

# Dasar-dasar Analisa Regresi

Tjipto Juwono, Ph.D.

May 2018



SURYA  
UNIVERSITY

# Sejarah Analisa Regresi

Istilah *regresi* pertama kali diperkenalkan oleh Francis Galton dalam salah satu papernya. Galton menemukan bahwa walaupun ada kecenderungan bahwa anak dari orang tua yang jangkung juga jangkung, dan anak dari orang tua yang pendek juga pendek, namun *tinggi rata – rata* dari anak yang lahir dari orang tua dengan tinggi tertentu akan *bergerak ke-arah (regress)* tinggi rata-rata dari populasi <sup>1</sup>.

<sup>1</sup>Francis Galton, Proceedings of Royal Society, London, vol. 40, 1886, pp. 4272.

# Sejarah Analisa Regresi

Temuan Galton dikonfirmasi oleh Karl Pearson. Ia mengumpulkan data tinggi badan dari ribuan keluarga <sup>2</sup>. Ia menemukan bahwa tinggi rata-rata anak-anak dari kelompok orang tua yang jangkung lebih pendek dari tinggi orang tua mereka. Sementara tinggi rata-rata dari anak-anak dari orang tua yang pendek lebih tinggi dari orang tua mereka. Jadi tinggi rata-rata dari anak-anak itu *bergerak ke arah (regress)* tinggi rata-rata dari populasi.

<sup>2</sup>K. Pearson and A. Lee, *Biometrika*, vol. 2, Nov. 1903, pp. 357462.

# Pengertian Analisa Regresi

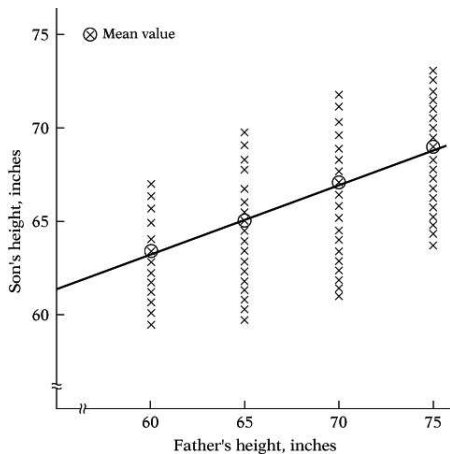
## Analisa Regresi:

Mempelajari ketergantungan satu variabel (yaitu variabel *dependent*) terhadap satu atau lebih variabel lainnya (yaitu variabel *independence* atau *explanatory*). Ketergantungan ini dipelajari dengan mengestimasi dan/atau memprediksi nilai mean/average dari populasi pada variabel dependent sebagai fungsi dari variabel(-variabel) independent yang diketahui, atau di-tetapkan (dalam hal sampling yang berulang).

# Contoh Analisa Regresi

- 1 Dalam riset yang dilakukan oleh Galton (dan Pearson), ia ingin mengetahui bagaimana tinggi rata-rata anak berubah, jika tinggi orang tua (ayah) diberikan.

# Contoh Analisa Regresi

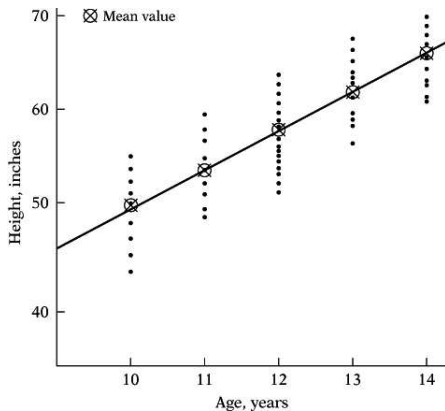


Gbr. 1: Hubungan antara tinggi anak dengan tinggi ayahnya

# Contoh Analisa Regresi

- 2 Misalkan kita ingin mempelajari hubungan antara tinggi seorang anak dengan usianya.

# Contoh Analisa Regresi



Gbr. 2: Hubungan antara tinggi anak dengan usianya



## Contoh Analisa Regresi

- 3. Mempelajari hubungan antara PCE (Personal Consumption Expenditure) dengan Income. Analisa ini berguna untuk mengestimasi MPC.
- 4. Monopolist dapat menetapkan harga atau output (tetapi tidak kedua-duanya). Untuk menentukan harga yang paling profitable maka perlu memperoleh bagaimana response dari demand terhadap perubahan harga.

# Variabel Random vs Variabel Tetap

**Variabel Random:** Adalah variabel yang mempunyai distribusi probabilitas. Variabel Random disebut juga *variabel stokastik*.

**Variabel Tetap:** Tidak mempunyai distribusi probabilitas.

# Hubungan Statistik vs Hubungan Deterministik

**Hubungan deterministik:** Dalam hubungan deterministik, semua variabel (baik itu dependen maupun independen) adalah variabel non-stokastik

**Hubungan statistik:** Dalam hubungan statistik, salah satu atau lebih (atau semua) variabelnya adalah variabel stokastik.

# Hubungan Statistik vs Hubungan Deterministik

**Contoh hubungan deterministik:** Jika hambatan suatu kabel diketahui, maka untuk suatu besar tegangan antara kedua ujung kabel, hanya ada satu angka arus listrik. Hubungan antara tegangan dan arus listrik adalah hubungan yang deterministik.

**Contoh hubungan statistik:** Untuk setiap tinggi orang tua tertentu, ada banyak kemungkinan tinggi anaknya. Tinggi anak untuk satu tinggi orang tua tertentu, merupakan suatu distribusi (Lihat Gbr. (1)). Variabel tinggi anak tersebut disebut variabel **random** atau **stokastik**.

# Hubungan Statistik vs Hubungan Deterministik

## Hubungan di Dalam Ekonometrika

Di dalam Ekonometrika, hubungan yang kita pelajari adalah hubungan statistik.

## Hubungan Statistik vs Hubungan Deterministik

Perhatikan kembali Gbr. (1). Untuk setiap tinggi orang tua yang tertentu, ada banyak kemungkinan tinggi anak yang membentuk suatu distribusi. Tinggi anak merupakan variabel random atau stokastik. Ingat definisi variabel random, yaitu: *variabel yang mempunyai distribusi probabilitas*. Dalam pembahasan kita, *variabel non-random (atau non-stokastik) disebut variabel tetap atau variabel fix (fixed variables)*.

## Regresi vs Hubungan Sebab-Akibat

### Regresi tidak berarti kausalitas

Misalkan kita menganalisa hubungan antara curah hujan dengan hasil panen. Secara statistik tidak ada alasan untuk mengatakan bahwa hasil panen ditentukan oleh curah hujan. Jika kita mengatakan yang sebaliknya, yaitu bahwa curah hujan ditentukan oleh hasil panen, maka itu sah-sah saja secara statistik. Jika kita mengatakan bahwa hasil panen ditentukan oleh curah hujan (dan bukan sebaliknya) maka itu berdasarkan pertimbangan non-statistik. (Misalnya: Kita tahu bahwa kita tidak dapat mengendalikan curah hujan dengan cara mengubah-ubah hasil panen).

# Regresi vs Hubungan Sebab-Akibat

## Regresi tidak berarti kausalitas

Hubungan statistik di dalam dirinya **tidak dapat** secara logis menunjukkan hubungan sebab-akibat. Untuk mengetahui hubungan sebab-akibat, seseorang harus mencarinya dari sumber di luar analisa statistik.



## Regresi vs Hubungan Sebab-Akibat

Dalam hal hubungan antara tinggi orang tua dengan tinggi anak, kita bisa saja membuat analisa di mana kita menetapkan tinggi anak sebagai variabel independen, dan menjadikan tinggi orang tua sebagai variabel dependen yang merupakan variabel stokastik. Pengetahuan bahwa tinggi anak dikendalikan oleh tinggi orang tua *bukan* berasal dari analisa statistik melainkan dari sumber lain (umpamanya biologi atau genetika).

## Apa perbedaan antara Korelasi dan Regresi?

**Korelasi:** Dalam korelasi, kita ingin mengetahui seberapa kuat hubungan linear antara dua variabel.

**Regresi:** Dalam regresi, kita ingin mengestimasi atau memprediksi nilai rata-rata dari suatu variabel (variabel dependen) berdasarkan nilai yang ditetapkan dari variabel yang lain (variabel independen).

## Apa perbedaan antara Korelasi dan Regresi?

- Korelasi:** Hubungan antara kedua variabel bersifat simetris, kita tidak membedakan antara kedua variabel. Selain itu kedua-duanya diperlakukan sebagai variabel random (stokastik)
- Regresi:** Hubungan antara kedua variabel bersifat a-simetris. Kita membedakan antara variabel independen dan variabel dependen. Selain itu, hanya variabel dependen yang diperlakukan sebagai variabel random (stokastik). Variabel independen diperlakukan sebagai variabel yang tetap (fixed, atau non-stokastik).

## Dependent Variable

- Explained variable
- Predictand
- Regressand
- Response
- Endogeneous
- Outcome
- Controlled variable

## Independent variable

- Explanatory Variable
- Predictor
- Regressor
- Stimulus
- Exogeneous
- Covariate
- Control variable

# Jenis-jenis Data Dalam Ekonometrik

Ada tiga macam data dalam ekonometrik

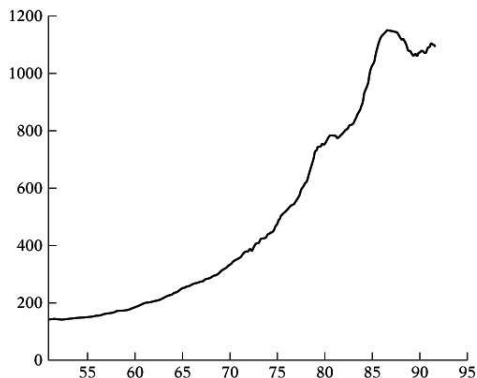
- ① Time Series
- ② Cross Section
- ③ Pooled Data (Kombinasi dari dua jenis data di atas).

# Jenis-jenis Data Dalam Ekonometrik

## ① Time Series

Satu set hasil observasi terhadap nilai-nilai dari suatu variabel, yang diperoleh pada berbagai waktu yang berbeda. Time series biasanya diambil pada interval waktu yang teratur, misalnya: harian, mingguan, bulanan, dst.

# Jenis-jenis Data Dalam Ekonometrik



Gbr. 3: Contoh data time series. Money supply, tahun 1950 - 1993

## ② Cross Section

Data cross-section adalah data dari satu atau lebih variabel yang diperoleh pada satu waktu tertentu. Dengan kata lain, semua data itu diperoleh pada waktu yang sama.



# Jenis-jenis Data Dalam Ekonometrik

- ③ **Pooled Data (Kombinasi dari dua jenis data di atas).**  
Di dalam data jenis ini, baik time series maupun cross section ada di dalamnya.
  - **Panel Data:** adalah contoh dari Pooled Data di mana unit cross-section yang sama (misalnya sebuah keluarga atau suatu perusahaan) diobservasi dalam satu jangka waktu tertentu.

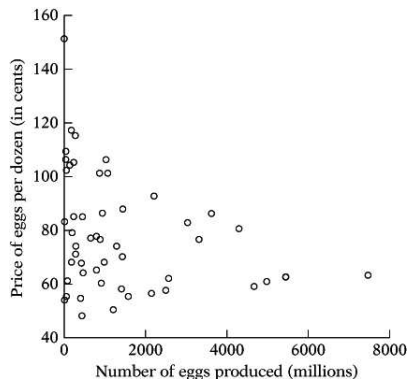
# Jenis-jenis Data Dalam Ekonometrik

| 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.

Gbr. 4: Pooled Data. Produksi Telur di Amerika tahun 1990 dan 1991

# Jenis-jenis Data Dalam Ekonometrik



Gbr. 5: Data Cross Section. Harga Telur vs Produksi Telur di Amerika tahun 1990

# Akurasi Data

Seringkali kualitas data-data yang diperoleh kurang bagus, alasannya antara lain:

- Kebanyakan data dalam ilmu sosial adalah noneksperimental. Karena itu, dapat terjadi error dalam observasi.
- Pada data-data yang diperoleh dari eksperimen, error dapat terjadi pada pengukuran, akibat pembulatan dan pendekatan.
- Pada data yang diperoleh dari questionnaire, dapat terjadi problem dari kurangnya response. Jika beruntung, seorang peneliti dapat memperoleh 40% response.
- Metode-metode sampling yang digunakan untuk memperoleh data bisa saja bervariasi dari satu sample ke sample lainnya, sehingga sulit untuk membandingkan satu sample dengan sample lainnya.
- Data ekonomi seringkali sangat teragregat, terutama dalam data-data makro.

Seorang peneliti harus selalu ingat bahwa kualitas hasil riset sangat bergantung pada kualitas data-nya.

## Jenis-jenis Variabel

- Skala Rasio:** Ada urutan, jarak antar satu variabel dengan variabel lain mempunyai makna, begitu pula rasio antar satu variabel dengan variabel lain.
- Skala Interval:** Ada urutan, jarak antar satu variabel dengan variabel lain penting, tetapi rasio tidak berarti.
- Skala Ordinal:** Ada urutan, tetapi jarak maupun rasio tidak bermakna. Memberi label pada kategori, di mana ada urutan pada kategori-kategori. Kategori yang satu lebih tinggi dari yang lain.
- Skala Nominal:** Hanya digunakan memberi label pada kategori. Tidak ada urutan, jarak, atau rasio.

# Tugas Kelas

## Tugas Kelas

- 1 Berikan contoh untuk skala rasio, skala interval, skala ordinal, skala nominal.
- 2 Dapatkah kita melakukan analisa regresi jika variabel independen-nya berupa variabel dengan skala nominal? Mengapa?
- 3 Berikan contoh dari kehidupan nyata data yang berupa data cross-section. Jelaskan!
- 4 Berikan contoh dari kehidupan nyata data yang berupa data time series. Jelaskan!
- 5 Berikan contoh variabel stokastik. Jelaskan!
- 6 Berikan contoh variabel tetap/non-stokastik. Jelaskan!
- 7 Berikan contoh analisa korelasi, berikan pula contoh analisa regresi. Jelaskan apa perbedaan antara kedua analisa itu.