

INOVASI PEMBELAJARAN ANALISIS REGRESI DENGAN R GUI

I Gede Nyoman Jaya Jaya
Departemen Statistika FMIPA UNPAD
jay.komang@gmail.com

ABSTRAK. Statistik bagi sebagian mahasiswa khususnya mahasiswa bukan-statistik menjadi phobia tersendiri karena dinilai sulit dimengerti. Rumus-rumus yang kompleks menambah ketakutan mahasiswa untuk berhadapan dengan matakuliah statistik. Salah satu ilmu statistik yang harus dipahami mahasiswa statistik maupun bukan statistik adalah analisis regresi. Analisis regresi banyak digunakan oleh mahasiswa khususnya mahasiswa bukan statistik dalam penyusunan tugas akhir. Seringkali mahasiswa hanya mengetahui bagaimana cara menghitung namun tidak memahami secara detail tentang model dan asumsi klasik dari analisis regresi itu sendiri. Hal ini tidak terlepas dari proses pembelajaran yang hanya mengedepankan cara menghitung dibandingkan pada penekanan filosofi model itu sendiri. Penelitian ini bertujuan mengembangkan metode interaktif pembelajaran analisis regresi dengan mengembangkan sebuah program Graphic User Interface (GUI) dengan program R. Program ini mempermudah dosen untuk memberikan pemahaman tentang asumsi-asumsi dalam analisis regresi. Melalui program ini diharapkan mahasiswa dapat memahami analisis regresi dengan lebih baik

Kata Kunci: *Regresi, R-GUI, Asumsi Klasik*

1. PENDAHULUAN

Statistik bagi sebagian mahasiswa khususnya mahasiswa bukan-statistik menjadi phobia tersendiri karena dinilai sulit dimengerti. Rumus-rumus yang kompleks menambah ketakutan mahasiswa untuk berhadapan dengan matakuliah statistik. Salah satu ilmu statistik yang harus dipahami mahasiswa statistik maupun bukan statistik adalah analisis regresi. Analisis regresi banyak digunakan oleh mahasiswa khususnya mahasiswa bukan statistik dalam penyusunan tugas akhir.

Dalam bidang sosial, metode analisis regresi menjadi salah satu metode paling banyak digunakan untuk menjelaskan hubungan antar fenomena dalam sebuah model linear. Untuk dapat memodelkan, menginterpretasikan dan menggunakan model regresi dalam penelitian maka mahasiswa dituntut memahami dengan baik asumsi-asumsi dari model regresi untuk menjamin bahwa model yang dibuat adalah model yang terbaik sehingga mahasiswa tidak salah dalam mengambil kesimpulan. Seringkali mahasiswa hanya mengetahui bagaimana cara menghitung namun tidak memahami secara detail tentang model dan asumsi klasik dari analisis regresi itu sendiri. Dalam bahasa regresi apakah model yang dibuat sudah memiliki parameter yang memenuhi sifat Best Linear Unbiased Estimation (BLUE) atau belum. Hal ini tidak terlepas dari proses pembelajaran yang hanya mengedepankan cara menghitung dibandingkan pada penekanan filosofi model itu sendiri. Penelitian ini bertujuan mengembangkan metode interaktif pembelajaran analisis regresi dengan mengembangkan sebuah GUI dengan program R. Program ini mempermudah dosen untuk memberikan pemahaman tentang asumsi-asumsi dalam

analisis regresi, apa itu error dan apa itu outlier. Melalui program ini diharapkan mahasiswa dapat memahami analisis regresi dengan lebih baik

2. METODE PENELITIAN

2.1. Analisis Regresi

Analisis regresi digunakan untuk menjelaskan atau memodelkan hubungan fungsional antara satu variabel Y yang disebut variabel respon, output, atau variabel respon, dan satu atau lebih variabel prediktor, masukan, independen atau variabel penjelas $X_1 \dots X_p$. Ketika $p = 1$ maka model regresi dikatakan model regresi sederhana. Jika $p > 1$ disebut dengan model regresi multiple. Namun jika variabel Y lebih dari satu, model regresi dinyatakan sebagai model regresi multivariat. Dalam analisis regresi, variabel respon harus merupakan variabel kontinu tetapi variabel penjelas dapat kontinu, diskrit atau kategori. Analisis Regresi memiliki beberapa tujuan (Faraway, 2002):

1. Memprediksikan nilai variabel respon
2. Menilai akibat atau hubungan antara variabel penjelas dengan variabel respon
3. Memberikan gambaran umum mengenai struktur data

Model dasar untuk analisis regresi adalah pasangan data (x_i, y_i) yang dihubungkan dengan persamaan :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.1)$$

dimana β_0 dan β_1 adalah parameter yang nilainya tidak diketahui dan akan kita taksir. Sedangkan ε adalah kekeliruan acak yang menyatakan penyimpangan dari data observasi dengan model regresi populasi dan tentunya nilai ini juga tidak diketahui dan ditaksir dengan e (*residual*). Metode yang umumnya digunakan untuk menaksir nilai β_0 dan β_1 adalah metode kuadrat terkecil. Metode ini berusaha mendapatkan nilai taksiran β_0 dan β_1

yaitu $\hat{\beta}_0$ dan $\hat{\beta}_1$ yang meminimumkan $\sum_{i=1}^n e_i^2$. Model taksiran regresi sederhana dapat

dituliskan : $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ dengan nilai $e_i = \hat{y}_i - y_i$ menunjukkan perbedaan antara nilai observasi dengan garis regresi sampel. Dengan menggunakan penyelesaian matematis maka diperoleh :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

dengan,

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.2)$$

Penaksir parameter regresi harus memiliki sifat **BLUE** (*Best Linear Unbiased Estimation*). Untuk mendapatkan taksiran yang *Best Linear Unbiased Estimation* (BLUE) asumsi klasik yang harus dipenuhi terkait dengan ε_i adalah (Jaya, 2016):

1. Harapan nilai kekeliruan sama dengan nol [$E(\varepsilon_i|X) = 0$]
2. Non Heteroskedastisitas [$var(\varepsilon_i|X) = \sigma^2$]

Asumsi ini menyatakan bahwa model regresi memiliki varians error yang konstan yaitu σ^2

3. Non Autokorelasi [$cov(\varepsilon_i, \varepsilon_j|X) = 0$]

Asumsi ini menyatakan bahwa kekeliruan dalam model regresi tidak saling berkorelasi.

4. Normality [$\varepsilon_i \sim N(0, \sigma^2)$]

Asumsi ini menyatakan bahwa kekeliruan berdistribusi normal.

Diluar keempat asumsi di atas, ada satu asumsi yang juga penting yaitu asumsi non multikolenieritas. Asumsi ini menyatakan bahwa tidak ada hubungan yang sempurna atau hampir sempurna antara variabel bebas X.

Jika asumsi ini terpenuhi, maka beberapa uji statistik dapat digunakan seperti uji F dan t . Asumsi BLUE akan dibahas lebih detail pada bagian estimasi.

Melalui pendekatan visualisasi, mahasiswa diajak memahami analisis regresi terkait dengan bagaimana dampak pelanggaran asumsi dengan lebih baik.

2.2 R-GUI

Software R hadir dengan lisensi gratis tetapi mempunyai kemampuan yang tidak terbatas karena sifatnya yang *open source*. *Software* R merupakan program aplikasi Statistika yang dirancang dengan kebutuhan minimal dan memiliki banyak fungsi-fungsi Statistik yang siap pakai sehingga memudahkan para pengguna untuk mempergunakannya, tidak seperti program komersial lain seperti SPSS, Statistika 7 atau SAS dimana fungsi-fungsi statistiknya *built in* dalam program utamanya. Karena R dirancang untuk dapat dikembangkan sendiri oleh para penggunanya yaitu dengan menulis fungsi Statistiknya sendiri sesuai dengan kebutuhan analisis statistik maka pengembangan software ini sangat cepat. Fungsi-fungsi statistik yang telah ada di simpan dalam sebuah *Package* yang tersimpan dalam *library* program R yang bisa di panggil ketika akan dipergunakan. Untuk mendapatkannya dapat di download secara gratis di <http://CRAN.R-project.org> (Jaya, 2010).

R-GUI adalah program sebuah interface yang memungkinkan pengguna R untuk berinteraksi melalui ikon grafis dan indikator visual. Beberapa package telah dikembangkan untuk memudahkan pengguna R membuat R-GUI. Salah satu yang paling banyak digunakan adalah package **gWidgets**.

3. HASIL PENELITIAN

3.1. SimReg Package

Dalam pengajaran Analisis Regresi penulis membuat sebuah package R dengan aplikasi GUI yang penulis beri nama R GUI. Package tersebut dibuat dalam bentuk function sebagai berikut:

```
> SimReg
function() {
  library(gWidgets)
  options("guiToolkit"="RGtk2")
  plotReg <- function(n,miu,s1,s2,b0,b1) {
    set.seed(1)
    e <- rnorm(svalue(n),0,svalue(s1))
    set.seed(100)
    x<-rnorm(svalue(n),svalue(miu),svalue(s2))
```

```

y<-svalue(b0)+svalue(b1)*x+e
fit<-lm(y~x)
.....hubungi
jay.komang@gmail.com.....

tbl[5,1] <- "Parameter of Intercept "
tbl[5,2,          expand=TRUE]          <-          (b0          <-
gslider(from=0,to=10,by=1,value=1, container=tbl,

handler=function(h,...)plotReg(n,miu,s1,s2,b0,b1))

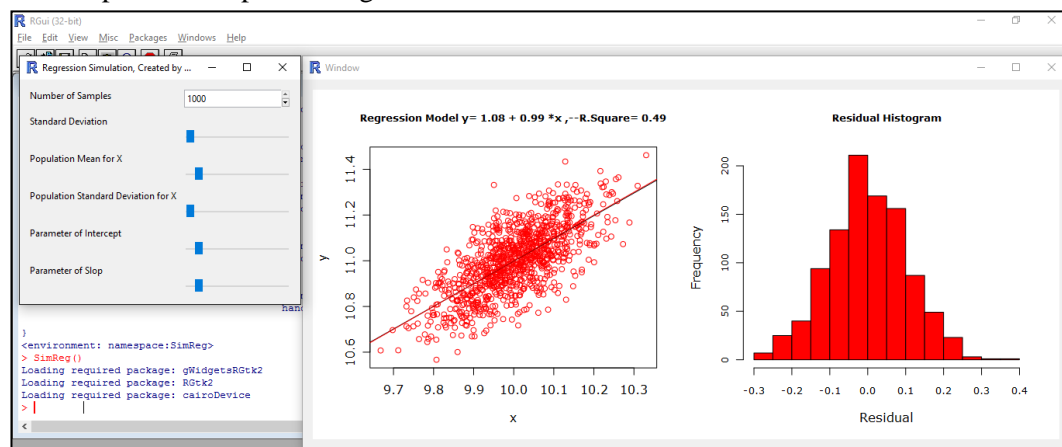
tbl[6,1] <- "Parameter of Slop"
tbl[6,2,          expand=TRUE]          <-          (b1          <-
gslider(from=0,to=10,by=1,value=1, container=tbl,
}

```

Package ini bisa dibuka pada R3.23. Untuk menjalankan package ini, terlebih dahulu package SimReg harus di install dulu dalam program R, dengan tahapan operasi sebagai berikut:

1. Install package SimReg.01
2. library(SimReg)
3. SimReg()

Akan diperoleh tampilan sebagai berikut:



Gambar 3.1. Visualiasi Package SimReg_0.1.0

Package ini dapat dimanfaatkan untuk mengajarkan regresi terkait dengan:

1. Bagaimana pengaruh ukuran sampel terhadap kenormalan residual
2. Bagaimana dampak jika error memiliki varians yang besar
3. Bagaimana dampak jika variabel independen memiliki varians yang besar

3.2. Multico Pacakge

Selain package SimReg penulis juga mengembangkan package Multico untuk mengajarkan kepada mahasiswa konsep dasar multikolenieritas yang meliputi apa itu multikolenieritas, apa dampaknya dan apakah benar menambah ukuran sampel mampu memperbaiki kondisi multikolenieritas.

```

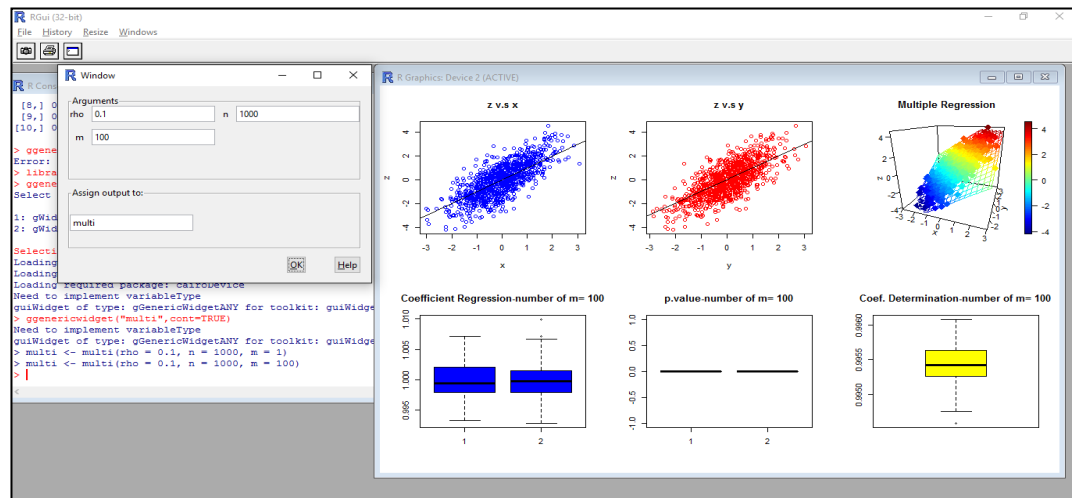
> multi
function (rho, n, m)
{
  coeff <- matrix(rep(0, 3 * m), m, 3)
  p.value <- matrix(rep(0, 3 * m), m, 3)
  R2 <- matrix(rep(0, m), m, 1)
  for (j in 1:m) {
    x <- rnorm(n, 0, 1)
    y <- rho * x + rnorm(n, 0, 1) * (1 - rho^2)^0.5
    e <- rnorm(n, 0, 0.1)
    z <- x + y + e
    data <- data.frame(x, y, z)
    fit <- lm(z ~ x + y, data = data)
    model <- summary(fit)
    par(mfrow = c(2, 3))
    .....hubungi
    jay.komang@gmail.com.....
  }
  list(coeff, p.value, R2)
}

```

Package ini bisa dibuka pada R3.23. Untuk menjalankan package ini, terlebih dahulu package Multico_0.1.0 harus di install dulu dalam program R, dengan tahapan operasi sebagai berikut:

1. Install package Multico_0.1.0
2. library(multi)
3. multi()

Akan diperoleh tampilan sebagai berikut:



Gambar 3.2. Visualiasi Package Multico_0.1.0

Paket ini dapat dimanfaatkan untuk mengajarkan regresi terkait dengan:

1. Bagaimana dampak multikoleniritas terhadap taksiran parameter regresi
2. Bagaimana dampak multikoleniritas terhadap uji signifikansi
3. Apakah menambah ukuran sampel menjadi solusi dari kasus multikolenieritas.

3.3. Hetero Pacakge

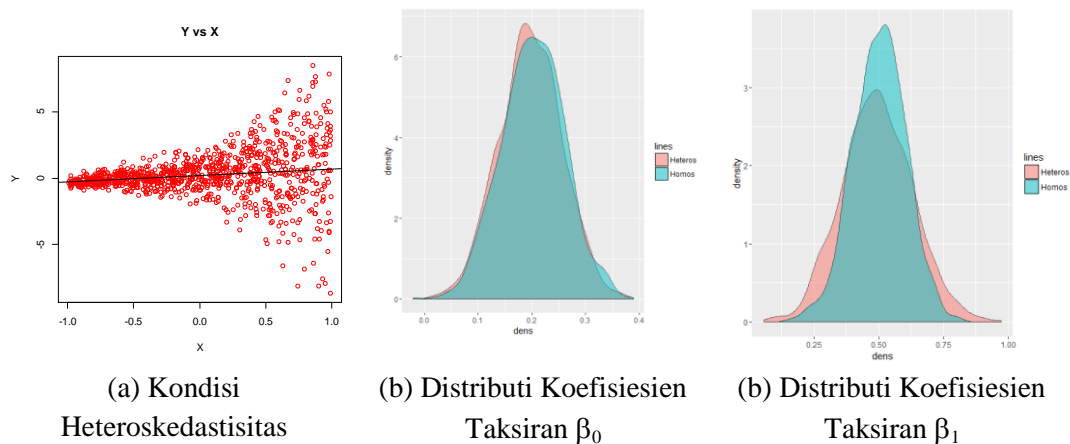
Selain package SimReg dan Multico penulis juga mengembangkan package Hetero untuk mengajarkan kepada mahasiswa konsep dasar heteroskedastisitas.

```
Hetero<-function(n,m,b0,b1,gamma){
set.seed(100484)
par.est.ncv<-matrix(NA,nrow=m,ncol=4)
sigma.est<-numeric(m)
X<-runif(n,-1,1)
for(i in 1:reps){
sigma0<-exp(X*gamma)
Y<-b0+b1*X+rnorm(n,0,exp(X*gamma))
model<-lm(Y~X)

.....hubungi
jay.komang@gmail.com.....

par.est.ncv[i,4]<-sqrt(diag(vcv)[2])
}
}
```

Akan diperoleh tampilan sebagai berikut:



Gambar 3.3. Visualiasi Package Hetero_0.1.0

Melalui package ini dapat disampaikan ke pada mahasiswa bahwa adanya heteroskedastisitas menyebabkan varians taksiran cenderung under estimate khusus untuk koefisien slop, seperti yang ditunjukkan pada Gambar 3.3.

4. SIMPULAN

1. Program R dapat dimanfaatkan untuk membantu dalam proses belajar mengajar statistik khususnya analisis regresi
2. Setiap pengajaran statistik harus didukung oleh program aplikasi yang interaktif sehingga menarik minat mahasiswa untuk belajar statistik dan balajar statistik tidak terlibat membosankan.

DAFAR PUSTAKA

- [1] Faraway, Julian J. 2002. Practical Regression and Anova Using R. (?)
- [2] Jaya, Mindra (2016), Analisis Regresi. Penerbit Departemen Statistika FMIPA UNPAD
- [3] Jaya, Mindra (2010), Komputasi Statistik dengan R. Penerbit Departemen Statistika FMIPA UNPAD
- [4] Verzani, John (2015). gWidgets API for building toolkit-independent, interactive GUIs. CRAN R
- [5] Weisber, Stanford. 2005. Applied Linear Regression. Third Edition. University of Minsesota : John Wiley