Review Materi

Praktikum Eksplorasi dan Visualisasi Data Kelas B

Pengambilan Sampel Random Sederhana

- Di dalam R, fungsi yang digunakan untuk mengambil sampel adalah fungsi sample()
- Ada 2 cara pengambilan sampel, yaitu Pengambilan Sampel dengan Pengembalian (Sampling with Replacement) dan tanpa pengembalian (Sampling without Replacement)

```
#Contoh pengambilan data sampel
pop <- 1:10
#pengambilan sampel dengan pengembalian
sample1 <- sample(pop,size=5,replace=T)
sample1

#pengambilan sampel tanpa pengembalian
sample2 <- sample(pop,size=5,replace=F)
sample2
```

Distribusi Peluang Normal

`pnorm()` = Untuk mencari probabilitas kumulatif dari distribusi normal. [$P(X \le x)$]

`dnorm()` = Untuk mencari densitas probabilitas dari distribusi normal. [$P(X = x)$]

`qnorm()` = Untuk mencari invers dari probabilitas kumulatif. [$x = P^{-1}(p), 0 \le p \le 1$]

Dalam R dapat diperoleh dengan syntax

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Distribusi Peluang Normal

UNIVERSITAS GADJAH MADA

Probabilitas Densitas

```
> #Probability density dari distribusi normal mean = 47.6
> #dan sd = 10.3
> #P(X=50)
> dnorm(50,mean=47.6,sd=10.3)
[1] 0.03769495
```

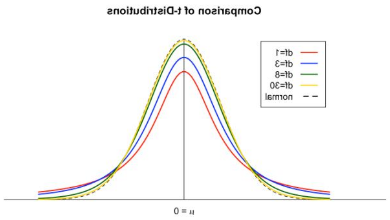
Probabilitas Kumulatif

```
> #Cumulative Distribution dari distribusi normal mean= 47.6 dan sd = 10.3
> #P(X>47.6)
> pnorm(47.6,mean=47.6,sd=10.3,lower.tail = F)
[1] 0.5
```

Invers Probabilitas Kumulatif

```
> #x?
> #Suatu distribusi Normal dengan mean= 47.6 dan sd = 10.3
> #P(X<=x)=1-0.85=0.15
> qnorm(0.15,mean=47.6,sd=10.3,lower.tail = T)
[1] 36.92474
> qnorm(0.85,mean=47.6,sd=10.3,lower.tail = F)
[1] 36.92474
```

Distribusi Peluang t



Usage

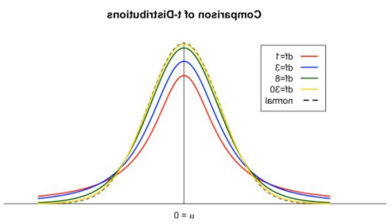
```
dt(x, df, ncp, log = FALSE)
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp)
```

Probabilitas Densitas

```
#b. calculate the probability density function at t=-4,2,0,2,4
with df=5
x<-seq(from=-4, to=4, by=2)
a<-dt(x,5)
a
```

```
> a
[1] 0.005123727 0.065090310 0.379606690 0.065090310 0.005123727
```

Distribusi Peluang t



Peluang Kumulatif

```
# Mencari P(X<=-2), P(X<=-0), dan P(X<=2) dengan df 5
df <- 5
ji <- c(-2,0,2)
pt(ji, df = df, lower.tail = TRUE)
```

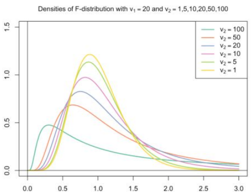
```
> pt(ji, df = df, lower.tail = TRUE)
[1] 0.05096974 0.50000000 0.94903026
```

Invers Probabilitas Kumulatif

```
#Mencari nilai x jika P(X<=x)=0.05096974 dengan df 5
qt(0.05096974,5)
```

```
> qt(0.05096974,5)
[1] -2
```

Distribusi Peluang F



Usage

```
df(x, df1, df2, ncp, log = FALSE)
pf(q, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2, ncp)
```

Probabilitas Densitas

```
df(1.2, df1 = 10, df2 = 20)
```

```
> df(1.2, df1 = 10, df2 = 20)
[1] 0.5626125
```

Distribusi Peluang F

Probabilitas Kumulatif

```
x = 1.5
v1 = 10
v2 = 20

# interval [0,1.5]
pf(x, df = v1, df2 = v2, lower.tail = TRUE)
> pf(x, df = v1, df2 = v2, lower.tail = TRUE)
[1] 0.7890535
```

Invers Probabilitas Kumulatif

```
#Nilai probabilitas
q <- c(0.25, 0.5, 0.75, 0.999)
#df 1
v1=10
#df2
v2=20
```

Distribusi Peluang F

```
#Mencari nilai x dari P(X<=x)=0.25 dengan df1 10 dan df2 20
qf(q[1], df1 = v1, df2 = v2, lower.tail = TRUE)
> qf(q[1], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 0.6563936
```

```
#
qf(q[2], df1 = v1, df2 = v2, lower.tail = TRUE)
> qf(q[2], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 0.9662639
```

```
#
qf(q[3], df1 = v1, df2 = v2, lower.tail = TRUE)
> qf(q[3], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 1.399487
```

Uji Hipotesis Rata-Rata Populasi Satu Angkatan

Prosedur Uji

1. Uji Normalitas Data
2. Uji Hipotesis Rata-rata Populasi Satu Angkatan

Dalam Uji Hipotesis Rata-rata Populasi Satu Angkatan umumnya digunakan Uji Z atau Uji T.

No	Uji Z	Uji T
1.	Variansi atau standar deviasi populasi (σ) diketahui.	Variansi atau standar deviasi populasi (σ) tidak diketahui.
2.	Variansi atau standar deviasi populasi dapat diacu dari penelitian sebelumnya.	Variansi atau standar deviasi populasi dapat diestimasi dari variansi atau standar deviasi sampel (s).
3.	Distribusi populasi tidak diketahui atau sembarang (bisa berdistribusi normal atau tidak), dan ukuran sampel (n) besar ($n \geq 30$).	Distribusi populasi sembarang. Ukuran sampel (n) besar ($n \geq 30$). Namun jika populasi berdistribusi normal, dapat digunakan untuk data kecil maupun besar.

Uji Hipotesis Rata-Rata Populasi Satu Angkatan

*Catatan :

- Jika terdapat kasus data tidak berdistribusi normal dan ukuran sampel kecil, maka data tersebut ditransformasi agar berdistribusi normal, sehingga dapat dilakukan uji T.

- Jika Variansi atau standar deviasi populasi (σ) tidak diketahui dan sampel besar lazimnya menggunakan uji Z. Namun, uji T juga bisa digunakan.

Uji Z

Dalam R library yang digunakan untuk melakukan uji z adalah library BSDA dengan fungsi z.test() dan zsum.test()

Uji Hipotesis Rata-Rata Populasi Satu Angkatan

Uji Z

```
z.test(
  x,
  y = NULL,
  alternative = "two.sided",
  mu = 0,
  sigma.x = NULL,
  sigma.y = NULL,
  conf.level = 0.95
)
```

```
zsum.test(
  mean.x,
  sigma.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  sigma.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  conf.level = 0.95
)
```

UNIVERSITAS GADJAH MADA

Uji Hipotesis Rata-Rata Populasi Satu Angkatan

Uji t

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

```
tsum.test(
  mean.x,
  s.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  s.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  var.equal = FALSE,
  conf.level = 0.95
)
```

Uji Hipotesis Rata-Rata Populasi Dua Angkatan

Uji hipotesis rata-rata populasi dua angkatan dibagi menjadi dua macam, yaitu :

- Uji hipotesis rata-rata populasi dua angkatan independen
- Uji hipotesis rata-rata populasi dua angkatan dependen (berpasangan)

Uji Rata-rata Populasi 2 Angkatan Independen	Uji Rata-rata Populasi 2 Angkatan Dependen
Kedua data berdistribusi **normal**. Kedua sampel berasal dari dua populasi yang **berbeda**.	Kedua data berdistribusi **normal**. Kedua sampel berasal dari populasi yang **sama**, namun diberi **perlakuan yang berbeda**.

Uji Hipotesis Rata-Rata Populasi Dua Angkatan Independen

Prosedur Uji

1. Uji Normalitas Data
2. Uji Kesamaan 2 Variansi

   Apabila nilai variansi yang tertinggi > 3 kali nilai variansi yang terendah, maka dianggap terdapat **perbedaan** variansi.

3. Uji Hipotesis

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

Note :

- paired = FALSE
- var.equal = TRUE, jika ada kesamaan variansi antara 2 sampel dan FALSE jika sebaliknya.

UNIVERSITAS GADJAH MADA

Uji Hipotesis Rata-Rata Populasi Dua Angkatan Dependen

Prosedur Uji

1. Uji Normalitas Data
2. Uji Hipotesis

```
t.test(x, y, alternative = c("two.sided", "less", "greater"),
mu = 0, paired = TRUE, conf.level = 0.95, ...)
```

UNIVERSITAS GADJAH MADA

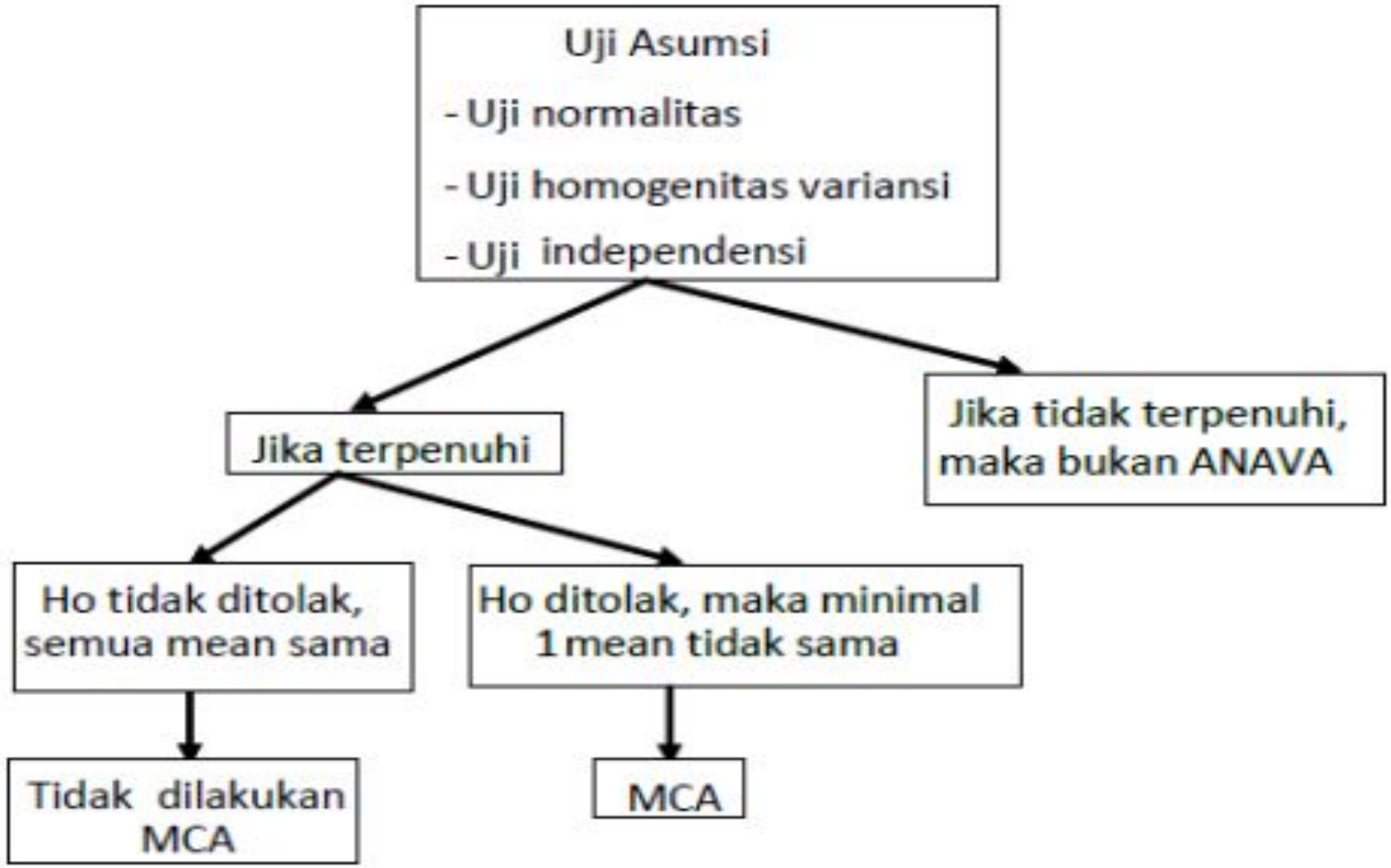
Catatan

1. Jika data tidak berdistribusi normal lakukan transformasi data berapapun jumlah sampelnya
2. Jika data hasil transformasi lebih buruk dari data asli, gunakan data asli, kecuali jika di soal terdapat keterangan normalitas data diasumsikan terpenuhi.

Analisis Variansi

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.

Analisis Variansi



Analisis Variansi

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.

2. Kesamaan Variansi

Masing-masing angkatan berasal dari populasi yang variansinya sama.

3. Independensi Data

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

Analisis Variansi

Uji Hipotesis One Way ANOVA

Apabila ingin diketahui perbedaan rata-rata

H0 : $\mu 1 = \mu 2 = \cdots = \mu k$ (tidak ada perbedaan mean)

H1 : $\mu i \neq \mu j$ (minimal ada sepasang mean yang berbeda)

Apabila ingin mengetahui adanya efek/pengaruh perlakuan

H0 : $\tau 1 = \tau 2 = \cdots = \tau k = 0$ (tidak ada efek/pengaruh)

H1 : minimal ada satu $\tau i \neq 0$, dengan i=1,2,...,k (ada efek/pengaruh)

Analisis Variansi

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 H0 ditolak pada pengujian uji hipotesis anova.

H0 : tidak ada perbedaan yang signifikan

H1 : terdapat perbedaan yang signifikan

UNIVERSITAS GADJAH MADA

Analisis Variansi

Uji MCA yang paling sering digunakan adalah 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.

Regresi Linear Sederhana

Analisis regresi merupakan suatu analisis yang terjadi di antara dua variabel, yaitu variabel independen (prediktor) dan variabel dependen (respon) dimana variabel prediktor diasumsikan mempengaruhi variabel respon secara linear, sehingga variabel respon dapat diduga dari variabel prediktor

Regresi Linear Sederhana

Alur Analisis Regresi Linear Sederhana

1. Uji Asumsi
   - Variabel dependen merupakan variabel kontinu
   - Linearitas

     adanya hubungan linear antara variabel dependen dengan variabel independen. Pengecekan asumsi ini dapat dilihat secara eksploratif dengan menggunakan scatter plot antara variabel dependen dan variabel independen.
   - Variabel dependen berdistribusi normal

     digunakan boxplot untuk melakukan uji asumsi normalitas.

Analisis Regresi Linear

2. Analisis Regresi Linear Sederhana

- Model Summary

  Untuk melihat ringkasan informasi dari model yang didapatkan. Untuk praktikum ini cukup nilai R, R-Square, Adjusted Square, dan Std. Error of the Estimate

- Uji Overall

  Untuk melihat apakah model layak digunakan atau tidak

- Uji Parsial :

  Uji Parsial Konstanta : untuk menentukan apakah konstanta layak masuk kedalam model

  Uji Parsial Koefisien : untuk menentukan apakah koefisien layak masuk kedalam model

Analisis Regresi Linear

3. Model Regresi

$$\widehat{Y} = \widehat{\beta 0} + \widehat{\beta 1}$$

Interpretasi :

- setiap penambahan/ pengurangan 1 satuan x maka nilai variabel y akan bertambah/ berkurang sebesar β dengan asumsi variabel lain konstan.
- Jika variabel independen bernilai nol, maka nilai variabel dependen adalah … (konstanta).

*Notes : Jika data yang diregresi adalah data hasil transformasi, maka model ‘dikembalikan’ ke data asli.

TERIMA KASIH

Sampai Jumpa di lain praktikum :)