

Praktikum Pemodelan Statistika terapan

Dosen Pengampu Ronny Sutsetyoko



Wahyu Ikbal Maulana

3323600056

D4 SDT B

Politeknik Elektronika Negeri Surabaya

Percobaan ke-1: Studi Kasus 1

```
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("iris")
datatable(iris)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
--------------	-------------	--------------	-------------	---------

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

A data.frame: 6 × 5

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
	<dbl>	<dbl>	<dbl>	<dbl>	<fct>
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

```
In [ ]: mvn(data = iris[, c(1:4)], multivariatePlot = 'qq') #hanya mengambil kolom varia
```

\$multivariateNormality A data.frame: 1 × 4

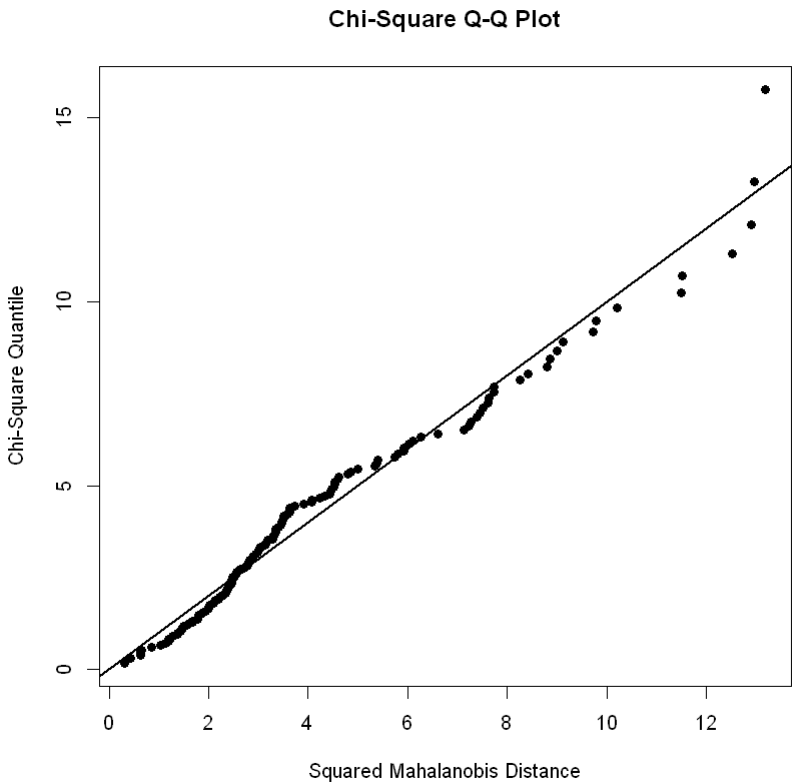
Test	HZ	p value	MVN
<chr>	<dbl>	<dbl>	<chr>
Henze-Zirkler	2.336394	0	NO


\$univariateNormality A data.frame: 4 × 5

	Test	Variable	Statistic	p value	Normality
	<l<chr>>	<l<chr>>	<l<chr>>	<l<chr>>	<l<chr>>
1	Anderson-Darling	Sepal.Length	0.8892	0.0225	NO
2	Anderson-Darling	Sepal.Width	0.9080	0.0202	NO
3	Anderson-Darling	Petal.Length	7.6785	<0.001	NO
4	Anderson-Darling	Petal.Width	5.1057	<0.001	NO

\$Descriptives A data.frame: 4 × 10

	n	Mean	Std.Dev	Median	Min	Max	
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<l<chr>>
Sepal.Length	150	5.843333	0.8280661	5.80	4.3	7.9	
Sepal.Width	150	3.057333	0.4358663	3.00	2.0	4.4	
Petal.Length	150	3.758000	1.7652982	4.35	1.0	6.9	
Petal.Width	150	1.199333	0.7622377	1.30	0.1	2.5	




 **Analisis :** Pengujian Asumsi Multivariate Normal Ketika menguji apakah variabel prediktor berdistribusi multivariate normal. Dengan menggunakan tes Henze-Zirkler untuk data multivariat dan tes Anderson-Darling untuk masing-masing variabel.

```
In [ ]: boxM(data = iris[, c(1:4)], grouping = iris[,5])
```


Box's M-test for Homogeneity of Covariance Matrices

data: iris[, c(1:4)]

Chi-Sq (approx.) = 140.94, df = 20, p-value < 2.2e-16

 **Analisis :** Nilai p-value lebih rendah dari nilai ambang batas yang ditetapkan. Hal ini mengindikasikan bahwa terdapat perbedaan signifikan dalam matriks kovariansi antar kelompok.

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

```
Call:
lda(Species ~ ., data = training_data)
```

Prior probabilities of groups:

setosa	versicolor	virginica
0.3392857	0.2946429	0.3660714

Group means:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	4.976316	3.405263	1.471053	0.2578947
versicolor	5.966667	2.784848	4.303030	1.3393939
virginica	6.585366	2.956098	5.534146	2.0097561

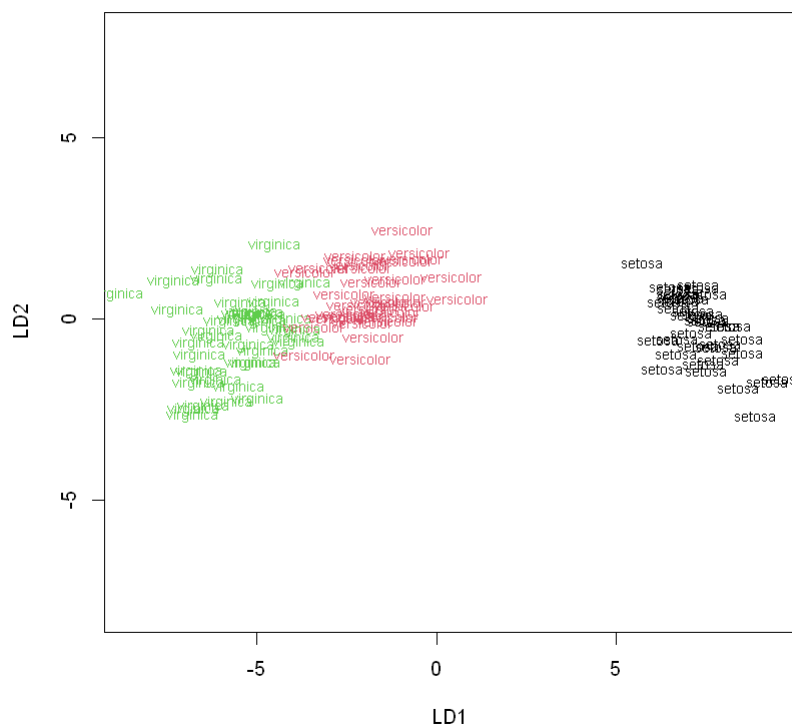
Coefficients of linear discriminants:

	LD1	LD2
Sepal.Length	0.8739864	-0.1767027
Sepal.Width	1.3356317	-1.8714548
Petal.Length	-2.1802288	1.1793326
Petal.Width	-2.7858940	-3.1801017

Proportion of trace:

LD1	LD2
0.993	0.007

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



```
In [ ]: library(dplyr)
#Hitung vektor rata-rata untuk setosa
setosa <- training_data %>%
  filter(Species == "setosa")
setosa_mean <- sapply(setosa[, -5], mean)
setosa_mean
```

Attaching package: 'dplyr'

The following object is masked from 'package:MASS':

select


The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

Sepal.Length: 4.97631578947368 **Sepal.Width:** 3.40526315789474 **Petal.Length:** 1.47105263157895 **Petal.Width:** 0.257894736842105

 **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
setosa_cov <- cov(setosa[, -5])
setosa_cov
```

A matrix: 4 × 4 of type dbl

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.13428876	0.108776671	0.014431010	0.01384068
Sepal.Width	0.10877667	0.159431010	0.008805121	0.01428165
Petal.Length	0.01443101	0.008805121	0.035085349	0.00712660
Petal.Width	0.01384068	0.014281650	0.007126600	0.01331437

```
In [ ]: #Hitung vektor rata-rata untuk versicolor
vcolor <- training_data %>% filter(Species == "versicolor")
vcolor_mean <- sapply(vcolor[, -5], mean)
vcolor_mean
```

Sepal.Length: 5.96666666666667 **Sepal.Width:** 2.78484848484848 **Petal.Length:** 4.3030303030303 **Petal.Width:** 1.33939393939394

 **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 [ ]: vcolor_cov <- cov(vcolor[, -5])
        vcolor_cov
```

A matrix: 4 × 4 of type dbl

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.23229167	0.06104167	0.14697917	0.04197917
Sepal.Width	0.06104167	0.10257576	0.07223485	0.03842803
Petal.Length	0.14697917	0.07223485	0.19092803	0.06581439
Petal.Width	0.04197917	0.03842803	0.06581439	0.04058712

```
In [ ]: #Hitung vektor rata-rata untuk virginica
        virgin <- training_data %>% filter(Species == "virginica")
        virgin_mean <- sapply(virgin[, -5], mean)
        virgin_mean
```

Sepal.Length: 6.58536585365854 **Sepal.Width:** 2.95609756097561 **Petal.Length:** 5.53414634146341 **Petal.Width:** 2.00975609756098

```
In [ ]: #Menghitung matriks kovarians untuk virginica[Date] 12
        virgin_cov <- cov(virgin[, -5])
        virgin_cov
```

A matrix: 4 × 4 of type dbl

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.46778049	0.11634146	0.36226220	0.05839634
Sepal.Width	0.11634146	0.11452439	0.08878659	0.04443902
Petal.Length	0.36226220	0.08878659	0.35130488	0.05640854
Petal.Width	0.05839634	0.04443902	0.05640854	0.06990244

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

A matrix: 4 × 4 of type dbl

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.27812031	0.09538660	0.17455746	0.03807206
Sepal.Width	0.09538660	0.12551039	0.05660885	0.03238290
Petal.Length	0.17455746	0.05660885	0.19243942	0.04311651
Petal.Width	0.03807206	0.03238290	0.04311651	0.04126798

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

```
sb0
sb1
```

A matrix: 1
× 1 of type
dbl
-78.23788

A matrix: 1 × 4 of type dbl

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
22.64578	20.76298	-15.76398	-14.46532

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

A matrix: 1
× 1 of type
dbl
-69.95959

A matrix: 1 × 4 of type dbl

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
14.95042	6.389414	5.041683	8.382122

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

A matrix: 1
× 1 of type
dbl
-99.92159

A matrix: 1 × 4 of type dbl

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
12.09427	3.188286	11.74888	22.76549

```
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

```

A matrix: 1

× 1 of type

dbl

72.68106

[1] "X1 adalah setosa"

```

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

```

A matrix: 1

× 1 of type

dbl

82.69588

[1] "X1 adalah versicolor"

```

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

```

Error in library(klaR): there is no package called 'klaR'
Traceback:

1. library(klaR)

```

In [ ]: install.packages("klaR")

```

Installing package into 'C:/Users/M S I/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)

also installing the dependencies 'miniUI', 'classInt', 'labelled', 'combinat', 'questionr'

Warning message in download.file(url, destfile, method, mode = "wb", ...):

"downloaded length 0 != reported length 0"

Warning message in download.file(url, destfile, method, mode = "wb", ...):

"URL 'https://cran.r-project.org/bin/windows/contrib/4.3/miniUI_0.1.1.1.zip': Timeout of 60 seconds was reached"

Error in download.file(url, destfile, method, mode = "wb", ...) :

download from 'https://cran.r-project.org/bin/windows/contrib/4.3/miniUI_0.1.1.1.zip' failed


```
Warning message in download.packages(pkgs, destdir = tmpd, available = available,
:
"download of package 'miniUI' failed"
Warning message in download.file(url, destfile, method, mode = "wb", ...):
"downloaded length 0 != reported length 0"
Warning message in download.file(url, destfile, method, mode = "wb", ...):
"URL 'https://cran.r-project.org/bin/windows/contrib/4.3/questionr_0.7.8.zip': Ti
meout of 60 seconds was reached"
Error in download.file(url, destfile, method, mode = "wb", ...) :
  download from 'https://cran.r-project.org/bin/windows/contrib/4.3/questionr_0.
7.8.zip' failed
Warning message in download.packages(pkgs, destdir = tmpd, available = available,
:
"download of package 'questionr' failed"
Warning message in download.file(url, destfile, method, mode = "wb", ...):
"downloaded length 0 != reported length 0"
Warning message in download.file(url, destfile, method, mode = "wb", ...):
"URL 'https://cran.r-project.org/bin/windows/contrib/4.3/klaR_1.7-3.zip': Timeout
of 60 seconds was reached"
Error in download.file(url, destfile, method, mode = "wb", ...) :
  download from 'https://cran.r-project.org/bin/windows/contrib/4.3/klaR_1.7-3.zi
p' failed
Warning message in download.packages(pkgs, destdir = tmpd, available = available,
:
"download of package 'klaR' failed"
package 'classInt' successfully unpacked and MD5 sums checked
package 'labelled' successfully unpacked and MD5 sums checked
package 'combinat' successfully unpacked and MD5 sums checked
```

The downloaded binary packages are in

C:\Users\M S I\AppData\Local\Temp\RtmpWoYifP\downloaded_packages

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

	predicted		
actual	setosa	versicolor	virginica
setosa	12	0	0
versicolor	0	16	1
virginica	0	0	9

 **Analisis :** Model yang didapat memiliki keakuratan sebesar 0.97 (jumlah prediksi & aktual benar / jumlah data). Namun, meskipun memiliki tingkat keakuratan yang tinggi, data yang digunakan tidak berdistribusi normal dan matriks kovariannya berbeda, sehingga akurasi tidak sepenuhnya akurat.

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


```
Call:
qda(Species ~ ., data = training_data)
```

```
Prior probabilities of groups:
      setosa versicolor virginica
0.3392857  0.2946429  0.3660714
```

```
Group means:
      Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa      4.976316   3.405263    1.471053    0.2578947
versicolor  5.966667   2.784848    4.303030    1.3393939
virginica    6.585366   2.956098    5.534146    2.0097561
```

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

	predicted		
actual	setosa	versicolor	virginica
setosa	12	0	0
versicolor	0	16	1
virginica	0	0	9

 **Analisis :** Pada hasil yang ditampilkan terdapat peningkatan performa model dari yang awalnya data aktual merah hanya benar 1, sekarang bertambah menjadi 2 data aktual merah yang benar. Ini membuktikan tentang seberapa baik model memprediksi dan meningkat setelah perhitungan kuadratik.

Percobaan ke-2: Studi Kasus 2