



Bahan Kuliah Ekonometrika I

Konsep Dasar Analisis Regresi

IMAM AWALUDDIN

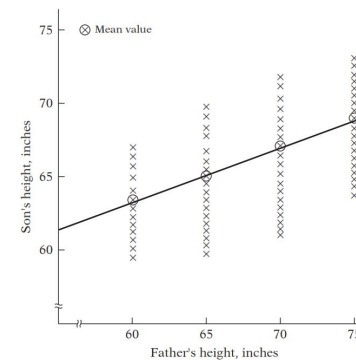
Mengenal Istilah Regresi

Istilah *regresi* pertama kali dikenalkan oleh Francis Galton. Pada salah satu tulisannya dia menyatakan bahwa “rata-rata tinggi badan anak yang lahir dari orang tua yang bertinggi badan tinggi tertentu, cenderung untuk bergerak atau ‘beregresi’ menuju rata-rata tinggi badan populasi secara keseluruhan”.

Interpretasi istilah regresi modern (saat ini) adalah:

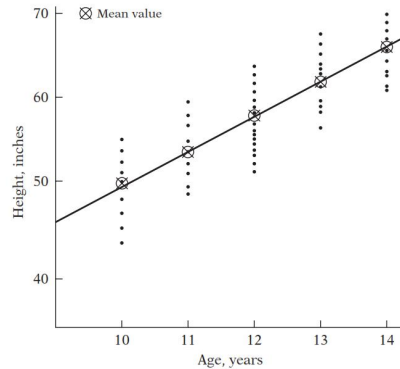
“Analisis regresi berkaitan dengan studi mengenai ketergantungan satu variabel, yaitu *variabel dependen*, terhadap satu atau lebih variabel lainnya, yaitu *variabel penjelas*, dengan tujuan untuk mengestimasi dan/atau memperkirakan nilai rerata (populasi) variabel dependen dari nilai yang diketahui atau nilai tetap dari variabel penjelas dalam sampling berulang.

Contoh 1: Tinggi Anak Laki-laki dan Tinggi Ayahnya



- ⊗ Memperkirakan tinggi anak laki-laki dari tinggi ayahnya
- ⊗ Pada setiap nilai tinggi ayah terdapat sebaran tinggi anak laki-laki
- ⊗ Secara rata-rata tinggi anak laki-laki meningkat seiring peningkatan tinggi ayah
- ⊗ Garis regresi:
 - Menghubungkan rata-rata tinggi anak dengan tinggi ayah

Contoh 2: Tinggi Anak Laki-laki dan Usianya



- Ø Menduga tinggi anak laki-laki dari umurnya
- Ø Pada setiap umur anak laki-laki terdapat sebaran tinggi
- Ø Secara rata-rata tinggi anak laki-laki meningkat dengan umur
- Ø Garis regresi:
 - Hubungan antara rata-rata tinggi dan umur

Contoh 3: Kurva Phillips



- Ø Untuk mengetahui tingkat perubahan upah berkaitan dengan tingkat pengangguran

Hubungan secara Deterministik vs Stokastik

- Ø Hubungan deterministik antar peubah apabila semua pasangan titik membentuk garis lurus
- Ø Hubungan stokastik:
 - Pasangan titik berada di sekitar (tidak tepat pada) garis
 - Adanya unsur random atau stokastik
 - Peubah random atau stokastik mempunyai sebaran peluang tertentu
- Ø Analisis regresi:
 - Adanya kemungkinan peubah lain yang tidak terukur yang juga menjelaskan peubah tak bebas (komponen stokastik)
 - Selain hubungan deterministik yang diasumsikan

Regresi vs Sebab Akibat

- Ø Hubungan sebab akibat tidak disimpulkan secara statistik
- Ø Hubungan sebab akibat harus dibentuk berdasarkan "*common sense*"
- Ø Penentuan hubungan sebab akibat secara "*a priori*"
- Ø Contoh: Berdasarkan bidang ilmu yang bersesuaian
 - Produksi dipengaruhi oleh curah hujan, bukan sebaliknya
 - Tinggi tubuh anak laki-laki dipengaruhi oleh umur, bukan sebaliknya
 - Konsumsi dipengaruhi oleh pendapatan, bukan sebaliknya

Regresi vs Korelasi

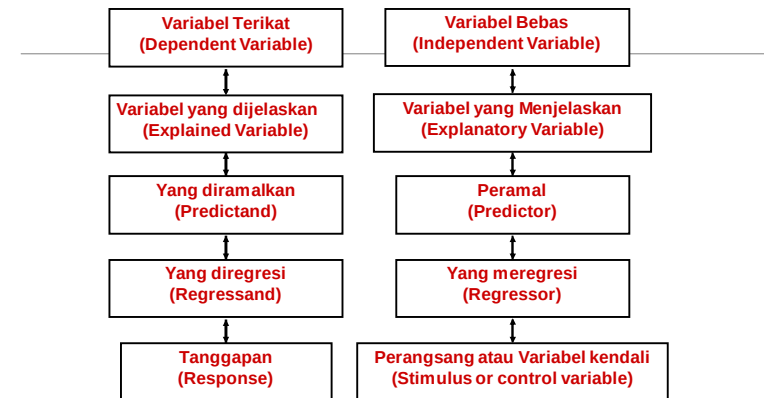
Regresi:

- Melibatkan unsur sebab akibat antara dua peubah (atau lebih)
- Peubah tak bebas dan peubah penjelas

Korelasi

- Hanya hubungan keamatan antar peubah
- Tanpa unsur sebab akibat

Istilah dan Notasi



Regresi Sederhana Vs Regresi Majemuk

} Analisis regresi sederhana (*simple regression analysis*)

- } Melibatkan dua peubah saja
- } Satu peubah tak bebas dan satu peubah penjelas

} Analisis regresi berganda/majemuk (*multiple regression analysis*)

- } Melibatkan lebih dari dua peubah
- } Satu peubah tak bebas dan beberapa peubah penjelas

Tipe Data untuk Analisis Ekonomi

Tipe data

- § *Time Series* data
- § *Cross section* data
- § *Pooled* data
- § *Panel, Longitudinal or Micropanel* data

Data Time Series (Runtun Waktu)

- Ø Sekumpulan pengamatan yang diamati pada satu/beberapa peubah pada waktu yang berbeda (Y_t)
 - § Data harian: harga saham, ramalan cuaca
 - § Data mingguan: supply uang
 - § Data bulanan: tingkat pengangguran, *Consumer Price Index* (CPI)
 - § Data tiga bulanan (quarterly): GDP (pendapatan nasional)
 - § Data tahunan: anggaran pemerintah
- Ø Hubungan stasioner mendasari analisis data *time series*
(di luar lingkup kuliah ini, akan dibahas di kuliah Ekonometrika II)

Data Cross-Section (Kerat-Lintang)

- Ø Hasil pengamatan pada satu atau beberapa peubah yang diperoleh pada satu waktu untuk beberapa individu (orang/negara/perusahaan)

Contoh:

- Produksi telur (Y_1) dan harga telur (X_1) untuk 50 negara bagian di US pada tahun 1990, $i=1, \dots, 50$
- Produksi telur (Y_2) dan harga telur (X_2) untuk 50 negara bagian di US pada tahun 1991, $i=1, \dots, 50$

TABLE 1.1 U.S. Egg Production

State	Y_1	Y_2	X_1	X_2	State	Y_1	Y_2	X_1	X_2
AL	2,206	2,186	92.7	91.4	MT	172	164	68.0	66.0
AK	0.7	0.7	151.0	149.0	NE	1,202	1,400	50.3	48.9
AZ	73	74	61.0	56.0	NV	2.2	1.8	53.9	52.7
AR	3,620	3,737	86.3	91.8	NH	43	49	109.0	104.0
CA	7,472	7,444	63.4	58.4	NJ	442	491	85.0	83.0
CO	788	873	77.8	73.0	NM	283	302	74.0	70.0
CT	1,029	948	106.0	104.0	NY	975	987	68.1	64.0
DE	168	164	117.0	113.0	NC	3,033	3,045	82.8	78.7
FL	2,586	2,537	62.0	57.2	ND	51	45	55.2	48.0
GA	4,302	4,301	80.6	80.8	OH	4,667	4,637	59.1	54.7
HI	227.5	224.5	85.0	85.5	OK	869	830	101.0	100.0
ID	187	203	79.1	72.9	OR	652	686	77.0	74.6
IL	793	809	65.0	70.5	PA	4,976	5,130	61.0	52.0
IN	5,445	5,290	62.7	60.1	RI	53	50	102.0	99.0
IA	2,151	2,247	56.5	53.0	SC	1,422	1,420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48.0	45.8
KY	412	483	67.7	73.5	TN	277	279	71.0	80.7
LA	273	254	115.0	115.0	TX	3,317	3,356	76.7	72.6
ME	1,069	1,070	101.0	97.0	UT	456	486	64.0	59.0
MD	885	898	76.6	75.4	VT	31	30	106.0	102.0
MA	235	237	105.0	102.0	VA	943	988	86.3	81.2
MI	1,406	1,396	58.0	53.8	WA	1,287	1,313	74.1	71.5
MN	2,499	2,697	57.7	54.0	WV	136	174	104.0	109.0
MS	1,434	1,468	87.8	86.7	WI	910	873	60.1	54.0
MO	1,580	1,622	55.4	51.5	WY	1.7	1.7	83.0	83.0

Note: Y_1 = eggs produced in 1990 (millions).
 Y_2 = eggs produced in 1991 (millions).
 X_1 = price per dozen (cents) in 1990.
 X_2 = price per dozen (cents) in 1991.

Jika kita hanya memperhatikan satu tahun saja (variabel Y_1 dan X_1) atau (Y_2 dan X_2), maka datanya adalah data cross-section

Jika keduanya digabung (tahun 1990 dan 1991), maka menjadi data pool

Data Pool (Gabungan)

- Ø Gabungan dari *time series* dan *cross section* data
- Ø Contoh: Produksi telur (Y_i) dan harga telur (X_i) untuk 50 negara bagian di US pada tahun 1990 dan tahun 1991
- Ø Perbedaan waktu tidak dipentingkan, dianggap sebagai ulangan
- Ø Total pengamatan $n = 50+50 = 100$
- Ø $i = 1, 2, \dots, 100$

Data Panel, Data Longitudinal, atau Data Mikropanel

Data yang diukur pada unit cross section yang sama dalam selang waktu tertentu

Unit cross section:

- Negara Bagian

Pengamatan

- Produksi telur
- Harga telur

Selang waktu: 1990 dan 1991

- Di setiap negara bagian diamati produksi telur dan harga telur pada dua periode waktu tersebut

Y_{it} : Produksi telur pada negara bagian i pada tahun t

X_{it} : Harga telur pada negara bagian i pada tahun t

$i = 1, \dots, 50, t = 1990, 1991$

Contoh Data Panel

TABLE 1.2 Investment Data for Four Companies, 1935-1954

Observation	I	F_{-1}	C_{-1}	Observation	I	F_{-1}	C_{-1}	Observation	I	F_{-1}	C_{-1}	Observation	I	F_{-1}	C_{-1}
GE				US				GM				WEST			
1935	33.1	1170.6	97.8	1935	209.9	1362.4	53.8	1935	317.6	3078.5	2.8	1935	12.93	191.5	1.8
1936	45.0	2015.8	104.4	1936	355.3	1807.1	50.5	1936	391.8	4661.7	52.6	1936	25.90	516.0	0.8
1937	77.2	2803.3	118.0	1937	469.9	2673.3	118.1	1937	410.6	5387.1	156.9	1937	35.05	729.0	7.4
1938	44.6	2039.7	156.2	1938	262.3	1801.9	250.2	1938	257.7	2792.2	209.2	1938	22.89	560.4	18.1
1939	48.1	2256.2	172.6	1939	230.4	1957.3	312.7	1939	330.8	4313.2	203.4	1939	18.84	519.9	23.5
1940	74.4	2132.2	186.6	1940	361.6	2202.9	254.2	1940	461.2	4643.9	207.2	1940	28.57	628.5	26.5
1941	113.0	1834.1	220.9	1941	472.8	2380.5	261.4	1941	512.0	4551.2	255.2	1941	48.51	537.1	36.2
1942	91.9	1588.0	287.8	1942	445.6	2168.6	298.7	1942	448.0	3244.1	303.7	1942	43.34	561.2	60.8
1943	61.3	1749.4	319.9	1943	361.6	1985.1	301.8	1943	499.6	4053.7	264.1	1943	37.02	617.2	84.4
1944	56.8	1687.2	321.3	1944	288.2	1813.9	279.1	1944	547.5	4379.3	201.6	1944	37.81	626.7	91.2
1945	93.6	2007.7	319.6	1945	258.7	1850.2	213.8	1945	561.2	4840.9	265.0	1945	39.27	737.2	92.4
1946	159.9	2208.3	346.0	1946	420.3	2067.7	232.6	1946	688.1	4900.0	402.2	1946	53.46	760.5	86.0
1947	147.2	1656.7	456.4	1947	420.5	1796.7	264.8	1947	568.9	3526.5	761.5	1947	55.56	581.4	111.1
1948	146.3	1604.4	543.4	1948	494.5	1625.8	306.9	1948	529.2	3245.7	922.4	1948	49.56	662.3	130.6
1949	98.3	1431.8	618.3	1949	405.1	1667.0	351.1	1949	555.1	3700.2	1020.1	1949	32.04	583.8	141.8
1950	93.5	1610.5	647.4	1950	418.8	1677.4	357.8	1950	642.9	3755.6	1099.0	1950	32.24	635.2	136.7
1951	135.2	1819.4	671.3	1951	588.2	2289.5	341.1	1951	755.9	4833.0	1207.7	1951	54.38	732.8	129.7
1952	157.3	2079.7	726.1	1952	645.2	2159.4	444.2	1952	891.2	4924.9	1430.5	1952	71.78	864.1	145.5
1953	179.5	2371.6	800.3	1953	641.0	2031.3	623.6	1953	1304.4	6241.7	1777.3	1953	90.08	1193.5	174.8
1954	189.6	2759.9	888.9	1954	459.3	2115.5	669.7	1954	1486.7	5593.6	2226.3	1954	68.60	1188.9	213.5

Notes: $Y = I$ = gross investment = additions to plant and equipment plus maintenance and repairs, in millions of dollars deflated by P_1 .
 $X_1 = F$ = value of the firm = price of common and preferred shares at Dec. 31 (or average price of Dec. 31 and Jan. 31 of the following year) times number of common and preferred shares outstanding plus total book value of debt at Dec. 31, in millions of dollars deflated by P_2 .
 $X_2 = C$ = cost of plant and equipment = accumulated sum of net additions to plant and equipment deflated by P_1 minus depreciation allowance deflated by P_1 in these definitions.
 P_1 = implicit price deflator of producers' durable equipment (1947 = 100).
 P_2 = implicit price deflator of GNP (1947 = 100).
 P_3 = depreciation expense deflator = 10-year moving average of wholesale price index of metals and metal products (1947 = 100).
Source: Reproduced from H. D. Vinod and Aman Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, New York, 1981, pp. 259-261.

Contoh Lain Panel Data

Sensus n rumah tangga untuk 5 tahun

Unit cross section:

- Rumah tangga

Setiap tahun, rumah tangga yang sama diwawancara, tentang jumlah pendapatan (Y)

Y_{it} : Pendapatan rumah tangga i pada tahun t

- $i = 1, \dots, n, t = 1, \dots, 5$

Tujuan:

- Untuk mempelajari perubahan secara finansial rumah tangga tsb sejak wawancara terakhir
- Mempelajari dinamika keuangan rumah tangga secara periodik

Sumber Data Ekonomi

Lembaga pemerintahan

- BPS
- SuSeNas: Sensus Ekonomi Pemerintah
- Bank Indonesia

Lembaga Internasional

- IMF
- Bank Dunia

Organisasi swasta lainnya

Internet

Metode Pengumpulan Data

Data yang dikumpulkan dari berbagai sumber dapat bersifat eksperimental maupun non eksperimental.

Data yang bersifat eksperimental sering dikumpulkan di bidang ilmu alam, peneliti ingin mengumpulkan data dengan menganggap beberapa factor adalah tetap, untuk menguji dampak dari beberapa factor atas fenomena tertentu.

Dalam ilmu social, data yang biasanya dijumpai adalah data-data yang secara alamiah bersifat non eksperimental, dalam arti data tersebut tidak dapat dikontrol atau dikendalikan oleh peneliti.

Sumber Bacaan

Gujarati, Damodar N. and Dawn C. Porter (2009). *Basic Econometrics*, Fifth Edition. McGraw-Hill/Irwin. New York. Chapter 1.