

Tugas 6

Analisis Multivariat

Tujuan:

Pada tugas ini, kalian akan melakukan refleksi terhadap pemahaman atas materi perkuliahan mengenai Multivariate Regression dengan melakukan analisis data dan membuat Laporan.

Due date: Hari Sabtu, 30 April 2022 pukul 23.59 WIB @emas 2

Anggota kelompok:

No	Nama	NPM	Kontribusi	Tingkat kontribusi
1	Evan Haryowidyatna	2006485011	Mencari, mengolah, dan menganalisis data lalu membuat laporan.	100%
2	Muhammad Jauhar Hakim	2006463982	Mencari, mengolah, dan menganalisis data lalu membuat laporan.	100%
3	Siskawati Simandalahi	2006572970	Mencari, mengolah, dan menganalisis data lalu membuat laporan.	100%



Laporan Tugas 6

Analisis Regresi Multivariat

Analisis Multivariat Kelas B

Kelompok L

Evan Haryowidyatna	2006485011
--------------------	------------

Muhammad Jauhar Hakim	2006463982
-----------------------	------------

Siskawati Simandalahi	2006572970
-----------------------	------------

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Indonesia

Depok

April 2022

I. Penjelasan Data

Data yang kami peroleh merupakan data kualitas udara. Kami mendapatkan data dari situs <https://archive.ics.uci.edu/ml/datasets/air+quality> Data kualitas udara terdiri dari 15 atribut yaitu:

1. Date (DD/MM/YYYY)
2. Time (HH.MM.SS)
3. True hourly averaged concentration CO in mg/m^3 (reference analyzer)
4. PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)
5. True hourly averaged overall Non Metanic HydroCarbons concentration in microg/m^3 (reference analyzer)
6. True hourly averaged Benzene concentration in microg/m^3 (reference analyzer)
7. PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)
8. True hourly averaged NOx concentration in ppb (reference analyzer)
9. PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NOx targeted)
10. True hourly averaged NO2 concentration in microg/m^3 (reference analyzer)
11. PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO2 targeted)
12. PT08.S5 (indium oxide) hourly averaged sensor response (nominally O3 targeted)
13. Temperature in $^{\circ}\text{C}$
14. Relative Humidity (%)
15. AH Absolute Humidity

II. Tujuan

Analisis regresi multivariat merupakan metode statistik yang memungkinkan dalam meneliti hubungan lebih dari dua variabel secara bersamaan (Ghozali, 2005). Dengan menggunakan teknik analisis ini maka dapat menganalisis pengaruh beberapa variabel terhadap variabel-variabel lainnya dalam waktu yang bersamaan.

III. Pemodelan

Pemodelan analisis regresi multivariat akan diuji lebih dari satu variabel dependen terhadap satu atau lebih variabel independen di mana pada kasus ini analisis regresi multivariat memiliki kelebihan yaitu setiap pengujiannya yang mempertimbangkan hubungan antar variabel dependen satu dengan yang lainnya dalam pembentukan suatu model.

Berikut ini adalah langkah kerja untuk analisis regresi multivariat:

1. Mengimpor data kualitas udara
2. Mendefinisikan data baru
3. Mengganti nama variabel
4. Analisis regresi secara individual variabel dependen Y
5. Analisis regresi full model secara multivariat variabel dependen Y
6. Analisis regresi parsial model secara multivariat variabel dependen Y untuk subset variabel independen (x2, x3, x4, x5, x6, x8, x9, x10)
7. Uji Hipotesis

IV. Analisis

Proses Komputasi R

Install package dan apply library

```
packages <- c("Hmisc", "matlib",  
"Matrix", "expm", "matrixcalc", "ellipsis", "Hotelling", "dplyr", "psych", "Rc  
mdrMisc", "Rcsdp", "mvnrmtest", "factoextra", "cluster", "ggplot2", "tree", "  
class", "car")  
  
if ( length(missing_pkgs <- setdiff(packages,  
rownames(installed.packages()))) > 0) {  
  message("Installing missing package(s): ", paste(missing_pkgs,  
collapse = ", "))  
  install.packages(missing_pkgs)  
}  
  
lapply(packages, library, character.only = TRUE)
```

1. Mengimpor data kualitas udara

```
data <- read.csv("AirQualityUCI.csv", header=TRUE, sep=";")
data <- as.data.frame(data)
head(data)
```

2. Mendefinisikan data baru

```
databaru=data[-1:-2]
head(databaru)
```

3. Mengganti nama variabel

```
colnames(databaru)

databaru <- rename(databaru,

                    x1 = CO.GT.,
                    x2 = PT08.S1.CO.,
                    x3 = NMHC.GT.,
                    x4 = C6H6.GT.,
                    x5 = PT08.S2.NMHC.,
                    x6 = NOx.GT.,
                    x7 = PT08.S3.NOx.,
                    x8 = NO2.GT.,
                    x9 = PT08.S4.NO2.,
                    x10 = PT08.S5.O3.)

head(databaru)
```

4. Analisis regresi secara individual variabel dependen Y

```
mmr <- lm(cbind(T, RH, AH) ~ x1 + x2 + x3 + x4
+ x5 + x6 + x7 + x8 + x9 + x10 , data = databaru)

summary(mmr)

head(resid(mmr))

head(fitted(mmr))

coef(mmr)

sigma(mmr)

vcov(mmr)
```

5. Analisis regresi full model secara multivariat variabel dependen Y

```
library(car)
```

```
Anova(mmr)
```

6. Analisis regresi parsial model secara multivariat variabel dependen Y untuk subset variabel independen (x2, x3, x4, x5, x6, x8, x9, x10)

```
mmr2 <- update(mmr, . ~ . - x1 - x7)
```

```
anova(mmr, mmr2)
```

```
lh.out <- linearHypothesis(mmr, hypothesis.matrix = c("x1 = 0", "x7 =  
0"))
```

```
Lh.out
```

7. Uji Hipotesis

```
E <- lh.out$SSPE
```

```
E
```

```
H <- lh.out$SSPH
```

```
H
```

Hasil Komputasi R

1. Mengimpor data kualitas udara

	Date	Time	CO.GT.	PT08.S1.CO.	NMHC.GT.	C6H6.GT.	PT08.S2.NMHC.	NOx.GT.	PT08.
	<chr>	<chr>	<dbl>	<int>	<int>	<dbl>	<int>	<int>	<int>
1	10/03/2004	18.00.00	2.6	1360	150	11.9	1046	166	1056
2	10/03/2004	19.00.00	2.0	1292	112	9.4	955	103	1174
3	10/03/2004	20.00.00	2.2	1402	88	9.0	939	131	1140
4	10/03/2004	21.00.00	2.2	1376	80	9.2	948	172	1092
5	10/03/2004	22.00.00	1.6	1272	51	6.5	836	131	1205
6	10/03/2004	23.00.00	1.2	1197	38	4.7	750	89	1337

A data.frame: 6 × 10

```
sum(is.na(data))
```

0

2. Mendefinisikan data baru

```
databaru=data[-1:-2]  
head(databaru)
```

	CO.GT.	PT08.S1.CO.	NMHC.GT.	C6H6.GT.	PT08.S2.NMHC.	NOx.GT.	PT08.S3.NOx.	NO2.GT.	PT08.
	<dbl>	<int>	<int>	<dbl>	<int>	<int>	<int>	<int>	<int>
1	2.6	1360	150	11.9	1046	166	1056	113	1692
2	2.0	1292	112	9.4	955	103	1174	92	1559
3	2.2	1402	88	9.0	939	131	1140	114	1555
4	2.2	1376	80	9.2	948	172	1092	122	1584
5	1.6	1272	51	6.5	836	131	1205	116	1490
6	1.2	1197	38	4.7	750	89	1337	96	1393

A data.frame: 6 × 10

3. Mengganti nama variabel

Untuk kelompok variabel dependen Y tidak diubah karena sudah sederhana yaitu

1. T → Temperature (Suhu)
2. RH → Relative Humidity (Kelembapan Relatif)
3. AH → Absolute Humidity (Kelembapan Mutlak)

Untuk kelompok variabel independen X akan diubah menjadi

1. x1 → CO.GT. (True hourly averaged concentration CO in mg/m³ (reference analyzer))
2. x2 → PT08.S1.CO. ((tin oxide) hourly averaged sensor response (nominally CO targeted))
3. x3 → NMHC.GT. (True hourly averaged overall Non Metanic HydroCarbons concentration in microg/m³ (reference analyzer))
4. x4 → C6H6.GT. (True hourly averaged Benzene concentration in microg/m³ (reference analyzer))
5. x5 → PT08.S2.NMHC. ((titania) hourly averaged sensor response (nominally NMHC targeted))
6. x6 → NOx.GT. (True hourly averaged NOx concentration in ppb (reference analyzer))
7. x7 → PT08.S3.NOx. ((tungsten oxide) hourly averaged sensor response (nominally NOx targeted))
8. x8 → NO2.GT. (True hourly averaged NO2 concentration in microg/m³ (reference analyzer))
9. x9 → PT08.S4.NO2. ((tungsten oxide) hourly averaged sensor response (nominally NO2 targeted))
10. x10 → PT08.S5.O3. ((indium oxide) hourly averaged sensor response (nominally O3 targeted))

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	T	RH
	<dbl>	<int>	<int>	<dbl>	<int>	<int>	<int>	<int>	<int>	<int>	<dbl>	<dbl>
1	2.6	1360	150	11.9	1046	166	1056	113	1692	1268	13.6	48.9
2	2.0	1292	112	9.4	955	103	1174	92	1559	972	13.3	47.7
3	2.2	1402	88	9.0	939	131	1140	114	1555	1074	11.9	54.0
4	2.2	1376	80	9.2	948	172	1092	122	1584	1203	11.0	60.0
5	1.6	1272	51	6.5	836	131	1205	116	1490	1110	11.2	59.6
6	1.2	1197	38	4.7	750	89	1337	96	1393	949	11.2	59.2

4. Analisis regresi secara individual variabel dependen Y

Response T :

Call:

```
lm(formula = T ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +  
    x10, data = databaru)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-33.127	-3.497	-0.477	3.377	22.759

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	35.6444866	1.0765154	33.11	<2e-16	***
x1	0.0008322	0.0010666	0.78	0.435	
x2	-0.0170039	0.0008270	-20.56	<2e-16	***
x3	-0.0091767	0.0005161	-17.78	<2e-16	***
x4	1.2135694	0.0066769	181.76	<2e-16	***
x5	-0.0211482	0.0008000	-26.43	<2e-16	***
x6	-0.0146951	0.0005621	-26.14	<2e-16	***
x7	-0.0107328	0.0004707	-22.80	<2e-16	***
x8	0.0228457	0.0009972	22.91	<2e-16	***
x9	0.0202109	0.0003389	59.64	<2e-16	***
x10	-0.0115345	0.0004278	-26.96	<2e-16	***

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

Residual standard error: 5.817 on 9346 degrees of freedom
Multiple R-squared: 0.9819, Adjusted R-squared: 0.9819
F-statistic: 5.067e+04 on 10 and 9346 DF, p-value: < 2.2e-16

Response RH :

Call:

```
lm(formula = RH ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +  
    x10, data = databaru)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-37.060	-9.281	-0.884	8.288	51.460

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	96.7199442	2.4229507	39.918	<2e-16	***
x1	0.0052105	0.0024007	2.170	0.0300	*
x2	0.0364726	0.0018614	19.594	<2e-16	***
x3	0.0019823	0.0011616	1.707	0.0879	.
x4	1.6032394	0.0150280	106.683	<2e-16	***
x5	-0.1562502	0.0018006	-86.776	<2e-16	***
x6	0.0442307	0.0012651	34.963	<2e-16	***
x7	-0.0192171	0.0010595	-18.137	<2e-16	***
x8	-0.0740584	0.0022444	-32.996	<2e-16	***
x9	0.0230033	0.0007627	30.159	<2e-16	***

```

x8          -0.0051878  0.0007939  -6.535 6.70e-11 ***
x9           0.0210593  0.0002698  78.061 < 2e-16 ***
x10         -0.0013970  0.0003406  -4.102 4.13e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.631 on 9346 degrees of freedom
Multiple R-squared:  0.987,    Adjusted R-squared:  0.9869
F-statistic: 7.074e+04 on 10 and 9346 DF,  p-value: < 2.2e-16

```

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

Residual standard error: 13.09 on 9346 degrees of freedom

Multiple R-squared: 0.9347, Adjusted R-squared: 0.9346

F-statistic: 1.338e+04 on 10 and 9346 DF, p-value: < 2.2e-16

Response AH :

Call:

```
lm(formula = AH ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +  
    x10, data = databaru)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-25.2583	-3.0392	0.2959	3.3746	28.1388

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	39.4871674	0.8570077	46.076	< 2e-16	***
x1	0.0028581	0.0008491	3.366	0.000766	***
x2	-0.0094867	0.0006584	-14.409	< 2e-16	***
x3	-0.0076665	0.0004108	-18.660	< 2e-16	***
x4	1.2646904	0.0053155	237.927	< 2e-16	***
x5	-0.0558128	0.0006369	-87.634	< 2e-16	***
x6	0.0020752	0.0004475	4.638	3.57e-06	***
x7	-0.0160390	0.0003748	-42.798	< 2e-16	***

```
head(resid(mmr))
```

	T	RH	AH
1	1.755275	0.5362015	1.963428
2	1.154520	4.8925357	3.380947
3	2.352900	2.9062604	3.614916
4	1.691391	5.2473381	2.348851
5	2.339903	3.3221514	2.023368
6	2.671187	3.8953512	2.604271

A matrix: 6 × 3 of type dbl

```
head(fitted(mmr))
```

	T	RH	AH
1	11.844725	48.36380	-1.205628
2	12.145480	42.80746	-2.655447
3	9.547100	51.09374	-2.864716
4	9.308609	54.75266	-1.562151
5	8.860097	56.27785	-1.234568
6	8.528813	55.30465	-1.819471

A matrix: 6 × 3 of type dbl

5. Analisis regresi full model secara multivariat variabel dependen Y

```
library(car)
Anova(mmr)
```

Warning message in cbind(x\$df, tests, pf(tests[ok, 2], tests[ok, 3], tests[ok, 4], :
"number of rows of result is not a multiple of vector length (arg 3)"

Type II MANOVA Tests: Pillai test statistic

	Df	test stat	approx F	num Df	den Df	Pr(>F)
x1	1	-0.0001	0	3	9344	0.638
x2	1	0.0002	1	3	9344	< 2.2e-16 ***
x3	1	-0.0502	-149	3	9344	< 2.2e-16 ***
x4	1	0.9370	46286	3	9344	1.053e-14 ***
x5	1	0.7041	7413	3	9344	< 2.2e-16 ***
x6	1	0.0073	23	3	9344	< 2.2e-16 ***
x7	1	-0.0060	-19	3	9344	0.638
x8	1	0.0412	134	3	9344	< 2.2e-16 ***
x9	1	-11.3348	-2862	3	9344	< 2.2e-16 ***
x10	1	0.0766	258	3	9344	1.053e-14 ***

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

T: 5.81708980572631 RH: 13.0927270358662 AH: 4.63095158598535

```
vcov(mmr)
```

	T:(Intercept)	T:x1	T:x2	T:x3	T:x4	T:x5
T:(Intercept)	1.158885e+00	1.591840e-05	-5.390961e-04	3.002965e-04	6.930255e-03	-3.068775e-04
T:x1	1.591840e-05	1.137701e-06	5.034156e-09	-5.191252e-08	-3.034643e-07	6.075872e-08
T:x2	-5.390961e-04	5.034156e-09	6.839756e-07	-1.664006e-07	-3.149216e-06	-4.404109e-08
T:x3	3.002965e-04	-5.191252e-08	-1.664006e-07	2.663391e-07	1.620236e-06	-3.459387e-08
T:x4	6.930255e-03	-3.034643e-07	-3.149216e-06	1.620236e-06	4.458139e-05	-2.203184e-06
T:x5	-3.068775e-04	6.075872e-08	-4.404109e-08	-3.459387e-08	-2.203184e-06	6.400215e-08
T:x6	1.369892e-04	-4.646136e-08	-7.119228e-08	7.488221e-08	1.084191e-06	-1.108389e-08
T:x7	-4.326129e-04	-3.782561e-09	8.984970e-08	-8.284048e-08	-2.618802e-06	1.657579e-08
T:x8	-2.011919e-04	-4.299766e-07	7.176172e-08	-1.054699e-07	-1.177694e-06	6.138940e-08
T:x9	1.690394e-05	-1.389970e-08	-4.775214e-08	-3.914123e-09	2.583364e-07	-1.647844e-08
T:x10	2.990513e-05	1.113432e-08	-1.605684e-07	7.781236e-09	1.195076e-07	-1.088750e-08

6. Analisis regresi parsial model secara multivariat variabel dependen Y untuk subset

```
mmr2 <- update(mmr, . ~ . - x1 - x7)
anova(mmr, mmr2)
```

	Res.Df	Df	Gen.var.	Pillai	approx F	num Df	den Df	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	9346	NA	37.55749	NA	NA	NA	NA	NA
2	9348	2	41.36782	0.252164	449.4077	6	18690	0

A anova: 2 × 8

variabel independen (x2, x3, x4, x5, x6, x8, x9, x10)

```
lh.out <- linearHypothesis(mmr, hypothesis.matrix = c("x1 = 0", "x7 = 0"))
lh.out
```

Sum of squares and products for the hypothesis:

	T	RH	AH
T	1538.6408	2698.1802	2274.6997
RH	-3258.3836	-5751.3974	-4833.3904
AH	-310.0926	-560.0352	-465.4882

Sum of squares and products for error:

	T	RH	AH
T	316254.94	-484284.16	72631.41
RH	-484284.16	1602086.66	34952.07
AH	72631.41	34952.07	200431.63

Multivariate Tests:

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	2	-0.0008578	-1.3355282	6	18690	0.47029
Wilks	2	1.0008575	-1.3346245	6	18688	0.47029
Hotelling-Lawley	2	-0.0008565	-1.3337210	6	18686	0.47029
Roy	2	0.0002705	0.8426716	3	9345	0.47029

7. Uji Hipotesis

Matriks E

```
E <- lh.out$SSPE  
E
```

	T	RH	AH
T	316254.94	-484284.16	72631.41
RH	-484284.16	1602086.66	34952.07
AH	72631.41	34952.07	200431.63

A matrix: 3 × 3 of type dbl

Matriks H

```
H <- lh.out$SSPH  
H
```

	T	RH	AH
T	1538.6408	2698.1802	2274.6997
RH	-3258.3836	-5751.3974	-4833.3904
AH	-310.0926	-560.0352	-465.4882

A matrix: 3 × 3 of type dbl

Uji Hipotesis Full Model :

$$H_0: \mathbf{B}_1 = \mathbf{0}$$

Di mana \mathbf{B}_1 merupakan semua elemen \mathbf{B} kecuali elemen pada baris pertama.

Uji Hipotesis untuk Subset X

$$H_0 : \mathbf{B}_d = \mathbf{0}$$

Di mana \mathbf{B}_d merupakan subset dari \mathbf{B}_1 yang akan dihapus atau tidak dipertahankan variabelnya.

V. Kesimpulan

Kami mengawali analisis dengan memastikan data yang kami miliki tidak memiliki missing value. Lalu kami memutuskan untuk tidak mengikut sertakan dua kolom pertama dari data airquality yang kami miliki karena kolom pertama hanya menunjukkan waktu dan pada proses analisis kali ini kami tidak tertarik untuk melihat pengaruh waktu. Setelah memilih data yang akan digunakan, kami merubah penamaan variabel karena penamaan yang terdapat pada data terlalu panjang terutama untuk variabel yang akan kami jadikan variabel independen atau X. Pada variabel Y atau dependen, kami tidak merubah namanya karena menurut kami sudah cukup singkat.

Kemudian kami melanjutkan proses analisis regresi multivariat dengan menggunakan ANOVA. Sehingga didapatkan response T, RH, dan AH secara individual. Pada response T variabel x1 tidak signifikan, pada response RH variabel x3 tidak signifikan, sedangkan pada response AH semua variabel independen x nya signifikan.

Lalu kami melakukan proses analisis regresi multivariat dengan full model, yaitu dengan menggunakan semua variabel independen X dan semua variabel dependen Y secara multivariat dan dipatkan hasilnya yaitu, pengaruh x1 dan x7 tidak terlalu signifikan terhadap kelompok variabel dependen Y secara multivariat.

Setelah itu kami melakukan uji hipotesis regresi multivariat dengan parsial model, yaitu dengan membuang variabel x1 dan x7, dan menetapkannya sebagai hipotesis matriks dengan nilai-nya 0. Karena hipotesis dengan menggunakan 4 statistik uji MANOVA p-value nya melebihi 0.05 semua, maka artinya hipotesis nol tidak ditolak yang artinya benar bahwa variabel x1 dan x7 tidak terlalu berpengaruh terhadap kelompok variabel dependen Y secara multivariat.