

ANALISIS REGRESI BERGANDA TERHADAP PENGARUH JUMLAH SISWA DAN GURU DI SEKOLAH SMA PROVINSI LAMPUNG PERIODE 2021-2024

Cahya Rahmawati

Universitas Koperasi Indonesia

Email : cahyarahma170@gmail.com

ABSTRAK

Analisis data ini bertujuan untuk menganalisis pengaruh jumlah siswa dan jumlah guru terhadap Sekolah Menengah Atas (SMA) di Provinsi Lampung pada periode 2021-2024 menggunakan metode regresi berganda. Data yang digunakan dalam tugas analisis data ini diperoleh dari Website Kaggle di website tersebut data ini diperoleh dari Data Pokok Pendidikan dan Kebudayaan, yang mencakup informasi jumlah siswa, guru dan sekolah di Provinsi Lampung dari Semester Ganjil hingga Genap.

Metode regresi berganda (multiple) digunakan untuk mengidentifikasi hubungan anatra variabel independen (jumlah siswa dan jumlah guru) dengan variabel dependen (jumlah sekolah). Hasil analisis menunjukkan bahwa baik jumlah siswa maupun jumlah guru memiliki pengaruh signifikan terhadap jumlah sekolah di Provinsi Lampung. Model regresi yang dihasilkan memiliki nilai R-Squared sebesar 0.8972 yang menunjukkan bahwa 89.72% variasi dalam jumlah sekolah dapat dijelaskan oleh jumlah siswa dan guru. Koefesien regresi untuk jumlah siswa adalah 0.0007012 dan untuk jumlah guru adalah 0.0336576 dengan nilai p masing-masing kurang dari 0.05, menandakan nilai p untuk jumlah siswa (X_1) adalah 0.4400 (tidak signifikans) dan nilai p untuk jumlah guru (x_2) adalah 0.0196 (signifikan statistik).

Studi ini juga menggaris bawahi pentingnya mempertimbangkan faktor jumlah siswa dan guru dalam penentuan kebutuhan dan distribusi sekolah.

1. PENDAHULUAN

Pendidikan merupakan aspek dalam pembangunan suatu daerah, yang tidak terlepas dari ketersediaan dan distrbusi sekolah. Di Provinsi Lampung jumlah Sekolah Menengah Atas (SMA) menjadi indikator penting dalam menilai aksesibilitas dan kualitas pendidikan.

Distribusi jumlah sekolah sangat mempengaruhi ketersediaan akses pendidikan bagi masyarakat. Pertumbuhan jumlah siswa dan kebutuhan tenaga pendidik menjadi faktor penting dalam merencanakan pembangunan sekolah yang efektif dan efesien.

Analisis data ini bertujuan untuk mengidentifikasi seberapa besar pengaruh jumlah siswa dan guru terhadap jumlah di SMA Provinsi Lampung. Hasil analisis data menggunakan metode Regresi Berganda (multiple) diharapkan dapat memberikan pandangan yang lebih mendalam bagi pengambil kebijakan pendidikan dalam merencanakan distribusi sumber daya pendidikan yang optimal.

Metode Regresi berganda dipilih untuk menganalisis hubungan simultan antara jumlah variabel independen (jumlah siswa dan guru) dengan variabel dependen (jumlah sekolah). Analisis ini akan dilakukan menggunakan aplikasi RStudio, yang memungkinkan pengolahan data yang efisien dan penyajian hasil analisis yang akurat. Dengan memahami faktor-faktor yang mempengaruhi jumlah sekolah SMA, diharapkan hasil analisis ini dapat memberikan kontribusi yang signifikan dalam perencanaan strategis pengembangan pendidikan di Provinsi Lampung.

2. TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Menurut Montgomery, peck, and Vining (Introduction to Linear Regression Analysis, 2012), analisis regresi berganda digunakan untuk memodelkan hubungan linear antara variabel dependen Y dengan dua atau lebih dari variabel independen X_1, X_2, \dots, X_k . Tujuannya adalah untuk menentukan sejauh mana masing-masing variabel independen mempengaruhi variabel dependen dan apakah ada interaksi antara variabel-variabel tersebut.

2.2 Regresi Linear Berganda

Bentuk umum model regresi linear berganda dengan k variabel independen adalah :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Dimana Y adalah variabel dependen, X_1, X_2, \dots, X_k adalah variabel-variabel independen, ε adalah galat acak (random error) dan $\beta_0, \beta_1, \dots, \beta_k$ adalah parameter-parameter populasi yang nilainya tidak diketahui.

Secara umum, diantar variabel-variabel X_1, X_2, \dots, X_k boleh merupakan fungsi dari variabel yang lain, tapi tidak boleh memuat parameter.

2.3 Uji Asumsi Klasik

- Uji Linearitas
Memastikan bahwa hubungan antara variabel independen (x) dan dependen (y) adalah linear. Maka ini penting karena regresi linear berganda mengasumsikan bahwa setiap variabel independen memiliki hubungan linear dengan variabel dependen.
- Uji Normalitas
Memastikan bahwa residual (kesalahan prediksi) dari model regresi berdistribusi normal ini penting untuk validitas inferensi statistik, seperti uji t dan F .
Shapiro-Wilk Test. Uji statistik yang digunakan untuk menguji hipotesis bahwa data berasal dari distribusi normal. Nilai p yang signifikan menunjukkan pelanggaran asumsi normalitas.
- Uji Heteroskedastisitas
jika varians residual tidak konstan (heteroskedastisitas) hasil regresi bias menjadi tidak valid.
Breusch-Pagan Test. Menguji apakah varians residual bergantung pada variabel independen, nilai p yang signifikan menunjukkan adanya heteroskedastisitas.

- **Uji Auto Korelasi**
Memastikan bahwa residual dari model tidak saling memiliki korelasi. Autokorelasi dapat menyebabkan estimasi dan inferensi statistik yang salah.
Durbin-Watson Test. Uji statistik untuk mendeteksi autokorelasi di residual. Nilai antara 1.5 dan 2.5 biasanya menunjukkan tidak adanya autokorelasi yang signifikan.
- **Asumsi Multikolinieritas**
Memastikan bahwa tidak ada hubungan linier yang kuat antara variabel independen. Multikolinieritas dapat membuat estimasi koefisien regresi menjadi tidak stabil dan sulit diinterpretasikan.
Variance Inflation Factor (VIF). Mengukur seberapa besar variabilitas koefisien regresi yang diperkirakan meningkat karena multikolinieritas. VIF lebih besar dari 10 mengindikasikan masalah multikolinieritas
Correlation Matrix. Korelasi tinggi misalnya, diatas 0.8 mengindikasikan multikolinieritas.

3 METODE PENELITIAN

Penelitian ini menggunakan metode Analisis Regresi Berganda menggunakan aplikasi RStudio untuk menginvestigasi pengaruh jumlah siswa dan guru terhadap jumlah sekolah. Di Sekolah Menengah Atas (SMA) Provinsi Lampung periode 2021-2024. Untuk variabel keputusan pada penelitian ini adalah Siswa yang dinotasikan sebagai X1, Guru yang dinotasikan sebagai X2, dan Sekolah yang dinotasikan sebagai Y.

Berikut adalah langkah-langkah metodologi yang akan digunakan :

1. Pengumpulan data :

Data yang digunakan berasal dari Data Pokok Pendidikan Kementerian Pendidikan dan Kebudayaan <https://www.kaggle.com/datasets/feryadiyulis/sma-di-provinsi-lampung-tahun-2021-2024> saya mengakses di website kaggle.

2. Pemrosesan Data :

Data akan dimuat dan diproses menggunakan aplikasi RStudio. Langkah ini mencakup pengimporan data dan persiapan data untuk analisis regresi.

3. Analisis Regresi Berganda :

Variable dependen dalam analisis ini adalah Sekolah SMA di Provinsi Lampung. Variabel independen adalah jumlah siswa dan jumlah guru. Regresi berganda akan dilakukan untuk mengidentifikasi hubungan simultan antara variabel independen dan variabel dependen.

4. Interpretasi dan Kesimpulan :

Evaluasi ini akan dilakukan berdasarkan koefisien regresi, p-nilai, dan interpretasi lainnya. Kesimpulan akan menguraikan temuan utama dari analisis regresi berganda dan menggambarkan implikasi potensialnya bagi pengambil kebijakan Pendidikan di Provinsi Lampung.

4 HASIL PENELITIAN DAN PEMBAHASAN

- Melakukan Analisis Regresi Berganda untuk menganalisis regresi variabel Y sebagai fungsi dari X1 dan X2 dapat digunakan seperti perintah lm() sebagai berikut :

```
> #analisis regresi
> g2 <- lm(Sekolah_Y ~ Siswa_X1 + Guru_X2, data=data2)
> g2

Call:
lm(formula = Sekolah_Y ~ Siswa_X1 + Guru_X2, data = data2)

Coefficients:
(Intercept)      Siswa_X1      Guru_X2
  2.2459370      0.0007012      0.0336576
```

Dalam hal ini Y adalah Sekolah variabel respons yakni Siswa X1 dan Guru X2.

Berdasarkan data diatas dapat disimpulkan bahwa :

$$\beta_0 = 2.2459370$$

$$\beta_1 = 0.0007012$$

$$\beta_2 = 0.0336576$$

- Melakukan summary dari hasil analisis untuk menampilkan hasil analisis sebelumnya dapat digunakan seperti sebagai berikut :

```
> summary(g2)

Call:
lm(formula = Sekolah_Y ~ Siswa_X1 + Guru_X2, data = data2)

Residuals:
    Min       1Q   Median       3Q      Max
-14.4391  -4.6413   0.5518   6.2727  10.0170

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.2459370   1.4856726   1.512   0.1342
Siswa_X1     0.0007012   0.0009038   0.776   0.4400
Guru_X2      0.0336576   0.0141509   2.378   0.0196 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.536 on 87 degrees of freedom
Multiple R-squared:  0.8972,    Adjusted R-squared:  0.8948
F-statistic: 379.5 on 2 and 87 DF,  p-value: < 2.2e-16

> anova(g2)
Analysis of Variance Table

Response: Sekolah_Y
      Df Sum Sq Mean Sq  F value    Pr(>F)
Siswa_X1    1  32186    32186  753.4413 < 2e-16 ***
```

```

Guru_X2      1      242      242      5.6572 0.01957 *
Residuals 87      3716      43
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Perintah ini dapat menghasilkan nilai min q1, med, q3, dan max pada residual data, table Anova, nilai F statistik sampai nilai R-squared (R^2)

Pendugaan Parameter : berdasarkan hasil summary pada data diatas adalah sebagai berikut :

$$Y = 2.2459370 + 0.0007012 X_1 + 0.0336576 X_2$$

Intersep : ini menunjukkan jumlah sekolah diperkirakan akan menjadi 2.2459 ketika jumlah siswa dan jumlah guru adalah nol.

Koefisien Siswa_X1 : ini menunjukkan bahwa setiap penambahan satu siswa akan meningkatkan kembali jumlah sekolah sebanyak 0.0007012, meskipun nilai ini tidak signifikan (p-value = 0.4400)

Koefisien Guru_X2 : ini menunjukkan bahwa setiap penambahan satu guru akan meningkatkan jumlah sekolah sebesar 0.0336576, dan nilai ini sangat signifikan pada tingkat signifikansi 5% (p-value = 0.0196)

Uji Simultan : nilai p-value (2.2×10^{-16}) $< \alpha$ (0.05) maka H_0 ditolak. Dapat disimpulkan bahwa terdapat pengaruh yang signifikan dari jumlah siswa dan jumlah guru terhadap jumlah sekolah di SMA Provinsi Lampung periode 2021-2024..

Uji Parsial : Untuk uji parsial (t-test) pada variabel independen Siswa_X1 dan Guru_X2 kita akan menguji hipotesis alternative (H_1) untuk masing-masing variabel. Pada tingkat signifikansi (α) 5% hipotesis tersebut dapat dirumuskan sebagai berikut :

Siswa_X1 : H_0 : koefisien regresi untuk Siswa_X1 adalah nol (tidak ada pengaruh yang signifikan dari jumlah siswa terhadap jumlah sekolah).

$$H_0: \beta_{Siswa_X1} = 0$$

H_1 : koefisien regresi untuk Siswa_X1 tidak nol (ada pengaruh yang signifikan dari jumlah siswa terhadap jumlah sekolah).

$$H_1: \beta_{Siswa_X1} \neq 0$$

Guru_X2 : H_0 : koefisien regresi untuk Guru_X2 adalah nol (tidak ada pengaruh yang signifikan dari jumlah guru terhadap jumlah sekolah).

$$H_0: \beta_{Guru_X1} = 0$$

H_1 : koefisien regresi untuk Guru_X2 tidak nol (ada pengaruh yang signifikan dari jumlah guru terhadap jumlah sekolah).

$$H_1: \beta_{Guru_X1} \neq 0$$

Keputusan berdasarkan nilai p-value sebagai berikut :

Siswa_X1 : Karena nilai p-value (0.4400) $> \alpha$ (0.05) maka H_0 di terima.

Guru_X2 : Karena nilai p-value (0.0196) $< \alpha$ (0.05) maka H_0 di tolak.

Kesimpulan :

Siswa_X1 H_0 diterima karena jumlah siswa tidak berpengaruh signifikan terhadap jumlah sekolah.

Guru_X2 H_0 ditolak karena jumlah guru berpengaruh signifikan terhadap jumlah sekolah

R-Squared :

Multiple R-Squared 0.8972, berarti sekitar 89,72% variasi jumlah sekolah dapat dijelaskan oleh variasi dalam jumlah siswa dan jumlah guru.

Adjusted R-Squared 0.08948, nilai R-Squared yang telah disesuaikan dengan jumlah variabel dalam model.

Residual :

Maka nilai residual menunjukkan sebaran kesalahan prediksi dari model, dengan minimum -14.4391 dan maksimum 10.0170.

Dengan demikian, model regresi berganda ini menunjukkan bahwa jumlah guru memiliki pengaruh yang signifikan terhadap jumlah sekolah di SMA Provinsi Lampung periode 2021-2024, sedangkan jumlah siswa tidak berpengaruh signifikan.

- Pendeteksian Multikolinieritas

```
> library(car)
Loading required package: carData
> vif(g2)
Siswa_X1  Guru_X2
76.4823   76.4824
```

Variance Inflation Factor (VIF) untuk mendeteksi multikolinieritas, nilai VIF yang tinggi menunjukkan adanya multikolinieritas.

Jika nilai VIF untuk 'Siswa_X1' dan 'Guru_X2' tetap sama besar, maka hal ini menunjukkan adanya korelasi yang sangat tinggi antara keduanya dalam data.

- Normalitas Galat

```
library(tseries)
> sisa<- residuals (g2)
> sisa
```

	1	2	3	4	5	6	
7	-4.5945215	9.2855876	9.1096316	3.8551023	8.2788020	2.2683462	0.8
746905							
14	-3.2285037	-7.2396865	-3.8174678	2.6102507	-0.5416213	6.3596104	-13.6
173081							
21	-5.4068714	-4.6569251	8.5444402	7.7325284	2.6005695	7.7704581	2.1
260065							
28	0.4112177	-3.5398256	-7.5755611	-3.9899610	2.4945553	-0.5563441	6.0
118313							
35	-14.4391081	-5.5842744	-4.9009489	9.3535988	8.6903202	2.5331330	8.2
051832							

```

36      37      38      39      40      41
42 2.2886765 0.6923844 -3.1429583 -7.4739026 -3.9268613 2.5155783 -1.3
074342
43      44      45      46      47      48
49 6.6562164 -14.3844480 -5.6340724 -4.7536960 9.6915804 8.2653088 0.8
402565
50      51      52      53      54      55
56 8.2093936 2.5425115 0.9244817 -3.0952740 -6.8259929 -4.1091708 2.6
712469
57      58      59      60      61      62
63 -1.0564039 7.0348654 -13.7197104 -6.5828908 -5.2228054 8.2707299 10.0
170439
64      65      66      67      68      69
70 0.9748663 9.4772365 3.8440155 0.1088810 -3.2242967 -7.1499562 -3.7
782231
71      72      73      74      75      76
77 1.9447026 -0.5908190 6.5845865 -13.6391055 -6.1572740 -4.8076958 7.1
585437
78      79      80      81      82      83
84 7.4545924 1.1095000 8.7696390 3.9337696 -1.8791992 -3.1527746 -6.9
935864
85      86      87      88      89      90
-4.0432750 2.4923390 -0.5894150 7.3671265 -13.8221161 -6.2036500
> shapiro.test(sisa)

Shapiro-wilk normality test

data: sisa
W = 0.95065, p-value = 0.001837

```

Ho : Galat menyebar normal

H1 : Galat tidak menyebar normal. Berdasarkan uji shapiro walk diatas, diperoleh nilai p-value (0.001837) > α (0.05) maka Ho ditolak. Dapat disimpulkan bahwa galat menyebar secara tidak normal, artinya uji asumsi normalitas tidak terpenuhi.

- Homogenitas Ragam Galat

```

bptest(g2)

studentized Breusch-Pagan test

data: g2
BP = 64.231, df = 2, p-value = 1.128e-14

```

Dari hasil analisis Galat menggunakan Uji Breusch-Pagan untuk Homogentias Ragam Galat yaitu: Nilai p-value sangat kecil (1.128^{-14}), jauh lebih kecil dari tingkat signifikasi umum (0.05). maka dari itu H0 ditolak. Terdapat bukti yang cukup untuk menyimpulkan bahwa terdapat pelanggaran terhadap asumsi galat dalam analisis regresi berganda ini.

- Non Autokorelasi Galat

```
dwtest(g2)

Durbin-Watson test

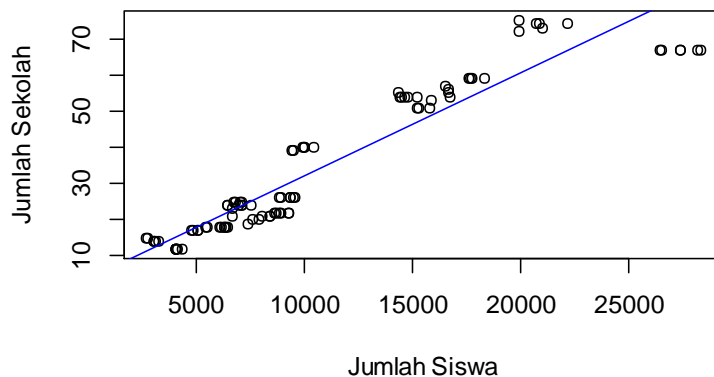
data: g2
DW = 1.5372, p-value = 0.01231
alternative hypothesis: true autocorrelation is greater than 0
```

Nilai p-value dari hasil Non Autokorelasi Galat (0.01231) lebih kecil dari 0.05. maka terdapat bukti yang cukup untuk menolak H_0 bahwa tidak ada autokorelasi positif bahwa dalam galat regresi atau dalam kata lain ada indikasi adanya autokorelasi positif dalam galat tersebut.

Plot Regresi Berganda

```
> #plot
> setwd("D:/semester 4/ANALISIS REGRESI")
> data2 <- read.csv("TUGAS.csv")
> model<-lm(Sekolah_Y~Siswa_X1+Guru_X2, data=data2)
> #membuat plot regresi berganda
> #plot antara variabel independen sekolah dan dependen guru&siswa
> plot(data2$Siswa_X1, data2$Sekolah_Y, main="Plot Regresi Berganda",
+       xlab="Jumlah Siswa", ylab="Jumlah Sekolah")
> abline(lm(Sekolah_Y~Siswa_X1, data=data2), col="blue")
```

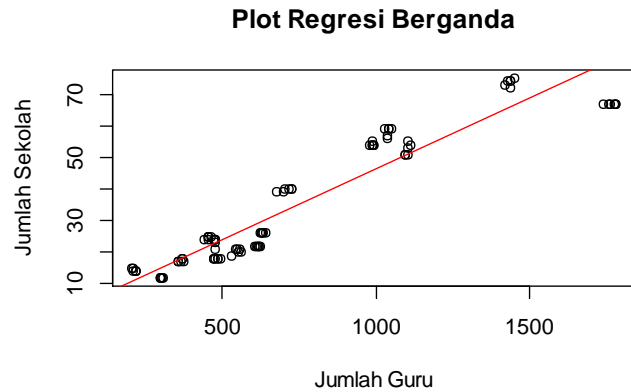
Plot Regresi Berganda



Interpretasi Hubungan antara Jumlah Siswa dan Jumlah Sekolah:

Garis regresi yang digambarkan menunjukkan tren linier positif, yang berarti terdapat kecenderungan bahwa semakin banyak jumlah siswa, maka semakin banyak jumlah sekolah. Ini bisa menunjukkan bahwa peningkatan jumlah siswa di suatu daerah diikuti oleh peningkatan jumlah sekolah yang diperlukan untuk mengakomodasi siswa tersebut.

```
> #plot antara variabel independen dan variabel dependen
> plot(data2$Guru_X2, data2$Sekolah_Y, main="Plot Regresi Berganda",
+       xlab="Jumlah Guru", ylab="Jumlah Sekolah")
> abline(lm(Sekolah_Y~Guru_X2, data=data2), col="red")
```

Interpretasi Hubungan antara Jumlah Guru dan Jumlah Sekolah:

Garis regresi yang digambarkan juga menunjukan tren linear positif, yang berarti semakin banyak jumlah guru, semakin banyak jumlah sekolah. Ini diartikan bahwa peningkatan jumlah guru di suatu daerah biasanya diikuti oleh peningkatan jumlah sekolah untuk menampung tenaga pendidid tersebut.

Artinya, kedua plot menunjukan bahwa terdapat hubungan ppositif antara variabel independen (Jumlah Siswa dan Jumlah Guru) dengan variabel dependen (Jumlah Sekolah). Maka peningkatan dalam jumlah siswa dan jumlah guru cenderung diikuti oleh peningkatan dalam jumlah sekolah. Analisis regresi berganda memberikan gambaran bahwa baik jumlah siswa maupun jumlah guru merupakan faktor penting yang mempengaruhi jumlah sekolah disuatu wilayah.

5 KESIMPULAN

Berdasarkan analisis berganda yang dilakukan, terdapat beberapa temuan penting mengenai pengaruh jumlah siswa dan guru terhadap jumlah sekolah di Provinsi Lampung pada periode 2021-2024.

Koefesien regresi untuk jumlah siswa adalah 0.0007012, namun nilai p value ($0.4400 > \alpha$ (0.05), menunjukan bahwa pengaruh jumlah siswa terhadap jumlah sekolah tidak signifikan secara statistic.

Koefesien regresi untuk jumlah guru adalah 0.0336576, dengan nilai p-value ($0.0196 < \alpha$ (0.05), menunjukan bahwa jumlah guru memiliki pengaruh yang signifikan terhadap jumlah sekolah.

Multiple R-Squared sebesar 0.8972 ini menunjukan bahwa 89.72% variasi jumlah sekolah dapat dijelaskan oleh variasi jumlah siswa dan guru.

Multikolinieritas. VIF yang sangat tinggi menunjukan adanya multikolinieritas antara jumlah siswa dan juga guru.

Normalitas Residual, Uji Shapiro-Wilk menunjukan residual tidak menyebar normal (p-value = 0.001837).

Heteroskedastisitas. Uji Breusch-Pagan menunjuka bahwa residual tidak menyebar normal ($p\text{-value} = 1.128^{-14}$).

Autokorelasi Residual. Uji Durbin-Watson yaitu adanya autokorelasi positif dalam residual ($p\text{-value} = 0.01231$).

6 DAFTAR PUSTAKA

Gareth James – Daniela Witten – Trevor Hastie – Robert Tibshirani – Jonathan Taylor | An Introduction to Statistical Learning with application in R. Springer. Retrieved From <https://www.statlearning>.

Rebecca Bevans | Published on February 25, 2020 | Linear Regression in R A step by step Guide and Examples. <https://www.scribbr.com/statistics/linear-regression-in-r/>