

Pengenalan Software R untuk Analisis Statistika

Departemen Statistika - FMIPA IPB

Kampus IPB Darmaga, 20 Maret 2018

Pengenalan R-environtment

Import Data Eksternal dan Melihat Isi Data

Menampilkan Statistik Deskriptif Numerik

Menampilkan Tabel Frekuensi dan Tabulasi Silang Variabel Kategorik



Menampilkan Distribusi Data Numerik

Visualisasi Sederhana

Uji-t Satu Populasi

Uji-t Dua Populasi Saling Bebas

One-Way ANOVA dan Post-hoc Test



Pendahuluan

Apa itu R?

R adalah implementasi sebuah komputasi dan pemrograman bahasa statistika

Instalasi

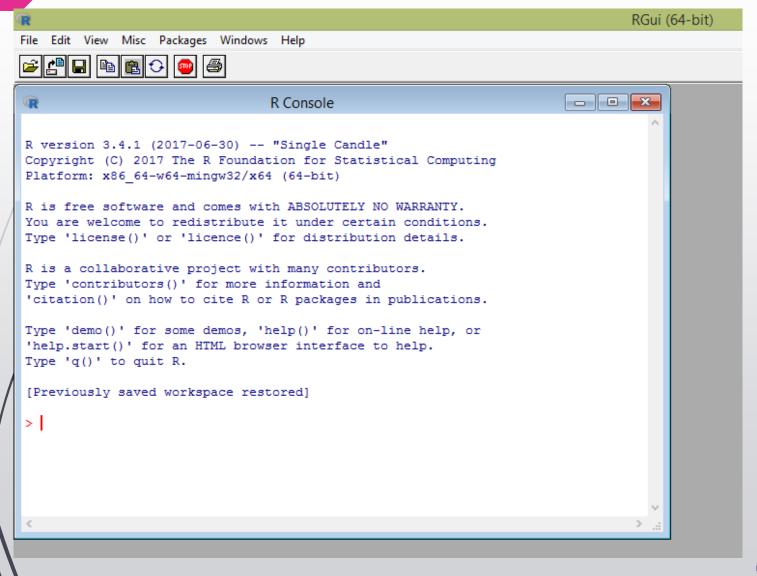
Dapat di unduh pada alamat:

https://cran.r-project.org/bin/windows/base/

Lalu, ikuti petunjuk untuk menginstall R pada komputer kita.



Tampilan R





Package R

- Terdapat fungsi-fungsi untuk mengimplementasikan metode statistik tingkat lanjut.
- Berisi koleksi fungsi R dan data yang dikompilasikan ke dalam format standar.
- ┝ Untuk melihat berbagai macam package:
- https://cran.r-project.org/web/views/
- Perlu koneksi internet untuk install packages



Operasi di R

Operator: +, -, *, /, ^, sqrt()

```
> (5*2)+7
[1] 17
```

Membuat objek sederhana menggunakan <--</p>

```
> x <- 6
> y <- 11
> x + y
[1] 17
```

Menampilkan bantuan untuk menulis keterangan: #

```
> x * sqrt(y) #x dikali dengan akar(y)
[1] 19.89975
```



Import Data Eksternal dan Melihat Isi Data



Mengimport file berekstensi csv ke R



Mengimport file txt atau yang lain ke R

```
> datasurvei1 <-
     read.table("Data_Training.txt",
     header=TRUE, sep="\t")</pre>
```



Mengimport file berekstensi xlsx ke R

```
> install.packages("xlsx", dependencies=TRUE)
> library(xlsx)
> datasurvei2 <-
        read.xlsx("Data_Training.xlsx",
        sheetName = "Sheet1")</pre>
```



Mencetak daftar variabel

> colnames(datasurvei)

```
> colnames(datasurvei)
[1] "JK" "Usia" "Asal" "Biaya"
[5] "Pekerjaan_Ortu" "Pend_Ayah" "Pend_Ibu" "Kiriman"
[9] "Pengeluaran" "Beasiswa"
```



Menampilkan 4 observasi pertama

> head(datasurvei, n=4)

```
> head(datasurvei, n=4)
                         Biaya Pekerjaan_Ortu Pend_Ayah
        JK Usia Asal
                                                                 Pend_Ibu
                                       PETANI SMA/Sederajat
1 PEREMPUAN
             18 ACEH 2400000
                                                                      SMP
2 LAKI-LAKI 18 JAMBI 2400000 KARYAWAN SWASTA
                                                    DIPLOMA SMA/Sederajat
             18 BOGOR 2400000
                                     PENSIUNAN SMA/Sederajat SMA/Sederajat
3 LAKI-LAKI
             18 PADANG 6000000
                                          PNS SMA/Sederajat
4 LAKI-LAKI
                                                                  DIPLOMA
  Kiriman Pengeluaran Beasiswa
1 1/250000
              900000
                           YA
  600000
              600000
                        TIDAK
  600000
              450000
                       TIDAK
  350000
              400000
                           YA
```



Mencetak seluruh data

- > print(datasurvei)
- > datasurvei

print((datasur	vei)
PI IIIC	(aa casa i	v /

/-		pi ilic(uacasi	ui veij				
		JK	Usia	Asal	Biaya	Pekerjaan_Ortu	Pend_Ayah
1	L	PEREMPUAN	18	ACEH	2400000	PETANI	SMA/Sederajat
2	2	LAKI-LAKI	18	JAMBI	2400000	KARYAWAN SWASTA	DIPLOMA
3	3	LAKI-LAKI	18	BOGOR	2400000	PENSIUNAN	SMA/Sederajat
4		LAKI-LAKI	18	PADANG	6000000		SMA/Sederajat
5		PEREMPUAN	17	PADANG	2400000	KARYAWAN SWASTA	SMA/Sederajat
6	5	/ LAKI-LAKI	18	TULUNGAGUNG	8000000	PNS	51
7	7/	PEREMPUAN	18	TUBAN	0	KARYAWAN SWASTA	SMP
ß	Ś	LAKI-LAKI	18	KEDIRI	6000000	PNS	S1
/9)	PEREMPUAN	17	SULAWESI SELATAN	2400000	WIRASWASTA	S1
1	LO	LAKI-LAKI	18	LAMPUNG	2400000	BURUH	SMP
	11		19	CIREBON	2400000	WIRASWASTA	SMP
	L2		18	KLATEN	2400000	PETANI	SMA/Sederajat
	L3		18	BANDUNG	2400000	WIRASWASTA	SMP
	L4		18	PEKALONGAN	2400000	TIDAK BEKERJA	SD
1	L5	LAKI-LAKI	18	RIAU	6000000	KARYAWAN SWASTA	SMA/Sederajat
1	L6	PEREMPUAN	18	BOYOLALI	8000000	GURU	51
1	L7	PEREMPUAN	17	JAKARTA	2400000	IBU RUMAH TANGGA	DIPLOMA
	L8		18	REMBANG	0	WIRASWASTA	
1	L9	PEREMPUAN	18	BEKASI	10000000		SMA/Sederajat
2	20	PEREMPUAN	18	BOGOR	0	WIRASWASTA	SMA/Sederajat



Menampikan data dalam bentuk spreadsheet

> /view(datasurvei)

	JK	Usia	Asal	Biaya	Pekerjaan_Ortu	Pend_Ayah
1	PEREMPUAN	18	ACEH	2400000	PETANI	SMA/Sederajat
2	LAKI-LAKI	18	JAMBI	2400000	KARYAWAN SWASTA	DIPLOMA
3	LAKI-LAKI	18	BOGOR	2400000	PENSIUNAN	SMA/Sederajat
4	LAKI-LAKI	18	PADANG	6000000	PNS	SMA/Sederajat
5	PEREMPUAN	17	PADANG	2400000	KARYAWAN SWASTA	SMA/Sederajat
6	LAKI-LAKI	18	TULUNGAGUNG	8000000	PNS	S1
7	PEREMPUAN	18	TUBAN	0	KARYAWAN SWASTA	SMP
8	LAKI-LAKI	18	KEDIRI	6000000	PNS	S1
9	PEREMPUAN	17	SULAWESI SELATAN	2400000	WIRASWASTA	S1
10	LAKI-LAKI	18	LAMPUNG	2400000	BURUH	SMP
11	LAKI-LAKI	19	CIREBON	2400000	WIRASWASTA	SMP
12	PEREMPUAN	18	KLATEN	2400000	PETANI	SMA/Sederajat
13	PEREMPUAN	18	BANDUNG	2400000	WIRASWASTA	SMP
14	PEREMPUAN	18	PEKALONGAN	2400000	TIDAK BEKERJA	SD
15	LAKI-LAKI	18	RIAU	6000000	KARYAWAN SWASTA	SMA/Sederajat
16	PEREMPUAN	18	BOYOLALI	8000000	GURU	S1
17	PEREMPUAN	17	JAKARTA	2400000	IBU RUMAH TANGGA	DIPLOMA
18	PEREMPUAN	18	REMBANG	0	WIRASWASTA	SMA/Sederajat
19	PEREMPUAN	18	BEKASI	10000000	PNS	SMA/Sederajat



Menampilkan struktur data

> str(datasurvei)

```
> str(datasurvei)
'data.frame': 363 obs. of 10 variables:
      : Factor w/ 2 levels "LAKI-LAKI", "PEREMPUAN": 2 1 1 1 2 1 2 1 2 1
$ JK
$ Usia : int 18 18 18 17 18 18 18 17 18 ...
$ Asal : Factor w/ 108 levels "ACEH". "BALI"...: 1 34 16 63 63 105 104 45
94 50 ...
$ Biaya
                : int 2400000 2400000 2400000 6000000 2400000 8000000 0 6000000 2
400000 2400000 ...
$ Pekerjaan_Ortu: Factor w/ 15 levels "BURUH", "DOSEN",...: 9 5 8 10 5 10 5 10 15 1
$ Pend_Ayah : Factor w/ 7 levels "DIPLOMA", "S1", ...: 6 1 6 6 6 2 7 2 2 7 ...
$ Pend_Ibu : Factor w/ 7 levels "DIPLOMA", "S1", ...: 7 6 6 1 6 2 7 6 2 6 ...
 $ Kiriman
                : int 1250000 600000 600000 350000 800000 1000000 800000 800000 4
00000 700000 ...
$ Pengeluaran
                : int 900000 600000 450000 400000 800000 690000 500000 800000 350
000 700000 ...
 $ Beasiswa : Factor w/ 2 levels "TIDAK", "YA": 2 1 1 2 2 1 2 1 2 2 ...
```



Mengakses variabel/kolom

- > datasurvei\$JK
- > datasurvei\$Biaya
- > datasurvei[,5]
- > datasurvei[,c(6,7)]
- > datasurvei\$JK
 - [1] PEREMPUAN LAKI-LAKI LAKI-LAKI LAKI-LAKI PEREMPUAN LAKI-LAKI PEREMPUAN [8] LAKI-LAKI PEREMPUAN LAKI-LAKI LAKI-LAKI PEREMPUAN PEREMPUAN
- [15] LAKI-LAKI PEREMPUAN PEREMPUAN PEREMPUAN PEREMPUAN PEREMPUAN LAKI-LAKI
- > datasurvei\$Biaya
 - 2400000 2400000 2400000 6000000 2400000 8000000 6000000 [1] 2400000 2400000 2400000 2400000 2400000 2400000 6000000 8000000 [17] 2400000 0 10000000 0 10000000 8000000 8000000 2400000
- > datasurvei[,c(6,7)]

```
Pend_Ayah Pend_Ibu

SMA/Sederajat SMP

DIPLOMA SMA/Sederajat

SMA/Sederajat SMA/Sederajat

SMA/Sederajat DIPLOMA

SMA/Sederajat SMA/Sederajat

SMA/Sederajat SMA/Sederajat

SMA/Sederajat SMA/Sederajat

SMA/Sederajat SMA/Sederajat

SMA/Sederajat SMA/Sederajat
```



Menampilkan Statistik Deskriptif Numerik



Rata-rata

- >/mean(datasurvei\$Biaya)
- > mean(datasurvei\$Biaya)
 [1] 5568595



Simpangan baku dan ragam

- > /sd(datasurvei\$Usia)
- > var(datasurvei\$Usia)

```
> sd(datasurvei$Usia)
[1] 0.5359686
> var(datasurvei$Usia)
[1] 0.2872624
```



Banyaknya observasi

```
> nrow(datasurvei) untuk data frame
> length(datasurvei$JK) untuk vektor
> nrow(datasurvei)
[1] 363
> length(datasurvei$JK)
[1] 363
```



Median

>/median(datasurvei\$Usia)

```
> median(datasurvei$Usia)
[1] 18
```



Menginstal package "psych"

(Untuk menjalankan beberapa fungsi statistika deskriptif)

- > install.packages("psych",
 dependencies=TRUE)
- > library(psych)



Skewness dan Kurtosis

- > /skew(datasurvei\$Usia)
- > kurtosi(datasurvei\$Usia)

```
> skew(datasurvei$Usia)
[1] -0.005029866
> kurtosi(datasurvei$Usia)
[1] 2.0735
```



Statistik 5 serangkai

- > quantile(datasurvei\$Usia)
- > summary(datasurvei\$Usia)
- > describe(datasurvei\$Usia)

```
> quantile(datasurvei$Usia)
      18
           18
 16
                18 20
> summary(datasurvei$Usia)
                         Mean 3rd Qu.
  Min. 1st Qu. Median
                                         Max.
               18.00
 16.00 18.00
                        17.99
                                18.00
                                        20.00
> describe(datasurvei$Usia)
                 sd median trimmed mad min max range skew kurtosis
         n mean
     1 363 17.99 0.54
                         18
                              17.99
                                      0 16 20
                                                   4 -0.01
```



Statistik berdasarkan grup (1)



Statistik berdasarkan grup (2)

```
> /aggregate(datasurvei$Pengeluaran,
     by=list(datasurvei$Pend_Ayah), FUN=mean)
  > aggregate(datasurvei$Pengeluaran,
              by=list(datasurvei$Pend_Ayah),
              FUN=mean)
          Group.1
          DIPLOMA 779166.7
               S1 862277.2
               52 1027777.8
               53 1000000.0
               SD 594444.4
    SMA/Sederajat 759727.9
              SMP 714583.3
```



Statistik berdasarkan grup (3)

```
> /aggregate(datasurvei$Pengeluaran,
      by=list(datasurvei$JK,
      datasurvei$Beasiswa), FUN=mean)
> aggregate(Pengeluaran ~ JK + Beasiswa,
      FUN=mean, data=datasurvei)
   > aggregate(datasurvei$Pengeluaran,
             by=list(datasurvei$JK, datasurvei$Beasiswa),
             FUN=mean)
      Group.1 Group.2
   1 LAKI-LAKI TIDAK 779108.9
   2 PEREMPUAN TIDAK 939142.9
                                         hasilnya
   3 LAKI-LAKI YA 640476.2
  4 PEREMPUAN YA 682500.0
                                         sama persis
   > aggregate(Pengeluaran ~ JK + Beasiswa,
             FUN=mean, data=datasurvei)
          JK Beasiswa Pengeluaran
   1 LAKI-LAKI
                TIDAK
                       779108.9
   2 PEREMPUAN TIDAK 939142.9
   3 LAKI-LAKI YA 640476.2
                 YA
                       682500.0
   4 PEREMPUAN
```



Menampilkan Tabel Frekuensi dan Tabulasi Silang Variabel Kategorik



Tabel frekuensi

```
> tabel1 <- table(datasurvei$Beasiswa)
> tabel1

> tabel1 <- table(datasurvei$Beasiswa)
> tabel1

TIDAK YA
241 122
TIDAK YA
```



Proporsi tabel frekuensi



Tabulasi silang



Proporsi terhadap total observasi

>/prop.table(tabel2)

```
> prop.table(tabel2)
Jenis Kelamin
Beasiswa LAKI-LAKI PEREMPUAN
TIDAK 0.2782369 0.3856749
YA 0.1157025 0.2203857
```



Proporsi berdasarkan baris (beasiswa)



Proporsi berdasarkan kolom (jenis kelamin)



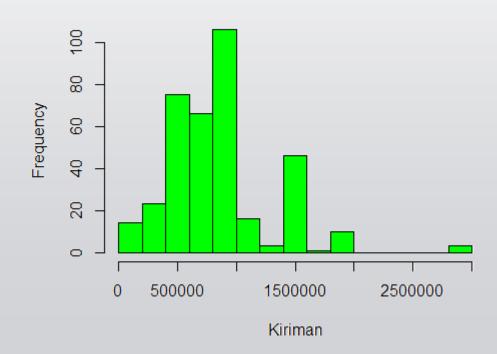
Menampilkan Distribusi Data Numerik



Histogram (1)

Pillihan warna yang digunakan

Histogram Kiriman

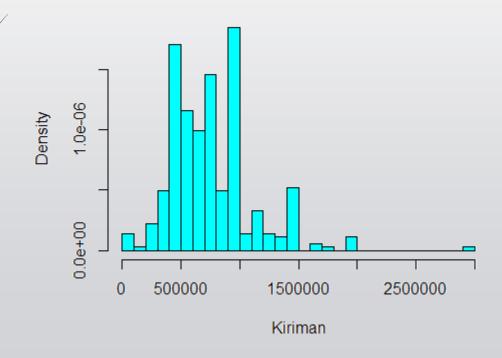




Histogram (2)

Histogram Kiriman

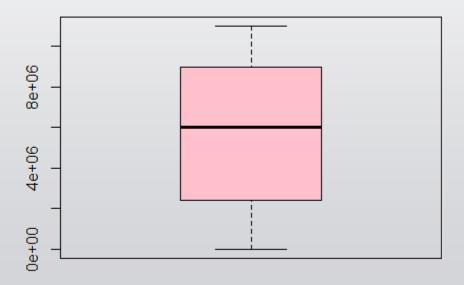
TRUE: menampilkan frekuensi FALSE: menampilkan density





Boxplot (1)

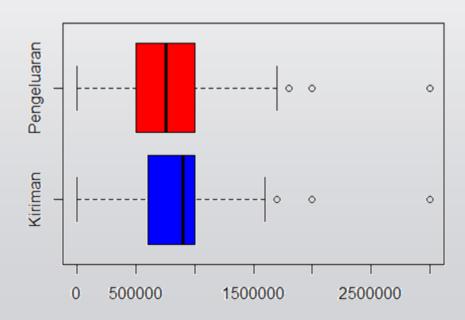
Boxplot Biaya





Boxplot (2)

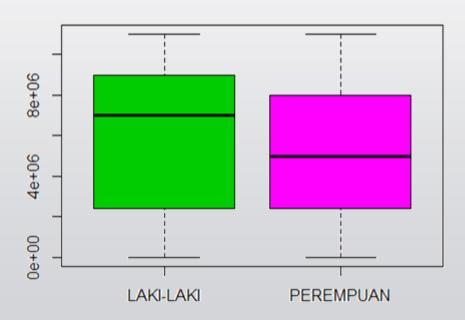
Boxplot Kiriman dan Pengeluaran





Boxplot (3)

Biaya Berdasarkan Jenis Kelamin



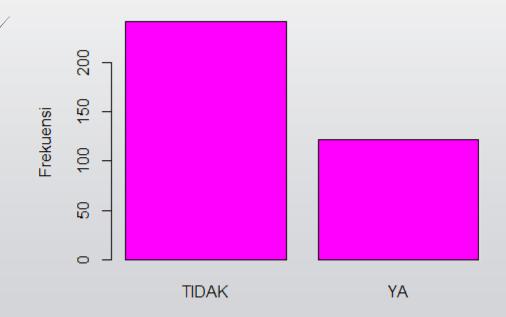


Visualisasi Sederhana



Barplot satu variabel

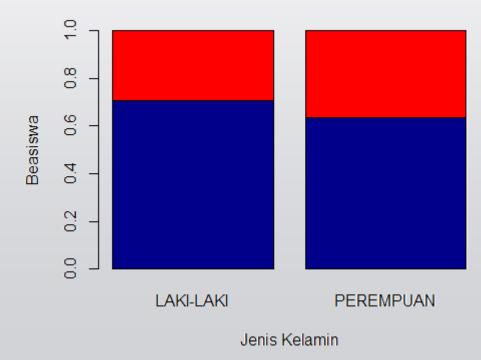
Beasiswa





Barplot tabulasi silang

Jenis Kelamin vs Beasiswa

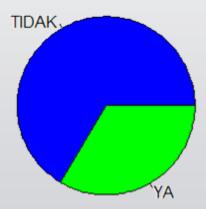




Pie-chart

> pie(tabel1, col=c("blue","green"),
 main="Beasiswa")

Beasiswa





Uji-t Satu Populasi



Uji-t dua arah, alpha=5%

```
> /t.test(datasurvei$Kiriman,
            alternative="two.sided", mu=950000,
            conf.level=0.95)
                                                          uji dua arah
alpha 5%
        > t.test(datasurvei$Kiriman, alternative="two.sided",
                 mu=950000, conf.level=0.95)
                One Sample t-test
                                                         p-value < 0,05
                                                         cukup bukti untuk
                                                         menolak H0
        data: datasurvei$Kiriman
        t = -2.0892, df = 362, p-value = 0.03739
        alternative hypothesis: true mean is not equal to 950000
        95 percent confidence interval:
        853737.5 947089.0
        sample estimates:
                                         selang kepercayaan 95%
        mean of x
         900413.2
```



Uji-t satu arah (H₁ lebih besar), alpha=10%

```
> t.test(datasurvei$Kiriman,
             alternative="greater", mu=880000,
             conf.level=0.90)
                                                           uii satu arah
alpha 10%
          > t.test(datasurvei$Kiriman, alternative="greater",
                   mu=880000, conf.level=0.90)
                  One Sample t-test
                                                         p-value > 0.10
                                                         belum cukup bukti
          data: datasurvei$Kiriman
                                                         untuk menolak H0
          t = 0.86005, df = 362, p-value = 0.1952
          alternative hypothesis: true mean is greater than 880000
          90 percent confidence interval:
                     Inf
           869940
          sample estimates:
          mean of x
           900413.2
```



Uji-t satu arah (H₁ lebih kecil), alpha=10%

```
> /t.test(datasurvei$Kiriman,
            alternative='less', mu=935000,
            conf.level=0.90)
                                                        uii satu arah
alpha 10%
        > t.test(datasurvei$Kiriman, alternative="less",
                 mu=935000, conf.level=0.90)
                One Sample t-test
                                                      p-value < 0,10
                                                      cukup bukti untuk
        data: datasurvei$Kiriman
                                                      menolak H0
        t = -1.4572, df = 362, p-value = 0.07296
        alternative hypothesis: true mean is less than 935000
        90 percent confidence interval:
             -Inf 930886.4
        sample estimates:
        mean of x
         900413.2
```



Uji-t Dua Populasi Saling Bebas



Asumsi ragam homogen

```
> lk <- datasurvei$Pengeluaran[
          datasurvei$JK=="LAKI-LAKI"]</pre>
```

> pr <- datasurvei\$Pengeluaran[
 datasurvei\$JK=="PEREMPUAN"]</pre>

mengambil data _ pengeluaran untuk laki-laki dan perempuan



Asumsi ragam homogen

```
>/ t.test(lk, pr, alternative="two.sided",
                 paired=FALSE, var.equal=TRUE, asumsi ragam populasi
hasilnya
                 conf.level=0.95)
sama persis
                                                               homogen (sama)
          > t.test(Pengeluaran ~ JK, data=datasurvei,
                 var.equal=TRUE, conf.level=0.95)
             t.test(Pengeluaran ~ JK, data=datasurvei, var.equal=TRUE,
                    conf.level=0.95)
                   Two Sample t-test
                                                               tolak H0
            data: Pengeluaran by JK
            t = -2.8174, df = 361, p-value = 0.005108
            alternative hypothesis: true difference in means is not equal to 0
            95 percent confidence interval:
             -182410.81 -32442.33
            sample estimates:
            mean in group LAKI-LAKI mean in group PEREMPUAN
                          738391.6
                                               845818.2
```



Asumsi ragam heterogen

```
> /t.test(lk, pr, alternative="two.sided",
            paired=FALSE, var.equal=FALSE
                                                           asumsi ragam
hasilnya
                                                            populasi heterogen
               conf.level=0.95)
sama persis
                                                            (berbeda)
         > t.test(Pengeluaran ~ JK, data=datasurvei,
                var.equal=FALSE, conf.level=0.95)
            > t.test(Pengeluaran ~ JK, data=datasurvei, var.equal=FALSE,
                    conf.level=0.95)
                   Welch Two Sample t-test
                                                              tolak H0
            data: Pengeluaran by JK
            t = -2.8377, df = 310.74, p-value < 0.004843
            alternative hypothesis: true difference in means is not equal to 0
            95 percent confidence interval:
             -181914.53 -32938.62
            sample estimates:
            mean in group LAKI-LAKI mean in group PEREMPUAN
                          738391.6
                                               845818.2
```

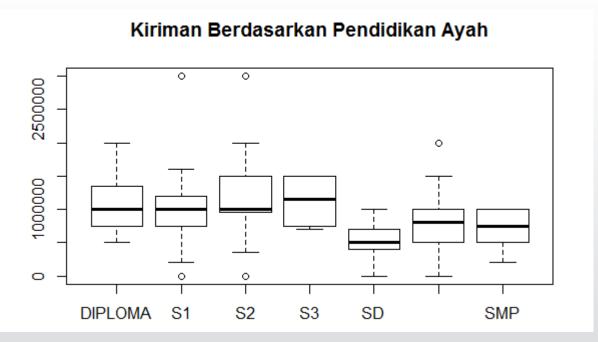
One-Way ANOVA dan Post-Hoc Test



One-way ANOVA (1)



One-way ANOVA (2)



One-way ANOVA (3)

- > anova2 <- aov(Kiriman ~ JK, data=datasurvei)
 #terima H0</pre>
- > summary(anova2)



Uji beda nyata terkecil (BNT)

> pairwise.t.test(datasurvei\$Kiriman, datasurvei\$Pend_Ayah, p.adj="none")

Pairwise comparisons using t tests with pooled SD

data: datasurvei\$Kiriman and datasurvei\$Pend_Ayah

	DIPLOMA	S1	S2	53	SD	SMA/Sederajat
S1	0.8115	-	-	-	-	-
52	0.0446	0.0027	-	-	-	-
53	0.6471	0.5544	0.5906	-	-	-
SD	7.1e-05	1.1e-06	2.5e-10	0.0107	-	-
SMA/Sederajat	0.0295	0.0011	9.7e-08	0.1514	0.0022	-
SMP	0.0127	0.0035	2.8e-06	0.0733	0.1499	0.2732
				7		

P value adjustment method: none

tolak H0

tak tolak H0



Uji Tukey/beda nyata jujur (BNJ)

> TukeyHSD(anova1)

> TukeyHSD(anova1)

Tukey multiple comparisons of means 95% family-wise confidence level

Fit: aov(formula = Kiriman ~ Pend_Ayah, data = datasurvei)

\$Pend_Ayah

_	diff	lwr	upr	p adj
/S1-DIPLOMA	-22813.53	-306238.721	260611.658	0.9999845
S2-DIPLOMA	223611.11	-105292.393	552514.61%	0.4063261
S3-DIPLOMA	104166.67	-569884.936	778218.269	0.9992988
SD-DIPLOMA	-474537.04	-824681.555	-124392.51	0.0013824
SMA/Sederajat-DIPLOMA	-202465.99	-477244.970	72312.998	0.3063071
SMP-DIPLOMA	-304166.67	-664462.403	56129.070	0.1611572
S2-S1	246424.64	4155.505	488693.780	0.0432489
S3-S1	126980.20	-509307.774	763268.170	0.9970275
SD-S1	-451723.51	-722126.915	-181320.097	0.0000232
SMA/Sederajat-S1	-179652.46	-340960.576	-18344.334	0.0180667
SMP-S1	-281353.14	-564778.324	2072.054	0.0531387
S3-S2	-119444.44	-777251.453	538362.564	0.9982471
SD-S2	-698148.15	-1015899.121	-380397.175	0.0000000
SMA/Sederajat-S2	-426077.10	-658171.878	-193982.317	0.0000020
SMP-S2	-527777.78	-856681.282	-198874.274	0.0000580
SD-S3	-578703.70	-1247384.292	89976.884	0.1396393
SMA/Sederajat-S3	-306632.65	-939116.673	325851.366	0.7810266
SMP-S3	-408333.33	-1082384.936	265718.269	0.5514074
SMA/Sederajat-SD	272071.05	10744.328	533397.773	0.0351352
SMP-SD	170370.37	-179774.148	520514.889	0.7780779
SMP-SMA/Sederajat	-101700.68	-376479.664	173078.304	0.9286190



tak tolak H0

tolak H0



Terima kasih.