

Percobaan ke-2: Studi Kasus 2

```
In [ ]: library(DT)      #Menampilkan tabel agar mudah dilihat di browser
library(MVN)      #Uji multivariate normal
library(MASS)     #Fungsi diskriminan analisis
library(biotools) #Melakukan uji Box-M
```

```
Warning message:
"package 'DT' was built under R version 4.3.2"
Warning message:
"package 'MVN' was built under R version 4.3.3"
Warning message:
"package 'biotools' was built under R version 4.3.3"
---
biotools version 4.2
```

```
In [ ]: data <- read.csv("rgb_dataset1.csv")
datatable(data)
```

```
Warning message in file(file, "rt"):
"cannot open file 'rgb_dataset1.csv': No such file or directory"
Error in file(file, "rt"): cannot open the connection
Traceback:

1. read.csv("rgb_dataset1.csv")
2. read.table(file = file, header = header, sep = sep, quote = quote,
.   dec = dec, fill = fill, comment.char = comment.char, ...)
3. file(file, "rt")
```

```
In [ ]: head(data)
```

A data.frame: 6 × 4

	R	G	B	Class
	<int>	<int>	<int>	<int>
1	205	92	92	1
2	240	128	128	1
3	250	128	114	1
4	233	150	122	1
5	255	160	122	1
6	220	20	60	1

```
In [ ]: mvn(data = data[, c(1:3)], multivariatePlot = 'qq') #hanya mengambil kolom varia
```

\$multivariateNormality

A data.frame: 1 × 4

Test	HZ	p value	MVN
<chr>	<dbl>	<dbl>	<chr>
Henze-Zirkler	1.65232	1.064249e-05	NO

\$univariateNormality

A data.frame: 3 × 5

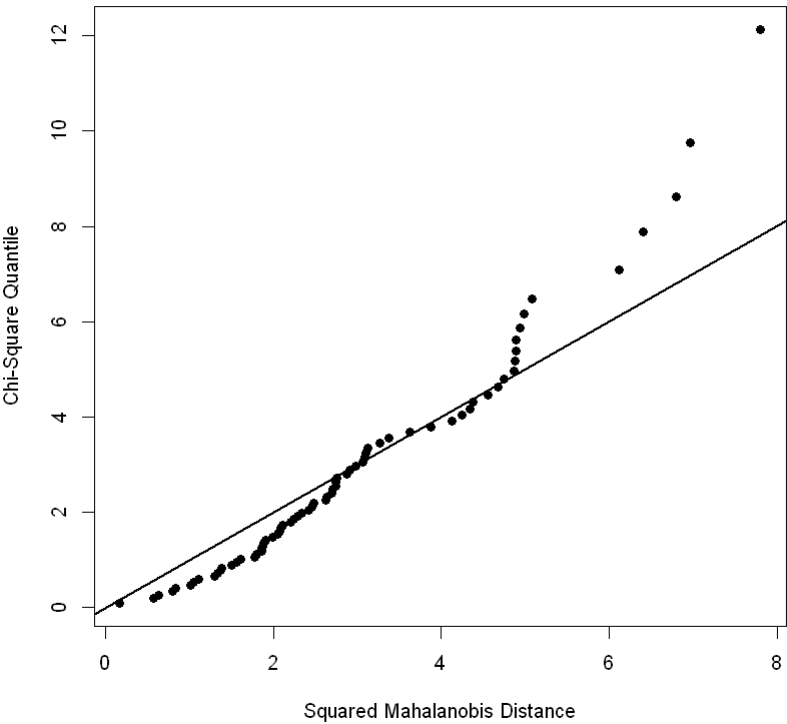
	Test	Variable	Statistic	p value	Normality
	<l<chr>>	<l<chr>>	<l<chr>>	<l<chr>>	<l<chr>>
1	Anderson-Darling	R	1.9242	0.0001	NO
2	Anderson-Darling	G	0.9229	0.0180	NO
3	Anderson-Darling	B	1.3914	0.0012	NO

\$Descriptives

A data.frame: 3 × 10

	n	Mean	Std.Dev	Median	Min	Max	25th	75th
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
R	72	160.2361	81.83933	169.0	0	255	119.75	230.75
G	72	131.7361	80.82149	133.5	0	255	78.75	191.25
B	72	120.2083	78.37288	131.5	0	255	46.50	183.25

Chi-Square Q-Q Plot



Analisis : Berdasarkan hasil plot, diketahui bahwa dataset tersebut tidak berdistribusi normal dikarenakan terdapat penyimpangan di jarak 5 pada garis

lurus 'chi-square qq-plot'.


```
In [ ]: boxM(data = data[, c(1:3)], grouping = data[,4])
```

Box's M-test for Homogeneity of Covariance Matrices

data: data[, c(1:3)]

Chi-Sq (approx.) = 61.604, df = 18, p-value = 1.124e-06

```
In [ ]: set.seed(123)
train_index <- sample(seq(nrow(data)), size = floor(0.75 * nrow(data)), replace
training_data <- data[train_index, ]
test_data <- data[-train_index, ]
```

 **Analisis :** Dalam regresi logistik multinomial memakai library nnet, menghasilkan model training yang mengiterasi optimisasi train model. Optimisasi tersebut bertujuan untuk meminimalkan fungsi objektivitas untuk menemukan model parameter yang optimal. Optimisasi berhenti setelah mencapai kriteria convergence, saat dimana improvisasi objek sudah kecil dan maksimum iterasi tercapai. Nilai AIC yang cenderung rendah menunjukkan bahwa model memiliki kualitas yang baik.

```
In [ ]: linearDA <- lda(formula = Class ~., data = training_data)
linearDA
```

Call:

lda(Class ~ ., data = training_data)

Prior probabilities of groups:

	1	2	3	4
	0.2222222	0.2407407	0.2962963	0.2407407

Group means:

	R	G	B
1	222.9167	73.66667	103.83333
2	159.6154	62.30769	194.07692
3	71.2500	199.25000	93.06250
4	204.3077	147.84615	95.38462

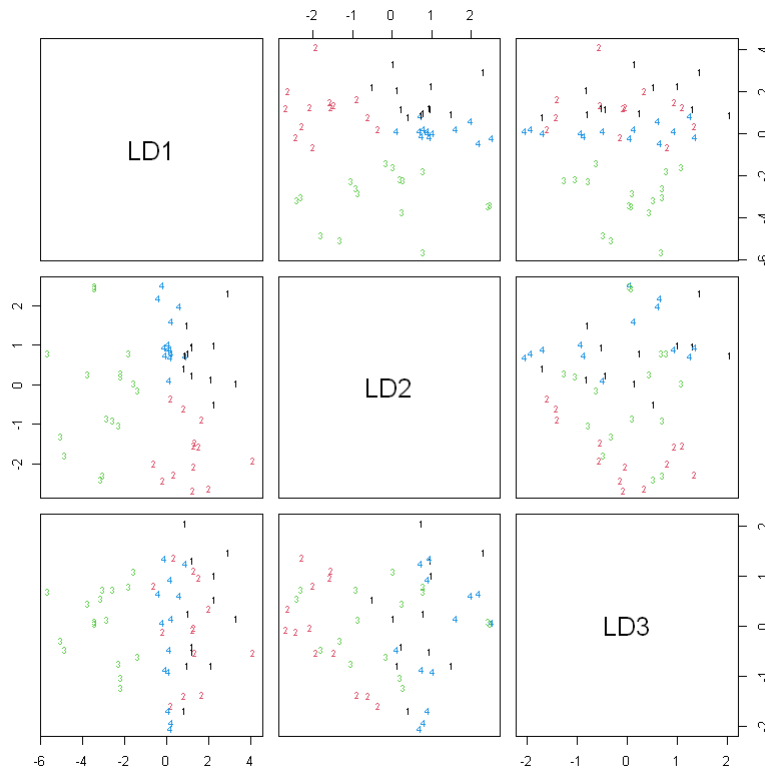
Coefficients of linear discriminants:

	LD1	LD2	LD3
R	0.01767158	0.013529646	-0.005108326
G	-0.01600958	0.007549927	-0.008127256
B	0.00474325	-0.016523544	-0.007870840

Proportion of trace:


	LD1	LD2	LD3
	0.7566	0.2383	0.0051

```
In [ ]: plot(linearDA, col = as.integer(training_data$Class))
```



```
In [ ]: library(dplyr)
#Hitung vektor rata-rata untuk setosa
kelassatu <- training_data %>%
filter(Class == 1)
setosa_mean <- sapply(kelassatu[, -4], mean)
setosa_mean
```

R: 222.916666666667 **G:** 73.6666666666667 **B:** 103.833333333333

 **Analisis :** Regresi multinomial memprediksi probabilitas pengamatan tertentu untuk menjadi bagian dari kelas tertentu. Kolom mewakili tingkat klasifikasi dan baris mewakili pengamatan. 20 baris pertama terklasifikasi sebagai carsinoma dan 20 baris terakhir tergolong sebagai adipose

```
In [ ]: #Menghitung matriks kovarians untuk setosa
kelassatu_cov <- cov(kelassatu[, -5])
kelassatu_cov
```

A matrix: 4 × 4 of type dbl

	R	G	B	Class
R	1401.356	1210.515	1454.348	0
G	1210.515	4377.879	3182.576	0
B	1454.348	3182.576	4511.606	0
Class	0.000	0.000	0.000	0

The Kernel crashed while executing code in the current cell or a previous cell.


Please review the code in the cell(s) to identify a possible cause of the failure.

Click

View Jupyter

```
In [ ]: #Hitung vektor rata-rata untuk class dua
kelasdua <- training_data %>% filter(Class == "2")
kelasdua_mean <- sapply(kelasdua[, -5], mean)
kelasdua_mean
```

R: 159.615384615385 **G:** 62.3076923076923 **B:** 194.076923076923 **Class:** 2

 **Analisis :** Didapat untuk akurasi skor dari tiap tiap konfigurasi dan semua plot yang. Didapat bahwa akurasinya sangat tinggi yaitu 98.68%. Disimpulkan bahwa model bagus dan stabil.

```
In [ ]: #Hitung vektor rata-rata untuk class tiga
kelastiga <- training_data %>% filter(Class == "3")
kelastiga_mean <- sapply(kelastiga[, -5], mean)
kelastiga_mean
```

R: 71.25 **G:** 199.25 **B:** 93.0625 **Class:** 3

```
In [ ]: #Hitung vektor rata-rata untuk class dua
kelasempat <- training_data %>% filter(Class == "4")
kelasempat_mean <- sapply(kelasempat[, -5], mean)
kelasempat_mean
```

R: 204.307692307692 **G:** 147.846153846154 **B:** 95.3846153846154 **Class:** 4

```
In [ ]: #Menghitung matriks kovarians gabungan(S-pooled)
Spooled <- (16 * setosa_cov + 16 * vcolor_cov + 16 * virgin_cov) / (16+16+16)
Spooled
```

Error in eval(expr, envir, enclos): object 'vcolor_cov' not found
Traceback:

```
In [ ]: #Linear Score Function untuk species setosa
#a. Intercept
sb0=-0.5*t(setosa_mean)%*%solve(Spooled) %*% setosa_mean
#b. Koeffisien
sb1=t(setosa_mean) %*% solve(Spooled)
sb0
sb1
```


```
In [ ]: #Linear Score Function untuk species vcolor
#a. Intercept
cb0=-0.5*t(vcolor_mean)%*%solve(Spooled)%*%vcolor_mean
#b. Koeffisien
cb1=t(vcolor_mean) %*% solve(Spooled)
```

```
cb0  
cb1
```

```
In [ ]: #Linear Score Function untuk species virginica  
#a. Intercept  
vb0=-0.5*t(virgin_mean)%%solve(Spoiled) %% virgin_mean  
#b. Koefisien  
vb1=t(virgin_mean) %% solve(Spoiled)  
vb0  
vb1
```

```
In [ ]: # Coba klasifikasi secara manual  
X1<-c(5,3.5,1.6,0.6)  
DL_s = sb0 + sb1[1]*X1[1] + sb1[2]*X1[2] + sb1[3]*X1[3] + sb1[4]*X1[4] + log(0.3)  
DL_c = cb0 + cb1[1]*X1[1] + cb1[2]*X1[2] + cb1[3]*X1[3] + cb1[4]*X1[4] + log(0.2)  
DL_v = vb0 + sb1[1]*X1[1] + sb1[2]*X1[2] + sb1[3]*X1[3] + sb1[4]*X1[4] + log(0.3)  
maks=DL_s  
if (DL_c>maks) maks=DL_c  
if (DL_v>maks) maks=DL_v  
maks  
if(maks==DL_s){print("X1 adalah setosa")  
}else if(maks==DL_c){ print("X1 adalah versicolor")}else {print("X1 adalah virg
```

```
In [ ]: # Coba klasifikasi secara manual  
X1<-c(6.7,3.1,4.4,1.4)  
DL_s = sb0 + sb1[1]*X1[1] + sb1[2]*X1[2] + sb1[3]*X1[3] + sb1[4]*X1[4] + log(0.3)  
DL_c = cb0 + cb1[1]*X1[1] + cb1[2]*X1[2] + cb1[3]*X1[3] + cb1[4]*X1[4] + log(0.2)  
DL_v = vb0 + sb1[1]*X1[1] + sb1[2]*X1[2] + sb1[3]*X1[3] + sb1[4]*X1[4] + log(0.3)  
maks=DL_s  
if (DL_c>maks) maks=DL_c  
if (DL_v>maks) maks=DL_v  
maks  
if(maks==DL_s){print("X1 adalah setosa")  
}else if(maks==DL_c){print("X1 adalah versicolor")}else {print("X1 adalah virgi
```


 **Analisis :** Klasifikasi manual dilakukan dengan menggunakan perhitungan matematis untuk menemukan koefisien yang menonjol dibandingkan yang lainnya.

```
In [ ]: library(klar)  
partimat(Species~., data=training_data, method="lda")
```

```
In [ ]: predicted<-predict(object = linearDA, newdata = test_data)  
table(actual=test_data$Species,predicted=predicted$class)
```

```
In [ ]: quadraticDA<-qda(formula= Species ~., data = training_data)  
quadraticDA
```

```
In [ ]: predicted<-predict(object=quadraticDA,newdata = test_data)  
table(actual=test_data$Species,predicted= predicted$class)
```

 **Analisis :** Dari hasil prediksi yang diperoleh, meskipun model Quadratic Discriminant Analysis (QDA) memberikan probabilitas 'Warna.Dasar' yang serupa dengan Linear Discriminant Analysis (LDA), QDA menunjukkan performa yang lebih

unggul dengan tingkat akurasi mencapai 0,97 (jumlah prediksi yang tepat dibandingkan dengan total data). Lebih lanjut, penggunaan dan perbandingan dengan metode Regresi Logistik Multinomial bisa dijadikan pertimbangan untuk mencapai hasil yang lebih akurat, terutama karena data yang digunakan memiliki distribusi yang tidak normal.