

PENERAPAN ANALISIS REGRESI *DUMMY* PADA DATA KUALITATIF KASUS EKONOMI

Roza Zelvia

IAIN Metro, Lampung

E-mail: zelviam_roza@yahoo.co.id

Alva Yenica Nandavita

IAIN Metro, Lampung

E-mail: zelviam_roza@yahoo.co.id

Hotman

IAIN Metro, Lampung

alvananda [E-mail: hotman@yahoo.co.id](mailto:hotman@yahoo.co.id)

Abstract

Regression analysis is basically the study of the dependence of one variable dependent (response/non-free) with one or more variables independent (explanatory / free), in order to estimate and / or predict the average population or the average value of the variable dependent based on the value independent variables that are known. In some economy cases, we usually would like to look at the relationship between variables, such as the relationship between the profit of the company with the types of company, income customers who finance with the amount of funding requested and the salary earned by the employees with education and work experience of the employee itself. However, dealing with the variables of economy cases, there are many variables that are ordinal or nominal scale called qualitative data. Therefore, it can not only be solved by using a simple regression analysis or multiple regressions. Moreover, Dummy regression analysis is one of the solutions in dealing with cases of economic data with qualitative independent variables. This paper will investigate the economic case data which have the categorical independent variables by using dummy regression analysis. The data used in this literature study is secondary data obtained from the data already

published to the Internet. Furthermore, based on data analysis, it was found that, in the dummy regression analysis, it was allowed more than one Independent variables, where the variables can interact or have no interaction each other. Those can be drawn by a scatterplot of the data.

Keywords: *dummy regression, qualitative data, regression analysis.*

Abstrak

Analisis regresi pada dasarnya adalah studi mengenai ketergantungan satu peubah dependen (respon/tak bebas) dengan satu atau lebih peubah independen (penjelas/bebas), dengan tujuan untuk mengestimasi dan/atau memprediksi rata-rata populasi atau nilai rata-rata peubah dependen berdasarkan nilai peubah independen yang diketahui. Dalam beberapa kasus ekonomi, sering kali kita ingin melihat hubungan antar peubah, seperti hubungan antara laba perusahaan dengan tipe perusahaan, gaji yang diperoleh karyawan dengan jenjang pendidikan dan pengalaman kerja karyawan tersebut. Namun, pada peubah-peubah dalam kasus ekonomi tersebut terdapat banyak peubah yang berskala ordinal atau nominal yang disebut juga data kualitatif. Sehingga dalam penyelesaian kasus tersebut tidak bisa dilakukan analisis regresi sederhana atau regresi berganda biasa. Analisis regresi dummy adalah salah satu solusi dalam menghadapi kasus data ekonomi dengan peubah bebas yang bersifat kualitatif. Tulisan ini akan menelusuri data kasus ekonomi yang memiliki peubah bebas bersifat kategorik dengan menggunakan analisis regresi dummy. Data yang digunakan dalam penelitian pustaka ini adalah data sekunder yang diperoleh dari data yang sudah di publish ke internet. Berdasarkan analisis data tersebut, peubah bebas pada analisis regresi dummy boleh lebih dari satu, dimana peubah-peubah tersebut bisa saling berinteraksi atau tidak saling berinteraksi. Hal ini ditunjukkan dari gambar scatterplot dari data tersebut.

Kata kunci: regresi dummy, data kualitatif, analisis regresi.

Pendahuluan

Analisis regresi dipergunakan untuk mengetahui hubungan fungsional antara peubah-peubah yang dinyatakan dalam bentuk persamaan matematik dan garis. Persamaan matematik dan garis yang didapat disebut dengan persamaan regresi yang dapat berbentuk garis lurus (linear) dan atau tidak lurus (non linear). Hubungan fungsional terdiri dari dua jenis peubah yaitu peubah bebas (*independent*) dan peubah tak bebas (*dependent*).¹ Dalam kehidupan sehari-hari, sering kali kita ingin melihat hubungan

¹Iman, Machali, *Statistik Manajemen Pendidikan*, (Yogyakarta: Kaukaba Dipantara (Anggota Ikapi), 2015), h. 444.

antarpeubah, seperti hubungan antara panjang dan bobot bayi, tingkat pendidikan ibu dan gizi balita dalam kasus sosial. Hubungan tersebut juga terdapat dalam kasus ekonomi, seperti ingin melihat hubungan antara laba perusahaan dengan tipe perusahaan, pendapatan nasabah yang melakukan pembiayaan dengan jumlah pembiayaan yang diminta.

Dalam menentukan mana yang menjadi peubah bebas dan peubah terikat tidaklah mudah, perlu dikaji berdasarkan teori-teori pendukung. Misalnya tipe perusahaan dapat dijadikan sebagai peubah bebas dan laba perusahaan sebagai peubah terikat, atau sebaliknya. Secara kuantitatif, hubungan antara peubah bebas dan peubah tak bebas tersebut dapat kita modelkan dalam suatu persamaan matematik, sehingga kita dapat menduga nilai suatu peubah tak bebas bila diketahui nilai dari peubah bebasnya. Persamaan matematik yang menggambarkan hubungan antara peubah bebas dan tak bebas sering disebut persamaan regresi.²

Persamaan regresi dengan satu peubah bebas dan satu peubah tak bebas disebut regresi sederhana, misalnya persamaan yang menggambarkan hubungan antara tinggi bapak dan tinggi anak laki-lakinya. Dalam contoh ini, tinggi bapak sebagai peubah bebas dan tinggi anak sebagai peubah tak bebas. Persamaan dengan beberapa peubah bebas dengan satu peubah tak bebas disebut regresi berganda, misalnya hubungan antara peubah bebas tingkat pendidikan, pendapatan, dan jumlah anak terhadap peubah tak bebas pengeluaran konsumsi keluarga.³

Hasil analisis regresi adalah berupa koefisien regresi untuk masing-masing peubah bebas. Koefisien ini diperoleh dengan cara memprediksi nilai peubah tak bebas dengan suatu persamaan. Koefisien regresi dihitung dengan dua tujuan sekaligus: Pertama, meminimumkan penyimpangan antara nilai aktual dan nilai estimasi peubah tak bebas; Kedua, mengoptimalkan korelasi antara nilai aktual dan nilai estimasi peubah tak bebas berdasarkan data yang ada. Teknik estimasi peubah tak bebas yang melandasi analisis regresi disebut *Ordinary Least Squares* (kuadrat terkecil).

² Ahmad, Ansori Mattjik, dan I Made, Sumertajaya, *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*, (Bogor: IPB Press, 2000), h. 173.

³ *Ibid.*

Regresi metode kuadrat terkecil merupakan regresi yang sering dipakai pada saat data pada peubah respon merupakan numerik (kuantitatif). Namun pada data riil, seringkali data dalam bentuk kategorik (kualitatif). Sehingga data tersebut tidak akan cocok jika dianalisis menggunakan regresi sederhana atau regresi berganda biasa. Maka, dalam tulisan ini akan dibahas mengenai analisis regresi yang digunakan untuk peubah bebas kualitatif yaitu analisis regresi *dummy* dengan data kasus ekonomi.

Sejarah Analisis Regresi

Istilah regresi pertama kali diperkenalkan oleh Sir Francis Galton pada tahun 1886. Galton menemukan adanya tendensi bahwa orang tua yang memiliki tubuh tinggi memiliki anak-anak yang tinggi, orang tua yang pendek memiliki anak-anak yang pendek pula. Kendati demikian, ia mengamati bahwa ada kecenderungan tinggi anak cenderung bergerak menuju rata-rata tinggi populasi secara keseluruhan. Dengan kata lain, ketinggian anak yang amat tinggi atau orang tua yang amat pendek cenderung bergerak ke arah rata-rata tinggi populasi. Inilah yang disebut hukum Galton mengenai regresi universal. Dalam bahasa Galton, ia menyebutkan sebagai regresi menuju mediokritas.

Hukum regresi semesta (*law of universal regression*) dari Galton diperkuat oleh temannya Karl Pearson, yang mengumpulkan lebih dari seribu catatan tinggi anggota kelompok keluarga. Ia menemukan bahwa rata-rata tinggi anak laki-laki kelompok ayah (yang) pendek lebih besar dari pada tinggi ayah mereka, jadi “mundurnya” (“regressing”) anak laki-laki yang tinggi maupun yang pendek serupa ke arah rata-rata tinggi semua laki-laki. Dengan kata lain Galton, ini adalah “kemunduran ke arah sedang”.

Selanjutnya istilah “*regression*” ini digunakan untuk menggambarkan garis yang menunjukkan adanya hubungan antara 2 peubah. Ada beberapa ahli statistik yang lebih menyukai menggunakan “*estimating line*” atau “garis taksiran” daripada menggunakan istilah garis regresi.⁴

⁴Noegroho, Boedijoewono, *Pengantar Statistika Ekonomi dan Bisnis, Jilid 1 (Deskriptif)*, (Yogyakarta: UPP STIM YKPN, 2016), h. 238

Secara umum, analisis regresi pada dasarnya adalah studi mengenai ketergantungan satu peubah dependen (tak bebas) dengan satu atau lebih peubah independen (variabel penjelas/bebas), dengan tujuan untuk mengestimasi dan/atau memprediksi rata-rata populasi atau nilai rata-rata peubah dependen berdasarkan nilai peubahindependen yang diketahui. Pusat perhatian adalah pada upaya menjelaskan dan mengevaluasi hubungan antara suatu peubah dengan satu atau lebih peubahindependen.

Skala Pengukuran

Berbicara statistika tentu akan berbicara tentang data. Data merupakan kumpulan dari karakteristik objek/individu yang diamati atau sering juga disebut sebagai kumpulan dari peubah-peubah (*variables*). Berdasarkan skala pengukurannya, peubah dapat dikategorikan menjadi empat, yaitu peubah berskala nominal, ordinal, interval, dan rasio.⁵

Kita akan membahas tiga pengukuran berbeda, ketiganya berbeda dalam hal sejauh mana pengamatan-pengamatan dapat dibandingkan. Dalam urutan terendah sampai tertinggi ditinjau dari derajat perbandingan dapat dilakukan, skala-skala itu ialah nominal, ordinal, dan interval.

1. Skala Nominal

Skala pengukuran nominal adalah sejumlah kategori yang bervariasi dalam hal kualitas tertentu namun tidak dalam hal besarnya. Setiap pengamatan dapat diklasifikasikan ke dalam satu dan hanya satu kategori, yang seringkali menggambarkan jenis berbeda dari suatu fenomena. Misalnya jenis kelamin mempunyai skala (taraf-taraf) pria dan wanita. Peubah-peubah lain yang diukur dengan skala nominal ialah agama yang dianut, ras, pekerjaan, metode pembelajaran, jenis perusahaan dan status pekerjaan. Peubah yang diukur pada skala nominal dinamakan peubah nominal. Peubah nominal sering juga dinamakan peubah kualitatif, sebab dua nilai yang berbeda bagi peubah demikian ini

⁵ Ahmad, An̄sori Mattjik, dan I Made, Sumertajaya, *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*, h. 7.

berbeda dalam hal kualitas, bukan kuantitas. Meskipun kategori berbeda dari skala nominal disebut taraf skala, namun tidak ada taraf yang dapat ditafsirkan lebih besar, lebih tinggi, atau lebih kecil daripada taraf lainnya. Nama atau label dapat digunakan untuk mengidentifikasi kategori-kategori suatu peubah nominal, namun nama itu tidak mewakili besaran perbedaan nilai peubah tersebut.⁶

Angka-angka yang disajikan pada skala nominal hanya sebagai nama penggolongan. Angka tersebut tidak mengukur besaran tetapi hanya sebagai lambang. Dalam hal ini, angka 1 tidak lebih besar dari 0, begitu pula angka 0 tidak lebih kecil dari 1. Pemberian kode 1 pada laki-laki dan 0 pada perempuan, tidak berarti laki-laki mempunyai nilai 1 dan perempuan nol. Angka-angka tersebut hanyalah membedakan antara laki-laki dan perempuan, dengan demikian kita juga bisa menukar laki-laki menjadi 0 dan perempuan menjadi 1 tanpa merubah makna.⁷

2. Skala Ordinal

Bila nilai-nilai suatu peubah dapat dibandingkan dalam hal besarnya (magnitudo), dengan nilai-nilai berbeda mewakili kuantitas-kuantitas berbeda pula, peubah demikian ini dinamakan kuantitatif. Setiap nilai yang mungkin dari suatu peubah kuantitatif selalu lebih besar atau lebih kecil dari nilai lainnya. Peubah kuantitatif yang paling kasar taraf-taraf pengukurannya diukur pada skala ordinal, dan dinamakan peubah ordinal. Ada urutan-urutan di antara nilai-nilai pada skala ordinal, namun jarak antara nilai-nilai tidak memiliki makna numerik yang pasti. Sebagian skala ordinal memberikan kepada setiap objek pengamatan peringkat yang terpisah (pertama, kedua, ketiga, dan seterusnya), misalnya tingkat keaktifan dalam kegiatan sosial, peringkat naik kelas, posisi pada hasil akhir suatu lomba, peringkat skor pada suatu ujian, dan kualitas produk. Untuk masing-masing peubah ini, kita dapat membandingkan dua

⁶ Bambang Juanda, *Ekonometrika Pemodelan dan Pendugaan*, (Bogor: IPB Press, 2009), h. 99-100.

⁷ Ahmad, Ansori Mattjik, dan I Made, Sumertajaya, *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*, h 8.

pengamatan dengan cara mengatakan mana yang lebih besar (lebih tinggi peringkatnya) pada peubah tersebut.

Yang lebih sering dijumpai di dalam penelitian ilmu sosial adalah skala ordinal, yang terdiri atas suatu kumpulan kategori-kategori yang terurut secara alami. Sebagai ilustrasi, kelas sosial dapat diklasifikasikan atas kelas atas, menengah, dan bawah; filsafat politik dapat diukur sebagai liberal, moderat, atau konservatif. Peubah-peubah ini dikatakan mempunyai skala ordinal. Untuk semua peubah ordinal, meskipun ada urutan kategori yang jelas, namun perbedaan jarak antara kategori-kategori itu tidak diketahui. Misalnya, kita dapat menyimpulkan bahwa seseorang yang dikategorikan sebagai moderat lebih liberal daripada seseorang yang dikategorikan sebagai konservatif, namun kita tidak dapat memberikan nilai numerik bagi sejauh mana keliberalan orang tersebut.

3. Skala Interval

Skala interval disamping dapat diurutkan, memiliki sifat bahwa ada perbedaan (jarak numerik tertentu) antara setiap pasang taraf atau nilai peubah. Oleh karenanya, kita dapat membandingkan tidak hanya dalam hal mana yang lebih besar atau lebih tua, namun juga seberapa lebih besar atau seberapa lebih tua. Di dalam skala interval, jarak antara semua taraf yang bersebelahan adalah sama. Teladan peubah interval adalah harga suatu barang, umur, tingkat kelahiran suatu bangsa, jumlah penduduk suatu kota, tingkat pendidikan (dalam tahun), dan pajak pendapatan yang dibayar wajib pajak pada tahun lalu. Pada skala interval, kita dapat membandingkan dua nilai dengan cara mengidentifikasi sifat-sifat berikut:

1. Apakah keduanya berbeda (sifat nominal)
2. Mana yang lebih besar (sifat ordinal)
3. Jarak atau perbedaan antara keduanya (sifat interval)

Misal, kita dapat mengatakan bahwa seseorang yang berumur 60 tahun memiliki umur berbeda dari seseorang yang berumur 20 tahun (secara nominal, keduanya dikatakan berbeda umurnya); dan bahwa yang pertama 40 tahun lebih tua (perbandingan interval).

Tabel 1 meringkaskan jenis-jenis perbandingan yang dapat dilakukan untuk ketiga taraf pengukuran. Sangat penting untuk menyadari perbedaan antara ketiganya, sebab masing-masing skala pengukuran memiliki sejumlah metode statistika yang paling cocok untuk skala bersangkutan.

Tabel 1. Cara Membandingkan Objek Pengamatan menurut Skala Pengukuran Peubah

JENIS DATA	SKALA PENGUKURAN	CARA MEMBANDINGKAN	TELADAN
Kuantitatif	Interval	Apakah berbeda? Mana yang lebih besar? Seberapa lebih besar?	Harga, Umur
Kualitatif	Ordinal	Apakah berbeda? Mana yang lebih besar?	Kelas sosial
	Nominal	Apakah berbeda?	Status

Tulisan ini lebih memfokuskan bagaimana memperlakukan suatu peubah bebas kualitatif atau yang skala pengukurannya nominal atau ordinal dalam suatu model regresi. Sebagai contoh, misalnya kita ingin menjelaskan fakta bahwa lingkungan sebuah rumah dapat dikategorikan dengan “biasa” dan “baik”. Untuk melakukan ini, kita buat suatu peubah artifisial atau peubah boneka yang biasa disebut sebagai peubah *dummy* yang nilainya 1, misalnya untuk lingkungan “baik” dan nilainya 0 untuk lingkungan “biasa”. Dalam teori statistika, peubah seperti ini sering disebut juga sebagai peubah indikator, yang mana dalam ilustrasi ini adalah peubah indikator untuk lingkungan “baik” karena diberi nilai 1.⁸

Regresi *Dummy*

Nama lain Regresi *Dummy* adalah Regresi Kategori. Regresi ini menggunakan prediktor kualitatif (yang bukan *dummy* dinamai prediktor kuantitatif). Pembahasan pada regresi ini hanya untuk satu macam peubah *dummy* dan dikhususkan pada estimasi

⁸Bambang, Juanda, *Ekonometrika Pemodelandan Pendugaan*, h. 100-101.

parameter dan kemaknaan pengaruh prediktor. Peubah (variabel) *dummy* disebut juga peubah kategorik, kualitatif, boneka atau peubah dikotomi. Prinsipnya adalah membandingkan karakteristik, misal jenis kelamin (pria dan wanita), tempat tinggal (desa dan kota), dll. Analisis regresi *dummy* merupakan analisis regresi berganda, hanya saja salah satu peubah bebasnya merupakan peubah kualitatif yang berskala nominal atau ordinal.

Metode yang digunakan adalah mengganti informasi yang bersifat kategori, misal untuk jenis kelamin, PRIA diwakili angka 1 dan WANITA diwakili angka 0, DESA diwakili angka 1 dan KOTA diwakili angka 0, dll. Jika peubah kualitatif tersebut lebih dari dua kategori, jumlah peubah *dummy* yang dibentuk harus sebanyak $n-1$, dimana n adalah banyaknya kategori peubah tersebut.⁹

1. Model Regresi *Dummy* dengan Peubah Bebas Kualitatif dengan 2 Kategori (Tanpa Interaksi dengan Peubah Bebas Lainnya)

Jika kita menggunakan data sampel yang berasal dari teknik *sampling* dan peubah tak bebas adalah Y , kemudian peubah bebas 1 (X_1) dan peubah bebas 2 (X_2) tersebut, maka model persamaan dapat ditulis sebagai berikut:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \varepsilon_i, i = 1, 2, \dots, n$$

Dengan ε_i adalah galat atau sisaan (perbedaan) antara nilai Y ke- i yang sebenarnya dengan nilai dugaannya.¹⁰

Persamaan regresi berganda peubah tak bebas \hat{Y} pada peubah bebas 1 dan peubah bebas 2 (X_1 & X_2) menggambarkan hubungan antara besarnya nilai peubah tak bebas dengan peubah bebasnya. Besarnya nilai \hat{Y} ditentukan oleh besarnya X_1 & X_2 dengan intersep β_0 dan gradient β_1 & β_2 yang sering juga disebut koefisien regresi. Persamaan regresi tersebut menyatakan bahwa

⁹ Norman, R. Draper, dan Harry, Smith, *Applied Regression Analysis, Third Edition*, (New York: John Wiley & Sons, Inc, 1998), h. 299.

¹⁰ Bambang, Juanda, *Ekonometrika Pemodelandan Pendugaan*, h. 101-102.

nilai \hat{Y} akan bertambah sebesar $\hat{\beta}_1 \& \hat{\beta}_2$ setiap kali nilai $X_1 \& X_2$ bertambah satu satuan, dan apabila bernilai nol (0), maka nilai \hat{Y} akan sebesar β_0 . $\beta_1 \& \beta_2$ menyatakan besarnya perubahan nilai \hat{Y} pada setiap satuan perubahan nilai $X_1 \& X_2$.

$$\hat{\beta}_1 = \frac{\Delta \hat{Y}}{\Delta X_1}$$

Dalam banyak kasus, koefisien $\hat{\beta}_0$ tidak diinterpretasikan sebagai besarnya nilai \hat{Y} apabila $X_1 \& X_2$ bernilai 0. Dalam banyak kasus, β_0 koefisien hanya berlaku sebagai koefisien penyesuaian.¹¹

Contoh kasus:

Seorang peneliti tertarik untuk memprediksi laba dua macam perusahaan (swasta asing dan swasta nasional) bila ditinjau dari besarnya biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya. (Untuk perusahaan swasta asing, laba yang diamati adalah laba yang diperoleh dari hasil penjualan produknya di wilayah Indonesia saja)

Tabel 2. Data Laba Perusahaans

NO.	LABA (Y)	TIPE (X2)	BIAYA IKLAN(X1)
1	9.17	1	10
2	8.54	1	12
3	7.68	1	12
..
..
..
46	6.35	0	8
47	4.66	0	6
48	6.98	0	9
49	4.33	0	8
50	8.13	0	8

Source :Deny Kurniawan @2007¹²

¹¹Asep, Saefufuddin, dkk, *Statistika Dasar*, (Jakarta: Grasindo, 2009), h. 113.

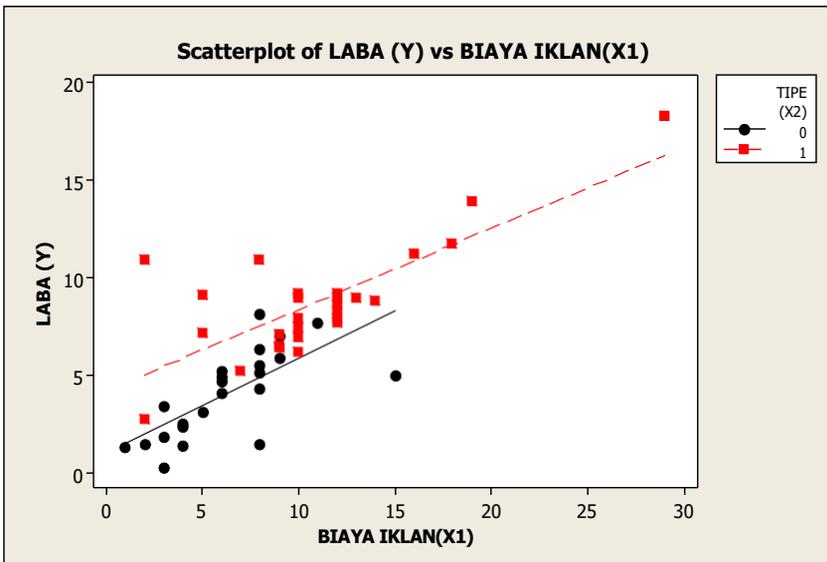
¹²<https://ineddeni.wordpress.com/2007/08/17/analisis-regresi-dengan-variabel-dummy/#more-20>, diakses tanggal 2 Januari 2017.

Selanjutnya akan dilakukan analisis pada data tersebut hingga memperoleh model umum dan model untuk setiap kategorik. Langkah-langkahnya sebagai berikut:

- Membuat variabel *dummy* bagi X_2 , misal Z_j

$$\text{Variabel } Z_j = \begin{cases} 0 & (\text{untuk peusahaan swasta asing}) \\ 1 & (\text{untuk perusahaan swasta nasional}) \end{cases}$$

- Apakah ada interaksi antara peubah X_1 dan X_2 ?



Gambar 1. Scatterplot Data Laba Perusahaan

Berdasarkan *Scatterplot* pada Gambar 1 menunjukkan bahwa tidak ada interaksi antara peubah X_1 dan X_2 . Hal tersebut ditunjukkan dengan 2 garis regresi yang tidak saling berpotongan, sehingga model persamaan regresinya yaitu:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 Z_j + \varepsilon_i, i = 1, 2, \dots, n; j = 0, 1$$

Model kategori untuk perusahaan swasta nasional, ($Z_j = 1$) :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 Z_j + \varepsilon_i$$

$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(1) + \varepsilon_i$$

$$= (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1 + \varepsilon_i$$

Model kategori untuk perusahaan swasta asing, ($Z_j = 0$) :

$$\ddot{Y} = \ddot{\beta}_0 + \ddot{\beta}_1 X_1 + \ddot{\beta}_2 Z_j + \varepsilon_i$$

$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(0) + \varepsilon_i$$

$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \varepsilon_i$$

- Perhitungan untuk memperoleh nilai $\hat{\beta}_0$, $\hat{\beta}_1$, dan $\hat{\beta}_2$ menggunakan software Minitab 16
Output pada Minitab:

Regression Analysis: LABA (Y) versus BIAYA IKLAN(X1), TIPE (X2)

The regression equation is

$$\text{LABA (Y)} = 1.33 + 0.431 \text{ BIAYA IKLAN(X1)} + 2.66 \text{ TIPE (X2)}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.3345	0.5078	2.63	0.012	

BIAYA IKLAN(X1) 0.43066 0.05588 7.71 0.000 1.278

TIPE (X2) 2.6627 0.5626 4.73 0.000 1.278

S = 1.75374 R-Sq = 75.9% R-Sq(adj) = 74.9%

- Setelah diperoleh nilai dugaan untuk tiap paramater tersebut, maka nilai tersebut disubstitusi kedalam model persamaan umum dan model untuk setiap kategori, sehingga diperoleh:

$$\ddot{Y} = 1.33 + 0.431 X_1 + 2.66 Z_j + \varepsilon_i \quad \dots (2)$$

Model kategori untuk perusahaan swasta nasional :

$$\ddot{Y} = 3.99 + 0.431 X_1 + \varepsilon_i$$

Model kategori untuk perusahaan swasta asing :

$$\ddot{Y} = 1.33 + 0.431 X_1 + \varepsilon_i$$

➤ Interpretasi

Berdasarkan hasil output Minitab 16 di atas dapat dilihat bahwa pada taraf uji nyata 5%, biaya iklan dan tipe perusahaan berpengaruh nyata (signifikan) terhadap laba perusahaan. Hal ini dapat dilihat dari Nilai-P yang lebih kecil dari 5%. Model persamaan umum untuk semua kategori yang diperoleh dikatakan model yang baik, hal ini dapat dilihat dari nilai R-Sq (*R-Square*) yang dihasilkan lebih dari 50%. Nilai R-Sq sebesar 75,9% memiliki arti bahwa 75,9% keragaman dari laba perusahaan dapat dijelaskan oleh biaya iklan dan tipe perusahaan. Berdasarkan model persamaan umum regresi *dummy*(1) di atas, dapat dinyatakan bahwa nilai dugaan laba perusahaan akan bertambah sebesar 0,431 dan 2,66 setiap kali nilai X_1 & X_2 bertambah satu satuan.

2. Model Regresi *Dummy* dengan Peubah Bebas Kualitatif dengan 2 Kategori (Yang Berinteraksi dengan Peubah Bebas Lainnya)

Data yang digunakan adalah data sampel contoh kasus sebelumnya dengan beberapa perubahan pada datum dengan peubah tak bebas adalah Y , kemudian peubah bebas 1 (X_1) dan peubah bebas 2 (X_2) dan terdapat interaksi antara X_1 dan X_2 , sehingga didapatkan model persamaan sebagai berikut:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + \varepsilon_i, i = 1, 2, \dots, n$$

Dengan ε_i adalah galat atau sisaan (perbedaan) antara nilai Y yang sebenarnya dengan nilai dugaannya.¹³

Contoh kasus:

Seorang peneliti tertarik untuk memprediksi laba 2 macam perusahaan (swasta asing dan swasta nasional) bila ditinjau dari besarnya biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya. (Untuk perusahaan swasta asing, laba yang diamati adalah laba yang diperoleh dari hasil penjualan produknya di wilayah Indonesia saja)

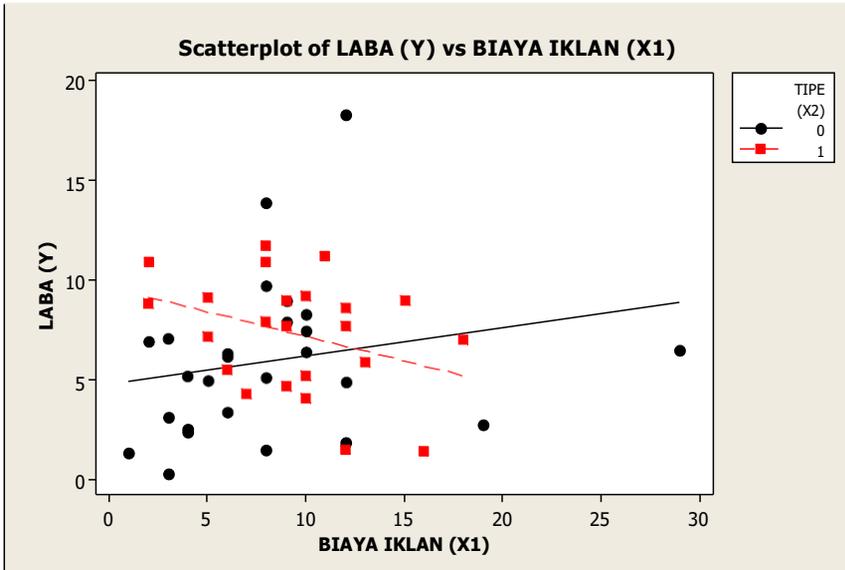
¹³Bambang, Juanda, *Ekonometrika Pemodelandan Pendugaan*, h. 103.

Tabel 2. Data Laba Perusahaan

NO.	LABA (Y)	TIPE (X2)	BIAYA IKLAN(X1)
1	9,17	1	10
2	1,32	0	1
3	8,54	1	12
..
..
..
46	4,33	1	7
47	5,2	1	10
48	6,96	0	2
49	2,78	0	19
50	13,9	0	8

Selanjutnya akan dilakukan analisis pada data tersebut hingga memperoleh model umum dan model untuk setiap kategorik. Langkah-langkahnya sebagai berikut:

- Membuat variabel *dummy* bagi X_2 , misal Z_j
 Variabel $Z_j = \begin{cases} 0 & (\text{untuk perusahaan swasta asing}) \\ 1 & (\text{untuk perusahaan swasta nasional}) \end{cases}$
- Apakah ada interaksi antara peubah X_1 dan X_2 ?



Gambar 2. Scatterplot Data Laba Perusahaan

Berdasarkan *Scatterplot* pada Gambar 2 menunjukkan bahwa ada interaksi antara peubah X_1 dan X_2 . Hal tersebut ditunjukkan dengan 2 garis regresi yang saling berpotongan, sehingga model persamaan regresinya yaitu:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 Z_j + \hat{\beta}_3 X_1 Z_j + \varepsilon_i, i = 1, 2, \dots, n; j = 0, 1$$

Model kategori untuk perusahaan swasta nasional, ($Z_j = 1$) :

$$\begin{aligned} \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 Z_j + \hat{\beta}_3 X_1 Z_j + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (1) + \hat{\beta}_3 (1) X_1 + \varepsilon_i \\ &= (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_3) X_1 + \varepsilon_i \end{aligned}$$

Model kategori untuk perusahaan swasta asing, ($Z_j = 0$) :

$$\begin{aligned} \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 Z_j + \hat{\beta}_3 X_1 Z_j + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (0) + \hat{\beta}_3 (1 \cdot 0) + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \varepsilon_i \end{aligned}$$

Berdasarkan *Scatterplot* pada Gambar 2 menunjukkan bahwa ada interaksi antara peubah X_1 dan X_2 , sehingga akan dibuat interaksi pada data dengan cara mengkalikan peubah yang berinteraksi.

Tabel 3. Data Laba Perusahaan dengan Interaksi Antara Biaya Iklan dan Tipe Perusahaan

NO.	LABA (Y)	TIPE (X2)	BIAYA IKLAN (X1)	BIAYA IKLAN (X1)*TIPE (X2)
1	9,17	1	10	10
2	1,32	0	1	0
3	8,54	1	12	12
..
..
..
46	4,33	1	7	7
47	5,2	1	10	10
48	6,96	0	2	0
49	2,78	0	19	0
50	13,9	0	8	0

- Perhitungan untuk memperoleh nilai $\hat{\beta}_0$, $\hat{\beta}_1$, dan $\hat{\beta}_2$ menggunakan software Minitab 16

Output pada Minitab:

Regression Analysis: LABA (Y) versus BIAYA IKLAN (X1), TIPE (X2), ...				
The regression equation is				
LABA (Y) = 4.76 + 0.141 BIAYA IKLAN (X1) + 4.85 TIPE (X2)				
- 0.385 BIAYA IKLAN (X1)*TIPE (X2)				
Predictor	Coef	SE Coef	T	P
Constant	4.757	1.176	4.04	0.000

BIAYA IKLAN (X1) 0.1410 0.1184 1.19 0.240

TIPE (X2) 4.852 2.209 2.20 0.033

BIAYA IKLAN (X1)*TIPE (X2) -0.3855 0.2179 -1.77 0.084

S = 3.45814 R-Sq = 10.4% R-Sq(adj) = 4.4%

- Setelah diperoleh nilai dugaan untuk tiap paramater tersebut, maka nilai tersebut disubtitusi kedalam model persamaan umum dan model untuk setiap kategori, sehingga diperoleh:

$$\hat{Y} = 4.76 + 0.141X_1 + 4.85Z_j - 0.385 X_1Z_j + \varepsilon_i \quad \dots (2)$$

Model kategori untuk perusahaan swasta nasional :

$$\hat{Y} = 9.61 - 0.244X_1 + \varepsilon_i$$

Model kategori untuk perusahaan swasta asing :

$$\hat{Y} = 4.76 + 0.141X_1 + \varepsilon_i$$

- Interpretasi

Berdasarkan hasil output Minitab 16 di atas dapat dilihat bahwa pada taraf uji nyata 5%, biaya iklan dan interaksi antara biaya iklan dan tipe perusahaan tidak berpengaruh nyata (signifikan) terhadap laba perusahaan. Hal ini dapat dilihat dari Nilai-P pada biaya iklan dan interaksi antara biaya iklan dan tipe perusahaan yang lebih besar dari 5%. Sementara tipe perusahaan berpengaruh nyata (signifikan) terhadap laba perusahaan. Hal ini dapat dilihat dari Nilai-P pada tipe perusahaan yang lebih kecil dari 5%. Model persamaan umum untuk semua kategori yang diperoleh dikatakan model yang kurang baik, hal ini dapat dilihat dari nilai R-Sq (*R-Square*) yang dihasilkan kurang dari 50%. Nilai R-Sq sebesar 10,4% memiliki arti bahwa 10,4% keragaman dari laba perusahaan dapat dijelaskan oleh biaya iklan dan tipe perusahaan serta interaksi antara biaya iklan dan tipe perusahaan. Berdasarkan model persamaan umum regresi *dummy*(2) di atas, dapat dinyatakan bahwa nilai dugaan laba perusahaan akan bertambah sebesar 4,76 dan 4,85 setiap kali nilai X_1 & X_2 bertambah satu satuan, dan akan berkurang sebesar 0,385 setiap kali nilai interaksi antara X_1 & X_2 bertambah satu satuan.

3. Model Regresi *Dummy* dengan Peubah Bebas Kualitatif dengan 3 Kategori atau Lebih

Sebagai ilustrasi, misalnya kita ingin mengembangkan suatu model untuk menduga gaji karyawan (\hat{Y}) berdasarkan pengalaman kerja (X_1) dan pendidikan (X_2) karyawan tersebut, dengan model sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i, i = 1, 2, \dots, n$$

Peubah pendidikan karyawan (X_2) misalnya dikategorikan ke dalam 3 kategori, yaitu lulusan "Sekolah Dasar (SD)", "Sekolah Lanjut (SL)", dan "Perguruan Tinggi (PT)".

Untuk pemodelan ini, misalnya telah dikumpulkan data ke-3 peubah dari n sampel responden karyawan yang diambil secara acak, sehingga model regresi sampelnya adalah sebagai berikut:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \varepsilon_i, i = 1, 2, \dots, n$$

Misalkan dilakukan 2 cara pemberian nilai untuk X_2 .

➤ Membuat variabel *dummy* bagi X_2 , misal Z_j

$$\text{Variabel } Z_j = \begin{cases} 0; & \text{untuk lulusan SD} \\ 1; & \text{untuk lulusan SL} \\ 2; & \text{untuk lulusan PT} \end{cases}$$

(X_2) Untuk model dengan cara pemberian nilai untuk peubah tersebut dapat diuraikan untuk masing-masing kategorinya, yaitu dengan memasukkan nilai peubah kategori tersebut.

$$\ddot{Y} = \ddot{\beta}_0 + \ddot{\beta}_1 X_1 + \ddot{\beta}_2 Z_j + \varepsilon_i, i = 1, 2, \dots, n; j = 0, 1, 2$$

Model kategori untuk lulusan "SD", ($Z_j = 0$):

$$\begin{aligned} \ddot{Y} &= \ddot{\beta}_0 + \ddot{\beta}_1 X_1 + \ddot{\beta}_2 Z_j + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(0) + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \varepsilon_i \end{aligned} \quad \dots(3)$$

Model kategori untuk lulusan "SL", ($Z_j = 1$):

$$\begin{aligned} \ddot{Y} &= \ddot{\beta}_0 + \ddot{\beta}_1 X_1 + \ddot{\beta}_2 Z_j + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(1) + \varepsilon_i \\ &= (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1 + \varepsilon_i \end{aligned} \quad \dots(4)$$

Model kategori untuk lulusan "PT", ($Z_j = 2$):

$$\ddot{Y} = \ddot{\beta}_0 + \ddot{\beta}_1 X_1 + \ddot{\beta}_2 Z_j + \varepsilon_i$$

$$\begin{aligned}
 &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(2) + \varepsilon_i \\
 &= (\hat{\beta}_0 + 2\hat{\beta}_2) + \hat{\beta}_1 X_1 + \varepsilon_i
 \end{aligned}
 \dots(5)$$

Dari ketiga model (3), (4), dan (5) ini terlihat jelas bahwa jika pengalaman karyawan sama, maka perbedaan gaji tiap jenjang pendidikan sama, yaitu sebesar nilai dugaan koefisien $\hat{\beta}_2$.

Dalam realitas, perbedaan gaji tiap jenjang pendidikan belum tentu sama. Oleh karena itu, sebaiknya menggunakan cara kedua yaitu dengan mengembangkan peubah *dummy* yang nilainya hanya ada dua kemungkinan, yaitu 0 dan 1. Aturan umumnya adalah “jika peubah bebas kualitatif (skala nominal atau ordinal) nilainya ada k kategori maka kita harus menggunakan k-1 peubah *dummy*.”

Untuk ilustrasi ini, ada 2 peubah *dummy* karena ada 3 kategori. Misalnya pemberian nilai untuk 2 peubah *dummy* tersebut adalah sebagai berikut :

Pendidikan	$D_1 = D_{SL}$	$D_2 = D_{PT}$
SD	0	0
SL	1	0
PT	0	1

Cara kedua dengan menggunakan peubah *dummy* ini dapat juga diungkapkan sebagai berikut:

$$\begin{aligned}
 D_1 = D_{SL} &= \begin{cases} 1, & \text{jika lulusan SL} \\ 0, & \text{jika selainnya} \end{cases} \\
 D_2 = D_{PT} &= \begin{cases} 1, & \text{jika lulusan PT} \\ 0, & \text{jika selainnya} \end{cases}
 \end{aligned}$$

Untuk pemodelan dengan menggunakan peubah *dummy* ini, misalnya telah dikumpulkan data dari n contoh responden karyawan yang diambil secara acak, sehingga model regresi contohnya adalah sebagai berikut :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 D_{SL} + \hat{\beta}_3 D_{PT} + \varepsilon_i, i = 1, 2, \dots, n$$

Model dengan menggunakan peubah *dummy* tersebut dapat diuraikan untuk masing-masing kategorinya, yaitu dengan memasukkan nilai peubah *dummy* tersebut.

Model kategori untuk lulusan "SD", ($D_{SL} = D_{PT} = 0$) :

$$\begin{aligned}\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 D_{SL} + \hat{\beta}_3 D_{PT} + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(0) + \hat{\beta}_3(0) + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \varepsilon_i\end{aligned}\quad \dots(6)$$

Model kategori untuk lulusan "SL", ($D_{SL} = 1, D_{PT} = 0$) :

$$\begin{aligned}\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 D_{SL} + \hat{\beta}_3 D_{PT} + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(1) + \hat{\beta}_3(0) + \varepsilon_i \\ &= (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1 + \varepsilon_i\end{aligned}\quad \dots(7)$$

Model kategori untuk lulusan "PT", ($D_{SL} = 0, D_{PT} = 1$) :

$$\begin{aligned}\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 D_{SL} + \hat{\beta}_3 D_{PT} + \varepsilon_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2(0) + \hat{\beta}_3(1) + \varepsilon_i \\ &= (\hat{\beta}_0 + \hat{\beta}_3) + \hat{\beta}_1 X_1 + \varepsilon_i\end{aligned}\quad \dots(8)$$

Dari ketiga model (6), (7) dan (8) ini terlihat jelas bahwa jika pengalaman karyawan sama, maka perbedaan gaji tiap jenjang pendidikan dapat berbeda. Interpretasi dari koefisien modelnya adalah sebagai berikut:

$\hat{\beta}_0$: rata-rata gaji karyawan lulusan SD yang belum berpengalaman

$\hat{\beta}_1$: rata-rata kenaikan gaji karyawan jika pengalamannya sama

$\hat{\beta}_2$: rata-rata perbedaan gaji karyawan lulusan SL dengan lulusan SD jika pengalamannya (X_1) sama

$\hat{\beta}_3$: rata-rata perbedaan gaji karyawan lulusan PT dengan lulusan SD jika pengalamannya (X_1) sama

$\hat{\beta}_2 - \hat{\beta}_3$: rata-rata perbedaan gaji karyawan lulusan PT dengan lulusan SL jika pengalamannya (X_1) sama¹⁴

Berdasarkan penjelasan beberapa model persamaan regresi *dummy* di atas bahwa analisis regresi *dummy* digunakan hanya untuk data yang salah satu atau beberapa peubah bebasnya berskala nominal atau ordinal. Namun dalam praktiknya lebih sering digunakan peubah berskala nominal untuk mendapatkan

¹⁴ Bambang, Juanda, *Ekonometrika Pemodelan dan Pendugaan*, h. 105-107.

hasil dugaan (data sampel) dengan keadaan sebenarnya (data populasi). Pada data kasus ekonomi banyak sekali ditemui peubah-peubah yang berskala nominal ataupun ordinal (data kualitatif), sehingga metode statistika yaitu analisis regresi *dummy* ini sering sekali digunakan sebagai alat bantu untuk menyelesaikan masalah atau melihat hubungan pada data-data ekonomi.

Simpulan

Analisis regresi adalah suatu analisis yang digunakan untuk melihat hubungan antar peubah yang diduga memiliki hubungan sebab akibat. Data yang dipakai minimal memiliki satu peubah tak bebas dan lebih dari satu peubah bebas. Dimana peubah tak bebas merupakan peubah yang dipengaruhi oleh peubah bebas. Pada analisis regresi *dummy* peubah bebas bersifat kategorik atau numerik dengan skala pengukuran nominal, ordinal atau interval. Peubah bebas pada analisis regresi *dummy* boleh lebih dari satu, dimana peubah-peubah tersebut bisa saling berinteraksi atau tidak saling interaksi. Hal ini ditunjukkan dengan menggambar *scatterplot* dari data tersebut. Analisis regresi *dummy* adalah salah satu analisis yang dipakai pada data ekonomi, yang dimana salah satu peubah bebas dari data tersebut berupa data kualitatif dengan skala pengukuran nominal atau ordinal.

Daftar Pustaka

- Ahmad, Ansori Mattjik, dan I Made, Sumertajaya, *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*, (Bogor: IPB Press, 2000).
- Asep, Saefufuddin, dkk, *Statistika Dasar*, (Jakarta: Grasindo, 2009).
- Bambang, Juanda, *Ekonometrika Pemodelandan Pendugaan*, (Bogor: IPB Press, 2009).
- <https://ineddeni.wordpress.com/2007/08/17/analisis-regresi-dengan-variabel-dummy/#more-20>, diakses tanggal 2 Januari 2017.

Imam, Machali, *Statistik Manajemen Pendidikan*, (Yogyakarta: Kaukaba Dipantara (Anggota Ikapi), 2015).

Noegroho, Boedijoewono, *Pengantar Statistika Ekonomi dan Bisnis, Jilid 1 (Deskriptif)*, (Yogyakarta: UPP STIM YKPN, 2016).

Norman, R. Draper, dan Harry, Smith, *Applied Regression Analysis, Third Edition*, (New York: John Wiley & Sons, Inc, 1998).