

# Regresi dengan Variabel Dummy

Junaidi

## 1. Pendahuluan

Dalam regresi, variabel dependent pada dasarnya tidak hanya dapat dipengaruhi oleh variabel independent kuantitatif, tetapi juga dimungkinkan oleh variabel kualitatif.. Lalu bagaimana cara memasukkan variabel independent kualitatif tersebut (yang tidak berbentuk angka) ke dalam model regresi ?

Variabel kualitatif tersebut harus dikuantitatifkan atributnya (cirinya). Untuk mengkuantitatifkan atribut variabel kualitatif, dibentuk variabel dummy dgn nilai 1 dan 0. Jadi, inilah yang dimaksud dengan variabel dummy tersebut. Nilai 1 menunjukkan adanya, sedangkan nilai 0 menunjukkan tidak adanya ciri kualitas tsb. Misalnya variabel jenis kelamin. Jika nilai 1 digunakan untuk laki-laki maka nilai 0 menunjukkan bukan laki-laki (perempuan), atau sebaliknya. (Kategori yg diberi nilai 0 disebut kategori dasar, dlm artian bahwa perbandingan dibuat atas kategori tsb.)

Jika variabel kualitatif tsb terdiri lebih dari dua kategori, jumlah variabel dummy yg dibentuk harus sebanyak  $n-1$ , dimana  $n$  adalah banyaknya kategori variabel tsb. Misalnya variabel jenis pekerjaan dgn tiga kategori yaitu pekerja kasar, setengah terampil dan pekerja terampil, maka dibentuk dua variabel dummy sbb:

Alternatif 1

$Dk1 = 1$  jika kasar,  $= 0$  jika lainnya

$Dk2 = 1$  jika setengah terampil,  $= 0$  jika lainnya

Pada alternatif 1,

$Dk1 = 0$ ,  $Dk2 = 0$  jika terampil

$Dk1 = 0$ ,  $Dk2 = 1$  jika  $\frac{1}{2}$  terampil

$Dk1 = 1$ ,  $Dk2 = 0$  jika kasar

Alternatif 2

$Dk1 = 1$  jika setengah terampil,  $= 0$  jika lainnya

Dk2 = 1 jika terampil, = 0 jika lainnya

Pada alternatif 2,

Dk1 = 0, Dk2 = 0 jika kasar

Dk1 = 0, Dk2 = 1 jika terampil

Dk1 = 1, Dk2 = 0 jika ½ terampil

## 2. Data Latihan

Untuk dasar latihan regresi dengan variabel dummy, diberikan data latihan sebagai berikut:

Jenis Kelamin (Ds)	Pendidikan (Pd)	Jenis Pekerjaan	Dk1	Dk2	Penghasilan (Y)
0	10	pekerja kasar	0	0	374938
0	15	pekerja kasar	0	0	474938
0	12	pekerja kasar	0	0	465375
0	12	pekerja kasar	0	0	462188
0	15	pekerja kasar	0	0	592875
0	12	pekerja kasar	0	0	455813
0	8	pekerja kasar	0	0	465375
0	15	pekerja kasar	0	0	449438
0	16	pekerja kasar	0	0	854250
1	8	pekerja kasar	0	0	621563
1	15	pekerja kasar	0	0	659813
1	15	pekerja kasar	0	0	745875
1	15	pekerja kasar	0	0	691688
1	15	pekerja kasar	0	0	745875
1	15	pekerja kasar	0	0	682125
1	12	pekerja kasar	0	0	867000
1	12	pekerja kasar	0	0	898875
1	15	pekerja kasar	0	0	739500
0	12	1/2 terampil	1	0	637500
0	11	1/2 terampil	1	0	631438
0	12	1/2 terampil	1	0	663438
0	12	1/2 terampil	1	0	653438
0	12	1/2 terampil	1	0	653438
1	12	1/2 terampil	1	0	673438
1	10	1/2 terampil	1	0	641438
1	17	1/2 terampil	1	0	2104688
1	13	1/2 terampil	1	0	673438

1	17	1/2 terampil	1	0	2214688
1	19	1/2 terampil	1	0	2868750
0	13	pekerja terampil	0	1	796125
0	11	pekerja terampil	0	1	781725
0	11	pekerja terampil	0	1	784125
0	16	pekerja terampil	0	1	1465500
1	16	pekerja terampil	0	1	1506000
1	10	pekerja terampil	0	1	772125
1	15	pekerja terampil	0	1	1453500
1	16	pekerja terampil	0	1	1530000
1	10	pekerja terampil	0	1	798450
1	16	pekerja terampil	0	1	2645625
1	18	pekerja terampil	0	1	1880625

### 3. Regresi Atas Satu Variabel Kualitatif (Dua Kategori)

Dari data yang diberikan, misalnya kita ingin meregresikan variabel kualitatif jenis kelamin (sex) terhadap penghasilan, dengan kuantifikasi laki-laki =1 dan perempuan =0

Model regresi yang kita bentuk :

$$Y = b_0 + b_1 D_s$$

Dimana : Y = penghasilan

$D_s$  = variabel dummy untuk seks

Berdasarkan pengolahan data, persamaan regresi yang terbentuk sebagai berikut:

$$\begin{array}{lll}
 Y = & 647884 + & 552801 D_s & R^2 = 0,205 \\
 & (130773) & (176335) & df = 38 \\
 t = & (4,95)** & (3,14)** & F = 9,82**
 \end{array}$$

Model ini memungkinkan untuk mengetahui perbedaan penghasilan berdasarkan jenis kelamin. Interpretasinya :  $b_0$  = rata-rata penghasilan pekerja perempuan,  $b_1$  = besarnya perbedaan penghasilan pekerja laki-laki dengan perempuan, dan  $b_0 + b_1$  merupakan rata-rata penghasilan pekerja laki-laki. Karena variabel  $D_s$  signifikan, artinya secara statistik terdapat perbedaan nyata penghasilan laki-laki dengan perempuan

Jika  $D_s=0$ , artinya rata-rata penghasilan pekerja perempuan adalah 647884 (sebesar  $b_0$ ).  
 Jika  $D_s=1$ , artinya rata-rata penghasilan pekerja laki-laki adalah  $647884 + 552801 = 1200685$  (sebesar  $b_0 + b_1$ )

#### 4. Regresi Atas Satu Variabel Kuantitatif dan Satu Variabel Kualitatif (dua kategori)

Dari data yang sudah diberikan, misalnya kita bentuk suatu model regresi sbb:

$$Y = b_0 + b_1 Ds + b_2 Pd + e_i$$

Dimana :  $Ds$  = dummy jenis kelamin

$Pd$  = pendidikan

Berdasarkan pengolahan datanya, ditampilkan hasilnya sebagai berikut :

$Y = -1039341 + 331928 Ds + 134978 Pd$	$R^2 = 0,52$
(378566) (152620) (27847)	$df = 37$
$t = (-2,90)** (2,27)* (4,91)**$	$F = 19,94**$

Interpretasi : Pada setiap tingkat pendidikan, terdapat perbedaan besarnya penghasilan laki-laki dan perempuan. Perbedaan tsb besar-nya adalah 331928 (sebesar  $b_1$ ) lebih tinggi untuk laki-laki

Misalnya jika tahun pendidikan ( $Pd$ ) = 15, maka

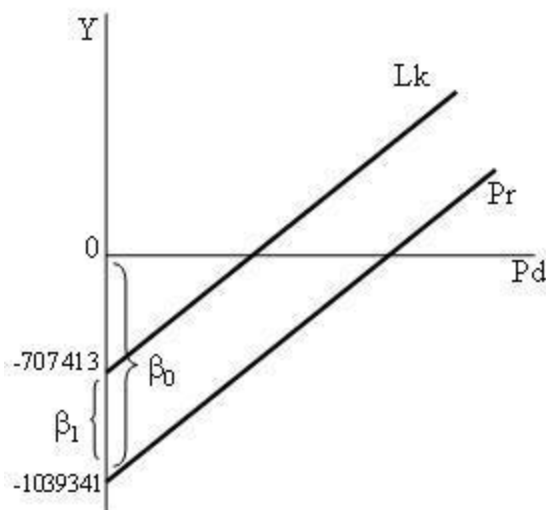
Jika  $Ds=0$ , rata-rata penghasilan perempuan adalah

$$Y = -1039341 + 331928 (0) + 134978 (15) = 985329$$

Jika  $Ds=1$ , rata-rata penghasilan laki-laki adalah

$$Y = -1039341 + 331928 (1) + 134978 (15) = 1317257$$

Secara grafik, hal ini dapat digambarkan sebagai berikut:



Ini memberikan arti fungsi penghasilan laki-laki dan perempuan sehubungan pendidikannya memiliki slope yg sama, tetapi intersep berbeda.

Dengan kata lain, rata-rata penghasilan laki-laki dan perempuan berbeda, tetapi tingkat perubahan penghasilan yg disebabkan oleh tahun pendidikan adalah sama.

#### 5. Regresi Atas Satu Variabel Kuantitatif dan Satu Variabel Kualitatif (lebih dua kategori)

Dari data yang diberikan, misalnya kita bentuk suatu model regresi sbb:

$$Y = b_0 + b_1 Dk1 + b_2 Dk2 + b_3 Pd + e_i$$

Dimana : Dk1 = dummy jenis pekerjaan

1 = setengah terampil

0 = lainnya

Dk2 = dummy jenis pekerjaan

1 = terampil

0 = lainnya

Pd = pendidikan

Berdasarkan pengolahan datanya, ditampilkan hasilnya sebagai berikut :

Y =	-1286520 + 475244Dk1 + 590905Dk2 + 145168Pd	R <sup>2</sup> = 0,62
	(314191) (22870) (145658) (146349)	df = 36
t =	(-4,09)** (6,35)** (32,26)** (4,04)**	F = 21,91**

Interpretasi :

Jika Dk1 = 0, Dk2 = 0 ® pekerja kasar

Jika Dk1 = 1, Dk2 = 0 ® pekerja setengah terampil

Jika Dk1 = 0, Dk2 = 1 ® pekerja terampil

Pada setiap tingkat pendidikan, terdapat perbedaan penghasilan pekerja kasar, setengah terampil dan terampil. Pekerja setengah terampil memiliki penghasilan 475244 (sebesar  $b_1$ ) lebih tinggi dibandingkan pekerja kasar. Pekerja terampil memiliki penghasilan 590905 (sebesar  $b_2$ ) lebih tinggi dibandingkan pekerja kasar.

Misalnya jika tahun pendidikan (Pd) = 15, maka

Jika Dk1=0 dan Dk2=0, rata-rata penghasilan pekerja kasar adalah:

$$Y = -1286520 + 475244 (0) + 590905 (0) + 145168 (15) = 891000$$

Jika Dk1=1 dan Dk2=0, rata-rata penghasilan pekerja ½ terampil :

$$Y = -1286520 + 475244 (1) + 590905 (0) + 145168 (15) = 1366244$$

Jika Dk1=0 dan Dk2=1, rata-rata penghasilan pekerja terampil:

$$Y = -1286520 + 475244 (1) + 590905 (1) + 145168 (15) = 1481905$$

## 6. Membandingkan Dua Regresi dengan Variabel Dummy

Secara implisit, teknik penggunaan variabel dummy (dari pembahasan sebelum ini), pada dasarnya mengandung asumsi bahwa variabel kualitatif mempengaruhi intersep tetapi tidak

mempengaruhi koefisien kemiringan dari berbagai regresi sub-kelompok. Kita dapat menghindari asumsi ini dengan menggunakan teknik “pooling” (mengelompokkan), dengan dasar pemikiran sbb: (catatan: banyak teknik yang lain, tetapi ini salah satunya)

Misalnya, dari model regresi terdahulu:

$$Y = b_0 + b_1 Ds + b_2 Pd + e_i$$

dimana :  $Ds$  = dummy jenis kelamin,  $Pd$  = pendidikan

Kita dapat merubah model ini dengan meregresi secara terpisah antara pekerja perempuan dan laki-laki sebagai berikut:

Regresi untuk pekerja perempuan

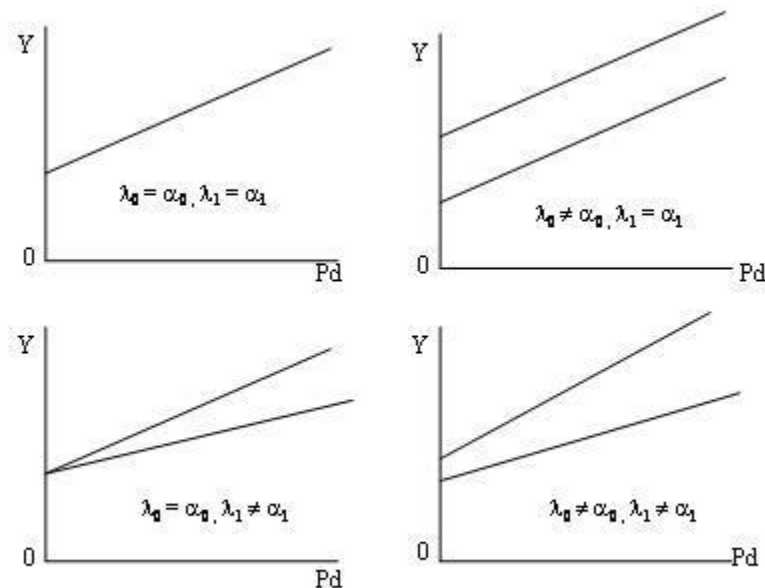
$$Y = l_0 + l_1 Pd + e_i$$

Regresi untuk pekerja laki-laki

$$Y = a_0 + a_1 Pd + e_i$$

(Catatan: banyaknya sampel untuk laki-laki dan perempuan tidak perlu sama)

Dua persamaan regresi tersebut, memberikan empat kemungkinan hasil sebagai berikut:



Semua kemungkinan tsb dapat diuji jika kita mengelompokkan (pool) semua sampel laki-laki dan perempuan, dengan memodifikasi model menjadi :

$$Y = b_0 + b_1 Ds + b_2 Pd + b_3 (Ds.Pd) + e_i$$

Dibandingkan model sebelumnya, model ini mempunyai variabel tambahan  $DsPd$

Implikasi model ini sebagai berikut:

Jika  $D_s=0$  (perempuan), maka model menjadi :

$$Y = b_0 + b_2 Pd + e_i, \text{ setara dgn } Y = l_0 + l_1 Pd + e_i$$

Jika  $D_s=1$  (laki-laki), maka model menjadi :

$$Y = (b_0 + b_1) + (b_2 + b_3)Pd + e_i, \text{ setara dgn } Y = a_0 + a_1 Pd + e_i$$

Dari data yg telah diberikan, kita meregresikan secara terpisah untuk pekerja perempuan dan laki-laki didapatkan :

$$\text{Regresi pekerja perempuan : } Y = 12971 + 50793 Pd$$

$$\text{Regresi pekerja laki-laki : } Y = -1242495 + 172830 Pd$$

Dengan teknik “pooling” kita modifikasi model tersebut, dan melalui perhitungan, persamaan regresinya menjadi :

$$Y = 12971 -1255466 D_s + 50793 Pd + 122036 D_s Pd$$

Jika  $D_s=0$  (perempuan), persamaan regresi tsb menjadi:

$$Y = 12971 -1255466 (0) + 50793 Pd + 122036 (0)Pd$$

$$Y = 12971 + 50793 Pd$$

Jika  $D_s=1$  (laki-laki), persamaan regresi tsb menjadi :

$$Y = 12971 -1255466 (1) + 50793 Pd + 122036 (1)Pd$$

$$Y = 12971 -1255466 + 50793 Pd + 122036 (1)Pd$$

$$Y = -1242495 + 172830 Pd$$

## REFERENCES

1. Amri A., Junaidi, Yulmardi. (2009). *Metodologi Penelitian Ekonomi dan Penerapannya*. Bogor. IPB Press
2. Gujarati DN. (2013). *Basic Econometrics*. New York. McGraw-Hill
3. Junaidi, J. (2014). *Regresi dengan Microsoft Office Excel*. Jambi. Fakultas Ekonomi dan Bisnis Universitas Jambi
4. Juanda, B. Junaidi. (2012). *Ekonometrika Deret Waktu: Teori dan Aplikasi*. Bogor. IPB Press
5. Junaidi, J. (2014). *Analisis Hubungan Deret Waktu untuk Peramalan*. Jambi. Fakultas Ekonomi dan Bisnis Universitas Jambi