

## KOMPUTASI GUI UNTUK INFERENSI VEKTOR MEAN DAN INFERENSI MATRIKS KOVARIANSI DENGAN MENGGUNAKAN *SOFTWARE R*

Yudha Subakti<sup>1</sup>, Moch. Abdul Mukid<sup>2</sup>, Hasbi Yasin<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

e-mail [yudha.subakti.91@gmail.com](mailto:yudha.subakti.91@gmail.com)

### ABSTRACT

Multivariate statistics is a branch of statistical science that discuss the analysis for multivariable case. Some analysis in multivariate statistics are discussing about inferences, there are inferences about mean vector and inferences of covariance matrices. Along with the development of technology, to support statistical analysis from both of inferences is requiring a statistical software, R is one of it with open source based. R is often used in statistical computing with command line interface (CLI) as the interface. In implementation, CLI requires the R user to remember names of used syntax and function. It makes less effective when the inferences have many related statistical analysis, so graphical user interface (GUI) needed to giving an easy way to accessing all of it. Testing for mean vectors of two populations will be performed using S. Rockiki's data about measures of oxygen consumption for 25 males and 25 females. Results about assumptions showing both populations are normal multivariate distributed and have different covariannce matrix. The conclusion from the testing for mean vectors of two populations has performed is both populations have different mean vectors. There are packages are used on construction of GUI in R, including gdata, tcltk2, and devtools with additional software like Rtools and ActivePerl. The GUI has four main menus such as File, Analysis, Plot, and Help. Based on GUI usage, the GUI has been able to processing the chosen analysis and showing valid output..

**Keywords:** Multivariate Statistics, Inferences about mean vector, Inferences of covariance matrices, R, GUI.

### 1. PENDAHULUAN

Dalam ilmu statistika sering dijumpai kasus data numerik dengan skala besar yang terdiri dari beberapa variabel. Beberapa analisis yang dibicarakan dalam statistika multivariat menyangkut inferensi, yaitu memperoleh kesimpulan yang sah berbasis informasi dalam sampel termasuk di dalamnya yaitu inferensi vektor mean dan inferensi matriks kovariansi (Johnson dan Wichern, 2007).

R merupakan salah satu *software* statistika berbasis *open source* yang sudah banyak dikenal dan digunakan untuk keperluan analisis dan komputasi statistika. Ada dua bentuk umum antarmuka pengguna dalam komputasi statistika yaitu *Command Line Interface* (CLI) dan *Graphical User Interface* (GUI). CLI biasanya terdiri dari sebuah konsol tekstual dimana pengguna mengetikkan serangkaian perintah dalam bentuk teks, salah satu contohnya yaitu R console pada R. Sedangkan GUI adalah sarana utama dalam berinteraksi dengan *desktop*, seperti Windows dan Mac OS, dan *software* statistika seperti JMP (SAS Institute Inc. 2007). Antarmuka ini berdasarkan pada paradigma WIMP (window, icon, menu, dan pointer) (Penners, 2005).

Penggunaan R untuk analisis statistika multivariat khususnya untuk inferensi vektor mean dan inferensi matriks kovariansi masih menggunakan CLI dalam R *console* yang membuat pengguna R harus mengingat nama fungsi dan *syntax* yang diperlukan. Hal ini dirasa kurang efektif jika dalam penggunaannya terdapat banyak analisis statistika sehingga GUI dibutuhkan untuk memudahkan pengguna dalam mengakses semua analisis yang akan digunakan. Berdasarkan permasalahan tersebut, penulis akan membuat antarmuka berupa GUI bagi pengguna R dalam keperluan analisis statistika multivariat mengenai inferensi vektor mean dan inferensi makriks kovariansi.

## 2. TINJAUAN PUSTAKA

### 2.1. Distribusi Normal Multivariat

Vektor random  $\mathbf{X}$  berdimensi  $p$  dengan  $\boldsymbol{\mu}$  adalah vektor nilai rata-rata dari vektor random  $\mathbf{X}$  berdimensi  $p \times 1$  dan  $\boldsymbol{\Sigma}$  adalah matriks varian-kovarian berdimensi  $p \times p$  berdistribusi normal multivariat dapat dinotasikan dengan  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Dalam notasi matriks dan vektor dapat ditulis

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

dengan  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ ,  $\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_p)$ , dan  $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$  adalah

matriks definit positif (Johnson dan Wichern, 2007)

### 2.2. Uji Asumsi Normal Multivariat

Pengujian asumsi normal multivariat dilakukan dengan membuat plot kuantil jarak mahalanobis ( $d^2_j$ ) yang didekati dengan kuantil *chi-square*. Jarak mahalanobis adalah ukuran yang menyatakan jarak nilai setiap kasus dari rata-rata seluruh kasus. Langkah-langkahnya sebagai berikut (Johnson dan Wichern, 2007):

- Menentukan Menentukan  $d^2_j = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$  dimana  $j = 1, 2, \dots, n$  dan  $\mathbf{S}$  adalah matriks varian-kovarian.
- Mengurutkan nilai  $d^2_j$  sesuai dengan urutan naik  $d^2_{(1)} \leq d^2_{(2)} \leq \dots \leq d^2_n$ .
- Menentukan nilai *chi-square*  $\chi^2_p((j-0.5)/n)$  yang merupakan kuantil ke- $100 * (j-0.5)/n$  dari distribusi *chi-square* dengan derajat bebas  $p$  dan  $j = 1, 2, \dots, n$ , dimana  $n =$  banyaknya observasi dan  $p =$  banyak variabel.
- Buat plot antara  $d^2_j$  dengan  $\chi^2_p((j-0.5)/n)$ .

Asumsi normal multivariat terpenuhi jika plot membentuk pola seperti garis lurus.

Menurut Daniel (1989), pengujian asumsi normal multivariat juga dapat dilakukan secara formal dengan menggunakan metode Kolmogorov-Smirnov dengan langkah-langkah pengujian hipotesis sebagai berikut:

Hipotesis:

$$H_0 : F(d^2_j) = F_0(d^2_j) \text{ (Data pengamatan berdistribusi normal multivariat)}$$

$$H_1 : F(d^2_j) \neq F_0(d^2_j) \text{ (Data pengamatan tidak berdistribusi normal multivariat)}$$

Tarif signifikansi:  $\alpha$

Statistik uji:

$$D = \text{Sup} |S(d^2_j) - F_0(d^2_j)|, \text{ dengan}$$

$$S(d^2_j) = \text{Proporsi jarak mahalanobis yang } \leq d^2_j$$

$$F_0(d^2_j) = \text{Fungsi peluang kumulatif dari distribusi } \textit{chi-square}$$

Kriteria uji:

$H_0$  ditolak jika  $D > W_{(1-\alpha)}$  dengan uji dua-sisi atau jika nilai *p-value*  $< \alpha$ , dimana  $W_{(1-\alpha)}$  merupakan kuantil  $1 - \alpha$  pada tabel Kolmogorov-Smirnov.

### 2.3. Uji Hipotesis Vektor Mean Satu Populasi

Misalkan  $X_1, X_2, \dots, X_n$  merupakan sampel random dari populasi normal  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Akan diperlihatkan pengujian  $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$  dibandingkan dengan  $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ , dimana  $\boldsymbol{\mu}_0$  adalah sebuah vektor yang diketahui. Uji hipotesis dilakukan dengan langkah-langkah sebagai berikut:

Hipotesis:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

$$H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Taraf signifikansi:  $\alpha$

Statistik uji:

$$T^2 = (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)' \left(\frac{\mathbf{S}}{n}\right)^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

Kriteria uji:

$$H_0 \text{ ditolak jika } T^2 > \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha) \text{ atau jika } p\text{-value} < \alpha.$$

Sementara itu kriteria uji untuk  $n$  besar, yaitu  $H_0$  ditolak jika  $T^2 > \chi_p^2(\alpha)$ .

#### 2.4. Uji Hipotesis Vektor Mean Dua Populasi

Akan dilakukan inferensi untuk  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ . Misal ingin diketahui apakah  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  atau  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$ . Kemudian bila ternyata  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq 0$ , ingin diketahui komponen mana saja yang berbeda.

Anggapan untuk struktur data

- Sampel  $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$  adalah sampel random berukuran  $n_1$  dari populasi  $p$  variat dengan vektor mean  $\boldsymbol{\mu}_1$  dan matriks kovariansi  $\boldsymbol{\Sigma}_1$ .
- Sampel  $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$  adalah sampel random berukuran  $n_2$  dari populasi  $p$  variat dengan vektor mean  $\boldsymbol{\mu}_2$  dan matriks kovariansi  $\boldsymbol{\Sigma}_2$ .
- Kedua sampel random bersifat independen.

Anggapan tersebut diperlukan untuk inferensi  $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$  dalam sampel besar. Adapun jika ukuran sampel kecil diperlukan tambahan anggapan, yaitu:

- Kedua populasi berdistribusi normal multivariat.
- $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

Rata-rata sampel dan kovariansi dari kedua sampel random, masing-masing ditunjukkan oleh  $\bar{\mathbf{X}}_1, \mathbf{S}_1, \bar{\mathbf{X}}_2$ , dan  $\mathbf{S}_2$ . Berdasarkan asumsi  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ , diperoleh dugaan gabungan dari matriks kovariansi yaitu  $\mathbf{S}_{gab}$ .

$$\begin{aligned} \mathbf{S}_{gab} &= \frac{\sum_{j=1}^{n_1} (\mathbf{X}_{1j} - \bar{\mathbf{X}}_1)(\mathbf{X}_{1j} - \bar{\mathbf{X}}_1)' + \sum_{j=1}^{n_2} (\mathbf{X}_{2j} - \bar{\mathbf{X}}_2)(\mathbf{X}_{2j} - \bar{\mathbf{X}}_2)'}{n_1 + n_2 - 2} \\ &= \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} \end{aligned}$$

Bila  $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$  adalah sampel random berukuran  $n_1$  dari  $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$  dan  $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$  adalah sampel random berukuran  $n_2$  dari  $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$  saling independen, maka uji hipotesis untuk membandingkan vektor mean dari dua populasi dilakukan dengan langkah-langkah sebagai berikut:

Hipotesis:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Taraf signifikansi:  $\alpha$

Statistik uji:

$$T^2 = ((\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left( \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{gab} \right)^{-1} ((\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))$$

Kriteria uji:

$$H_0 \text{ ditolak jika } T^2 > \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1} \text{ atau jika } p\text{-value} < \alpha.$$

Ketika dalam pengujian asumsi, kedua populasi berdistribusi normal tetapi matriks kovariansinya berbeda, maka statistik ujinya pun berbeda dengan statistik uji ketika kedua

asumsi terpenuhi, begitu juga untuk kriteria uji yang digunakan. Ada pun langkah-langkah pengujian hipotesisnya sebagai berikut:

Hipotesis:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Taraf signifikansi:  $\alpha$

Statistik uji:

$$T^2 = \left( (\bar{X}_1 - \bar{X}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right)' \left( \frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^{-1} \left( (\bar{X}_1 - \bar{X}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right)$$

Kriteria uji:

$H_0$  ditolak jika  $T^2 > \frac{vp}{v-p+1} F_{p,v-p+1}$  atau jika  $p\text{-value} < \alpha$ , dengan

$$v = \frac{p+p^2}{\sum_{i=1}^2 \frac{1}{n_i} \left\{ \text{tr} \left[ \left( \frac{1}{n_i} \mathbf{S}_i \left( \frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^{-1} \right)^2 \right] + \left( \text{tr} \left[ \frac{1}{n_i} \mathbf{S}_i \left( \frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^{-1} \right] \right)^2 \right\}}$$

dimana  $\min(n_1, n_2) \leq v \leq n_1 + n_2$ .

## 2.5. Uji Perbandingan Berpasangan

Anggaplah bahwa pengukuran  $p$  diambil dari masing-masing unit sehingga responnya adalah  $X_{1ji}$  dan  $X_{2ji}$ , dimana  $X_{1ji}$  adalah pengukuran ke- $i$  dari  $j$  unit untuk perlakuan 1, dan  $X_{2ji}$  adalah pengukuran ke- $i$  dari  $j$  unit untuk perlakuan 2. Perbedaan untuk kedua perlakuan adalah  $D_{ji} = X_{1ji} - X_{2ji}$  dan  $\mathbf{D}_j = (D_{j1}, \dots, D_{jp})'$ .

Asumsikan bahwa  $E(\mathbf{D}_j) = \boldsymbol{\delta} = (\delta_1, \dots, \delta_p)'$  dan  $\text{cov}(\mathbf{D}_j) = \boldsymbol{\Sigma}_d$ . Jika diasumsikan lebih lanjut  $\mathbf{D}_j \sim N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ , maka dapat dibuat inferensi mengenai  $\boldsymbol{\delta}$  menggunakan statistik  $T^2$  Hotelling dengan  $T^2 = n(\bar{\mathbf{D}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\delta})$ , dimana  $\bar{\mathbf{D}} = \frac{1}{n} \sum_{j=1}^n \mathbf{D}_j$  dan  $\mathbf{S}_d^2 = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{D}_j - \bar{\mathbf{D}})(\mathbf{D}_j - \bar{\mathbf{D}})'$ .

Apabila selisih  $\mathbf{D}_1, \dots, \mathbf{D}_n$  merupakan sampel random dari sebuah populasi  $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ , maka  $T^2 = n(\bar{\mathbf{D}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\delta})$  didistribusikan sebagai variabel acak  $[(n-1)p/(n-p)] F_{p,n-p}$ . Jika  $n$  dan  $n-p$  keduanya besar,  $T^2$  kira-kira didistribusikan sebagai variabel acak  $\chi_p^2$ .

Misalkan sampel acak terdiri dari  $\mathbf{d}_1, \dots, \mathbf{d}_n$  dan populasinya adalah  $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ . Uji hipotesisnya dilakukan dengan langkah-langkah sebagai berikut:

Hipotesis:

$$H_0 : \boldsymbol{\delta} = \mathbf{0}$$

$$H_1 : \boldsymbol{\delta} \neq \mathbf{0}$$

Taraf signifikansi:  $\alpha$

Statistik uji:

$$T^2 = n \bar{\mathbf{d}}' \mathbf{S}_d^{-1} \bar{\mathbf{d}}$$

Kriteria uji:

$H_0$  ditolak jika  $T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$  atau jika  $p\text{-value} < \alpha$ .

## 2.6. Uji Kesamaan Matriks Kovariansi

Salah satu asumsi yang dibuat ketika membandingkan dua atau lebih vektor mean yaitu bahwa matriks kovariansi dari populasi yang berbeda adalah sama. Dengan anggapan  $g$  populasi dan  $p$  variabel, hipotesis nolnya adalah  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$ , dimana  $\boldsymbol{\Sigma}_\ell$  adalah matriks kovariansi untuk populasi ke- $\ell$ , dengan  $\ell = 1, 2, \dots, g$  dan  $\boldsymbol{\Sigma}$  dianggap matriks kovariansi secara umum. Hipotesis alternatifnya yaitu bahwa paling

sedikit terdapat dua matriks kovariansi yang tidak sama. Jenis uji yang sering digunakan adalah uji *Box's M* (*Box's M-test*) dengan langkah-langkah pengujian hipotesisnya sebagai berikut:

Hipotesis:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$$

$$H_1 : \Sigma_i \neq \Sigma_j \text{ untuk beberapa } 1 \leq i \neq j \leq g$$

Taraf signifikansi:  $\alpha$

Statistik uji:

$$C = (1 - u) \left\{ \left[ \sum_{\ell=1}^g (n_{\ell} - 1) \right] \ln(|S_{gab}|) - \sum_{\ell=1}^g [(n_{\ell} - 1) \ln(|S_{\ell}|)] \right\},$$

$$\text{dengan } u = \left[ \sum_{\ell=1}^g \frac{1}{(n_{\ell}-1)} - \frac{1}{\sum_{\ell=1}^g (n_{\ell}-1)} \right] \left[ \frac{2p^2+3p-1}{6(p+1)(g-1)} \right]$$

Kriteria uji:

$$H_0 \text{ ditolak jika } C > \chi_{p(p+1)(g-1)/2}^2(\alpha) \text{ atau jika } p\text{-value} < \alpha.$$

## 2.7. Software R

R merupakan salah satu *software* yang digunakan untuk keperluan komputasi statistika yang berbasis *open source*. Yang menjadi salah satu keunggulan dari *software* ini adalah banyaknya pilihan analisis statistika yang tersedia dalam bentuk paket R dimana versi R yang teranyar saat ini adalah R versi 3.2.1 (<http://www.r-project.org/>). Versi awal R dibuat pada tahun 1992 oleh Robert Gentleman dan Ross Ihaka di Universitas Auckland, New Zealand.

## 2.8. Tcl/Tk dan Paket R *tcltk2*

Paket *tcltk* dikembangkan oleh Peter Dalgaard dan telah diikutsertakan dengan basis R sejak versi 1.1.0. *Software* R dapat berinteraksi dengan Tcl/Tk melalui paket R yang disebut *tcltk*. Paket ini telah digunakan dalam banyak proyek GUI untuk R, terutama *Rcmdr* GUI. Paket *tcltk* membuat dan menggunakan *tcl/tk interpreter* dari dalam R, sedemikian hingga *interpreter* ini dapat mengeksekusi perintah-perintah yang dikenal dalam bahasa R.

## 3. METODE PENELITIAN

### 3.1. Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder yang berasal dari buku yang berjudul *Applied Multivariate Statistical Analysis* edisi keenam karangan Richard A. Johnson dan Dean W. Wichern (2007). Data tersebut merupakan data milik S. Rokicki mengenai penelitian yang mengevaluasi fungsi paru-paru pada 25 pria dan 25 wanita yang sehat dengan melakukan pengukuran pengonsumsi oksigen setelah mereka diminta untuk lari di treadmill sampai kelelahan.

### 3.2. Metode Analisis

Adapun langkah-langkah dalam pelaksanaan penelitian dan analisis data yaitu diawali dengan menentukan data yang akan digunakan, dilanjutkan dengan mengidentifikasi analisis statistika multivariat yang sesuai dengan kriteria data. Langkah selanjutnya yaitu mengidentifikasi hal-hal yang dibutuhkan dalam pembuatan GUI termasuk di dalamnya proses kerja GUI saat dijalankan. Kemudian membuat GUI dengan menggunakan *software* R dan melakukan analisis statistika multivariat menggunakan GUI yang telah dibuat sehingga *output* dari analisis dapat ditampilkan.

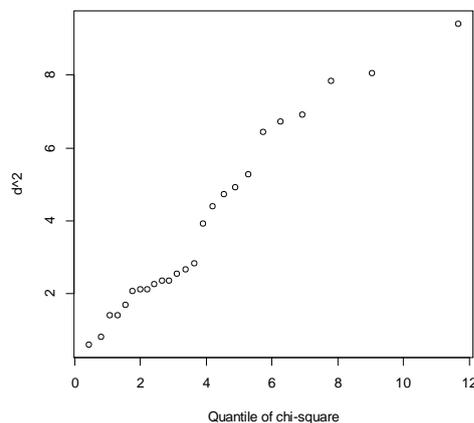
## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

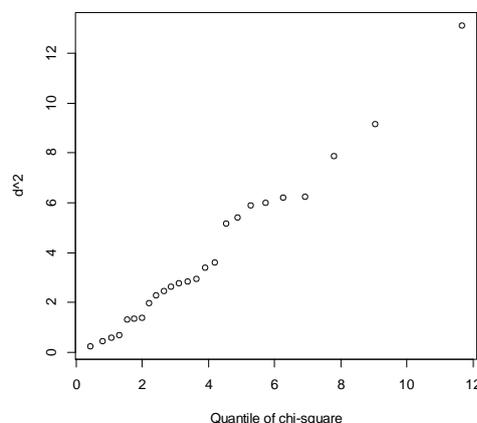
Data yang digunakan merupakan data sekunder yang sifatnya sebagai percobaan, dengan tujuan untuk pengujian validitas GUI terhadap algoritma pembuatannya. *Software* yang digunakan untuk menyimpan data, yaitu *Microsoft Office Excel 2007* sehingga ekstensi data yang dibutuhkan GUI berupa *xlsx* atau pun *xls*. Data yang sudah tersimpan dengan ekstensi *xlsx* atau *xls* nantinya akan melalui proses impor saat GUI dijalankan. Tabulasi data yang terbentuk dengan menggunakan software *Microsoft Office Excel 2007* yaitu terdiri dari 8 kolom dan 26 baris.

### 4.2. Pemilihan Analisis Statistika Multivariat

Tahap selanjutnya akan dilakukan analisis statistika multivariat yang sesuai dengan kriteria data sehingga analisis yang digunakan yaitu analisis vektor mean untuk dua populasi. Analisis ini digunakan untuk mengetahui apakah vektor mean kedua populasi berbeda atau tidak. Karena ukuran sampel yang kecil maka kedua populasi akan diuji asumsi normal multivariat dan kesamaan matriks kovariansinya. Berikut plot Q-Q untuk kedua populasi:



**Gambar 1.** Plot Q-Q pada Populasi Pria



**Gambar 2.** Plot Q-Q pada Populasi Wanita

Dapat dilihat pada Gambar 1 dan Gambar 2 yang menunjukkan kedua pola plot cenderung mengikuti pola garis lurus, untuk memastikan bahwa kedua populasi

berdistribusi normal multivariat maka dilakukan pengujian dengan metode Kolmogorov-Smirnov. Berikut hasil analisis dari pengujian uji normalitas pada populasi pria:

Hipotesis:

$H_0$  : Data pengukuran pengonsumsi oksigen pria berdistribusi normal multivariat

$H_1$  : Data pengukuran pengonsumsi oksigen pria tidak berdistribusi normal multivariat

Taraf signifikansi:  $\alpha = 5\%$

Statistik uji:  $D = 0,14508$

Kriteria uji: Tolak  $H_0$  jika  $D > W_{(0,95)} = 0,264$ , dengan  $n = 25$ , untuk uji dua-arah

Kesimpulan:

Karena  $D = 0,14508 < W_{(0,95)} = 0,264$ , maka pada taraf signifikansi 5%  $H_0$  gagal ditolak. Sehingga dapat dikatakan bahwa data pengukuran pengonsumsi oksigen pria berdistribusi normal.

Kemudian dilanjutkan dengan pengujian normalitas pada populasi wanita dengan hasil analisis sebagai berikut:

Hipotesis:

$H_0$  : Data pengukuran pengonsumsi oksigen wanita berdistribusi normal multivariat

$H_1$  : Data pengukuran pengonsumsi oksigen wanita tidak berdistribusi normal multivariat

Taraf signifikansi:  $\alpha = 5\%$

Statistik uji:  $D = 0,12707$

Kriteria uji: Tolak  $H_0$  jika  $D > W_{(0,95)} = 0,264$ , dengan  $n = 25$ , untuk uji dua-arah

Kesimpulan:

Karena  $D = 0,12707 < W_{(0,95)} = 0,264$ , maka pada taraf signifikansi 5%  $H_0$  gagal ditolak. Sehingga dapat dikatakan bahwa data pengukuran pengonsumsi oksigen wanita berdistribusi normal.

Hasil pengujian yang dilakukan memberikan kesimpulan bahwa kedua populasi berdistribusi normal multivariat, maka asumsi normal multivariat dari data yang digunakan telah terpenuhi. Setelah asumsi normal multivariat terpenuhi, langkah selanjutnya yaitu menguji kesamaan matriks kovariansi dari kedua populasi dengan hasil analisis sebagai berikut:

Hipotesis:

$H_0 : \Sigma_1 = \Sigma_2$

$H_1 : \Sigma_1 \neq \Sigma_2$

Tingkat signifikansi:  $\alpha = 5\%$

Statistik uji:  $C = 58,0011$

Kriteria uji: Tolak  $H_0$  jika  $C > \chi_{10}^2(0,05) = 18,31$

Kesimpulan:

Karena  $C = 58,0011 > \chi_{10}^2(0,05) = 18,31$ , maka pada taraf signifikansi 5%  $H_0$  ditolak. Artinya matriks kovariansi dari data pengukuran pengonsumsi oksigen pria berbeda dengan matriks kovariansi populasi wanita.

Hasil dari pengujian kesamaan matriks kovariansi menunjukkan bahwa matriks kovariansi kedua populasi berbeda, kemudian dilanjutkan dengan menguji vektor mean kedua populasi dengan hasil analisis sebagai berikut:

Hipotesis:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Tingkat signifikansi:  $\alpha = 5\%$

Statistik uji:  $T^2 = 96,3732$

Kriteria uji: Tolak  $H_0$  jika  $T^2 > \frac{vp}{v-p+1} F_{p,v-p+1}$ , dengan  $F_{(4),(42,253)}(0,05) = 2,593$

Kesimpulan:

Karena  $T^2 = 96,3732 > 4,284 F_{(4),(42,253)}(0,05) = 2,593$ , maka pada taraf signifikansi 5%  $H_0$  ditolak. Artinya vektor mean pengukuan pengonsumsi oksigen pria berbeda dengan vektor mean populasi wanita.

### 4.3. Persiapan dan Kebutuhan dalam Pembuatan GUI

Sebelum membuat GUI, diperlukan proses identifikasi mengenai hal-hal apa saja yang diperlukan dalam GUI tersebut. Proses ini bertujuan untuk memahami hal-hal yang menjadi bagian dari *interface* itu sendiri, antara lain *input*, proses, dan *output*. Yang menjadi *input* untuk GUI yaitu *file* data yang disimpan dalam bentuk ekstensi *xls* maupun *xlsx* menggunakan *Microsoft Office Excel 2007*. Sedangkan bagian proses dan *output* pada GUI, sepenuhnya menggunakan R versi 3.1.2 dan Rtools versi 3.1.0.1942. Agar file data berupa *xls* dan *xlsx* bisa digunakan di R, maka digunakan *software* tambahan yaitu *ActivePerl*. Adapun paket R yang sangat dibutuhkan dalam proses pembuatan GUI ini yaitu *gdata*, *tcltk2*, dan *devtools*.

### 4.4. Mendesain *Interface* Utama GUI

Ketika semua kebutuhan yang akan digunakan dalam membuat GUI sudah tersedia, maka tahap selanjutnya yaitu mendesain *interface* utama dari GUI. Semua *syntax* yang digunakan untuk membuat *interface* utama GUI disimpan dalam bentuk *file* dengan ekstensi R dengan nama “*mltvrt\_gui*”. Dengan menggunakan perintah `mltvrt_gui()` pada R *Console* yang terdapat dalam *file* `mltvrt_gui.R`, maka akan muncul sebuah *interface* yang ditunjukkan oleh Gambar 3.



Gambar 3. *Interface* Utama GUI

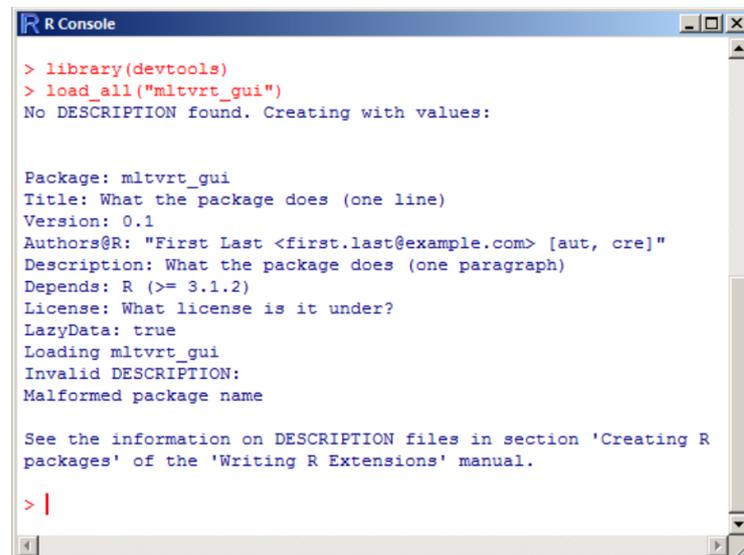
### 4.5. Menyusun dan Menghubungkan Perintah dalam GUI

Dalam pembuatan GUI ini, semua paket yang dibutuhkan terinstal secara *online* saat *software* R dijalankan sehingga paket yang memiliki ketergantungan terhadap paket lainnya akan terunduh secara otomatis. Semua *syntax* yang ada di GUI ini dibagi menjadi 20 *file* R dengan tujuan memudahkan dalam pengelompokan tugas masing-masing *syntax*.

Setelah semua *file* R siap untuk digunakan, maka tahap selanjutnya yaitu menyiapkan sebuah direktori kerja baru atau yang biasa disebut dengan *folder*. Dalam proses pembuatan GUI ini, nama *folder* yang digunakan adalah “*mltvrt\_gui*” beserta sub-direktori

dengan nama “R”, yang akan dijadikan tempat untuk menempatkan semua *file* dengan ekstensi R.

Setelah direktori kerja baru dibuat, semua *file* dengan ekstensi R yang telah dibuat selanjutnya disalin ke direktori tersebut. Dilanjutkan dengan penggunaan paket *devtools* dengan *syntax* yang dimuat pada gambar 5.



```
> library(devtools)
> load_all("mltvrt_gui")
No DESCRIPTION found. Creating with values:

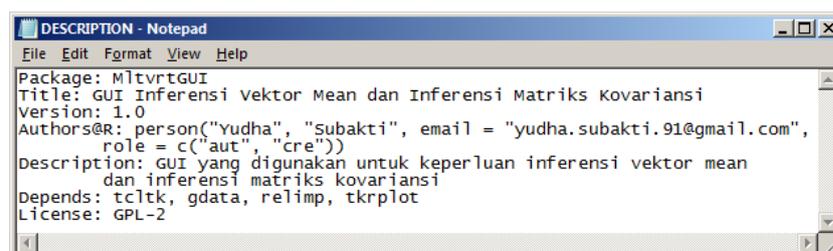
Package: mltvrt_gui
Title: What the package does (one line)
Version: 0.1
Authors@R: "First Last <first.last@example.com> [aut, cre]"
Description: What the package does (one paragraph)
Depends: R (>= 3.1.2)
License: What license is it under?
LazyData: true
Loading mltvrt_gui
Invalid DESCRIPTION:
Malformed package name

See the information on DESCRIPTION files in section 'Creating R
packages' of the 'Writing R Extensions' manual.

> |
```

**Gambar 4.** Penggunaan Paket *devtools* untuk Memuat Semua *File* R

Dari Gambar 4 terdapat peringatan bahwa tidak ditemukan *file DESCRIPTION* pada saat proses memuat semua *file* R sehingga paket *devtools* secara otomatis akan membuat *file DESCRIPTION* di dalam folder “mltvrt\_gui”. Setelah *file DESCRIPTION* tersedia dalam direktori, langkah selanjutnya yaitu mengubah konten yang ada dalam *file* tersebut dengan menggunakan aplikasi *Notepad* di *Windows*. Dalam pembuatan GUI ini, *file DESCRIPTION* diubah sedemikian rupa seperti Gambar 5.



```
DESCRIPTION - Notepad
File Edit Format View Help
Package: MltvrtGUI
Title: GUI Inferensi vektor Mean dan Inferensi Matriks kovariansi
Version: 1.0
Authors@R: person("Yudha", "Subakti", email = "yudha.subakti.91@gmail.com",
  role = c("aut", "cre"))
Description: GUI yang digunakan untuk keperluan inferensi vektor mean
  dan inferensi matriks kovariansi
Depends: tcltk, gdata, relimp, tkrplot
License: GPL-2
```

**Gambar 5.** Konten *File DESCRIPTION* yang Digunakan untuk GUI

Setelah mengubah konten *file DESCRIPTION*, *interface* utama GUI dapat ditampilkan dengan menggunakan perintah `mltvrt_gui()`. Jika tidak terdapat *error* pada salah satu *file* R yang digunakan, maka *interface* utama GUI akan muncul seperti yang ditunjukkan oleh Gambar 3.

#### 4.6. Penggunaan dan Hasil dari GUI

Berdasarkan analisis statistika multivariat yang dilakukan sebelumnya, Gambar 6 dan Gambar 7 merupakan *output* yang ditunjukkan GUI untuk menguji asumsi normal multivariat secara formal menggunakan metode Kolmogorov-Smirnov.



## DAFTAR PUSTAKA

- Dalgaard, P., 2001, *A Primer on the R-Tcl/Tk Package*, R News Vol. 1/3.
- Dalgaard, P., 2001, *The R-Tcl/Tk interface*, DSC 2001 Proceedings of the 2nd International Workshop on Distributed Statistical Computing.
- Dalgaard, P., 2002, *Changes to the R-Tcl/Tk Package*, R News Vol. 2/3.
- Daniel, W.W., 1978, *Applied Nonparametric Statistics*, Houghton Mifflin Company, Boston..
- Johnson, R.A. and Dean, W.W., 2007, *Applied Multivariate Statistical Analysis* 6<sup>th</sup> Edition, Pearson Prentice Hall, New Jersey.
- Lawrence, M.F. and Verzani, J., 2012, *Programming Graphical User Interface in R*, CRC Press, Florida.
- Penners, R., GTK-WIMP (Windows Impersonator), URL: <http://gtk-wimp.sourceforge.net/>
- SAS Institute Inc., *JMP 11: Statistical Discovery Software*, URL: <http://www.jmp.com/>
- Wickham, H., *devtools: Tools to make an R developer's life easier*, URL: <https://github.com/hadley/devtools>
- Wickham, H., 2015, *R Packages*, O'Reilly Media, Inc., California.