

Analisis Deskriptif menggunakan R paket R-commander

Untuk tahap ini akan membahas penggunaan **R-Commander** untuk membuat statistik deskriptif dari suatu kumpulan data, khususnya pembuatan ringkasan (*summary*) data dan pembuatan tabel. Ringkasan data difokuskan pada pembuatan statistik deskriptif, yaitu ukuran-ukuran **pemusatan, penyebaran, kemiringan, keruncingan, dan lokasi data-data numerik (metrik)**. Sedangkan pembuatan tabel difokuskan pada data-data nonnumerik.

Paket R menyediakan beberapa fasilitas berkaitan dengan pembuatan ringkasan dari data numerik dan nonnumerik. Secara lengkap fasilitas yang berkaitan dengan ringkasan data

- **Ringkasan numerik**
- **Distribusi Frekuensi**
- **Hitung observasi hilang**
- **Tabel statistika**
- **Matriks korelasi**
- **Uji korelasi**
- **Uji kenormalan Shapiro-Wilk**

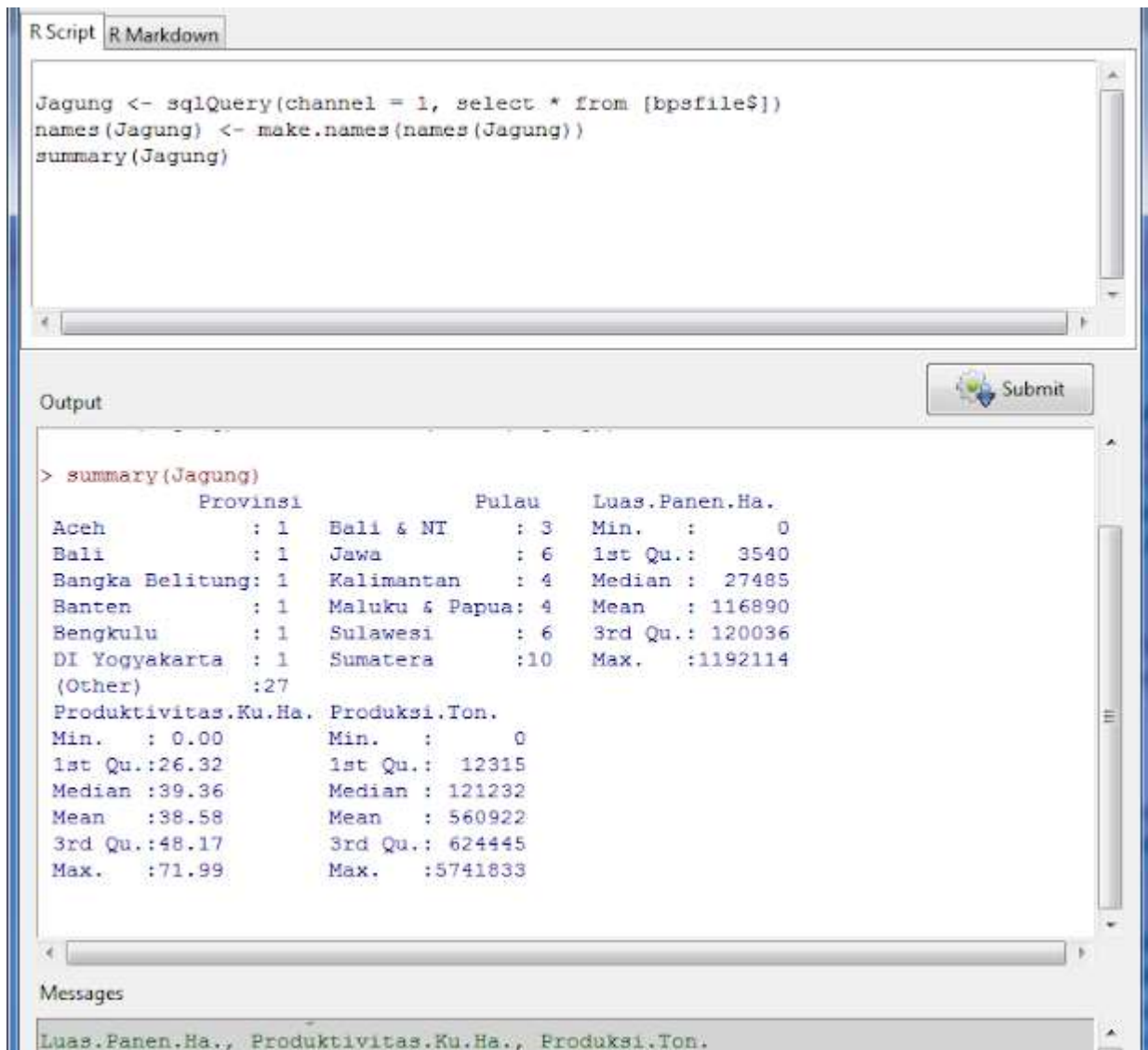
Sebelum memulai tutorialnya. langkah pertama yaitu persiapkan data yang digunakan kemudian masukkan/ import data ke program R. Untuk datanya bisa didownload [disini \(Dropbox\)](#). Oh iya, sebelum tutorial bisa import sendiri ya, tapi kalau belum tau silahkan kesini [Input Data pada Program R](#). berikut sedikit statisik deskriptif

1. Ringkasan Numerik (Summary)

Pada bagian ini akan dijelaskan cara perhitungan ringkasan deskriptif dari data dengan menggunakan berbagai metode statistika deskriptif. Secara umum ada dua macam data yang akan dibuat ringkasan numeriknya, yaitu **data numerik** (*skala interval atau rasio*) dan **data kategorik** (*skala nominal atau ordinal*). R menyediakan dua macam cara untuk menampilkan ringkasan numerik dari variabel-variabel yang ada pada data, yaitu menampilkan ringkasan numerik dari semua variabel yang ada, dan menampilkan ringkasan numerik hanya dari variabel tertentu saja.

Perhitungan Ringkasan Numerik dari semua variabel dapat dilakukan dengan **R Commander**, yaitu gunakan menu **Statistics**, pilih **Summaries**, pilih **Aktive Data set**. Pada data numerik, ringkasan numerik akan menampilkan beberapa besaran statistik yaitu **Mean, Min, Max, Kuartil 1, Median, dan Kuartil 3**. Sedangkan pada

data kategorik, ringkasan numerik hanya menampilkan jumlah atau **frekuensi** pada masing masing kategori yang ada.



The screenshot shows the R Commander interface. The top pane, labeled 'RScript', contains the following R code:

```
Jagung <- sqlQuery(channel = 1, select * from [bpsfile$])
names(Jagung) <- make.names(names(Jagung))
summary(Jagung)
```

The bottom pane, labeled 'Output', shows the result of the `summary(Jagung)` command. It displays a summary of the 'Jagung' dataset, including the number of observations for each province and island, and the distribution of 'Luas.Panen.Ha.' and 'Produksi.Ton.'.

Summary of Jagung

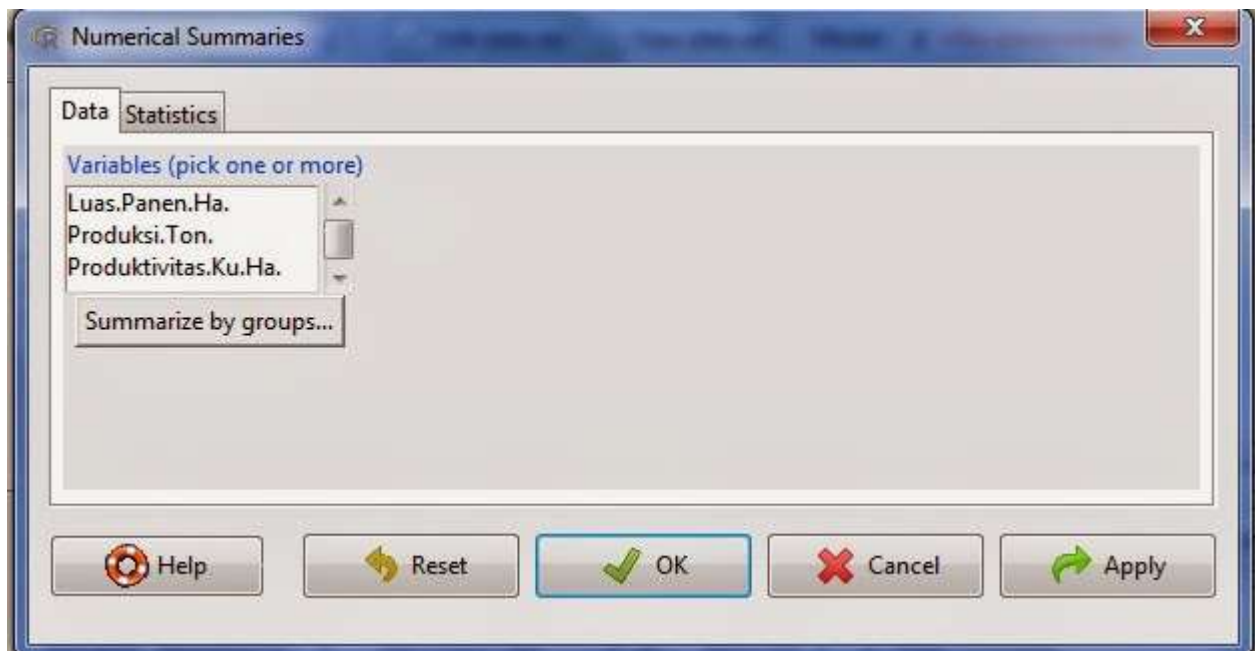
Provinsi	Pulau	Luas.Panen.Ha.
Aceh : 1	Bali & NT : 3	Min. : 0
Bali : 1	Jawa : 6	1st Qu.: 3540
Bangka Belitung: 1	Kalimantan : 4	Median : 27485
Banten : 1	Maluku & Papua: 4	Mean : 116890
Bengkulu : 1	Sulawesi : 6	3rd Qu.: 120036
DI Yogyakarta : 1	Sumatera : 10	Max. : 1192114
(Other) : 27		

Produktivitas.Ku.Ha. and Produksi.Ton.

Produktivitas.Ku.Ha.	Produksi.Ton.
Min. : 0.00	Min. : 0
1st Qu.: 26.32	1st Qu.: 12315
Median : 39.36	Median : 121232
Mean : 38.58	Mean : 560922
3rd Qu.: 48.17	3rd Qu.: 624445
Max. : 71.99	Max. : 5741833

The 'Messages' pane at the bottom shows the command: `Luas.Panen.Ha., Produktivitas.Ku.Ha., Produksi.Ton.`

Perhitungan Ringkasan Numerik khusus untuk variabel numrik dapat dilakukan dengan R-Commander, yaitu gunakan menu **Statistics**, pilih **Summaries**, pilih **Numerical summaries**. Setelah itu akan terlihat jendela informasi tentang variabel metrik dari dataset yang akan ditampilkan ringkasan numeriknya



Misalkan akan dibuat ringkasan numerik untuk variabel **Produksi** , maka pada jendela dialog yang muncul, klik **Produksi** pada pilihan Peubah. Kemudian klik besaran-besaran statistik yang akan ditampilkan ringkasannya. Setelah itu, klik OK untuk menampilkan output ringkasan numeriknya, sehingga diperoleh output pada jendela keluaran seperti berikut ini.

```
> numSummary(Jagung[, "Produksi.Ton."], statistics=c("mean", "sd", "IQR",
+ "quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd    IQR 0%   25%   50%   75%   100%  n
560922.3 1133941 612130  0 12315 121232 624445 5741833 33
```

R juga menyediakan fasilitas untuk menampilkan ringkasan numerik dari variabel numerik berdasarkan group atau faktor tertentu. Misalkan akan dibuat ringkasan untuk variabel **Produksi** berdasarkan **Pulau**, maka gunakan menu **Statistics**, pilih **Summaries**, pilih **Numerical summaries**, dan isikan pilihan seperti sebelumnya, yaitu variabel **Produksi** pada pilihan Peubah yang muncul. Kemudian klik Summaries by Groups, dan pilih variabel **Pulau** dari daftar Peubah kelompok.



Selanjutnya klik **OK**, maka akan diperoleh output ringkasan numerik pada jendela keluaran seperti pada output berikut ini.

```
> numSummary(Jagung[, "Produksi.Ton."], groups=Jagung$Pulau,
+ statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

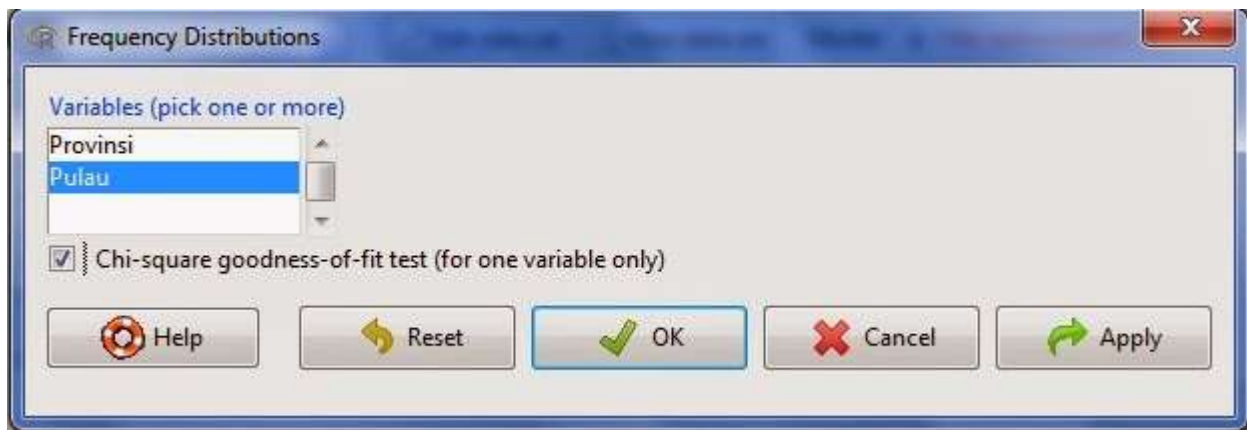
	mean	sd	IQR	0%	25%	50%
Bali & NT	464559.00	354796.72	326662.0	57954	341199.50	624445.0
Jawa	1696831.50	2291833.23	2483226.5	0	76860.50	692419.5
Kalimantan	70452.25	75929.26	110519.8	7283	8189.75	56447.0
Maluku & Papua	12064.00	10947.23	10281.5	1710	5741.25	9700.0
Sulawesi	481198.00	524419.69	491752.5	69137	126000.00	289783.5
Sumatera	371851.60	568316.33	413519.2	818	27074.75	119134.0

	75%	100%	data:n
Bali & NT	667861.50	711278	3
Jawa	2560087.00	5741833	6
Kalimantan	118709.50	161632	4
Maluku & Papua	16022.75	27146	4
Sulawesi	617752.50	1440003	6
Sumatera	440594.00	1725727	10

Dari output tersebut dapat dijelaskan bahwa rata-rata Produksi jagung tertinggi berdasarkan pulau yaitu pulau jawa sedangkan yang terendah yaitu papua dan maluku.

2. Distribusi Frekuensi

Pembuatan Distribusi Frekuensi untuk variabel kategorik dapat dilakukan dengan R-Commander, yaitu gunakan menu **Statistics**, pilih **Summaries**, pilih **Frequency Distribution**. Setelah itu akan terlihat jendela pilihan tentang variabel kategorik.



Misalkan akan dibuat distribusi frekuensi untuk variabel **Pulau**, maka pada jendela dialog pilihan variabel yang muncul, klik **Pulau** pada pilihan Peubah. Sebagai tambahan, pada menu pilihan Distribusi Frekuensi. ini dapat juga dilakukan Uji Kecocokan untuk mengevaluasi apakah probabilitas masing-masing kategori sesuai dengan yang dihipotesakan. Misalkan akan diuji apakah persentase negara pada masing-masing **Pulau** adalah sama, yaitu $1/6$, maka pilih **chi-square goodness of fit** (hanya untuk satu peubah) sehingga diperoleh jendela pilihan probabilitas yang menjadi hipotesis seperti gambar berikut ini.



Klik OK untuk menampilkan output hasil pengujian, sehingga diperoleh output pada jendela keluaran seperti berikut ini.

Output

```
> .Table <- table(Jagung$Pulau)
> .Table # counts for Pulau
      Bali & NT      Jawa      Kalimantan Maluku & Papua      Sulawesi      Sumatera
           3           6           4           4           6           10
> round(100*.Table/sum(.Table), 2) # percentages for Pulau
      Bali & NT      Jawa      Kalimantan Maluku & Papua      Sulawesi      Sumatera
      9.09      18.18      12.12      12.12      18.18      30.30
> .Probs <- c(0.166666666666667,0.166666666666667,0.166666666666667,0.166666666666667,0.166666666666667,0.166666666666667)
> chisq.test(.Table, p=.Probs)

      Chi-squared test for given probabilities

data:  .Table
X-squared = 5.7273, df = 5, p-value = 0.3337
```

Hasil tersebut menunjukkan bahwa nilai p-value pengujian adalah 0.3337. Sehingga jika digunakan $\alpha=0.05$ dapat disimpulkan bahwa pengujian menunjukkan gagal tolak H_0 . Hal ini dikarenakan nilai p-value lebih besar dari α . Dengan demikian dapat disimpulkan bahwa proporsi negara di masing-masing **Pulau** adalah sama yaitu 1/6.

3. Tabel Statistik

R juga menyediakan fasilitas untuk membuat ringkasan statistik dalam tabel untuk suatu variabel numerik (metrik) berdasarkan variabel nonmetrik (kategorik) atau faktor tertentu. Pada R-Commander, gunakan menu Statistika, pilih Ringkasan, pilih Tabel statistika



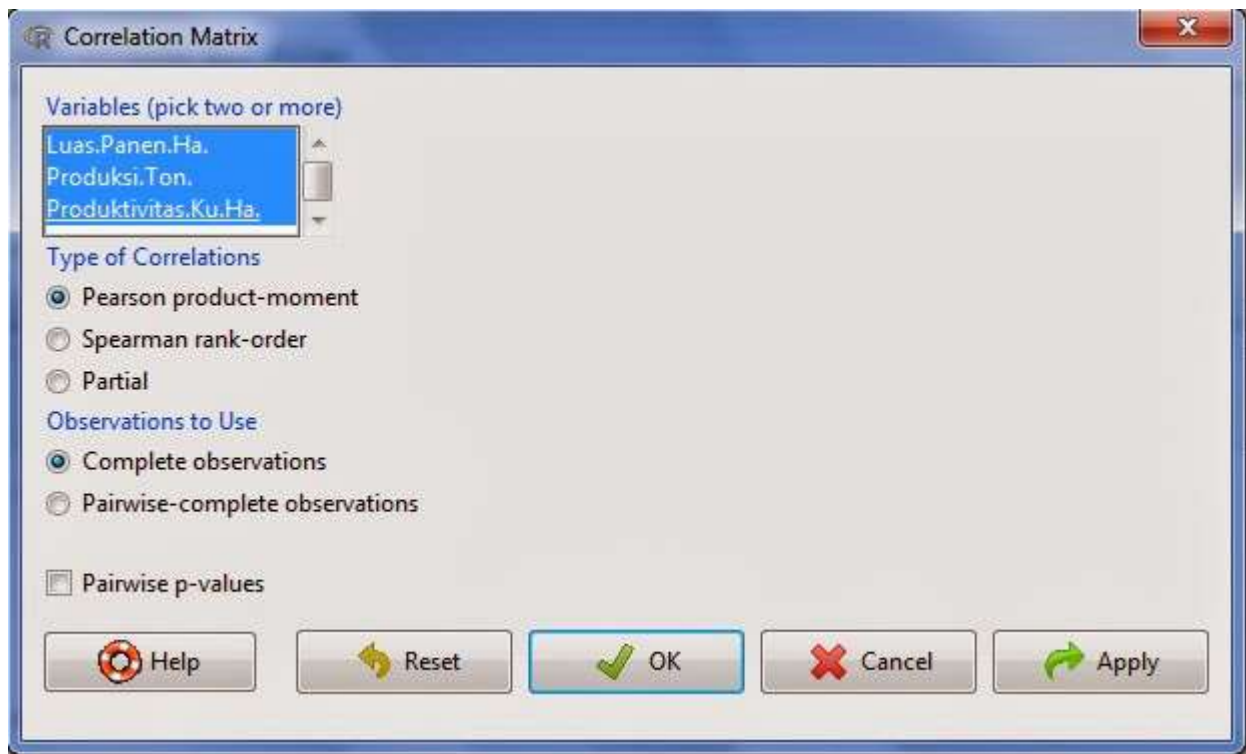
Misalkan akan dibuat tabel statistika untuk variabel **Produksi** berdasarkan variabel **Pulau**, maka pada jendela dialog pilihan Faktor yang muncul, klik **Produksi** dan klik **Pulau** pada pilihan Peubah respon. Kemudian pilih besaran statistik yang akan ditampilkan pada tabel statistika yang akan dibuat. Setelah itu, klik OK untuk menampilkan output tabel statistiknya, sehingga diperoleh output pada jendela keluaran seperti berikut ini.

```
> tapply(Jagung$Produksi.Ton., list(Pulau=Jagung$Pulau), mean, na.rm=TRUE)
```

```
Pulau
      Bali & NT      Jawa      Kalimantan Maluku & Papua      Sulawesi      Sumatera
464559.00    1696831.50      70452.25      12064.00    481198.00    371851.60
```

4. Matriks korelasi

Perhitungan Matriks Korelasi untuk variabel-variabel metrik dan nonmetrik dapat dilakukan dengan R-Commander, yaitu gunakan menu **statistics**, pilih **Summaries**, pilih Correlation matriks, Setelah itu akan terlihat jendela pilihan tentang variabelvariabel dari dataset yang akan ditampilkan matriks korelasinya seperti pada Gambar berikut ini.

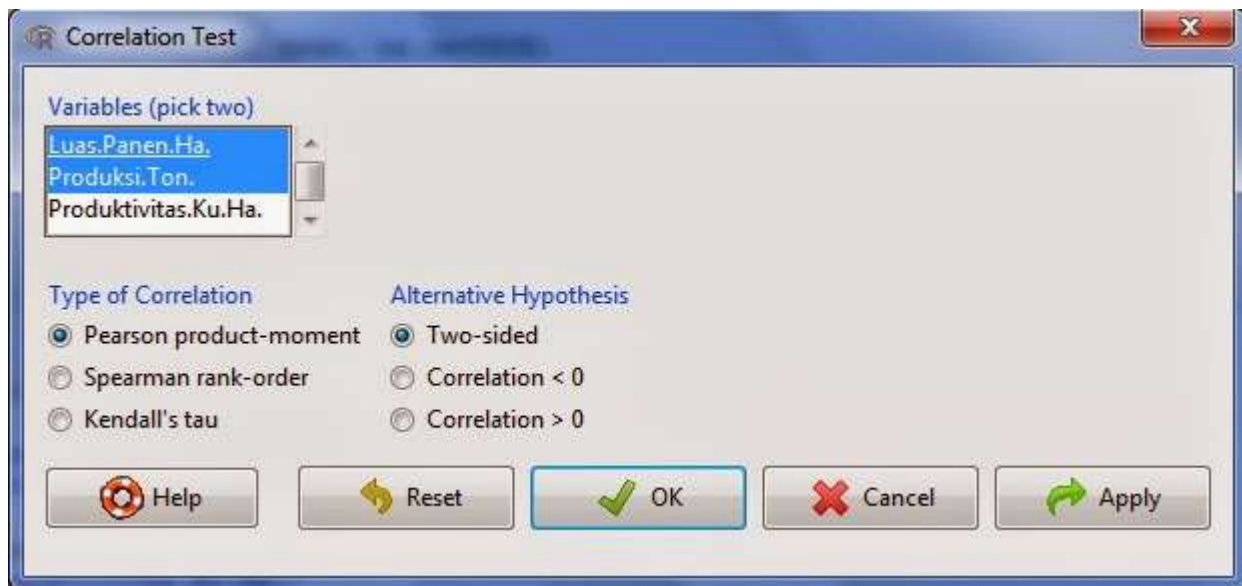


Misalkan akan dibuat matriks korelasi untuk variabel **Luas lahan, Produksi dan Produktivitas**, maka pada jendela dialog pilihan Peubah yang muncul, klik ketiga variabel tersebut. Kemudian pilih jenis korelasi (dalam kasus ini pilih **Produk-Momen Pearson**) yang akan ditampilkan pada matriks korelasi yang akan dibuat. Setelah itu, klik OK untuk menampilkan output matriks korelasi, sehingga diperoleh output pada jendela keluaran seperti berikut ini.

```
> cor(Jagung[,c("Luas.Panen.Ha.", "Produksi.Ton.", "Produktivitas.Ku.Ha.")], use="complet
Luas.Panen.Ha.  Produksi.Ton.  Produktivitas.Ku.Ha.
Luas.Panen.Ha.  1.0000000    0.9910503    0.3340857
Produksi.Ton.   0.9910503    1.0000000    0.3925239
Produktivitas.Ku.Ha. 0.3340857  0.3925239    1.0000000
```

5. Uji korelasi

Perhitungan Uji Korelasi, baik untuk korelasi **Produk-Momen Pearson** ataupun korelasi **Rank-Order Spearman** dapat dilakukan dengan **R-Commander**, yaitu gunakan menu **Statistics**, pilih **Summaries**, pilih **Correlation Matrix**.



Misalkan akan dilakukan pengujian korelasi untuk variabel **Produksi** dan **Luas Panen**, maka pada jendela dialog pilihan Peubah yang muncul, klik kedua variabel tersebut. Kemudian pilih jenis korelasi (dalam kasus ini pilih **Produk-Momen Pearson**) yang akan diuji. Setelah itu, klik OK untuk menampilkan output pengujian pada jendela keluaran seperti berikut ini.

```
> cor.test(Jagung$Luas.Panen.Ha., Jagung$Produksi.Ton., alternative="two.sided", method="spearman")

Pearson's product-moment correlation

data: Jagung$Luas.Panen.Ha. and Jagung$Produksi.Ton.
t = 41.3363, df = 31, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9817781 0.9956148
sample estimates:
      cor 
0.9910503
```

Hasil tersebut menunjukkan bahwa nilai p-value pengujian adalah $2.2e-16$. Sehingga jika digunakan $\alpha=0.05$ dapat disimpulkan bahwa pengujian menunjukkan tolak H_0 . Hal ini dikarenakan nilai p-value lebih kecil dari α . Dengan demikian dapat disimpulkan bahwa ada korelasi positif antara **Luas Panen** dan **Produksi**.