

<p><b>Nama:</b> Muhammad Rayyan Naufal</p> <p><b>NIM:</b> 065002300024</p>	 <p>UNIVERSITAS TRISAKTI</p> <p>PRAKTIKUM STATISTIKA</p>	<p><b>MODUL 10</b> <b>STATISTIKA</b></p> <p><b>Nama Dosen:</b> Dedy Sugiarto</p> <p><b>Nama Aslab:</b></p> <ol style="list-style-type: none"><li>1. Tarum Widyasti P (064002200027)</li><li>2. Kharisma Maulida S (064002200024)</li></ol>
<p><b>Hari/Tanggal:</b> Rabu, 5 Juni 2024</p>		

## MODUL 10

### Analysis of Variance (ANOVA)

---

#### Teori Singkat

Analisis ragam atau analysis of variance (ANOVA) merupakan teknik statistik yang dapat digunakan untuk menguji perbedaan rata-rata antar lebih dari 2 grup sampel. Teknik ANOVA sesungguhnya terbagi menjadi beberapa jenis antara lain ANOVA satu arah (one-way ANOVA), ANOVA dua arah (two-way ANOVA), ANOVA tiga arah (three-way ANOVA). Pada sesi ini hanya akan dibahas mengenai teknik ANOVA satu arah. Teknik ANOVA juga dipakai dalam kasus analisis data eksperimen untuk meneliti pengaruh dari baik dari satu faktor (variabel bebas) maupun beberapa faktor terhadap suatu variabel respon (variabel terikat).

#### Lab Setup

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.
2. Menjalankan R Studio.
3. Menjalankan Jupyter

#### 4. Menjalankan Excel

### ELEMEN KOMPETENSI I

Tiga macam metode pencegahan terhadap korosi dari suatu produk, dicoba efektivitasnya. Hasilnya berupa kedalaman korosi (dalam 0.001 inch) adalah sebagai berikut :

Metode A	77	54	67	74	71
Metode B	60	41	59	65	62
Metode C	49	52	69	47	56

Dengan menggunakan alpha 0.05, ujilah bahwa ketiga metode tersebut mempunyai pengaruh yang sama terhadap pencegahan korosi :

#### Ubah data menjadi data bertumpuk pada excel

A	B
metode (x)	korosi (y)
a	77
a	54
a	67
a	74
a	71
b	60
b	41
b	59
b	65
b	62
c	49
c	52
c	69
c	47
c	56

#### Script R:

```
df_nama = read.delim("clipboard")
View(df_nama)
Head(df_nama)
```

```
R 4.0.2 · ~ / ~
> df_rayyan = read.delim("clipboard")
> View(df_nama)
> head(df_nama)
  metode.x korosi.y.
1          a      77
2          a      54
3          a      67
4          a      74
5          a      71
6          b      60
> |
```

## Analisis Anova

Analisis keseluruhan terhadap ada atau tidaknya perbedaan pada metode

### Script R:

```
> model <- aov(korosi.y ~ metode.x, data=df_nama)
> summary(model)
```

```
> model <- aov(korosi.y ~ metode.x, data = df_rayyan)
> summary(model)
  Df Sum Sq Mean Sq F value Pr(>F)
metode.x     2  548.8  274.40   3.348 0.0699 .
Residuals   12  983.6   81.97
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> |
```

## Interpretasi

Berikut adalah langkah-langkah yang dilakukan. Pertama, membuat variabel dataframe baru bernama `df\_rayyan`, kemudian membaca data ke dalam dataframe tersebut. Fungsi `head()` digunakan untuk menampilkan header dari dataframe. Fungsi `aov()` digunakan untuk melakukan analisis varians (ANOVA). Dalam konteks ini, fungsi tersebut digunakan untuk menguji apakah ada perbedaan signifikan dalam variabel respon `korosi.y` berdasarkan variabel prediktor `metode.x` dalam dataframe `df\_vira`. Terakhir, menampilkan ringkasan atau hasil dari perhitungan ANOVA dengan menggunakan fungsi `summary()`.

## Analisis Tukey test

Analisis ada atau tidaknya perbedaan antara 2 metode

**Script R:**

```
tukey.test <- TukeyHSD(model)
tukey.test
```

```
> tukey.test <- TukeyHSD(model)
> tukey.test
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = korosi.y ~ metode.x, data = df_rayyan)

$metode.x
   diff      lwr      upr      p adj
b-a -11.2 -26.47609  4.076092 0.1657711
c-a -14.0 -29.27609  1.276092 0.0736936
c-b  -2.8 -18.07609 12.476092 0.8778031
```

**Interpretasi**

Fungsi untuk melakukan Uji Tukey adalah metode post-hoc yang digunakan setelah ANOVA untuk melakukan perbandingan ganda antara grup-grup guna menentukan grup mana yang berbeda secara signifikan satu sama lain. Ini sangat berguna jika hasil ANOVA menunjukkan adanya perbedaan signifikan antara grup secara keseluruhan. Hasil uji ini akan menampilkan perbandingan dalam hal difference, lower, upper, dan p-value adjustment.

**Screnshoot full screen**

## Muhammad Rayyan Naufal - 065002300024

The screenshot shows the RStudio interface. The left pane displays the R console output:

```
R> names(df_rayyan)
[1] "metode.x" "korosi.y"
> head(df_rayyan)
#>   metode.x korosi.y
#> 1       a        77
#> 2       a        54
#> 3       a        67
#> 4       a        74
#> 5       a        71
#> 6       b        60
> model <- aov(korosi.y ~ metode.x, data = df_rayyan)
> summary(model)
#> 
#>             Df Sum Sq Mean Sq F value Pr(>F)
#> metode.x    2  548.8  274.40  3.348 0.0699 .
#> Residuals 12  983.6   81.97
#> 
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #analisis turkey test
> tukey.test <- TukeyHSD(model)
> tukey.test
#> Tukey multiple comparisons of means
#> 95% family-wise confidence level
#> 
#> Fit: aov(formula = korosi.y ~ metode.x, data = df_rayyan)
#> 
#> $metode.x
#>   diff      lwr      upr   p adj
#> b-a -11.2 -26.47609  4.076092 0.1657711
#> c-a -14.0 -29.27609  1.276092 0.0736936
#> c-b -2.8 -18.07609 12.476092 0.8778031
```

The right pane shows the RStudio environment browser with the following objects listed:

- df\_rayyan\_clean: 8 obs. of 2 variables
- manajemen\_data: 45 obs. of 4 variables
- model: List of 13
- model\_reg: List of 12
- Prak6rayyan: 16 obs. of 1 variable
- rayyan: 8 obs. of 1 variable
- rayyan2: 10 obs. of 1 variable
- test\_result: List of 10
- tukey.test: List of 1

## Python

### Source code:

```
import numpy as np
import scipy.stats as stats

# Data waktu produksi dari ketiga mesin
metode_A = np.array([77, 54, 67, 74, 71])
metode_B = np.array([60, 41, 59, 65, 62])
metode_C = np.array([49, 52, 69, 47, 56])

# Gabungkan data ke dalam satu array
data_tarum = [metode_A, metode_B, metode_C]

# Hitung ANOVA menggunakan scipy.stats
f_statistic, p_value = stats.f_oneway(metode_A, metode_B, metode_C)

# Tampilkan hasil
print("Nilai F: {f_statistic}")
print("Nilai p: {p_value}")
```

### Output:

```
import numpy as np
import scipy.stats as stats

# Data waktu produksi dari ketiga mesin
metode_A = np.array([77, 54, 67, 74, 71])
metode_B = np.array([60, 41, 59, 65, 62])
metode_C = np.array([49, 52, 69, 47, 56])

# Gabungkan data ke dalam satu array
data_tarum = [metode_A, metode_B, metode_C]

# Hitung ANOVA menggunakan scipy.stats
f_statistic, p_value = stats.f_oneway(metode_A, metode_B, metode_C)

# Tampilkan hasil
print(f"Nilai F: {f_statistic}")
print(f"Nilai p: {p_value}")

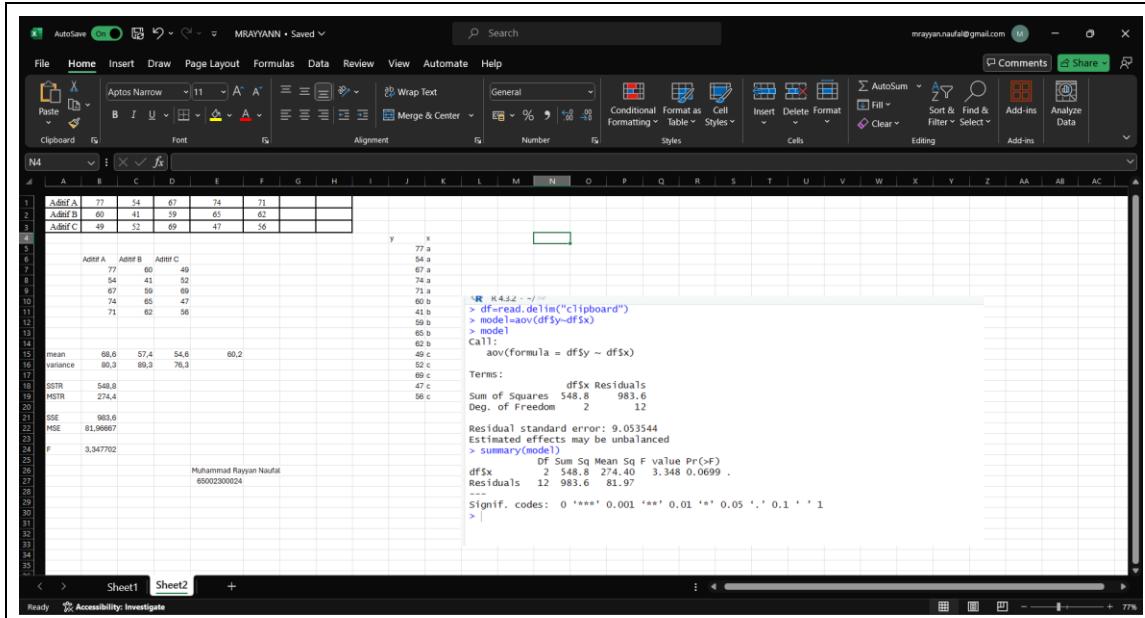
[1] ✓ 5.4s
...
Nilai F: 3.347702318015454
Nilai p: 0.06993237223084404
```

## Interpretasi

Langkah-langkah dalam analisis data ini dimulai dengan mengimpor library, yaitu numpy untuk operasi array dan scipy.stats untuk fungsi statistik. Kemudian, tiga array NumPy yang berisi waktu produksi dari tiga mesin yang berbeda (metode\_A, metode\_B, metode\_C) dibuat. Data waktu produksi dari ketiga mesin tersebut kemudian disusun ke dalam satu list (data\_tarum). Setelah itu, fungsi `f\_oneway` dari scipy.stats digunakan untuk melakukan ANOVA satu arah pada ketiga set data, yang menghasilkan nilai statistik F dan nilai p (p-value). Terakhir, nilai F dan nilai p dicetak ke konsol untuk menentukan apakah ada perbedaan signifikan dalam waktu produksi antara ketiga mesin.

## Excel:

## Muhammad Rayyan Naufal - 065002300024



### Interpretasi

Setelah memasukkan data ke dalam Excel, hitunglah mean menggunakan fungsi `AVERAGE()` dan varians menggunakan fungsi `VAR.S()`. Kemudian, hitung SSTR berdasarkan nilai mean yang telah diperoleh. Setelah itu, hitung MSTR dengan membagi SSTR dengan 2. Selanjutnya, hitung SSE berdasarkan nilai varians yang ada, dan kemudian hitung MSE dengan membagi SSE dengan jumlah data. Akhirnya, hitung nilai F dengan membagi MSTR dengan MSE.

## ELEMEN KOMPETENSI II

Gunakan dataset plant growth yang telah tersedia di R untuk meneliti pengaruh beberapa treatment (perlakuan) terhadap tingkat pertumbuhan tanaman.

### Menampilkan Data

#### Script R:

```
df_nama=PlantGrowth
View(df_rayyan)
head(df_rayyan)
```

#### Output:

```
> df_rayyan=PlantGrowth  
> View(df_rayyan)  
> head(df_rayyan)  
  weight group  
1 4.17  ctrl  
2 5.58  ctrl  
3 5.18  ctrl  
4 6.11  ctrl  
5 4.50  ctrl  
6 4.61  ctrl  
>
```

## Analisis Anova

### Script R:

```
model <- aov(weight ~ group, data = df_rayyan)  
summary(model)
```

### Output:

```
> model <- aov(weight ~ group, data = df_rayyan)  
> summary(model)  
   Df Sum Sq Mean Sq F value Pr(>F)  
group        2 3.766  1.8832  4.846 0.0159 *  
Residuals    27 10.492  0.3886  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Interpretasi

Langkah-langkahnya adalah sebagai berikut: Pertama, kita akan membuat sebuah variabel dataframe baru yang dinamakan `df\_vira`. Kemudian, kita akan menggunakan fungsi `read` untuk membaca dataframe. Setelah itu, kita akan menggunakan fungsi `head` untuk menampilkan header dari dataframe tersebut. Selanjutnya, kita akan menggunakan fungsi `aov()` untuk melakukan analisis varians (ANOVA). Dalam konteks ini, kita ingin menguji apakah ada perbedaan signifikan dalam variabel respon korosi (`y`) berdasarkan variabel prediktor metode (`x`) dalam dataframe `df\_nama`. Terakhir, kita akan menampilkan summary atau hasil dari perhitungan ANOVA tersebut.

## Analisis Tukey test

### Source code:

```
tukey.test <- TukeyHSD(model)  
tukey.test
```

**Output:**

```
> tukey.test <- TukeyHSD(model)
> tukey.test
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = weight ~ group, data = df_rayyan)

$group
      diff      lwr      upr      p adj
trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
trt2-ctrl  0.494 -0.1972161 1.1852161 0.1979960
trt2-trt1  0.865  0.1737839 1.5562161 0.0120064
```

**Interpretasi**

Setelah dilakukan analisis ANOVA untuk menentukan apakah terdapat perbedaan signifikan antara grup-grup, Uji Tukey merupakan metode post-hoc yang digunakan untuk melakukan perbandingan ganda antara grup-grup tersebut. Tujuannya adalah untuk menentukan grup mana yang memiliki perbedaan signifikan satu sama lain. Uji Tukey menjadi khususnya bermanfaat apabila analisis ANOVA menunjukkan bahwa terdapat perbedaan signifikan secara keseluruhan antara grup-grup tersebut. Hasil dari Uji Tukey akan menampilkan perbedaan antar grup (difference), batas bawah (lower), batas atas (upper), dan nilai p yang sudah disesuaikan (p adjustment) untuk perbandingan antar grup tersebut.

**Screnshoot full screen**

## Muhammad Rayyan Naufal - 065002300024

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The main area contains a console window with R code and its output. The code performs an ANOVA test on the 'weight' variable grouped by 'ctrl'. It also includes a Tukey HSD post-hoc test. The right side of the interface shows the Environment pane listing various objects like 'df\_rayyan\_clean', 'model', and 'tukey.test'. The bottom status bar shows the date and time.

```
R 4.3.2 -->
> df_rayyan<-PlantGrowth
> view(df_rayyan)
> head(df_rayyan)
#> #> weight group
#> 1 5.17 ctrl1
#> 2 5.58 ctrl1
#> 3 5.18 ctrl1
#> 4 6.11 ctrl1
#> 5 4.50 ctrl1
#> 6 4.61 ctrl1
>
> model <- aov(weight ~ group, data = df_rayyan)
> summary(model)
Df Sum Sq Mean Sq F value Pr(>F)
group          2  3.766  1.8832  4.846 0.0159 *
Residuals     27 10.492  0.3886
...
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> tukey.test <- TukeyHSD(model)
> tukey.test
Tukey multiple comparisons of means
 95% family-wise confidence level
Fit: aov(formula = weight ~ group, data = df_rayyan)

$group
    diff      lwr      upr   p adj
trt1-ctrl -0.371 -1.0622161  0.3202161 0.3908711
trt2-ctrl  0.494 -0.1972161  1.1852161 0.1979960
trt2-trt1  0.865  0.1737839  1.5562161 0.0120064
>
```

## Python: Source code:

```
import numpy as np
import scipy.stats as stats
# Data waktu produksi dari ketiga mesin
plant_a = np.array ([4.17, 5.58, 5.18, 6.11, 4.50, 4.61, 5.17, 4.53, 5.33, 5.14])
plant_b = np.array ([4.81, 4.17, 4.41, 3.59, 5.87, 3.83, 6.03, 4.89, 4.32, 4.69])
plant_c = np.array ([6.31, 5.12, 5.54, 5.50, 5.37, 5.29, 4.92, 6.15, 5.80, 5.26])
# Gabungkan data ke dalam satu array
data_rayyan = [plant_a, plant_b, plant_c]
# Hitung ANOVA menggunakan scipy.stats
f_statistic, p_value = stats.f_oneway(plant_a, plant_b, plant_c)
# Tampilkan hasil
print(f"Nilai F: {f_statistic}")
print(f"Nilai p: {p_value}")
```

## Output:

## Muhammad Rayyan Naufal - 065002300024

```
▶ [1]
import numpy as np
import scipy.stats as stats
# Data waktu produksi dari ketiga mesin
plant_a = np.array ([4.17, 5.58, 5.18, 6.11, 4.50, 4.61, 5.17, 4.53, 5.33, 5.14])
plant_b = np.array ([4.81, 4.17, 4.41, 3.59, 5.87, 3.83, 6.03, 4.89, 4.32, 4.69])
plant_c = np.array ([6.31, 5.12, 5.54, 5.50, 5.37, 5.29, 4.92, 6.15, 5.80, 5.26])
# Gabungkan data ke dalam satu array
data_rayyan = [plant_a, plant_b, plant_c]
# Hitung ANOVA menggunakan scipy.stats
f_statistic, p_value = stats.f_oneway(plant_a, plant_b, plant_c)
# Tampilkan hasil
print(f"Nilai F: {f_statistic}")
print(f"Nilai p: {p_value}")

[1]
...
Nilai F: 4.846087862380136
Nilai p: 0.015909958325622895
```

### Interpretasi

`import numpy as np` dan `import scipy.stats as stats` digunakan untuk mengimpor modul yang diperlukan dalam analisis data dan perhitungan statistik. Data waktu produksi dari tiga mesin disimpan dalam array numpy `plant\_a`, `plant\_b`, dan `plant\_c`. Data dari ketiga mesin tersebut digabungkan ke dalam satu array `data\_rayyan`. Fungsi `stats.f\_oneway` dari modul scipy digunakan untuk menghitung nilai F dan p-value dari uji ANOVA. Hasil nilai F dan p-value kemudian ditampilkan menggunakan fungsi `print`.

### Excel:

The screenshot shows an Excel spreadsheet with data and R code output. The data is organized into three columns: plant\_a, plant\_b, and plant\_c, with each column containing 10 data points. An R code output is displayed below the data, showing the results of an ANOVA analysis. The R code includes the following lines:

```
> model <- aov(weight ~ group, data = df_rayyan)
> summary(model)
```

The output shows the following summary statistics:

	DF	Sum Sq	Mean Sq	F value
group	2	3.766	1.8832	4.846
Residuals	27	10.492	0.3886	
group		0.0159	*	
Residuals				

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

The R code also includes the following lines at the bottom:

```
mean variance
SSTR MSTR
SSE MSE
MSE F
```

The R code output ends with the name "Muhammad Rayyan Naufal" and the ID "65002300024".

### Interpretasi

Setelah memasukkan data ke dalam Excel, hitunglah mean menggunakan fungsi `AVERAGE()` dan varians menggunakan fungsi `VAR.S()`. Kemudian, hitung SSTR berdasarkan nilai mean yang telah diperoleh. Setelah itu, hitung MSTR dengan membagi SSTR dengan 2. Selanjutnya, hitung SSE berdasarkan nilai varians yang ada, dan kemudian hitung MSE dengan membagi SSE dengan jumlah data. Akhirnya, hitung nilai F dengan membagi MSTR dengan MSE.

Sumber :

<http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

<https://rpubs.com/aaronsc32/post-hoc-analysis-tukey>

Github

<https://github.com/rayyan-naufal/Uni/tree/main/S2/ProbabilitasStatistika/Praktikum/Modul10>

### CEK LIST (✓)

1. Memahami analisis ragam. (✓)

### KESIMPULAN

Pada Praktikum ini, kita mempelajari dan menggunakan uji Tukey serta uji ANOVA melalui perangkat lunak R, Python, dan Excel, yang memberikan pemahaman yang komprehensif tentang analisis data dan statistik di berbagai platform pemrograman.

### FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian (menit)
Memahami analisis ragam.	3	3	33

#### Keterangan Tingkat Kesulitan

- 1: Sangat Mudah
- 2: Mudah
- 3: Biasa
- 4: Sulit
- 5: Sangat Sulit

#### Keterangan Tingkat Ketertarikan

- 1: Tidak Tertarik
- 2: Cukup Tertarik
- 3: Tertarik
- 4: Sangat Tertarik