

Prosedur Analisis Regresi dengan Variabel Dummy

Oleh Wahyu Widhiarso | Fakultas Psikologi UGM
Manuskrip Tidak Dipublikasikan, Tahun 2010

Mengapa kita melibatkan variabel kualitatif dalam penelitian kita (sebagai variabel independen)?¹ Kita tertarik pada pengaruh variabel independen kualitatif, misalnya apakah jenis kelamin dapat memprediksi ketahanan kerja seseorang? 2) Kita ingin memasukkan variabel kualitatif tersebut untuk meningkatkan kualitas penelitian kita. Misalnya memasukkan variabel seperti jenis kelamin, ras, dll sejajar dengan variabel independen lainnya. 2) Kita ingin menghilangkan pengaruh mereka terhadap prediktor kita yang utama karena dapat menyebabkan bias estimasi. Misalnya memasukkan pangkat/jabatan dalam regresi yang menguji prediksi gaji terhadap kepuasan kerja. Berikut ini akan dibahas prosedur analisis regresi dengan menggunakan variabel dummy.

SISTEM PENGKODEAN DUMMY

Pengkodean data kategorikal memerlukan pengkategorian eksklusif. Artinya satu subjek/sampel adalah masuk dalam satu kategori, tidak boleh dua kategori. Sampel A misalnya, tidak boleh masuk ke dalam kategori laki-laki dan perempuan. Si B tidak boleh masuk ke dalam kategori PNS dan wiraswasta meskipun kedua profesi tersebut dijalannya. Aturan ini berlaku variabel dummy. Sebuah variabel dengan kategori sebanyak k akan membutuhkan seperangkat $k - 1$ variabel dummy untuk menjangkau semua informasi yang terkandung didalamnya. Jadi misalnya saya memiliki variabel tingkat pendidikan dari SD hingga PT (4 kategori) maka jumlah variabel kategori yang sama buat ada 3 kategori. Menggunakan pola koding biner (0,1), variabel dummy selalu variabel berbentuk dikotomi.

Semua responden yang menjadi anggota kategori yang diberi kode 1 sedangkan responden tidak dalam kategori tersebut dikode dari 0. Dengan cara seperti ini maka setiap responden akan memiliki kode 1 pada kategori yang sesuai dengannya dan kode 0 pada kategori yang tidak sesuai dengannya. Kode biner dapat dianggap sebagai mirip ke saklar listrik: kode A 1 sinyal bahwa kategori yang diberikan adalah "on" untuk responden (misalnya, dia adalah anggota dari kelompok tertentu, atau karakteristik tertentu hadir), karena bukan anggota, variabel dummy yang menunjukkan kategori yang diaktifkan "off" (yaitu, karakteristik ini tidak ada).

CONTOH DESAIN PENELITIAN

Sebuah penelitian hendak menguji prediksi ketahanan kerja terhadap kepuasan kerja. Karena dari teori dinyatakan bahwa kepuasan kerja turut dipengaruhi oleh jabatan dan bidang kerja, maka peneliti turut memasukkannya dalam penelitian. Formulasi informasi variabelnya sebagai berikut :

- Variabel dependen : kepuasan kerja (Y)
- Variabel independen : ketahanan kerja (X₁), jabatan (X₂), bidang kerja (X₃)

Variabel kepuasan kerja dan ketahanan kerja adalah variabel kuantitatif karena bersifat kontinum karena didapatkan dari pengukuran melalui skala. Jabatan dan bidang kerja adalah variabel kualitatif karena bersifat kategorikal. Dalam hal ini jabatan ada 2 kategori : pelaksana & manajerial. Bidang kerja ada 3 yaitu suplai, produksi, dan marketing.

CONTOH PENYAJIAN DATA

Untuk menyajikan data dengan variabel yang berbentuk dummy, kita mengkodinya dengan kode 1 dan 0. Misalnya jabatan yang terdiri dari 2 kategori, maka kita membuat 1 variabel baru yang isinya adalah kode 1 dan 0, misalnya 1 untuk manajer dan 0 untuk pelaksana. Agar memudahkan pemaknaan terhadap hasil berilah angka 1 pada kategori yang secara teoritik akan memiliki nilai prediksi lebih tinggi. Dalam hal ini saya memberikan kode 1 pada manajer. Data ini dapat diunduh [di sini](#)

Jika kategorinya lebih dari 2, maka kita juga melakukan hal yang sama. Variabel bidang kerja saya jabarkan menjadi 3 variabel baru, dengan kode 1 dan 0. Tampilannya bisa dilihat di Tabel 1.

puas	motif	jabatan	Manajer	Pelaksana	bidang	suplai	produksi	marketing
3	4	1	1	0	1	1	0	0
4	5	1	1	0	1	1	0	0
3	6	1	1	0	1	1	0	0
5	6	1	1	0	2	0	1	0
4	5	1	1	0	2	0	1	0
5	6	1	1	0	2	0	1	0
6	6	1	1	0	3	0	0	1
5	7	1	1	0	3	0	0	1
6	8	1	1	0	3	0	0	1
4	8	1	1	0	3	0	0	1
5	9	2	0	1	1	1	0	0
8	9	2	0	1	1	1	0	0
9	8	2	0	1	1	1	0	0
8	7	2	0	1	2	0	1	0
7	6	2	0	1	2	0	1	0
8	5	2	0	1	2	0	1	0
9	9	2	0	1	3	0	0	1
9	8	2	0	1	3	0	0	1
8	8	2	0	1	3	0	0	1
7	9	2	0	1	3	0	0	1

Tabel 1. Tampilan Data dengan Variabel Dummy

Terlihat ada variabel bidang yang terdiri dari 3 kategori (1 untuk suplai, 2 untuk produksi dan 3 untuk marketing). Lalu saya jabarkannya menjadi 3 variabel yang memiliki nama sesuai dengan kategorinya. Misalnya untuk variabel suplai, kode 1 saya berikan pada subjek pada bagian suplai, dan kode 0 untuk variabel produksi dan marketing. Setelah siap, data bisa kita masukkan ke SPSS.

ANALISIS DATA

Untuk menganalisisnya, kita bisa menggunakan regresi berganda (*multiple regression*) atau regresi berjenjang (*hierarchical regression*). Jika kita hendak mencari secara langsung prediksi semua variabel independen kita bisa memiliki regresi berganda, akan tetapi jika kita ingin melihat prediksi tiap variabel independen secara berurutan kita menggunakan regresi berjenjang. Kali ini kita akan menggunakan regresi berganda.

Dalam analisis regresi berganda dengan variabel dummy, tidak semua kategori dalam variabel dummy dilibatkan analisis regresi jika jumlah kategorinya. Jika jumlah kategori lebih dari 2 maka jumlah kategori yang kita libatkan adalah $k-1$ (k =jumlah kategori). Misalnya jumlah kategori bidang kerja adalah 3, maka jumlah kategori yang kita libatkan adalah $3-1= 2$ kategori. Untuk desain kita seperti ini, kita masukkan variabel kepuasan kerja (Y), ketahanan kerja (X1), jabatan (X2) dan 2 kategori pada bidang kerja (X3) yaitu produksi dan marketing. Kategori suplai tidak kita masukkan karena menjadi basis interpretasi variabel bidang kerja nantinya.

LATIHAN 1

Sebelum kita membahas regresi dengan variabel dummy saya mengajak untuk menganalisis dengan regresi variabel kuantitatif yang biasa kita lakukan. Kita akan memprediksi kepuasan kerja (Y) berdasarkan ketahanan kerja (X1).

	puas	tahan	manajer	suplai	produksi
1	3	4	0	1	0
2	4	5	0	1	0
3	3	6	0	1	0
4	5	6	0	0	1
5	4	5	0	0	1
6	5	6	0	0	1
7	6	6	0	0	0
8	5	7	0	0	0
9	6	8	0	0	0
10	4	8	0	0	0
11	5	9	1	1	0
12	8	9	1	1	0
13	9	8	1	1	0
14	8	7	1	0	1
15	7	6	1	0	1
16	8	5	1	0	1
17	9	9	1	0	0
18	9	8	1	0	0
19	8	8	1	0	0
20	7	9	1	0	0

Tabel 2. Tampilan Data dengan Variabel Dummy

Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel puas (Y) pada DEPENDENT, trus tahan (X1) pada kolom INDEPENDENT. Klik OK.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.546 ^a	.298	.259	1.750

a. Predictors: (Constant), tahan

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.406	1	23.406	7.640	.013 ^a
	Residual	55.144	18	3.064		
	Total	78.550	19			

a. Predictors: (Constant), tahan
b. Dependent Variable: puas

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.243	1.818		.684	.503
	tahan	.706	.255	.546	2.764	.013

a. Dependent Variable: puas

Tabel 3. Hasil Analisis Regresi dengan Variabel Dummy

Hasil analisis menunjukkan bahwa variabel ketahanan mampu memprediksi kepuasan kerja ($F=7.640$; $p<0.05$) yang terlihat pada kolom ANOVA dengan sumbangan efektif sebesar 29.8 persen yang terlihat dari tabel Model Summary. Artinya ketahanan dapat memprediksi sebesar 29.8 persen kepuasan individu, sisanya diprediksi oleh hal-hal selain ketahanan. Gak apa, 29.8 persen termasuk sumbangan yang sedang kok. Dari tabel koefisien kita dapat menurunkan menjadi persamaan regresi sebagai berikut :

$$Y = B_0 + B_1(\text{TAHAN}) = 1.243 + 0.706(\text{TAHAN})$$

Dari persamaan ini kita dapat memprediksi jika seseorang memiliki skor ketahanan mental sebesar 2 maka kepuasannya sebesar $Y = 1.243 + 0.706(2) = 2.655$. Dan jika ia memiliki ketahanan sebesar 0 maka kepuasannya sebesar 1.243. Inilah yang berkaitan dengan hal-hal lain selain ketahan yang mempengaruhi kepuasan individu, atau $100\% - 29.8\% = 70.20\%$

LATIHAN 2

Kali ini kita berlatih untuk memprediksi kepuasan kerja (Y) berdasarkan jabatan (X2). Variabel jabatan telah di dummy dengan kode kategori 1 untuk manajer dan 0 untuk pelaksana.

Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel puas (Y) pada DEPENDENT, trus jabatan (X2) pada kolom INDEPENDENT Klik OK.

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.833 ^a	.693	.676	1.157	

a. Predictors: (Constant), manajer

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	54.450	1	54.450	40.668	.000 ^a
	Residual	24.100	18	1.339		
	Total	78.550	19			

a. Predictors: (Constant), manajer
b. Dependent Variable: puas

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.500	.366		12.298	.000
	manajer	3.300	.517	.833	6.377	.000

a. Dependent Variable: puas

Tabel 4. Hasil Analisis Regresi dengan Variabel Dummy

Terlihat pada tabel ANOVA bahwa jabatan mampu memprediksi kepuasan secara signifikan ($F=40.668$; $p<0.05$). Pada tabel Coefficient terlihat bahwa ada perbedaan kepuasan antara jabatan manajer dan pelaksana ($b=3.300$; $p<0.05$). Persamaan regresi untuk analisis ini adalah sebagai berikut :

$$Y = B_0 + B_1(\text{MANAJER})$$

$$Y = 4.50 + 3.30(\text{MANAJER})$$

Misalnya kita ditanya berapakah kepuasan individu PELAKSANA ? Kita masukkan ke dalam persamaan di atas, PELAKSANA adalah sama dengan MANAJER = 0 (ingat pengkodean di muka)

$$Y = 4.50 + 3.30(0) = 3.30$$

Berapa kepuasan MANAJER ? (MANAJER = 1)

$$Y = 4.50 + 3.30(1) = 7.80$$

Jabatan manajer memiliki kepuasan lebih tinggi dibanding dengan pelaksana (selisih 4.5 poin)

LATIHAN 3

Sekarang sedikit kompleks, kita berlatih untuk memprediksi kepuasan kerja (Y) berdasarkan bidang kerja (X₃) yang memiliki banyak kategori.

Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel PUAS (Y) pada DEPENDENT, trus masukkan kategori SUPLAI dan PRODUKSI pada kolom INDEPENDENT Klik OK. Variabel MARKETING tidak kita masukkan karena ia akan menjadi referensi seperti halnya variabel PELAKSANA sebelumnya.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.296 ^a	.088	-.020	2.053

a. Predictors: (Constant), produksi, suplai

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.883	2	3.442	.816	.459 ^a
	Residual	71.667	17	4.216		
	Total	78.550	19			

a. Predictors: (Constant), produksi, suplai
b. Dependent Variable: puas

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.750	.726		9.299	.000
	suplai	-1.417	1.109	-.328	-1.278	.219
	produksi	-.583	1.109	-.135	-.526	.606

a. Dependent Variable: puas

Tabel 5. Hasil Analisis Regresi dengan Variabel Dummy

Terlihat pada tabel ANOVA bahwa bidang kerja tidak mampu memprediksi kepuasan ($F=.816$; $p>0.05$). Sumbangan efektif prediksi bidang kerja hanya sekitar 0.88 persen. Sedikit sekali ya, 99,2 persen adalah prediksi oleh variabel yang lain. Kita masuk ke tabel COEFFICIENT. Sama seperti pada analisis berdasarkan jabatan di muka, prediksi kepuasan dapat kita ketahui melalui persamaan regresi. Berikut ini persamaannya.

$$Y = B_0 + B_1(\text{SUPLAI}) + B_2(\text{PRODUKSI})$$

$$Y = 6.750 - 1.417(\text{SUPLAI}) - 0.583(\text{PRODUKSI})$$

ingat MARKETING tidak kita masukkan karena menjadi kategori yang dijadikan referensi, seperti halnya PELAKSANA

Jika kita ditanya berapa KEPUASAN bidang SUPLAI ?

$$Y = 6.750 - 1.417(1) - 0.583(0) = 4.75$$

Berapa KEPUASAN bidang MARKETING ?

$$Y = 6.750 - 1.417(0) - 0.583(0) = 6.75$$

Nilai signifikansi di atas menunjukkan perbedaan antara kategori yang ditampilkan dengan MARKETING sebagai referensi. Terlihat bahwa tidak ada perbedaan KEPUASAN antara bidang MARKETING dan SUPLAI ($t=-1.278$; $p>0.05$) dan juga tidak ada perbedaan antara bidang MARKETING dengan PRODUKSI ($t=-0.526$; $p>0.05$)

Trus, bagaimana perbedaan KEPUASAN antara PRODUKSI dan SUPLAI? Anda dapat menganalisis kembali regresi seperti yang saya contohkan dengan menggunakan kategori PRODUKSI atau SUPLAI sebagai referensinya.

LATIHAN 4

Pada bagian ini kita akan menganalisis semua variabel yang kita miliki.

Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel PUAS (Y) pada DEPENDENT, trus tahan (X_1), MANAJER (X_2), dan 2 kategori bidang kerja (X_3), yaitu SUPLAI dan PRODUKSI pada INDEPENDENT(S) lalu klik OK. Cara membacanya sama seperti yang saya contohkan.

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.884 ^a	.781	.722	1.071	

a. Predictors: (Constant), produksi, manajer, suplai, tahan

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	61.336	4	15.334	13.361	.000 ^a
	Residual	17.214	15	1.148		
	Total	78.550	19			

a. Predictors: (Constant), produksi, manajer, suplai, tahan
b. Dependent Variable: puas

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.022	1.825		2.752	.015
	tahan	.011	.252	.009	.044	.966
	manajer	3.281	.643	.828	5.107	.000
	suplai	-1.405	.635	-.325	-2.212	.043
	produksi	-.561	.774	-.130	-.724	.480

a. Dependent Variable: puas

Tabel 6. Hasil Analisis Regresi dengan Variabel Dummy

Terlihat pada tabel ANOVA bahwa secara simultan (bersama-sama) semua prediktor mampu memprediksi KEPUASAN secara signifikan ($F=13.361$ $p<0.05$). Sumbangan efektif prediksi bidang kerja hanya sekitar 78.1 persen. Kita masuk ke tabel COEFFICIENT, prediksi KEPUASAN dapat kita ketahui melalui persamaan regresi. Berikut ini persamaannya :

$$Y = B_0 + B_1(\text{TAHAN}) + B_2(\text{MANAJER}) + B_2(\text{SUPLAI}) + B_3(\text{PRODUKSI})$$

$$Y = 5.022 - 0.011(\text{TAHAN}) + 3.281(\text{MANAJER}) - 1.405(\text{SUPLAI}) + 0.561(\text{PRODUKSI})$$

Jika di tanya berapakah KEPUASAN pekerja pada jabatan PELAKSANA dalam bidang SUPLAI jika KETAHANAN mereka skornya 2 ?

$$Y = 5.022 - 0.011(2) + 3.281(0) - 1.405(1) + 0.561(0) = 3.95$$

Nilai signifikansi di atas menunjukkan bahwa ketahanan tidak mampu memprediksi secara signifikan ($B=0.011$; $p>0.05$). Ada perbedaan kepuasan antara MANAJER dan PELAKSANA ($B=3.281$; $p<0.05$) dan seterusnya, sama membacanya dengan cara sebelumnya.

Terlihat bahwa ada perbedaan antara hasil analisis secara terpisah dengan analisis secara simultan. Meski secara umum dalam hal signifikansi tidak ada perbedaan, akan tetapi dari sisi prediksi ada perbedaan. Memang demikian adanya, yang dikarenakan ada varian-varian di data turut dilibatkan. Analisis secara simultan memiliki kelebihan dibanding secara terpisah, karena manusia adalah integral, kita tidak hanya melihat satu sisi darinya namun harus melihat sisi yang lainnya pula. Dalam hal ini kita tidak melihat KEPUASAN individu tidak hanya dari KETAHANANNYA saja akan tetapi dari JABATAN dan BIDANG KERJANYA.

Analisis regresi dengan menggunakan variabel dummy memiliki kompleksitas, hasil analisis ini memiliki kemiripan dengan analisis kovarian, namun memiliki sistematika komputasi yang sedikit berbeda.

Jena, November 2010
Wahyu Widhiarso