

# KAJIAN KORELASI ANTAR PEUBAH BEBAS DALAM REGRESI LINIER BERGANDA TERHADAP KOEFISIEN DETERMINASINYA

Hernales Sunardi<sup>1</sup>, Sigit Nugroho<sup>2</sup>, dan Baki Swita<sup>2</sup>

*1 Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu*

*2 Staf Pengajar Matematika Fakultas MIPA Universitas Bengkulu*

## ABSTRAK

Penelitian ini bertujuan mengkaji dan mempelajari tentang hubungan korelasi yang terjadi antar peubah bebas dalam regresi linier berganda serta mengetahui pengaruh korelasi tersebut terhadap koefisien determinasinya. Metode yang digunakan dalam penulisan ini adalah studi literatur dan data yang digunakan adalah data *trial and error* yang dibuat menggunakan program komputer Microsoft EXCEL. Hasil penelitian menunjukkan bahwa korelasi yang terjadi antar peubah bebas dalam regresi linier berganda akan berpengaruh terhadap nilai koefisien determinasi yang dihasilkan dan model regresi yang baik dihasilkan jika antar peubah bebasnya tidak saling berkorelasi.

*Kata Kunci : Korelasi, Regresi Linier Berganda, Koefisien Determinasi.*

## 1. PENDAHULUAN

Pada proses pengolahan data peneliti akan selalu berkepentingan untuk menentukan hubungan antara dua peubah atau lebih dan seberapa kuat hubungan antara peubah-peubah tersebut. Jika ingin mengetahui keeratan antar peubah maka digunakan analisis korelasi, sedangkan jika ingin mengetahui bentuk hubungan dua peubah atau lebih, digunakan analisis regresi.

Korelasi dan Regresi keduanya mempunyai kaitan yang sangat erat. Setiap regresi pasti ada korelasinya, sedangkan korelasi belum tentu dilanjutkan dengan regresi. Hubungan antara peubah-peubah yaitu antara peubah yang telah diketahui dengan peubah yang akan diramalkan, diformulasikan dalam bentuk persamaan matematis. Persamaan matematis yang memungkinkan untuk meramalkan nilai-nilai suatu peubah tak bebas dari nilai-nilai satu atau lebih peubah bebas disebut persamaan regresi. Peubah yang nilainya ingin diduga berdasarkan persamaan regresi disebut peubah tak bebas, dan peubah yang digunakan sebagai dasar untuk membuat pendugaan disebut peubah bebas.

Secara umum regresi merupakan metode statistika yang digunakan untuk mengetahui ada tidaknya hubungan linier antara satu atau beberapa peubah bebas terhadap satu peubah tak bebas. Jika nilai peubah tak bebas diduga berdasarkan satu peubah bebas saja, maka dinamakan regresi sederhana (*Simple Regression*). Jika nilai peubah tak bebas diduga berdasarkan dua atau lebih peubah bebas, maka dinamakan regresi linier berganda (*Multiple Regression*).

Analisis regresi juga dapat digunakan untuk melakukan pendugaan terhadap parameter yang ada dalam model serta untuk kepentingan peramalan. Jika tujuan utama dari pembentukan model regresi adalah untuk meramalkan peubah tak bebas dari peubah-peubah bebas tertentu, maka model yang terbentuk haruslah model regresi terbaik, yaitu model yang mencerminkan pola hubungan yang sesungguhnya antara peubah bebas dengan peubah tak bebas. Setiap pembentukan model regresi terbaik

dapat dilakukan dengan melihat berbagai kriteria, diantaranya adalah koefisien determinasi ( $R^2$ ), koefisien determinasi yang disesuaikan ( $\bar{R}^2$ ), uji analisis varians, uji parsial (uji-t), dan kuadrat tengah galat.

Demikian juga dalam pembentukan model regresi linier berganda, dapat dilihat dengan salah satu dari kriteria tersebut, yaitu dengan koefisien determinasi ( $R^2$ ), dan koefisien determinasi yang disesuaikan ( $\bar{R}^2$ ). Dalam regresi linier berganda, kebaikan model dari  $R^2$  dan pengujian hipotesis dilihat dari koefisien regresi. Jika nilai  $R^2$  dekat dengan satu maka makin baik kecocokan model dengan data dan sebaliknya, jika nilai  $R^2$  dekat dengan nol maka kecocokan model dengan data kurang baik.

Akan tetapi pada regresi linier berganda sering terjadi korelasi yang tinggi antar peubah-peubah bebasnya yang dikenal dengan istilah multikolinieritas. Multikolinieritas berarti adanya hubungan linier yang sempurna atau pasti diantara beberapa atau semua peubah bebas dari model regresi berganda. Jika multikolinieritas terjadi, maka pendugaan menggunakan metode kuadrat terkecil akan menghasilkan penduga yang masih tetap tak bias dan konsisten, tetapi tidak efisien sehingga varian dari koefisien regresi menjadi tidak minimum.

## 2. REGRESI LINIER BERGANDA

### 2.1 Model Regresi Linier Berganda

Regresi linier berganda merupakan perluasan langsung dari regresi linier sederhana. Pada regresi linier sederhana hanya satu peubah bebas ( $X$ ) yang digunakan, sedangkan regresi linier berganda ada sebanyak  $k$  peubah bebas yang digunakan yaitu  $X_1, X_2, \dots, X_k$  (Paulson, 2007).

Persamaan regresi linier berganda yang mengandung  $k$  peubah bebas dan  $n$  pengamatan secara umum dapat dirumuskan dalam bentuk berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

Model Regresi Linier Berganda pada persamaan diatas dapat dituliskan dalam bentuk matriks. Dengan menggunakan lambang matriks model regresi linier berganda dapat ditulis sebagai model regresi linier umum, yaitu:

$$Y = X\beta + \varepsilon$$

### 2.2 Asumsi Regresi Linier Berganda

Asumsi yang digunakan pada model regresi linier berganda pada dasarnya sama dengan asumsi untuk model regresi linier sederhana. Manurung *et al.*, (2005) dalam Naftali mengatakan bahwa pada regresi ada sepuluh asumsi yang harus dipenuhi yang disebut dengan asumsi linier klasik. Adapun sepuluh asumsi tersebut yaitu:

1. Model regresi linier,
2. Nilai  $X$  tetap dalam sampel yang dilakukan berulang-ulang,
3.  $E(\varepsilon_i) = 0$ ,
4. Homokedastisitas atau  $\text{var}(\varepsilon_i)$  tetap untuk semua pengamatan,
5. Tidak ada Autokorelasi,
6.  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  untuk  $i \neq j$ ,

7.  $n > k + 1$  (jumlah pengamatan harus lebih besar dari jumlah koefisien yang diduga),
8. Nilai  $X$  bervariasi,
9. Spesifikasi model harus benar,
10. Tidak ada multikolinieritas.

### 2.3 Metode Kuadrat Terkecil Pada Regresi Linier Berganda

Metode yang paling umum digunakan untuk menduga koefisien regresi adalah metode kuadrat terkecil. Metode ini dipilih karena mudah digunakan dan dibawah asumsi-asumsi tertentu memiliki sifat-sifat yang dapat menghasilkan penduga yang baik.

Rencher & Schaalje (2008) mengatakan nilai penduga untuk koefisien  $\beta_0, \beta_1, \dots, \beta_k$  pada model regresi persamaan (1) atau (3) dapat diduga menggunakan metode kuadrat terkecil yaitu dengan cara meminimumkan jumlah kuadrat galat:

$$JKG = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_k X_{ik})^2$$

Peminimuman  $JKG$  dapat diperoleh dengan cara menurunkan  $JKG$  secara parsial terhadap koefisien  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  kemudian menyamakan hasil turunan tersebut dengan nol. Sehingga diperoleh persamaan normal sebagai berikut:

$$(X'X)\hat{\beta} = (X'Y)$$

Jika  $(X'X)$  tidak singular atau  $(X'X)$  merupakan matriks dengan  $\det(X'X) \neq 0$  sehingga matriks  $(X'X)$  mempunyai invers, maka persamaan normal mempunyai jawab yang tunggal yaitu:

$$\hat{\beta} = (X'X)^{-1} (X'Y)$$

Penduga kuadrat terkecil  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  bersifat *best linier unbiased estimator (BLUE)*. *Best* maksudnya adalah varian minimum dan *linier* menunjukkan bahwa penduga adalah fungsi linier dari  $Y$  (Rencher dan Schaalje, 2008).

### 2.4 Pengujian Hipotesis Koefisien Regresi Linier Berganda

Untuk melihat pengaruh variable bebas terhadap variable tak bebas baik secara keseluruhan maupun secara individual maka dilakukan pengujian koefisien regresi.

### 2.5 Koefisien Determinasi

Setelah menaksir persamaan regresi dari data, masalah berikutnya yang dihadapi ialah menilai baik buruknya kecocokan model regresi yang digunakan dengan data. Untuk menilai kecocokan model dengan data, diperlukan suatu ukuran yang dinamakan dengan koefisien determinasi atau koefisien penentu yang dilambangkan dengan  $R^2$ .

Koefisien determinasi biasanya dinyatakan dalam bentuk persen, karena koefisien determinasi merupakan persentase keragaman peubah bebas yang dapat dijelaskan oleh model persamaan regresi. Nilai  $R^2$  persamaan regresi yang makin mendekati 100% menunjukkan bahwa makin banyak keragaman peubah bebas yang

dapat dijelaskan dari persamaan regresi tersebut. Koefisien Determinasi memiliki dua sifat penting yaitu :

1. Koefisien determinasi ( $R^2$ ) merupakan besaran non negatif artinya selalu bernilai positif, batasnya adalah  $0 \leq R^2 \leq 1$ .
2. Jika  $R^2$  bernilai 1 berarti suatu kecocokan sempurna, sedangkan  $R^2$  bernilai 0 berarti tidak ada hubungan antara peubah tak bebas dengan peubah yang menjelaskannya.

Rumus yang digunakan untuk mencari nilai Koefisien Determinasi yaitu :

$$R^2 = \frac{JKR}{JKT}$$

### 3. KOEFISIEN KORELASI ANTAR PEUBAH BEBAS DALAM REGRESI LINIER BERGANDA

#### 3.1 Koefisien Korelasi

Korelasi merupakan suatu hubungan antar peubah. Korelasi ada yang positif dan negatif. Uji statistik yang mengukur hubungan antar peubah disebut analisis korelasi. Indeks yang mengukur hubungan antar peubah disebut koefisien korelasi.

Koefisien korelasi digunakan untuk mengukur hubungan keeratan antar dua peubah. Pada proses perhitungannya korelasi tidak menggunakan model meskipun hubungan yang diukur bersifat linier. Dengan kata lain besaran koefisien korelasi tidak menggambarkan hubungan sebab akibat antar dua peubah atau lebih tetapi hanya menggambarkan hubungan linier antar peubah.

Koefisien korelasi bivariat yang paling lama dan banyak digunakan yaitu korelasi yang dikembangkan oleh *Karl Pearson* yang disebut dengan koefisien korelasi momen hasil kali Pearson atau sering disingkat dengan koefisien korelasi Pearson. Perhitungan dalam korelasi ini didasarkan pada data sebenarnya (peubah asli).

Misalkan  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  merupakan pasangan data yang diperoleh dari dua peubah  $X$  dan  $Y$ . Koefisien korelasi antar  $X$  dan  $Y$  dinotasikan dengan  $r_{xy}$  dan dirumuskan sebagai berikut :

$$r = \frac{\text{cov}(X_i, Y_i)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

atau

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Jika penyebut dan pembilang dari persamaan diatas dibagi dengan  $n$ , maka  $r$  menjadi:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \left(\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2\right)}}$$

Jika antara kedua peubah bebas yang dihitung nilai korelasinya ( $X$  dan  $Y$ ) di bentuk kedalam model maka akan membentuk suatu model regresi linier sederhana. Kecocokan model regresinya dapat dilihat dari kuadrat koefisien korelasi antara  $X$  dan  $Y$  ( $r^2$ ). Pada regresi linier sederhana nilai  $r^2$  tersebut sama dengan nilai koefisien determinasinya ( $R^2$ ).

Akan tetapi hubungan  $r^2 = R^2$  tersebut hanya benar untuk dua peubah karena koefisien korelasi hanya mengukur hubungan linier antar dua peubah, sedangkan  $R^2$  berlaku untuk banyak peubah karena definisi  $R^2$  tidak tergantung pada  $X$ .  $R$  dapat dipandang sebagai perluasan dari korelasi sederhana  $r$  untuk banyak peubah.

Koefisien korelasi mempunyai standar error ( $S_r$ ) yang dapat dihitung menggunakan persamaan berikut:

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

Pada analisis korelasi tidak ada asumsi statistika yang diperlukan dalam menghitung koefisien korelasi, tetapi ada asumsi tentang uji hipotesis dan selang kepercayaan dalam koefisien korelasi. Hipotesis untuk koefisien korelasi dapat dilihat pada tabel berikut:

**Tabel 1. Hipotesis koefisien korelasi**

$H_0$	$H_1$	Daerah penolakan $H_0$
$r = 0$	$r \neq 0$	$ t  \geq t_{\alpha, n-2}$

Sedangkan untuk pengujian koefisien korelasi dapat digunakan uji-t. Statistik uji yang digunakan yaitu:

$$t = \frac{r}{S_r}$$

### 3.2 Multikolinieritas

Pada prakteknya jarang ditemukan adanya peubah bebas yang berhubungan secara sempurna atau data yang tidak memuat beberapa komponen galat. Akan tetapi jika terdapat beberapa atau semua peubah yang korelasi (koliner) satu sama lain merupakan suatu peristiwa yang umum dalam praktek regresi linier berganda. Koliner berarti terdapat beberapa peubah bebas ( $X_i$ ) yang berkorelasi satu sama lain, sehingga menghasilkan data dalam kondisi yang buruk (*ill-conditioned*).

Pada regresi linier berganda korelasi antar peubah bebas dikenal dengan istilah multikolinieritas yang pertama kali diperkenalkan oleh Ragnar Frisch pada tahun 1934, yang menyatakan bahwa multikolinieritas terjadi jika adanya hubungan linier yang sempurna (*perfect*) atau pasti (*exact*) diantara beberapa atau semua peubah bebas dari

model regresi berganda (Rahardiantoro, 2008). Jadi dalam regresi linier berganda multikolinieritas merupakan masalah yang selalu terjadi ketika dua atau lebih peubah bebas saling berkorelasi satu sama lain.

Kasus pada multikolinieritas yang sering terjadi ada dua yaitu multikolinieritas sempurna dan multikolinieritas tak sempurna. Multikolinieritas sempurna terjadi jika suatu peubah bebas bergantung sepenuhnya pada peubah bebas yang lain, sedangkan multikolinieritas tak sempurna dapat terjadi bila dua atau lebih peubah dalam model saling berkaitan.

Menurut Montgomery & Peck *dalam* Naftali (2007) adanya multikolinieritas dalam regresi linier berganda disebabkan oleh berbagai hal antara lain metode pengumpulan data yang digunakan, kendala model pada populasi yang diamati, spesifikasi model, dan penentuan jumlah peubah bebas yang lebih banyak dari jumlah pengamatan. Oleh karena itu, dalam suatu penelitian harus benar-benar diperhatikan metode, model, spesifikasi model dan jumlah peubah bebas yang digunakan.

### 3.2.1 Pengaruh Multikolinieritas

Sembiring (2003) menyatakan jika korelasi terjadi antara dua peubah atau lebih dalam suatu persamaan regresi maka pendugaan koefisien dari peubah yang bersangkutan tidak lagi tunggal, melainkan tidak terhingga banyaknya sehingga tidak mungkin lagi menduganya. Hal ini disebabkan  $(X'X)$  menjadi singular sehingga matriks  $(X'X)^{-1}$  tidak dapat dihitung karena satu atau lebih kolom merupakan kombinasi linier dari kolom-kolom lainnya sehingga  $|X'X| = 0$ . Akibatnya nilai dugaan yang diperoleh tidak unik atau tunggal lagi (Draper & Smith, 1992).

Pendugaan yang umum suatu koefisien regresi berfungsi untuk mengukur perubahan nilai harapan dari peubah tak bebas jika diberikan peubah bebas yang meningkat oleh satu unit sementara semua peubah bebas yang lain dianggap tetap. Akan tetapi hal ini tidak sepenuhnya bisa diterapkan jika terjadi multikolinieritas. Demikian juga pendugaan menggunakan metode kuadrat terkecil akan menghasilkan penduga yang masih tak bias dan konsisten tetapi tidak efisien lagi karena varian yang dihasilkan menjadi lebih besar atau tidak minimum. Varian yang seperti itu tidak sesuai dengan sifat penduga kuadrat terkecil, sehingga pendugaan dengan metode kuadrat terkecil untuk regresi linier berganda ini kurang efisien dan kurang tepat lagi.

Multikolinieritas juga menyebabkan koefisien regresi untuk suatu peubah bergantung pada nilai peubah bebas yang lain. Koefisien regresi yang akan diduga cenderung mempunyai variabilitas sampel yang besar dan bertukar-tukar secara luas dari satu sampel ke sampel yang berikutnya. Akibatnya hasil yang diperoleh merupakan informasi yang tidak tepat mengenai koefisien regresi yang sebenarnya sehingga koefisien regresi tidak mencerminkan pengaruh apapun dari peubah bebas terhadap peubah tak bebas tetapi hanya pengaruh secara parsial (sebagian).

Montgomery & Hines (1990) *dalam* Rahardiantoro (2008) menjelaskan bahwa dampak multikolinieritas dapat mengakibatkan koefisien regresi yang dihasilkan oleh analisis regresi berganda menjadi sangat lemah atau tidak dapat memberikan hasil analisis yang mewakili sifat atau pengaruh dari peubah bebas yang bersangkutan. Masalah multikolinieritas dapat menyebabkan uji-t menjadi tidak signifikan padahal jika masing-masing peubah bebas diregresikan secara terpisah dengan peubah tak bebas (*simple regression*) uji-t menunjukkan hasil yang signifikan.

### 3.2.2 Mendeteksi Multikolinier

Ada beberapa pendekatan umum yang dapat digunakan untuk mendeteksi dan mengevaluasi kolinier antara peubah bebas baik secara informal maupun formal. Nachrowi dan Usman (2006) dalam Naftali (2007) menyatakan multikolinieritas dapat dideteksi dengan adanya koefisien determinasi ( $R^2$ ) yang tinggi dan uji-F yang signifikan. Cara ini merupakan pendeteksian adanya multikolinieritas secara informal, sedangkan secara formal, dapat dilihat dari matriks korelasi antar peubah bebas, nilai *Variance Inflation Factor (VIF)*, dan nilai eigen matriks ( $X'X$ ).

Cara mendeteksi adanya multikolinier yang serius dapat dilakukan dengan cara informal berikut:

1. Perubahan besar pada koefisien regresi yang diduga jika peubah bebas ditambah atau dihapus.
2. Tidak ada hasil yang signifikan pada pengujian koefisien regresi secara individu untuk peubah bebas yang penting.
3. Koefisien regresi yang di duga dengan tanda secara aljabar yang adalah kebalikan dari nilai yang diharapkan dari pertimbangan teoritis atau pengalaman sebelumnya.
4. Koefisien yang besar dari korelasi sederhana antar pasangan peubah bebas pada matriks korelasi.
5. Interval kepercayaan yang luas untuk koefisien regresi yang mewakili peubah bebas yang penting.

Metode informal hanya yang dideskripsikan mempunyai pembatasan penting dan tidak memberikan pengukuran yang kualitatif dari dampak multikolinier dan tidak akan mengidentifikasi sifat alami dari multikolinier. Sebagai contoh, jika peubah bebas  $X_1, X_2$  dan  $X_3$  mempunyai korelasi berpasangan yang rendah, maka pengujian untuk koefisien korelasi sederhana tidak akan memperlihatkan keberadaan dari hubungan antar kelompok peubah bebas sehingga korelasi yang tinggi antara  $X_1$  dan kombinasi linier dari  $X_2$  dan  $X_3$ . Pembatasan yang lain dari metode informal yaitu kadang-kadang perilaku yang diamati boleh terjadi tanpa adanya multikolinier.

Suatu cara yang lebih formal dan sering digunakan untuk mendeteksi korelasi antar peubah bebas yaitu dengan cara menghitung nilai *variance inflation factor (VIF)*. *variance inflation factor* yang mengukur berapa banyak varian dari penduga koefisien regresi yang dibandingkan jika peubah bebas tidak berhubungan secara linier. Nilai VIF dapat dihitung menggunakan persamaan berikut:

$$VIF = \frac{1}{1 - R^2}$$

di mana  $R^2$  adalah koefisien determinasi. Jika nilai VIF suatu peubah melebihi 10 dan nilai  $R^2$  melebihi 0,90 maka suatu peubah dikatakan berkorelasi sangat tinggi dan dapat dipastikan adanya multikolinieritas terjadi.

Cara formal lain yang digunakan untuk mendeteksi dan mengevaluasi multikolinieritas yaitu dengan cara menghitung nilai eigen ( $\lambda$ ) dari matriks ( $X'X$ ). Semakin kecil nilai eigen yang diperoleh maka makin tinggi korelasi yang terjadi antar peubah bebas (Sembiring, 2003). Jumlah dari nilai eigen yang diperoleh akan sama dengan banyaknya nilai eigen. Nilai eigen ( $\lambda$ ) diperoleh dengan cara menghitung:

$$|A - \lambda I| = 0$$

Keterangan:

$A$  : Matriks kuadrat ( $X'X$ )

$I$  : Matriks identitas

$\lambda$  : Nilai eigen

Nilai eigen memiliki beberapa metode penting untuk mendeteksi multikolinieritas, yaitu:

### 1. Index Kondisi (*Condition Index*)

CI dapat dihitung untuk masing-masing nilai eigen ( $\lambda$ ). Pertama, nilai eigen diurutkan dari nilai yang besar ke nilai yang kecil, di mana  $\lambda_{maks}$  = nilai eigen yang paling besar dan  $\lambda_j$  = nilai eigen yang paling kecil. CI diperoleh dengan cara membagi nilai eigen yang terbesar ( $\lambda_{maks}$ ) dengan masing-masing nilai eigen yang lain, sehingga diperoleh:

$$CI = \frac{\lambda_{maks}}{\lambda_j}, j = 1, 2, \dots, k$$

### 2. Bilangan Kondisi (*Condition Number*)

Bilangan kondisi (CN) merupakan rasio antara nilai eigen yang terbesar dan nilai eigen yang terkecil, sehingga diperoleh:

$$CN = \frac{\lambda_{maks}}{\lambda_{min}}$$

Jika  $CN < 100$  berarti tidak ada multikolinier, jika  $100 \leq CN \leq 1000$  maka terjadi kolinier yang wajar, sedangkan jika  $CN > 1000$  maka terjadi kolinier yang berarti. Belsley *et al.* (1980) dalam Paulson (2007) mengatakan bahwa jika  $\sqrt{CN} > 30$  maka terjadi kolinier yang berarti dan mengarah ke kolinier yang serius.

### 3. Proporsi Varian

Cara lain yang bermanfaat yaitu menggunakan proporsi varian untuk masing-masing peubah bebas ( $X_i$ ) dimana proporsi total varian dari penduga untuk komponen utama tertentu. Nilai eigen merupakan varian dari komponen utama. Analisis komponen utama hanya suatu nilai peubah bebas yang baru merupakan kombinasi linier dari peubah bebas yang asli. Nilai eigen yang mendekati nol menunjukkan adanya kolinier. Jika nilai eigen meningkat maka nilai CI akan meningkat. Proporsi varian merupakan jumlah total variabilitas yang dijelaskan oleh nilai eigen komponen utama dari peubah bebas. Baris yang kumulatif adalah kontribusi beberapa atau semua nilai eigen untuk memasukkan peubah bebas yang spesifik. Jumlah dari semua proporsi varian akan sama dengan 1 dan jumlah dari nilai eigen merupakan banyaknya nilai eigen.

### 4. TELADAN PENERAPAN

Teladan penerapan ini bertujuan agar penjelasan tentang hubungan antara korelasi antar peubah bebas yang terjadi dalam regresi linier berganda terhadap koefisien determinasinya dapat dipahami dengan mudah. Dalam hal ini, data yang digunakan adalah data yang dibentuk dengan metode *trial & error* menggunakan program Microsoft Excel.

Data yang dibentuk adalah untuk model regresi linier berganda yang terdiri dari dua peubah bebas ( $X_1$  dan  $X_2$ ) dan peubah tak bebas ( $Y$ ). Berdasarkan data tersebut akan diamati korelasi untuk pasangan peubah-peubah yang digunakan. Pasangan peubah yang akan diamati tersebut yaitu  $(X_1, X_2), (X_1, Y), (X_2, Y)$ .

Dari pasangan peubah tersebut terdapat beberapa kemungkinan bentuk data yang mungkin terjadi. Kemungkinan bentuk tersebut dapat dilihat pada tabel berikut:

**Tabel 2. Bentuk Kemungkinan Data**

Kemungkinan	$(X_1, X_2)$	$(X_1, Y)$	$(X_2, Y)$
1	✘	✘	✘
2	✘	✘	✓
3	✘	✓	✘
4	✘	✓	✓
5	✓	✓	✓
6	✓	✓	✘
7	✓	✘	✓
8	✓	✘	✘

Keterangan:

- $X_i$  : Peubah bebas ke- $i$ ,  $i = 1, 2$
- $Y$  : Peubah yang dianggap sebagai peubah tak bebas
- ✘ : Menyatakan bahwa tidak ada korelasi antar peubah
- ✓ : Menyatakan bahwa ada korelasi antar peubah

Berdasarkan kemungkinan bentuk data tersebut akan diketahui bagaimana pengaruh korelasi yang terjadi antar peubah bebas terhadap koefisien determinasinya

#### 4.1 Perhitungan dan Uji Korelasi terhadap Pasangan Peubah

Seperti penjelasan diatas, sebelum melakukan perhitungan koefisien korelasi terhadap pasangan peubah yang digunakan terlebih dahulu dibentuk data dengan metode *trial & error* menggunakan program Microsoft Excel. Data yang dibentuk adalah untuk model regresi linier berganda yang terdiri dari dua peubah bebas ( $X_1$  dan  $X_2$ ) dan satu peubah tak bebas ( $Y$ ). Secara umum, model regresi berganda untuk kasus diatas dapat ditulis sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, 20$$

Langkah-langkah yang dilakukan dalam membentuk data tersebut adalah sebagai berikut:

- Bentuk terlebih dahulu data  $X_1$  dan  $X_2$ , kemudian bentuk data tersebut menjadi matriks.
- Lakukan pengujian terhadap  $X_1$  dan  $X_2$  menggunakan uji-t
- Buat hipotesis sebagai berikut:

$$H_0 : r(X_1, X_2) = 0$$

$$H_1 : r(X_1, X_2) \neq 0$$

- Hitung korelasi antar  $X_1$  dan  $X_2$  menggunakan rumus:

$$r(X_i, Y_i) = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\left[ \left( n \sum_{i=1}^n X_i^2 - \sum_{i=1}^n (X_i)^2 \right) \left( n \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n (Y_i)^2 \right) \right]^{\frac{1}{2}}}$$

Atau menggunakan fungsi yang ada pada program microsoft excel yaitu fungsi CORREL yang dapat digunakan untuk menghitung  $r(X_i, Y_i)$ .

- Hitung standar error dari koefisien korelasi tersebut menggunakan rumus:

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

- Hitung nilai  $t$  hitung menggunakan rumus:

$$t_{hitung} = \frac{r}{S_r}$$

- Bandingkan  $t_{hitung}$  dengan  $t_{tabel}$ , dimana  $t_{tabel} = t_{\alpha; n-2}$ , dengan kriteria pengujian yaitu Tolak  $H_0$  jika  $|t_{hitung}| \geq t_{tabel}$
- Tentukan kesimpulan yang diperoleh dari hasil perbandingan antara  $t_{hitung}$  dengan  $t_{tabel}$ .

Setelah data  $X_1$  dan  $X_2$  telah dibentuk dan memenuhi kondisi yang diinginkan, langkah selanjutnya yang harus dilakukan yaitu membentuk data  $Y$ . Data  $Y$  yang dibentuk tersebut harus dilakukan pengujian menggunakan uji-t terlebih dahulu agar terpenuhi kondisi yang diinginkan. Langkah yang dilakukan sama dengan langkah yang dilakukan untuk peubah  $X_1$  dan  $X_2$ .

Berikut ini salah satu contoh dari simulasi data menggunakan program Microsoft Excel yang telah memenuhi kondisi yang ditentukan:

**Tabel 3. Kemungkinan Data yang dihasilkan**

n	$X_0$	$X_1$	$X_2$	Y
1	1	31	-18	-13
2	1	-1	-17	15
3	1	-1	3	-2
4	1	5	-7	-7
5	1	11	-5	0
6	1	-15	-15	6
7	1	7	-6	14
8	1	9	10	-7
9	1	10	-3	-8
10	1	11	-6	-9
11	1	6	5	2
12	1	0	9	-3
13	1	2	3	13
14	1	11	5	9
15	1	-6	-3	-11
16	1	13	3	3
17	1	1	4	14
18	1	-10	-7	-8
19	1	-3	-3	-14
20	1	-6	-7	18

Setelah setiap bentuk kemungkinan data berhasil di bentuk, langkah selanjutnya adalah menghitung nilai koefisien determinasi ( $R^2$ ) model. Langkah-langkah yang harus dilakukan adalah:

- Namakan matriks yang memuat semua peubah dengan matriks A
- Lakukan pemfaktoran QR terhadap matriks A, sehingga di dapat matriks Q dan matriks R.
- Hitung Jumlah kuadrat Regresi ( $JKR$ ), Jumlah Kuadrat Galat ( $JKG$ ) dan Jumlah Kuadrat Total ( $JKT$ ) berdasarkan matriks R.
- Hitung nilai koefisien determinasi ( $R^2$ ) model regresi dengan rumus:

$$R^2 = \frac{JKR}{JKT}$$

Berikut salah satu contoh hasil perhitungan nilai korelasi antar peubah berdasarkan data yang telah dibentuk:

**Tabel 4. Hasil Perhitungan**

Pasangan Peubah	(X1,X2)	(X1,Y)	(X2,Y)
Nilai Korelasi ( r )	0.01448	-0.23276	0.01641
Kesalahan dari r (Sr)	0.23568	0.22923	0.23567
t-hitung	0.06146	-1.01539	0.06962
t-tabel	2.10092		
Kesimpulan	Ho diterima	Ho diterima	Ho diterima
$R^2$ Model regresi	0.0579		

#### 4.2 Pembahasan

Data yang diamati pada penelitian ini yaitu data untuk model regresi linier berganda yang terdiri dari dua peubah bebas ( $X_1$  dan  $X_2$ ) dan peubah tak bebas ( $Y$ ). Dari peubah tersebut akan dibentuk pasangan antar peubah bebas dan tak bebas maupun sesama peubah bebas sehingga terbentuk pasangan  $(X_i, X_j)$  dan  $(X_i, Y)$ .

Untuk model regresi dengan dua peubah bebas, pasangan peubah yang akan diamati tersebut yaitu  $(X_1, X_2), (X_1, Y), (X_2, Y)$  dan banyaknya pasangan data yang mungkin terjadi yaitu sebanyak  $2^3$  atau delapan kemungkinan.

Secara umum banyak pasangan data yang akan diamati dapat diperoleh dari persamaan  $C_2^m + m$ , dimana  $C$  menyatakan kombinasi dan  $m$  menyatakan banyaknya peubah bebas yang digunakan. Sedangkan untuk banyaknya kemungkinan data yang akan diamati diperoleh dari persamaan  $2^{C_2^m + m}$ . Misalkan  $Q = C_2^m + m$ , sehingga diperoleh banyaknya kemungkinan data yang mungkin terjadi yaitu  $2^Q$ .

Berdasarkan pada pasangan data dan kemungkinan bentuk data tersebut akan diamati korelasi yang terjadi antar peubah. Korelasi yang terjadi antar peubah dipengaruhi oleh faktor langsung dan faktor tak langsung. Faktor langsung yaitu dipengaruhi langsung oleh peubah yang saling berkorelasi sedangkan faktor tidak langsung yaitu faktor yang berasal dari peubah lain.

Hasil korelasi yang diamati tersebut kemudian dibandingkan dengan nilai koefisien determinasi ( $R^2$ ) model regresi yang digunakan. Pada model regresi nilai  $R^2$  tersebut digunakan untuk menilai kecocokan data dengan model regresi. Semakin tinggi nilai  $R^2$  maka kecocokan data dengan model yang digunakan semakin baik, dan sebaliknya.

Berdasarkan hasil pengamatan yang dilakukan dapat dilihat bahwa korelasi yang terjadi antar peubah akan berpengaruh nilai  $R^2$  yang dihasilkan. Demikian juga jika terjadi korelasi antar peubah bebas akan mempengaruhi nilai korelasi masing-masing peubah bebas terhadap peubah tak bebas.

Dari hasil pengamatan yang dilakukan diketahui bahwa persamaan regresi yang baik jika antar peubah bebas tidak berkorelasi satu sama lain, sedangkan antara peubah bebas dan tak bebas berkorelasi tinggi.

#### KESIMPULAN

Berdasarkan hasil simulasi data dan pembahasan sebelumnya, maka dapat diambil kesimpulan sebagai berikut :

1. Korelasi yang terjadi antar peubah bebas dalam regresi linier berganda akan mempengaruhi nilai  $R^2$ .
2. Model regresi yang baik dihasilkan dari kemungkinan data yang antar peubah bebasnya tidak berkorelasi.

## DAFTAR PUSTAKA

- Christensen, R. 1996. *Analysis of Variance, Design and Regression Applied Statistical Method*. Chapman & Hall. London.
- Draper, N. R. dan H. Smith. 1992. *Analisis Regresi Terapan, edisi kedua (terjemahan)*. PT. Gramedia Pustaka Utama. Jakarta.
- Montgomery, D.C. 1976. *Design and Analysis of Experiments*. Jhon Wiley & Sons. New York.
- Naftali, Y. 2007. *Regresi dan Multikolinieritas*.  
<http://yohanli.wordpress.com/2007/12/18/multikolinieritas-dalam-regresi/>
- Kutner, *et al.*, 2005. *Applied Linier Statistical Models*. McGraw-Hill. New York.
- Paulson, D. S. 2007. *Handbook of Regression and Modeling Aplification for the Clinical and Pharmaceutical Industries*. Chapman & Hall. London.
- Rahardiantoro, D. 2008. *Principal Component Analysis (PCA) sebagai Metode Jitu untuk Mengatasi Masalah Multikolinieritas*.  
<http://dickyRahardiantoro.blogspot.com/2006/12/Principal-Component-Analysis-pca.htm>
- Rencher, A. C. dan G. B. Schaalje. 2007. *Linier Model in Statistics*. John Willey & Sons. London.
- Sembiring, R. K. 2003. *Analisis Regresi*. ITB Bandung. Bandung.
- Walpole, R.E dan R.H. Myers 1995. *Ilmu Peluang dan Statistik untuk Insinyur dan Ilmuwan, Terjemahan R.K Sembiring, Edisi Keempat*. ITB. Bandung.
- Weisberg, S. 2007. *Applied Linear Regression*. John Wiley & Sons. New York.