size=2, vjust=-  
  1)  
pl1 <- pl1 + theme_classic()  
pl1 <- pl1 + labs(title = "Mean Plot with 95% CI")  
pl1 <- pl1 + labs(x = "Group", y = "Barthel Index Score Mean")  
  
pl1
```



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

Jawab:

```
st_wide_aov <- aov(stroke_wide$Bart_diff~factor(stroke_wide$Group))
summary(st_wide_aov)
```

Hasil:

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

Pada uji ANOVA, hipotesis null-nya adalah “rata-rata (mean) pada semua populasi adalah sama”. Dari uji ANOVA yang telah dilakukan, terlihat bahwa p-value >0,05 yang artinya hipotesis null diterima. Sehingga disimpulkan bahwa rata-rata nilai perubahan Functional Ability pada 3 grup intervensi adalah sama.

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:

```
mlr1 <- lm(Ability ~ Time + Group + Time*Group, data = stroke_long)
summary(mlr1)
```

Hasil:

```
Call:
lm(formula = Ability ~ Time + Group + Time * Group, data = stroke_long)

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 ***
Time         6.32440    1.14342   5.531 1.07e-07 ***
GroupF       3.34821    8.16569   0.410  0.6823
GroupG      -0.02232    8.16569  -0.003  0.9978
Time:GroupF  -1.99405    1.61705  -1.233  0.2191
Time:GroupG  -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:

```
mlr2 <- lm(Ability ~ Time + Group, data = stroke_long)
summary(mlr2)
```

Hasil:

```
Call:
lm(formula = Ability ~ Time + Group, data = stroke_long)

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 ***
Time         4.7644    0.6619   7.198 1.42e-11 ***
GroupF      -5.6250    3.7147  -1.514  0.13164
GroupG     -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:

```
models <- list(mlr1, mlr2)
```

```
mod.names <- c('ability.time.group.timegroup', 'ability.time.group')
```

```
aictab(cand.set = models, modnames = mod.names)
```

Hasil:

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcwt	Cum.Wt	LL
ability.time.group	5	1720.50	0.00	0.65	0.65	-855.09
ability.time.group.timegroup	7	1721.74	1.24	0.35	1.00	-853.56

Nilai AIC pada kedua grup terlihat pada kolom AICc. Huruf “c” kecil adalah indikasi bahwa AIC sudah dikalkulasi dari AIC yang dikoreksi pada jumlah sampel kecil. Model yang memiliki nilai AIC terendah selalu ditulis pertama. Pada hasil terlihat bahwa nilai AIC pada model “mlr2” lebih kecil dibandingkan dengan model “mlr1” yang terjadi pada saat interaksi antara time dan group dihilangkan. Sehingga dapat dikatakan bahwa model “mlr2” lebih ‘fit’ (best fitting model).

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:

Model yang lebih baik adalah model nomor 9 (mlr2) karena nilai AIC yang lebih kecil jika dibandingkan dengan model nomor 8 (mlr1). Sehingga dapat disimpulkan bahwa dengan tidak mengikutsertakan variabel interaksi antara Time dengan Group dapat membuat model tersebut lebih baik jika dibandingkan dengan model yang mengikutsertakan variabel interaksi. Dari hasil tersebut dapat diinterpretasikan bahwa dengan meningkatnya waktu sebesar 4,7 kali akan meningkatkan Barthel Index Score, dan apabila berada dalam Grup intervensi G maka akan menurunkan Barthel 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:

```
rnd <- lme(Ability~as.numeric(Time) + as.factor(Group), data = stroke_long,  
          random=~1|Subject)  
summary(rnd)
```

Hasil:

```
Linear mixed-effects model fit by REML  
Data: stroke_long  
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.numeric(Time) + as.factor(Group)  
              Value Std.Error DF   t-value p-value  
(Intercept)   36.84152  7.307749 167   5.041432  0.0000  
as.numeric(Time)  4.76438  0.282241 167  16.880541  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  
Correlation:  
              (Intr) as.(T) a.(G)F  
as.numeric(Time) -0.174  
as.factor(Group)F -0.696  0.000  
as.factor(Group)G -0.696  0.000  0.500  
  
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
- Auto regressive
- Unstructure

Jawab:

a. Exchangeable

```
geel <- geeglm(Ability~as.numeric(Time) + as.factor(Group) ,family=gaussian,  
              data=stroke_long,          id=as.factor(Subject),          wave=as.numeric(Time),  
              corst="exchangeable")  
summary(geel)
```

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

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

b. Auto regressive

```
gee2 <- geeglm(Ability~as.numeric(Time) + as.factor(Group),family=gaussian,
              data=stroke_long,id=as.factor(Subject),wave=as.numeric(Time),corst="ar1")
summary(gee2)
```

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

```
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
```

c. Unstructured

```
gee3 <- geeglm(Ability~as.numeric(Time) + as.factor(Group), family=gaussian,
data=stroke_long,          id=as.factor(Subject),          wave=as.numeric(Time),
corst="unstructured")
summary(gee3)
```

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:

a. Exchangeable

```
exch <- corCompSymm(form = ~ 1 | Subject)
gls1 <- gls(Ability~as.numeric(Time) + as.factor(Group),
            data=stroke_long,
            correlation=exch)
summary(gls1)
```

```
Generalized least squares fit by REML
Model: Ability ~ as.numeric(Time) + as.factor(Group)
Data: stroke_long
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.numeric(Time)  4.8      0.28   16.88  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

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

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
```

b. Auto regressive

```
ar1 <- corAR1(form = ~ 1 | Subject)
gls2 <- gls(Ability~as.numeric(Time) + as.factor(Group),
            data=stroke_long,
            correlation=ar1)
summary(gls2)
```

```

Generalized least squares fit by REML
Model: Ability ~ as.numeric(Time) + as.factor(Group)
Data: stroke_long
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.numeric(Time)  4.6     0.49    9.41  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

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

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

```

c. Unstructured

```

un <- corSymm(form = ~ 1 | Subject)
gls3 <- gls(Ability~as.numeric(Time) + as.factor(Group),
            data=stroke_long,
            correlation=un)
summary(gls3)

```



```

Generalized least squares fit by REML
Model: Ability ~ as.numeric(Time) + as.factor(Group)
Data: stroke_long
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.numeric(Time)  4.5      0.71   6.30 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

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

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:

$aic = AIC(mlr3, gls1, gls2, gls3)$

Hasil:

	df	AIC
mlr3	5	1720
gls1	6	1468
gls2	6	1325
gls3	33	1343

Berdasarkan hasil pada tabel, nilai AIC yang paling kecil adalah nilai pada “gls2” yaitu model GLS dari struktur korelasi auto regressive. Sehingga dapat disimpulkan bahwa model dari struktur korelasi tersebut lebih ‘fit’ jika dibandingkan dengan model yang lain.