

Praktikum Eksplorasi dan Visualisasi Data

Pertemuan 8

Analisis Variansi

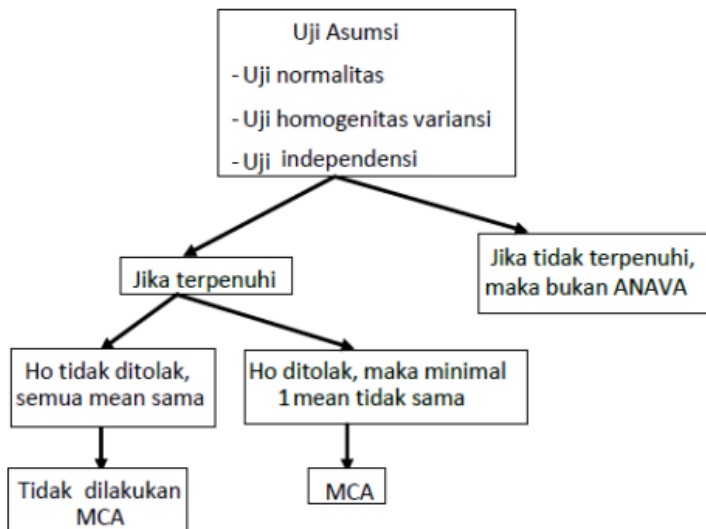
Analisis variansi atau sering disebut dengan ANOVA (Analysis of Variance) adalah salah satu metode analisis inferensi statistika yang digunakan untuk menguji hipotesis rata-rata lebih dari dua angkatan. Dalam analisis variansi dikenal adanya variabel bebas/perlakuan (bersifat kategorik) dan variabel tak bebas/terikat (bersifat numerik).

Ada berbagai jenis analisis variansi, salah satunya adalah Analisis Variansi Satu Arah (OneWay ANOVA) yaitu analisis variansi yang hanya memperhitungkan satu faktor saja. Tujuan dari Analisis Variansi Satu Arah adalah untuk mengetahui pengaruh atau efek terhadap variabel dependen yang disebabkan oleh suatu faktor yang terdiri dari beberapa level faktor yang jumlahnya berhingga. Disebut Anava Satu Arah karena hanya ada satu faktor yang dipelajari dalam penelitian. Adapun Anava Satu Arah memiliki dua model, yakni model efek tetap dan model efek random. Pada kesempatan ini akan dipelajari model efek tetap, yang bertujuan mengambil kesimpulan untuk level-level faktor/angkatan yang ada.

Tabel hasil observasi dengan k perlakuan :

1	2	...	i	...	k
x_{11}	x_{21}	...	x_{i1}	...	x_{k1}
x_{12}	x_{22}	...	x_{i2}	...	x_{k2}
...
x_{1n_1}	x_{2n_2}	...	x_{in_1}	...	x_{kn_1}

Alur Analisis



Uji Asumsi

Dalam analisis variansi, terdapat 3 asumsi sifat data yang harus dipenuhi, yaitu :

1. Distribusi Normal,

k kelompok perlakuan secara umum berasal dari populasi berdistribusi normal.

(Pengujian dilakukan dengan melihat bentuk dari boxplot beberapa angkatan. Jika asumsi data berdistribusi normal belum terpenuhi, maka dilakukan transformasi beberapa angkatan sesuai prosedur yang sudah dipelajari di pertemuan sebelumnya.)

2. Kesamaan Variansi

Masing-masing angkatan berasal dari populasi yang variansinya sama.

Secara Deskriptif :

Pengujian dilakukan dengan menghitung variansi data tiap angkatan, lalu dilihat apakah nilai variansi tertinggi nilainya 3 kali lebih besar dari nilai variansi yang terendah. Langkah-langkah pengujiannya sama seperti yang dipelajari di pertemuan sebelumnya.

Secara Inferensi :

Pengujian dilakukan dengan menggunakan Levene's Test

sintaks :

```
leveneTest(formula, dataset)
```

Parameters:

formula: a formula of the form values ~ groups

dataset: a matrix or data frame

3. Independensi Data

Pengambilan sampel dilakukan secara random/acak (tidak saling memberikan pengaruh).

(Pengujian dilakukan dengan melihat apakah diketahui pengambilan sampel dilakukan secara acak atau random. Jika tidak diketahui pada soal, maka dapat diasumsikan sudah terpenuhi.)

Uji Hipotesis One Way ANOVA

Adapun, hipotesis yang digunakan dalam analisis variansi satu arah model tetap yaitu :

Apabila ingin diketahui perbedaan rata-rata

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ (tidak ada perbedaan mean)

$H_1 : \mu_i \neq \mu_j$ (minimal ada sepasang mean yang berbeda)

Apabila ingin mengetahui adanya efek/pengaruh perlakuan

$H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$ (tidak ada efek/pengaruh)

$H_1 : \text{minimal ada satu } \tau_i \neq 0, \text{ dengan } i=1,2,\dots,k$ (ada efek/pengaruh)

Note:

Perlu dipahami bawah pemilihan hipotesis hanya **berdasarkan konteks persoalan** yang ada pada kasus yang dihadapi. Tujuan dari kedua hipotesis yang ada sama, yakni melihat dan membandingkan menggunakan nilai rata-rata data.

Didalam R akan digunakan fungsi aov,

```
aov(formula, data = NULL, projections = FALSE, qr = TRUE,  
     contrasts = NULL, ...)
```

Arguments

<code>formula</code>	A formula specifying the model.
<code>data</code>	A data frame in which the variables specified in the formula will be found. If missing, the variables are searched for in the standard way.
<code>projections</code>	Logical flag: should the projections be returned?
<code>qr</code>	Logical flag: should the QR decomposition be returned?
<code>contrasts</code>	A list of contrasts to be used for some of the factors in the formula. These are not used for any <code>Error</code> term, and supplying contrasts for factors only in the <code>Error</code> term will give a warning.
<code>...</code>	Arguments to be passed to <code>lm</code> , such as <code>subset</code> or <code>na.action</code> . See 'Details' about <code>weights</code> .

Langkah-Langkah Uji Hipotesis

a. Hipotesis

$H_0 : \dots$

$H_1 : \dots$

(Diisi sesuai konteks yang ada pada soal)

Dengan :

$\tau_1/\mu_1 : \dots$

$\tau_2/\mu_2 : \dots$

\dots

$\tau_k/\mu_k : \dots$

b. Tingkat signifikansi :

$\alpha = \dots$ (Jika tidak diketahui pada soal, maka digunakan $\alpha = 0.05$)

c. Statistik uji :

P-Value = \dots

d. Daerah kritik :

H_0 ditolak jika $P\text{-Value} < \alpha = \dots$

e. Kesimpulan :

Karena $P\text{-Value} = \dots < \alpha = \dots$ maka H_0 **ditolak**. Artinya, dengan tingkat konfidensi $\dots\%$ disimpulkan bahwa \dots

Interpretasi

(Interpretasi dari Uji Hipotesis Anava Satu Arah)

Uji Perbandingan Ganda (Multiple Comparison Analysis (MCA))

Dilakukan setelah dilakukannya uji ANOVA untuk melihat rata-rata populasi mana yang benar-benar berbeda.

Syarat dapat dilakukannya MCA : jumlah level faktor lebih dari dua dan diperoleh kesimpulan H_0 ditolak pada pengujian uji hipotesis.

H_0 : tidak ada perbedaan yang signifikan

H_1 : terdapat perbedaan yang signifikan

Beberapa Metode yang dapat digunakan

1. TukeyHSD

```
TukeyHSD(x, which, ordered = FALSE, conf.level = 0.95, ...)
```

Lebih jelasnya silahkan Run, `help("TukeyHSD")`

Akan terdapat kolom

- diff: perbedaan rata-rata diantara dua grup
- lwr, upr: batas bawah dan atas confidence interval at 95% (default)
- p adj: p-value setelah disesuaikan.

2. "bonferroni", "holm", "hochberg", "hommel", "BH", "BY", "fdr", "none"

```
pairwise.t.test(x, g, p.adjust.method = p.adjust.methods,  
                pool.sd = !paired, paired = FALSE,  
                alternative = c("two.sided", "less", "greater"),  
                ...)
```

Arguments

<code>x</code>	response vector.
<code>g</code>	grouping vector or factor.
<code>p.adjust.method</code>	Method for adjusting p values (see p.adjust).
<code>pool.sd</code>	switch to allow/disallow the use of a pooled SD
<code>paired</code>	a logical indicating whether you want paired t-tests.
<code>alternative</code>	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". Can be abbreviated.
<code>...</code>	additional arguments to pass to <code>t.test</code> .

```
pairwise.t.test(variabel_dependen, variabel_independen, p.adj =  
"bonferroni", paired=F)
```

Contoh Kasus

1. Tiga proses pengerasan yang berbeda sebagai hasil perendaman air-garam, perendaman minyak, dan perendaman air murni pada sampel- sampel campuran logam tertentu yang diambil secara acak. Hasil ditunjukkan pada tabel di bawah ini,

Tabel Hasil Pengerasan pada Satuan Tertentu

Rendaman air garam	Rendaman minyak	Rendaman air murni
145	152	150
150	150	149
153	147	152
148	155	151
141	140	142
152	146	141
146	158	145
154	152	152
139	151	153
148	143	142

Apakah ada pengaruh jenis rendaman terhadap hasil pengerasan ?

Penyelesaian:

Diketahui:

Tiga proses pengerasan yang berbeda sebagai hasil perendaman air-garam, perendaman minyak, dan perendaman air murni pada sampel- sampel campuran logam tertentu yang diambil secara acak Data hasil pengerasan pada satuan tertentu.

Ditanya:

Apakah ada pengaruh jenis rendaman terhadap hasil pengerasan?

Jawab:

Untuk mengetahui apakah ada pengaruh jenis rendaman terhadap hasil pengerasan, dapat dilakukan uji one way anova, karena hanya ada satu faktor saja yaitu jenis rendaman. Sebelumnya, data harus memenuhi asumsi yang berlaku kemudian barulah dilakukan uji hipotesis untuk mengetahui ada tidaknya pengaruh jenis rendaman terhadap hasil pengerasan.

Berikut syntax dalam software R untuk menyelesaikan permasalahan ini.

```
#memanggil library
```

```
library(openxlsx)
```

```
library(tidyverse)
```

```
library(AID)
```

```
library(MASS)
```

```
library(car)
```

```
library(plotly)
```

#Menyiapkan data

```
anova<-read.xlsx("D:\\KULIAH\\Asprak\\P-8\\Pertemuan 8.xlsx")
```

```
anova1<-anova%>%
```

```
  gather(key="jenis.rendaman",value="hasil.pengerasan")
```

#Cek asumsi

#1. Normalitas data

```
cek_nor<-boxplot(anova[1:3],col=c("green","yellow","pink"),
                 border="blue",
                 xlab="jenis rendaman",
                 ylab="satuan tertentu",
                 main="Cek Normalitas")
```

#Transformasi Data

#1. Transformasi Box-cox dengan library AID

```
out = boxcoxnc(anova1$hasil.pengerasan, method = "mle",
               lambda = seq(-5,5,0.001), verbose = F, plot = F)
out$lambda.hat
```

#2. Transformasi Box-cox dengan library MASS

```
out      =      boxcox(anova1$hasil.pengerasan~1,      lambda      =
seq(-5,5,0.0001), plotit = F)
out$x[which.max(out$y)]
```

#3. Transformasi Box-cox dengan library car

```
out      =      powerTransform(anova1$hasil.pengerasan,      family      =
"bcPower")
out$lambda
```

#setiap dantum dipangkatkan dengan 5

```
transformasi.bc=function(data, lambda){
  for(i in 1:length(data)){
    data[i]=data[i]^lambda
  }
  return(data)
}
```

#data hasil transformasi

```
anova2=anova1 %>%
```

```
mutate(transformasi=transformasi.bc(anova1$hasil.pengerasan,5))%
>%
dplyr::select(-hasil.pengerasan)
```

#cek normalitas setelah data ditransformasi

```
cek_norm1<-ggplot(anova2, aes(x=jenis.rendaman,y=transformasi))+
  geom_boxplot(fill="blue",color="pink")+
  ggtitle("Cek Normalitas")+
  xlab("Jenis Rendaman")+
  ylab("Satuan tertentu")
ggplotly(cek_norm1)
```

#2. Asumsi Kesamaan Variansi

```
cek_var<-anova2 %>%
  group_by(jenis.rendaman) %>%
  summarise(var(transformasi))
```

#jika var tertinggi > 3 variansi rendah, maka terjadi heterogenitas variansi

```
max(cek_var$`var(transformasi)`)
```

>

```
3*min(cek_var$`var(transformasi)`)
```

#3. Independensi data, diasumsikan terpenuhi

#Uji Anova

```
anova3<-aov(transformasi~jenis.rendaman,data=anova2)
anova3
```

```
summary(anova3)
```

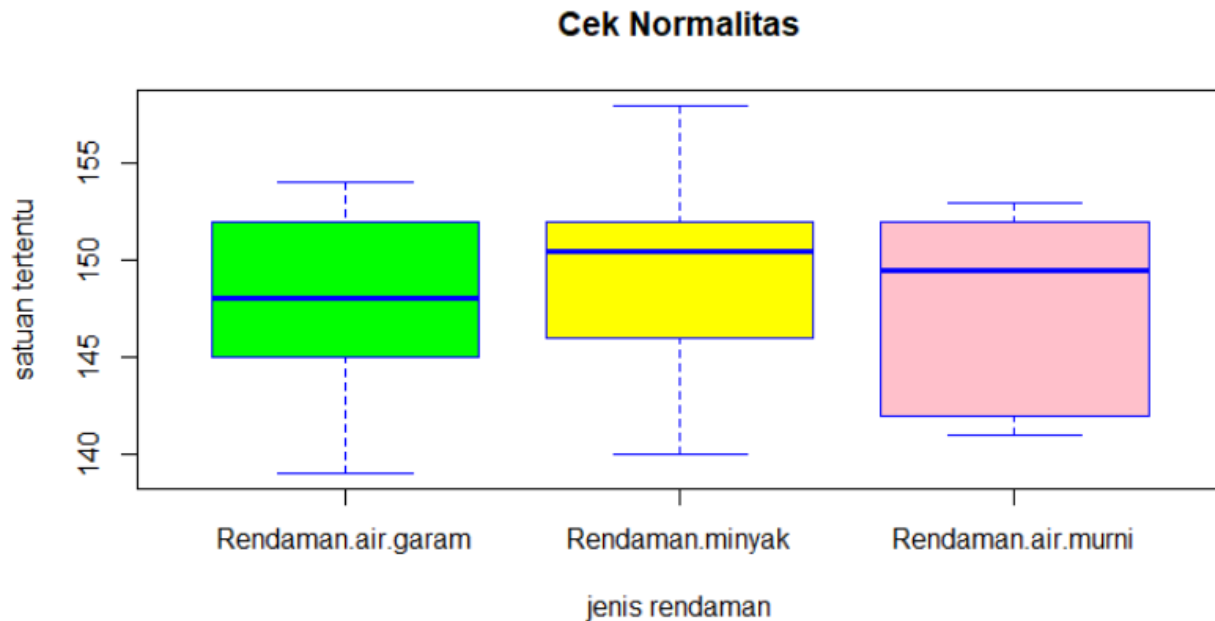
#F tabel

```
qf(0.95,2,27)
```

1) Uji Asumsi

a) Normalitas Dilakukan uji asumsi normalitas menggunakan boxplot.

Ouput boxplot sebagai berikut,



Interpretasi:

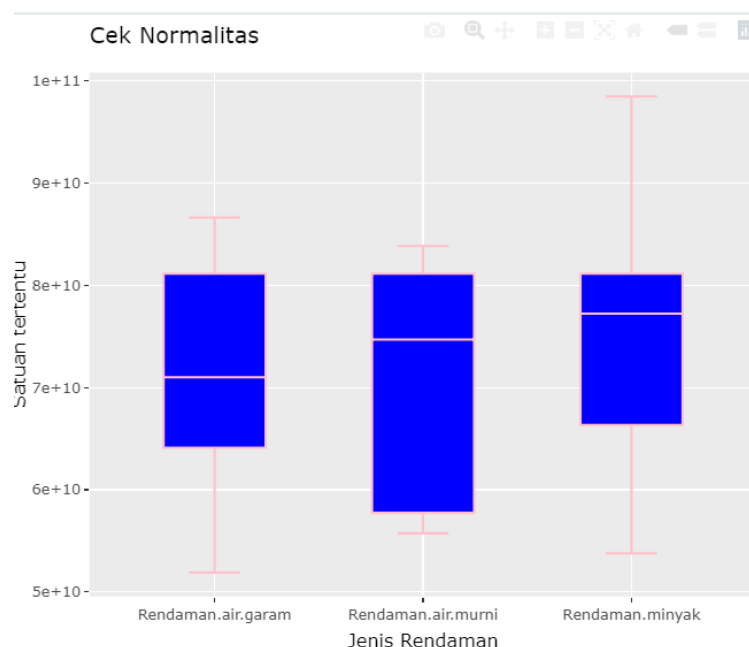
Dari boxplot di atas, dapat dilihat distribusi data hasil pengerasan dari rendaman air garam, rendaman minyak, dan rendaman air murni. Pada boxplot rendaman air garam, dapat dilihat letak median berada mendekati di tengah. Hal ini menunjukkan sebaran nilai atas dan nilai bawah mendekati merata. Selain itu, tidak terdapat outlier. Oleh karena itu dapat disimpulkan bahwa data rendaman air garam berdistribusi mendekati normal. Sedangkan pada boxplot data rendaman minyak, dapat dilihat bahwa median lebih dekat dengan kuartil atas atau dengan kata lain nilai rendah lebih menyebar. Hal ini menunjukkan bahwa distribusi data rendaman minyak menjurai ke bawah. Selain itu, tidak ada outlier. Selanjutnya, pada boxplot rendaman air murni, terlihat bahwa median lebih dekat dengan kuartil atas atau dengan kata lain nilai rendah lebih menyebar. Hal ini menunjukkan bahwa distribusi data rendaman air murni menjurai ke bawah. Walaupun boxplot untuk data rendaman air garam sudah berdistribusi normal, namun tetap perlu dilakukan transformasi karena setiap level faktor harus mengikuti distribusi normal. Selanjutnya dilakukan transformasi menggunakan metode box-cox.

Transformasi menggunakan metode box-cox

```
> #Transformasi Data
> #1. Transformasi Box-cox dengan library AID
> out = boxcoxnc(anova1$hasil.pengerasan, method = "mle",
+               lambda = seq(-5,5,0.001), verbose = F, plot = F)
> out$lambda.hat
[1] 4.638
> #2. Transformasi Box-cox dengan library MASS
> out = boxcox(anova1$hasil.pengerasan~1, lambda = seq(-5,5,0.0001), plotit
+ F)
> out$x[which.max(out$y)]
[1] 4.6382
>
> #3. Transformasi Box-cox dengan library car
> out = powerTransform(anova1$hasil.pengerasan, family = "bcPower")
> out$lambda
anova1$hasil.pengerasan
4.638202
```

Dari boxcox di atas diperoleh nilai lamda mendekati 5. Transformasi dilakukan dengan cara memangkatkan setiap datum pada data dengan 5.

Berikut boxplot hasil transformasi boxcox,



Interpretasi:

Dari boxplot di atas, dapat dilihat distribusi data hasil pengerasan dari rendaman air garam, rendaman minyak, dan rendaman air murni. Pada boxplot rendaman air garam, dapat dilihat letak median dekat dengan tengah. Hal ini menunjukkan sebaran nilai atas dan nilai bawah hampir merata. Selain itu, tidak terdapat outlier. Oleh karena itu dapat disimpulkan bahwa data rendaman air garam mendekati distribusi normal. Sedangkan pada boxplot data rendaman minyak, dapat dilihat bahwa median lebih dekat dengan kuartil atas atau dengan kata lain nilai

rendah lebih menyebar. Hal ini menunjukkan bahwa distribusi data rendaman minyak menjurai ke bawah. Selain itu, tidak ada outlier. Selanjutnya, pada boxplot rendaman air murni, terlihat bahwa median lebih dekat dengan kuartil atas atau dengan kata lain nilai rendah lebih menyebar. Hal ini menunjukkan bahwa distribusi data rendaman air murni menjurai ke bawah. Dapat disimpulkan bahwa dengan dilakukannya transformasi, tidak mengubah juraian dari boxplot sebelumnya. **Akan tetapi, dengan mempertimbangkan bahwa tidak ditemukan outlier pada seluruh boxplot sehingga kondisi ini memungkinkan untuk mengasumsikan data mendekati distribusi normal. Sehingga diasumsikan asumsi normalitas terpenuhi. Untuk keperluan analisis selanjutnya, digunakan data hasil transformasi.**

b). Kesamaan Variansi

```
> #2. Asumsi Kesamaan Variansi
> cek_var<-anova2 %>%
+   group_by(jenis.rendaman) %>%
+   summarise(var(transformasi))
> cek_var
# A tibble: 3 x 2
  jenis.rendaman   `var(transformasi)`
  <chr>           <dbl>
1 Rendaman.air.garam 1.34e20
2 Rendaman.air.murni 1.21e20
3 Rendaman.minyak    1.83e20
```

Interpretasi:

Dari tabel tersebut terlihat bahwa variansi tertinggi yaitu transformasi_rendaman minyak sebesar 1.83e20. Dimana nilai tersebut tidak lebih tinggi dari 3 kali variansi terendah yaitu pada transformasi_rendaman air murni yaitu $3 \times 1.21e20 = 3.63e20$ sehingga dapat disimpulkan bahwa terdapat kesamaan variansi antara ketiga angkatan. Maka asumsi kesamaan variansi terpenuhi.

Bisa memanfaatkan type data logical di R

```
> #jika var tertinggi > 3 variansi rendah, maka terjadi heterogenitas variansi
> max(cek_var$`var(transformasi)` > 3*min(cek_var$`var(transformasi)`))
[1] FALSE
```

FALSE artinya variansi tertinggi < 3 kali variansi terendah, maka asumsi kesamaan variansi terpenuhi.

c) Independensi

Karena sampel diambil secara acak, maka asumsi independensi atau kerandoman data terpenuhi.

Karena asumsi normalitas, kesamaan variansi, dan independensi terpenuhi, selanjutnya dapat dilakukan analisis variansi satu arah untuk mengetahui ada tidaknya pengaruh jenis rendaman terhadap hasil pengerasan.

2) One way Anova

a. Uji Hipotesis

$H_0: \tau_1 = \tau_2 = \tau_3$ (tidak ada pengaruh jenis rendaman terhadap hasil pengerasan)

H_1 : minimal ada $\tau_i \neq 0$, $i=1,2,3$ (ada pengaruh jenis rendaman terhadap hasil pengerasan) dengan

τ_1 : perlakuan ke-1 yaitu rendaman air garam

τ_2 : perlakuan ke-2 yaitu rendaman minyak

τ_3 : perlakuan ke-3 yaitu rendaman air murni

(Diambil hipotesis ini karena ingin mengetahui efek perlakuan)

b. Tingkat signifikansi

$\alpha = 0.05$

c. Statistik uji

```
> #Uji Anova
```

```
> anova3<-aov(transformasi~jenis.rendaman,data=anova2)
```

```
> anova3
```

```
Call:
```

```
  aov(formula = transformasi ~ jenis.rendaman, data = anova2)
```

```
Terms:
```

	jenis.rendaman	Residuals
Sum of Squares	1.336832e+20	3.936908e+21
Deg. of Freedom	2	27

```
Residual standard error: 12075239226
```

```
Estimated effects may be unbalanced
```

```
>
```

```
> summary(anova3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
jenis.rendaman	2	1.337e+20	6.684e+19	0.458	0.637
Residuals	27	3.937e+21	1.458e+20		

p-value = 0.637

F = 0.458

d. Daerah kritik

H_0 ditolak jika $p\text{-value} < \alpha$ atau $F > F_{\alpha,k-1,n-k}$

```
> #F tabel
```

```
> qf(0.95,2,27)
```

```
[1] 3.354131
```

e. Kesimpulan

Karena $p\text{-value} = 0.637 > \alpha = 0.05$ atau $F = 0.458 < F_{\text{crit}} = 3.354131$ maka H_0 tidak ditolak, atau dengan kata lain **tidak ada pengaruh jenis rendaman terhadap hasil pengerasan.**

Interpretasi:

Untuk mengetahui apakah ada pengaruh perlakuan dalam hal ini jenis rendaman terhadap hasil pengerasan, dapat dilakukan analisis variansi satu arah. Karena ketiga asumsi sudah terpenuhi, maka dapat dilakukan uji hipotesis. Diambil hipotesis nol yang menyatakan bahwa tidak ada pengaruh jenis rendaman terhadap hasil pengerasan, dan hipotesis alternatifnya ada pengaruh jenis rendaman terhadap hasil pengerasan. Tingkat signifikansi sebesar 5% atau 0.05. Daerah kritik pada uji ini adalah H_0 ditolak jika $p\text{-value} < \alpha$ atau $F > F_{\alpha, k-1, n-k}$. Karena $p\text{-value} = 0.637 > \alpha = 0.05$ atau $F = 0.458 < F_{\text{crit}} = 3.354131$ maka H_0 tidak ditolak, atau dengan kata lain **tidak ada pengaruh jenis rendaman terhadap hasil pengerasan.**

Latihan

1. Digunakan data *chickwts* dengan dua variabel *weight* dan *feed*. Peneliti ingin mencari apakah terdapat perbedaan pakan ayam terhadap pertumbuhan berat badan anak ayam. Apa yang bisa disimpulkan untuk membantu peneliti tersebut?

Catatan

- untuk menjawab soal ini digunakan data *chickwts*, data bawaan R
 - Data pada kasus ini telah diuji dan berdistribusi normal, gunakan data asli untuk menjawab permasalahan ini
 - Pada kasus ini lakukan MCA dengan metode TukeyHSD
2. Suatu pabrik tekstil mempunyai sejumlah besar perkakas tenun. **Setiap perkakas tenun dianggap memberikan hasil kain per menit yang sama.** Untuk menyelidiki anggapan tersebut, lima perkakas tenun dipilih secara acak dan keluaran dari kelima perkakas tersebut dicatat pada waktu yang berbeda-beda. Dalam upaya penyelidikan, kelima perkakas tersebut dioperasikan pada kondisi dan tempat serta waktu yang sama. Dengan demikian, rancangan dasar yang dipakai adalah RAL (Rancangan Acak Lengkap). Data yang diperoleh sebagai berikut :

Perkakas	Hasil Kain (pon/kain)				
	1	2	3	4	5
1	14	14.1	14.2	14	14.1
2	13.9	13.8	13.9	14	14
3	14.1	14.2	14.1	14	13.9
4	13.6	13.8	14	13.9	13.7
5	13.8	13.6	13.9	13.8	14

Lakukan analisis dengan taraf signifikansi 5%.

3. Diadakan penelitian untuk menentukan pengaruh nutrisi pada perhatian anak SD. Diambil 15 anak secara random pada 3 pola sarapan pagi : tanpa sarapan, sarapan ringan, dan full sarapan. Perhatian mereka terhadap pelajaran dalam menit dicatat dalam tabel berikut :

Tabel Lama Perhatian siswa saat pembelajaran (menit)

Tanpa sarapan	Sarapan ringan	Full Sarapan
8	14	10
7	16	12
9	12	16
13	17	15
10	11	12

Lakukan Anava-One Way untuk data di atas dengan $\alpha = 0.05$.

4. A chemical engineer wants to compare the hardness of four blends of paint. Six samples of each paint blend were applied to a piece of metal. The pieces of metal were cured. Then each sample was measured for hardness. In order to test for the equality of means and to assess the differences between pairs of means, the analyst uses one-way ANOVA with multiple comparisons.

Paint	Hardness
Blend 2	14.9
Blend 3	13
Blend 4	15
Blend 1	17
Blend 1	13.9
Blend 3	16.4
Blend 1	10.4
Blend 2	3.2
Blend 2	1.9
Blend 2	7.3
Blend 4	17.8
Blend 2	9.6
Blend 4	22.9
Blend 4	17.4
Blend 3	13.3
Blend 1	19.3
Blend 1	16.1
Blend 3	6.5
Blend 1	11.7
Blend 4	16.9
Blend 4	18.4
Blend 3	11.9
Blend 2	14.5
Blend 3	16.8