Tugas Mata Kuliah Analisa R

Final Exam

Disusun oleh:

Amila Hanifan Muslimah

Prodi Magister Epidemiologi

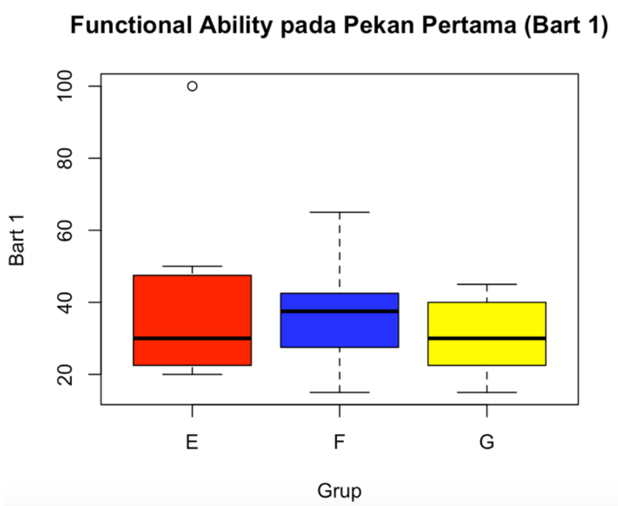
Fakultas Kedokteran Universitas Padjadjaran

2023

Tugas

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.

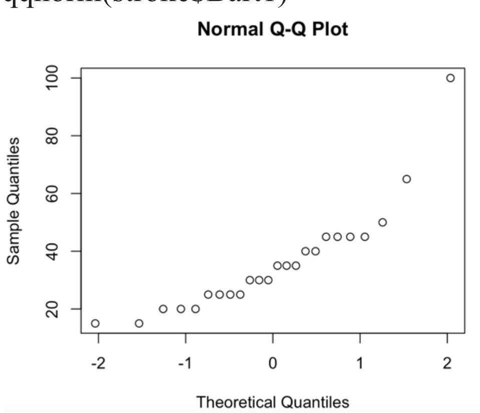
   boxplot(stroke$Bart1~stroke$Group,xlab="Grup",ylab="Bart 1",col=c("red","blue","yellow"),main="Functional Ability pada Pekan Pertama (Bart 1)")



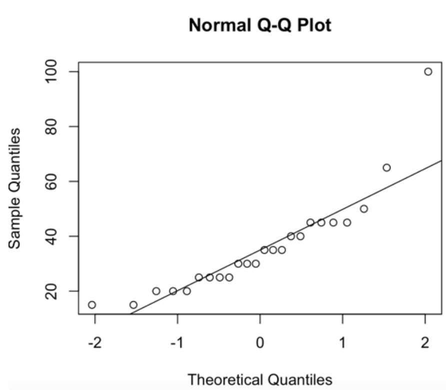
   Grup F memiliki nilai Bart 1 yang lebih bervariasi, dengan range yang lebih lebar antara nilai minimum dan maksimum, dibandingkan grup E dan F.

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

   #grafik normalitas
   qqnorm(stroke$Bart1)



   qqline(stroke$Bart1)



g11

#normality test :
shapiro.test(stroke$Bart1)

Shapiro-Wilk normality test
data: stroke$Bart1
W = 0.82449, p-value = 0.0007617

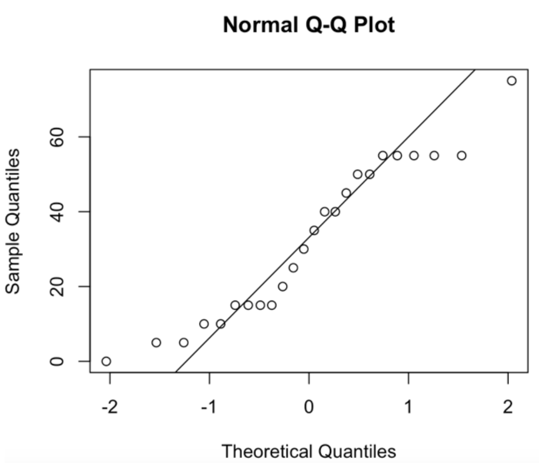
Berdasarkan uji normalitas Shapiro-Wilk hasil p value : 0.0007617 (< 0.05) sehingga data tidak berdistribusi normal

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

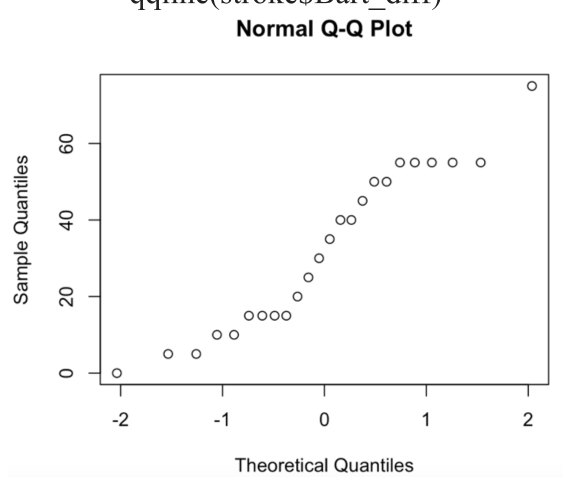
   stroke$Bart_diff<-stroke$Bart8-stroke$Bart1

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.

   #grafik normalitas
   qqnorm(stroke$Bart_diff)



qqline(stroke$Bart_diff)



#normality test

shapiro.test(stroke$Bart_diff)

Shapiro-Wilk normality test
data: stroke$Bart_diff
W = 0.92817, p-value = 0.08875

Berdasarkan uji normalitas Shapiro-Wilk hasil p value : 0.08875 (> 0.05) sehingga data berdistribusi normal

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

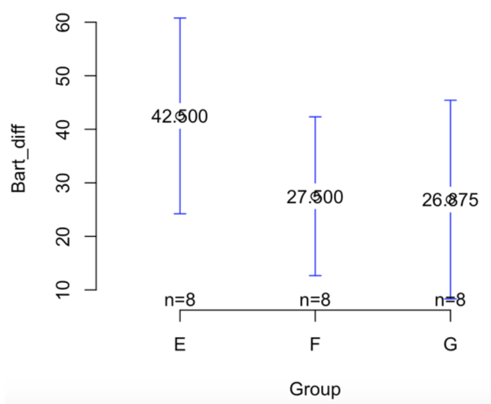
   bartlett.test(stroke$Bart_diff, stroke$Group)

   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

   Data perubahan dari Functional Ability antara grup intervensi menunjukan kesamaan variance yang homogen

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

   plotmeans(Bart_diff ~ Group, data = stroke, frame = FALSE, mean.labels = TRUE, connect = FALSE)



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

   summary(aov(Bart_diff~Group, data = stroke))

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

   Pada uji anova didaptkan hasil yang tidak signifikan untuk data mean nilai perubahan Functional Ablity antara 3 grup intervensi

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.

   model1 <- lm(ability ~ as.numeric(time) + Group + as.numeric(time)*Group, data = stroke_long)

model1	list [13] (S3: lm)	List of length 13
coefficients	double [6]	29.8214 6.3244 3.3482 -0.0223 -1.9940 -2.6860
residuals	double [192]	8.85 2.53 -3.79 -10.12 18.56 12.23 ...
effects	double [192]	-725.66 151.26 2.81 -68.50 -9.74 -34.81 ...
rank	integer [1]	6
fitted.values	double [192]	36.1 42.5 48.8 55.1 61.4 67.8 ...
assign	integer [6]	0 1 2 2 3 3
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	186
contrasts	list [1]	List of length 1
xlevels	list [1]	List of length 1
call	language	lm(formula = ability ~ as.numeric(time) + Group ...
terms	formula	ability ~ as.numeric(time) + Group + as.numeric(ti...
model	list [192 x 3] (S3: data.frame)	A data.frame with 192 rows and 3 columns

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

   model2 <- lm(ability ~ as.numeric(time) + Group, data = stroke_long)

model2	list [13] (S3: lm)	List of length 13
coefficients	double [4]	36.84 4.76 -5.63 -12.11
residuals	double [192]	3.39 -1.37 -6.13 -10.90 19.34 14.57 ...
effects	double [192]	-725.66 151.26 2.81 -68.50 20.18 15.40 ...
rank	integer [1]	4
fitted.values	double [192]	41.6 46.4 51.1 55.9 60.7 65.4 ...
assign	integer [4]	0 1 2 2
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	188
contrasts	list [1]	List of length 1
xlevels	list [1]	List of length 1
call	language	lm(formula = ability ~ as.numeric(time) + Group, ...
terms	formula	ability ~ as.numeric(time) + Group
model	list [192 x 3] (S3: data.frame)	A data.frame with 192 rows and 3 columns

10. Menghitung AIC model no 8 dan 9, serta menginterpretasikan perbandingan nilai AIC nya.
   glance(model1) : AIC 1721

```
# A tibble: 1 × 12
  r.squ…¹ adj.r…² sigma stati…³  p.value    df logLik   AIC   BIC devia…⁴ df.re…⁵  nobs
    <dbl>   <dbl> <dbl>   <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>   <dbl>   <int> <int>
1   0.261   0.241  21.0    13.2 5.70e-11     5  -854. 1721. 1744.  81709.     186   192
# … with abbreviated variable names ¹r.squared, ²adj.r.squared, ³statistic, ⁴deviance,
#   ⁵df.residual
```

```
glance(model2) : AIC 1720
# A tibble: 1 × 12
  r.squ…¹ adj.r…² sigma stati…³  p.value    df logLik   AIC   BIC devia…⁴ df.re…⁵  nobs
    <dbl>   <dbl> <dbl>   <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>   <dbl>   <int> <int>
1   0.249   0.237  21.0    20.8 1.08e-11     3  -855. 1720. 1736.  83016.     188   192
# … with abbreviated variable names ¹r.squared, ²adj.r.squared, ³statistic, ⁴deviance,
#   ⁵df.residual
```

AIC model 2 lebih kecil dari AIC model 1

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)

    Memilih model no 9 karena lebih sesuai (fitted)

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.

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

mixmodel	list [18] (S3: lme)	List of length 18
modelStruct	list [1] (S3: lmeStructInt, lmeS	List of length 1
dims	list [5]	List of length 5
contrasts	list [1]	List of length 1
coefficients	list [2]	List of length 2
varFix	double [4 x 4]	5.34e+01 -3.58e-01 -5.18e+01 -5.18e+01 -3.58e-01 7.97e-02 3.85e-…
sigma	double [1]	8.960882
apVar	double [2 x 2]	2.50e-02 -7.43e-05 -7.43e-05 2.99e-03
logLik	double [1]	-727.7796
numIter	NULL	Pairlist of length 0
groups	list [192 x 1] (S3: data.frame)	A data.frame with 192 rows and 1 column
call	language	lme.formula(fixed = ability ~ as.numeric(time) + as.factor(Group), data = …
terms	formula	ability ~ as.numeric(time) + as.factor(Group)
method	character [1]	'REML'
fitted	double [192 x 2]	41.6 46.4 51.1 55.9 60.7 65.4 46.9 51.7 56.5 61.2 66.0 70.8 …
residuals	double [192 x 2]	3.39 -1.37 -6.13 -10.90 19.34 14.57 -1.94 -6.71 -11.47 -16.24 14.00 …
fixDF	list [2]	List of length 2
na.action	NULL	Pairlist of length 0
data	list [192 x 8] (S3: tbl_df, tbl, c	A tibble with 192 rows and 8 columns

13. Melakukan ulang analisis dengan Functional Ability (Bartlet) sebagai outcome(y) dan explanatory variables meliputi: waktu(time/week), grup intervensi (group) dengan General Estimating Equation (GEE) dengan correlation structure:
    1. Exchangeable

```
gee_exch <-
geeglm(ability~as.factor(Group)+as.numeric(time)+as.factor(Group)*as.numeric(
time),family=gaussian,

data=stroke_long,id=as.factor(Subject),wave=as.numeric(time),corst="exchangea
ble")
exch<-corCompSymm(form = ~ 1 | Subject)
gls.exch<-gls(ability~as.factor(Group)+as.numeric(time)+
         as.factor(Group)*as.numeric(time), data=stroke_long,
         correlation=exch)
summary(gls.exch)

Generalized least squares fit by REML
  Model: ability ~ as.factor(Group) + as.numeric(time) + as.factor(Group) *
as.numeric(time)
  Data: stroke_long
       AIC      BIC    logLik
  1452.715 1478.521 -718.3573

Correlation Structure: Compound symmetry
 Formula: ~1 | Subject
 Parameter estimate(s):
    Rho
0.84671

Coefficients:
                                       Value Std.Error   t-value p-value
(Intercept)                        29.821429  7.497378  3.977581  0.0001
as.factor(Group)F                   3.348214 10.602894  0.315783  0.7525
as.factor(Group)G                  -0.022321 10.602894 -0.002105  0.9983
as.numeric(time)                    6.324405  0.467228 13.536016  0.0000
as.factor(Group)F:as.numeric(time) -1.994048  0.660760 -3.017809  0.0029
as.factor(Group)G:as.numeric(time) -2.686012  0.660760 -4.065033  0.0001

 Correlation:
                                   (Intr) as.(G)F as.(G)G as.n() a.(G)F:
as.factor(Group)F                  -0.707
as.factor(Group)G                  -0.707  0.500
as.numeric(time)                   -0.280  0.198   0.198
as.factor(Group)F:as.numeric(time)  0.198 -0.280  -0.140  -0.707
as.factor(Group)G:as.numeric(time)  0.198 -0.140  -0.280  -0.707  0.500
```

```
Standardized residuals:
       Min         Q1        Med         Q3        Max 
-2.1857469 -0.6199072 -0.2425206  0.6097030  2.9190912 

Residual standard error: 21.87467 
Degrees of freedom: 192 total; 186 residual
```

2. Auto regressive

```
gee_ar1 <-
geeglm(ability~as.factor(Group)+as.numeric(time)+as.factor(Group)*as.numeric(
time),family=gaussian,

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

Generalized least squares fit by REML
  Model: ability ~ as.factor(Group) + as.numeric(time) + as.factor(Group) * 
as.numeric(time) 
  Data: stroke_long 
       AIC      BIC    logLik
  1320.321 1346.127 -652.1607

Correlation Structure: AR(1)
 Formula: ~1 | Subject 
 Parameter estimate(s):
      Phi 
0.9495754 

Coefficients:
                                     Value Std.Error   t-value p-value
(Intercept)                       33.39312  7.937178  4.207178  0.0000
as.factor(Group)F                 -0.11518 11.224865 -0.010262  0.9918
as.factor(Group)G                 -6.22568 11.224865 -0.554632  0.5798
as.numeric(time)                   6.07484  0.843600  7.201091  0.0000
as.factor(Group)F:as.numeric(time) -2.14085  1.193030 -1.794467  0.0744
as.factor(Group)G:as.numeric(time) -2.23826  1.193030 -1.876112  0.0622

 Correlation: 
                                   (Intr) as.(G)F as.(G)G as.n() a.(G)F:
as.factor(Group)F                  -0.707                              
as.factor(Group)G                  -0.707  0.500                       
as.numeric(time)                   -0.478  0.338   0.338               
```

```
as.factor(Group)F:as.numeric(time)  0.338 -0.478  -0.239  -0.707
as.factor(Group)G:as.numeric(time)  0.338 -0.239  -0.478  -0.707  0.500

Standardized residuals:
       Min         Q1        Med         Q3        Max
-2.1430431 -0.5861291 -0.2259572  0.6532219  2.8251592

Residual standard error: 21.42606
Degrees of freedom: 192 total; 186 residual
```

3. Unstructure

```
gee_un <-
geeglm(ability~as.factor(Group)+as.numeric(time)+as.factor(Group)*as.numeric(
time),family=gaussian,

data=stroke_long,id=as.factor(Subject),wave=as.numeric(time),corst="unstructur
ed")
un<-corSymm(form = ~ 1 | Subject)
gls.un<-gls(ability~as.factor(Group)+as.numeric(time)+
             as.factor(Group)*as.numeric(time), data=stroke_long,
           correlation=un)
summary(gls.un)

Generalized least squares fit by REML
  Model: ability ~ as.factor(Group) + as.numeric(time) + as.factor(Group) *
as.numeric(time)
  Data: stroke_long
       AIC      BIC    logLik
  1338.118 1451.019 -634.0591

Correlation Structure: General
 Formula: ~1 | Subject
 Parameter estimate(s):
 Correlation:
  1     2     3     4     5     6     7
2 0.931
3 0.868 0.931
4 0.789 0.875 0.952
5 0.708 0.819 0.892 0.913
6 0.576 0.731 0.815 0.855 0.965
7 0.426 0.606 0.693 0.782 0.886 0.945
8 0.319 0.522 0.609 0.707 0.840 0.908 0.975

Coefficients:
                               Value Std.Error   t-value p-value
(Intercept)                 35.71491  7.944761  4.495404  0.0000
```

```
as.factor(Group)F                  -5.51048 11.235588 -0.490449  0.6244
as.factor(Group)G                 -11.30440 11.235588 -1.006125  0.3157
as.numeric(time)                    6.69319  1.166119  5.739712  0.0000
as.factor(Group)F:as.numeric(time)  -3.23684  1.649141 -1.962742  0.0512
as.factor(Group)G:as.numeric(time)  -3.85733  1.649141 -2.338991  0.0204

 Correlation:
                                   (Intr) as.(G)F as.(G)G as.n() a.(G)F:
as.factor(Group)F                  -0.707
as.factor(Group)G                  -0.707  0.500
as.numeric(time)                   -0.760  0.537   0.537
as.factor(Group)F:as.numeric(time)  0.537 -0.760  -0.380  -0.707
as.factor(Group)G:as.numeric(time)  0.537 -0.380  -0.760  -0.707  0.500

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.47104067 -0.50835339 -0.03544122  0.82609510  2.70723745

Residual standard error: 21.27331
Degrees of freedom: 192 total; 186 residual
```

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)

    aic = AIC(gls.exch,gls.ar1,gls.un)

	df	AIC
gls.exch	8	1452.715
gls.ar1	8	1320.321
gls.un	35	1338.118

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.

Model	AIC
Exchangeable	1452
Auto regressive	1320
Unstructure	1338

AIC untuk model auto regressive memiliki nilai yang paling kecil