

## REGRESI LINIER

OLEH: JONATHAN SARWONO

### 1.1 Pengertian

Apa yang dimaksud dengan regresi linier? Istilah regresi pertama kali dalam konsep statistik digunakan oleh Sir Francis Galton dimana yang bersangkutan melakukan kajian yang menunjukkan bahwa tinggi badan anak-anak yang dilahirkan dari para orang tua yang tinggi cenderung bergerak (*regress*) kearah ketinggian rata-rata populasi secara keseluruhan. Galton memperkenalkan kata regresi (*regression*) sebagai nama proses umum untuk memprediksi satu variabel, yaitu tinggi badan anak dengan menggunakan variabel lain, yaitu tinggi badan orang tua. Pada perkembangan berikutnya hukum Galton mengenai regresi ini ditegaskan lagi oleh Karl Pearson dengan menggunakan data lebih dari seribu. Pada perkembangan berikutnya, para ahli statistik menambahkan istilah regresi berganda (*multiple regression*) untuk menggambarkan proses dimana beberapa variabel digunakan untuk memprediksi satu variabel lainnya.

Regresi dalam pengertian moderen menurut Gujarati (2009) ialah sebagai kajian terhadap ketergantungan satu variabel, yaitu variabel tergantung terhadap satu atau lebih variabel lainnya atau yang disebut sebagai variabel – variabel eksplanatori dengan tujuan untuk membuat estimasi dan / atau memprediksi rata – rata populasi atau nilai rata-rata variabel tergantung dalam kaitannya dengan nilai – nilai yang sudah diketahui dari variabel eksplanasinya. Selanjutnya menurut Gujarati meski analisis regresi berkaitan dengan ketergantungan atau dependensi satu variabel terhadap variabel – variabel lainnya hal tersebut tidak harus menyiratkan sebab – akibat (*causation*). Dalam mendukung pendapatnya ini, Gujarati mengutip pendapat Kendal dan Stuart yang diambil dari buku mereka yang berjudul “*The Advanced Statistics*” yang terbit pada tahun 1961 yang mengatakan bahwa, ” suatu hubungan statistik betapapun kuat dan sugestifnya tidak akan pernah dapat menetapkan hubungan sebab akibat (*causal connection*); sedang gagasan mengenai sebab akibat harus datang dari luar statistik, yaitu dapat berasal dari teori atau lainnya”.

Sedang menurut Levin & Rubin (1998:648), regresi digunakan untuk menentukan sifat – sifat dan kekuatan hubungan antara dua variabel serta memprediksi nilai dari suatu variabel yang belum diketahui dengan didasarkan pada observasi masa lalu terhadap variabel tersebut dan variabel-variabel lainnya. Selanjutnya dalam regresi kita akan mengembangkan persamaan estimasi (*estimating equation*), yaitu rumus matematika yang menghubungkan variabel-variabel yang diketahui dengan variabel-variabel yang tidak diketahui. Setelah dipelajari pola hubungannya, kemudian kita dapat mengaplikasikan analisis korelasi (*correlation analysis*) untuk menentukan tingkatan dimana variabel – variabel tersebut berhubungan. Kesimpulannya, analisis korelasi mengungkapkan seberapa benar persamaan estimasi sebenarnya menggambarkan hubungan tersebut. Lebih lanjut Levin & Rubin mengatakan bahwa: “ Kita sering menemukan hubungan sebab akibat antar variabel – variabel; yaitu variabel bebas ‘menyebabkan’ variabel tergantung berubah. Sekalipun demikian mereka melanjutkan bahwa: “penting untuk kita perhatikan bahwa yang kita anggap hubungan (*relationship*) yang diketemukan melalui regresi sebagai hubungan asosiasi (*relationship of association*) tetapi tidak selalu harus sebab dan akibat (*cause and effect*). Kecuali kita mempunyai alasan – alasan khusus untuk percaya bahwa (perubahan pada) nilai – nilai variabel tergantung disebabkan oleh nilai –

nilai variabel (variabel) bebas; jangan menyimpulkan (*infer*) hubungan sebab akibat dari hubungan yang diketemukan dalam regresi.

Karena Levin & Rubin dalam mendefinisikan regresi juga menggunakan istilah “analisis korelasi”, maka sebaiknya dalam bagian ini penulis perlu menjelaskan perbedaan antara regresi dan korelasi. Menurut Gujarati (2009: 20) analisis korelasi bertujuan untuk mengukur kekuatan (*strength*) atau tingkatan (*degree*) hubungan linier (*linear association*) antara dua variabel. Untuk mengukur kekuatan hubungan linier ini digunakan koefisien korelasi. Sebaliknya dalam regresi kita tidak melakukan pengukuran seperti itu. Dalam regresi kita membuat estimasi atau memprediksi nilai rata-rata satu variabel didasarkan pada nilai – nilai tetap variabel – variabel lain. Perbedaan yang mendasar antara regresi dan korelasi ialah dalam regresi terdapat (hubungan) asimetri dalam kaitannya dengan perlakuan terhadap variabel tergantung dan variabel bebas. Variabel tergantung diasumsikan statistikal, acak atau stokhastik, yaitu mempunyai distribusi probabilitas. Sedang variabel bebas / prediktornya diasumsikan mempunyai nilai – nilai tetap. Sebaliknya dalam korelasi kita memperlakukan dua variabel atau variabel – variabel apa saja secara simetris, yaitu tidak ada perbedaan antara variabel bebas dan variabel tergantung. Sebagai contoh korelasi antara nilai ujian matematik dan statistik sama dengan korelasi nilai ujian statistik dan matematik. Lebih lanjut dalam korelasi kedua variabel diasumsikan random.

Regresi linier mempunyai persamaan yang disebut sebagai persamaan regresi. Persamaan regresi mengekspresikan hubungan linier antara variabel tergantung / variabel kriteria yang diberi simbol Y dan salah satu atau lebih variabel bebas / prediktor yang diberi simbol X jika hanya ada satu prediktor dan  $X_1, X_2$  sampai dengan  $X_k$ , jika terdapat lebih dari satu prediktor (Crammer & Howitt, 2006:139). Persamaan regresi akan terlihat seperti di bawah ini:

- Untuk persamaan regresi dimana Y merupakan nilai yang diprediksi, maka persamaannya ialah:

$$Y = a + \beta_1 X_1 \text{ (untuk regresi linier sederhana)}$$

$$Y = a + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \text{ (untuk regresi linier berganda)}$$

- Untuk persamaan regresi dimana Y merupakan nilai sebenarnya (observasi), maka persamaan menyertakan kesalahan (*error term / residual*) akan menjadi:

$$Y = a + \beta_1 X_1 + e \text{ (untuk regresi linier sederhana)}$$

$$Y = a + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e \text{ (untuk regresi linier berganda)}$$

Dimana:

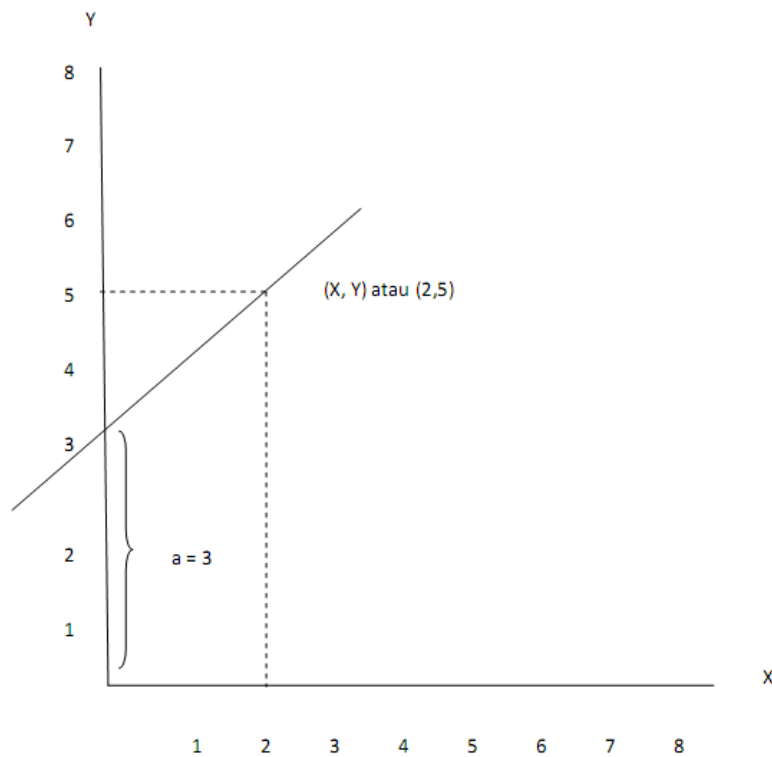
- X: merupakan nilai sebenarnya suatu kasus (data)
- $\beta$ : merupakan koefisien regresi jika hanya ada satu prediktor dan koefisien regresi parsial jika terdapat lebih dari satu prediktor. Nilai ini juga mewakili mewakili koefisien regresi baku (*standardized*) dan koefisien regresi tidak baku (*unstandardized*). Koefisien regresi ini merupakan jumlah perubahan yang terjadi pada Y yang disebabkan oleh perubahan nilai X. Untuk menghitung perubahan ini dapat dilakukan dengan cara mengkalikan nilai prediktor sebenarnya (observasi) untuk kasus (data) tertentu dengan koefisien regresi prediktor tersebut.
- a: merupakan intercept yang merupakan nilai Y saat nilai prediktor sebesar nol.

Sedang garis regresi didefinisikan sebagai garis lurus yang ditarik dari titik – titik diagram pencar (*scattered diagram*) dari nilai variabel tergantung dan variabel bebas sehingga garis tersebut menggambarkan hubungan linier antara variabel-variabel tersebut. Jika nilai-nilai ini merupakan garis regresi nilai baku maka garis ini sama dengan garis korelasi. Garis ini disebut juga sebagai garis kecocokan yang sempurna dimana garis lurus tersebut berada pada posisi terdekat pada titik-titik diagram pencar. Garis ini dapat digambarkan dari nilai-nilai persamaan regresi dalam bentuk yang paling sederhana yaitu:

Nilai yang diprediksi = intercept + (koefesien regresi x nilai prediktor)

Sumbu vertikal dari diagram pencar digunakan untuk menggambarkan nilai-nilai variabel tergantung sedang sumbu horizontal menggambarkan nilai prediktor. Intercept merupakan titik sumbu vertikal yang merupakan nilai variabel tergantung yang diprediksi saat nilai prediktor atau variabel bebas sebesar nol. Nilai yang diprediksi akan sebesar akan sebesar 0 jika koefesien regresi baku digunakan. Itulah sebabnya saat menggunakan IBM SPSS keluaran yang digunakan dalam koefesien regresi menggunakan keluaran pada kolom “*unstandardized coefficient*”.

Jika digambarkan akan nampak seperti di bawah ini:



**Gambar 1.1 Garis Regresi**

Persamaannya ialah

$$Y = a + \beta_1 X_1$$

Dengan:

Y= variabel tergantung / variabel kriteria

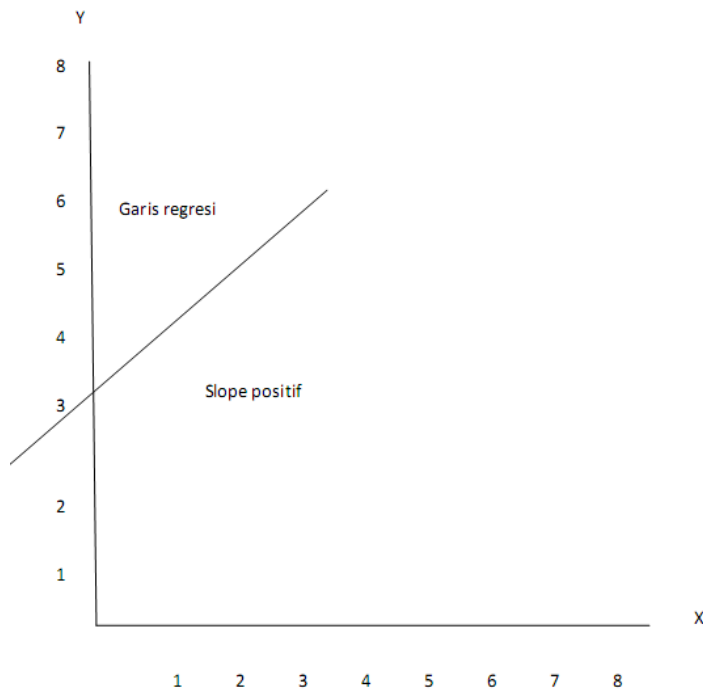
a= intercept Y

$\beta$  = kemiringan (slope)

X= variabel bebas

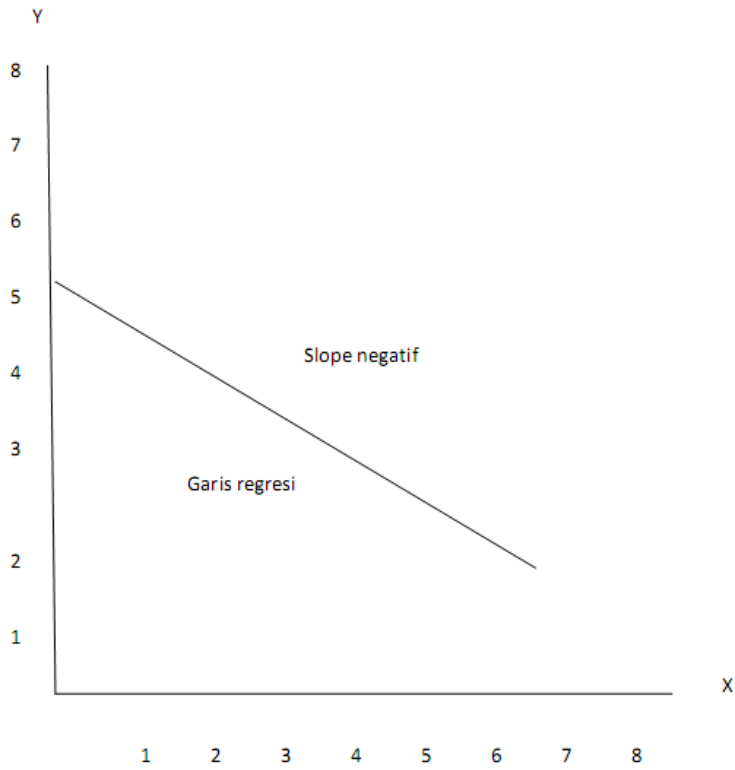
Garis regresi mempunyai 3 (tiga) kemungkinan yaitu: 1) hubungan linier positif, 2) hubungan linier negatif, dan 3) tidak ada hubungan linier. Gambarnya seperti di bawah ini:

### 1) Hubungan Linier Positif



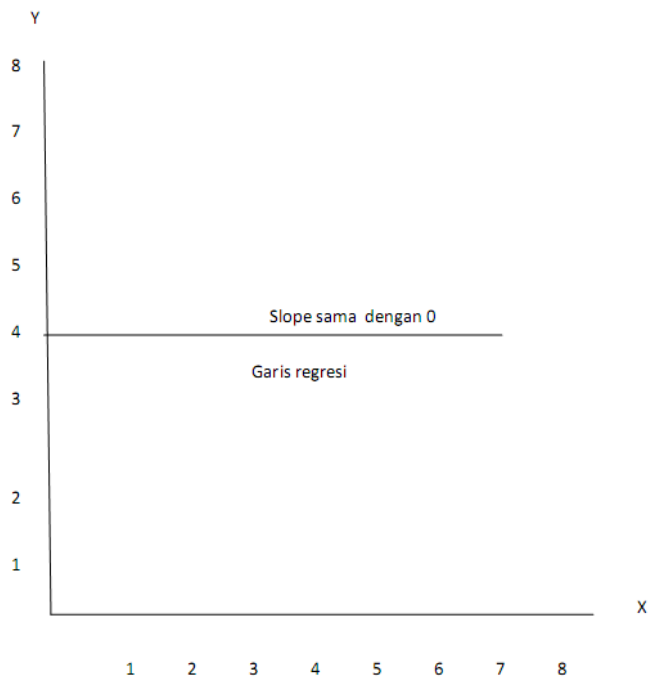
**Gambar 1.2 Hubungan Linier Positif**

### 2) Hubungan Linier Negatif



**Gambar 1.3 Hubungan Linier Negatif**

3) Tidak Ada Hubungan Linier



## Gambar 1.4 Tidak Ada Hubungan Linier

Agar kita memperoleh kejelasan dalam penggunaan istilah, maka di bawah ini diberikan istilah-istilah yang mewakili pengertian variabel bebas dan variabel tergantung dalam regresi. Gujarati memberikan istilah sebagai berikut:

- Variabel tergantung (*dependent variable*): disebut juga sebagai variabel yang dijelaskan (*explained variable*) / variabel yang diprediksi (*predictand*) / regresan (*regressand*) / variabel yang merespon (*response*) / *endogenous* / keluaran (*outcome*) / variabel yang dikontrol (*controlled variable*).
- Variabel yang menerangkan (*explanatory variable*): disebut juga sebagai variabel tergantung (*dependent variable*) / variabel yang memprediksi (*predictor*) / regresor (*regressor*) / variabel stimulus (*stimulus*) / *exogenous* / kovariat (*covariate*) / variabel kontrol (*control variable*).

### 1.2 Tujuan

Tujuan menggunakan analisis regresi ialah

- Membuat estimasi rata-rata dan nilai variabel tergantung dengan didasarkan pada nilai variabel bebas.
- Menguji hipotesis karakteristik dependensi
- Untuk meramalkan nilai rata-rata variabel bebas dengan didasarkan pada nilai variabel bebas diluar jangkauan sampel.

### 1.3 Asumsi Penggunaan Regresi

Penggunaan regresi linear sederhana didasarkan pada asumsi diantaranya sbb:

- Model regresi harus linier dalam parameter
- Variabel bebas tidak berkorelasi dengan *disturbance term* (Error) .
- Nilai *disturbance term* sebesar 0 atau dengan simbol sebagai berikut:  $(E(U / X) = 0$
- Varian untuk masing-masing *error term* (kesalahan) konstan
- Tidak terjadi otokorelasi
- Model regresi dispesifikasi secara benar. Tidak terdapat bias spesifikasi dalam model yang digunakan dalam analisis empiris.
- Jika variabel bebas lebih dari satu, maka antara variabel bebas (*explanatory*) tidak ada hubungan linier yang nyata

### 1.4 Syarat-Syarat

Model kelayakan regresi linear dalam IBM SPSS didasarkan pada hal-hal sebagai berikut:

- a. Model regresi dikatakan layak jika angka signifikansi pada ANOVA sebesar  $< 0.05$
- b. Predictor yang digunakan sebagai variabel bebas harus layak. Kelayakan ini diketahui jika angka *Standard Error of Estimate*  $< Standard Deviation$

- c. Koefesien regresi harus signifikan. Pengujian dilakukan dengan uji t. Koefesien regresi signifikan jika  $t \text{ hitung} > t \text{ table}$  (nilai kritis). Dalam IBM SPSS dapat diganti dengan menggunakan nilai signifikansi (sig) dengan ketentuan sebagai berikut:
  - Jika  $\text{sig} < 0,05$ ; koefesien regresi signifikan
  - Jika  $\text{sig} > 0,05$ ; koefesien regresi tidak signifikan
- d. Tidak boleh terjadi multikolinieritas, artinya tidak boleh terjadi korelasi antar variabel bebas yang sangat tinggi atau terlalu rendah. Syarat ini hanya berlaku untuk regresi linier berganda dengan variabel bebas lebih dari satu. Terjadi multikolinieritas jika koefesien korelasi antara variable bebas  $> 0,7$  atau  $< - 7$
- e. Tidak terjadi otokorelasi jika:  $- 2 \leq DW \leq 2$
- f. Keselerasan model regresi dapat diterangkan dengan menggunakan nilai  $r^2$  semakin besar nilai tersebut maka model semakin baik. Jika nilai mendekati 1 maka model regresi semakin baik. Nilai  $r^2$  mempunyai karakteristik diantaranya: 1) selalu positif, 2) Nilai  $r^2$  maksimal sebesar 1. Jika Nilai  $r^2$  sebesar 1 akan mempunyai arti kesesuaian yang sempurna. Maksudnya seluruh variasi dalam variabel tergantung (variabel Y) dapat diterangkan oleh model regresi. Sebaliknya jika  $r^2$  sama dengan 0, maka tidak ada hubungan linier antara variabel bebas (variabel X) dan variabel tergantung (variabel Y).
- g. Terdapat hubungan linier antara variabel bebas (X) dan variabel tergantung (Y)
- h. Data harus berdistribusi normal
- i. Data berskala interval atau rasio
- j. Terdapat hubungan dependensi, artinya satu variabel merupakan variabel tergantung yang tergantung pada variabel (variabel) lainnya.

### 1.5 Konsep Linieritas Dalam Regresi

Ada dua macam linieritas dalam analisis regresi, yaitu linieritas dalam variabel dan linieritas dalam parameter. Yang pertama, linier dalam variabel merupakan nilai rata-rata kondisional variabel tergantung yang merupakan fungsi linier dari variabel (variabel) bebas. Sedang yang kedua, linier dalam parameter merupakan fungsi linier parameter dan dapat tidak linier dalam variabel.

### 1.6 Uji Hipotesis

Pengujian hipotesis dapat didasarkan dengan menggunakan dua hal, yaitu: tingkat signifikansi atau probabilitas ( $\alpha$ ) dan tingkat kepercayaan atau *confidence level*. Didasarkan tingkat signifikansi pada umumnya orang menggunakan 0,05. Kisaran tingkat signifikansi mulai dari 0,01 sampai dengan 0,1. Yang dimaksud dengan tingkat signifikansi adalah probabilitas melakukan kesalahan tipe I, yaitu kesalahan menolak hipotesis ketika hipotesis tersebut benar. Tingkat kepercayaan pada umumnya ialah sebesar 95%, yang dimaksud dengan tingkat kepercayaan ialah tingkat dimana sebesar 95% nilai sample akan mewakili nilai populasi dimana sample berasal. IBM SPSS menggunakan istilah interval kepercayaan (*confidence interval*). Interval kepercayaan yang sering juga disebut *margin of error* merupakan nilai yang mencerminkan kurang atau lebih, misalnya interval kepercayaan 5 dan 50% dalam pengambilan sampel mempunyai makna bahwa sampel yang kita pilih akan menjawab pertanyaan-pertanyaan yang kita berikan dalam kisaran antara 45% (50% - 5%) dan 55% (50% + 5%).

Untuk melakukan uji hipotesis diperlukan hipotesis riset. Hipotesis riset merupakan dugaan mengenai sifat fakta-fakta yang memungkinkan. Hipotesis ini dianggap sebagai titik awal penyelidikan. Sebagai contoh, peneliti berpendapat bahwa promosi secara besar-besaran berpengaruh terhadap penjualan. Sedang dalam statistik, terdapat dua hipotesis, yaitu hipotesis nol ( $H_0$ ) dan hipotesis alternatif ( $H_1 / H_a$ ). Hipotesis nol dalam statistik merupakan hipotesis yang sebenarnya kita uji. Hipotesis ini merupakan pernyataan peneliti yang mengatakan bahwa variabel bebas tidak mempunyai pengaruh terhadap variabel tergantung (dalam regresi) dan tidak ada hubungan antara dua variabel (dalam korelasi). Dalam hipotesis nol kita berasumsi bahwa sampel – sampel berasal dari populasi yang sama. Sedang hipotesis alternatif merupakan hipotesis yang memberikan alternatif berbeda dengan hipotesis nol-nya. Dengan demikian dalam kaitannya dengan hipotesis nol di atas; maka peneliti mengatakan bahwa variabel bebas mempunyai pengaruh terhadap variabel tergantung (dalam regresi) dan ada hubungan antara dua variabel (dalam korelasi).

Contoh uji hipotesis misalnya rata-rata produktivitas pegawai sama dengan 10 ( $\mu x = 10$ ), maka bunyi hipotesisnya ialah:

- $H_0$ : Rata-rata produktivitas pegawai sama dengan 10
- $H_1$ : Rata-rata produktivitas pegawai tidak sama dengan 10

Hipotesis statistiknya:

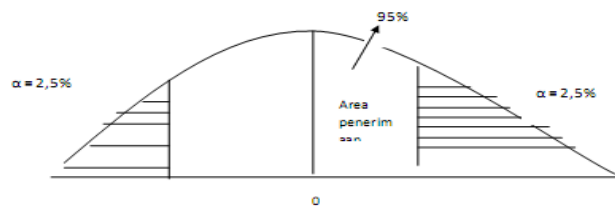
- $H_0: \mu x = 10$
- $H_1: \mu x > 10$  Untuk uji satu sisi (one tailed) atau
- $H_1: \mu x < 10$
- $H_1: \mu x \neq 10$  Untuk uji dua sisi (two tailed)

Beberapa hal yang harus diperhatikan dalam uji hipotesis ialah;

- Untuk pengujian hipotesis kita menggunakan data sample.
- Dalam pengujian akan menghasilkan dua kemungkinan, yaitu pengujian signifikan secara statistik jika kita menolak  $H_0$  dan pengujian tidak signifikan secara statistik jika kita menerima  $H_0$ .
- Jika kita menggunakan nilai t atau F, maka jika nilai t atau F yang semakin besar atau menjauhi 0, kita akan cenderung menolak  $H_0$ ; sebaliknya jika nilai t atau F semakin kecil atau mendekati 0 kita akan cenderung menerima  $H_0$ .

Jika kita menggunakan kurva untuk menguji hipotesis, maka hal itu dapat digambarkan sebagai berikut:

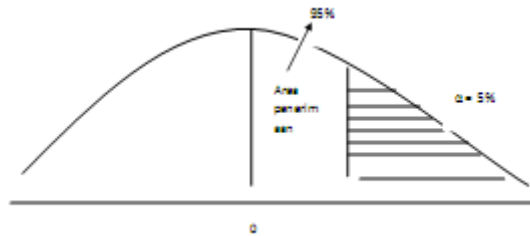
a) **Untuk uji dua sisi**





**Gambar 1.5 Uji Dua Sisi**

b) Untuk uji satu sisi sebelah kanan



**Gambar 1.6 Uji Satu Sisi Kanan**

c) Untuk uji satu sisi sebelah kiri



**Gambar 1.7 Uji Satu Sisi Kiri**

### 1.7 Karakteristik Model yang Baik

Model dikatakan baik menurut Gujarati (2009), jika memenuhi beberapa kriteria seperti di bawah ini:

- Parsimoni: Suatu model tidak akan pernah dapat secara sempurna menangkap realitas; akibatnya kita akan melakukan sedikit abstraksi ataupun penyederhanaan dalam pembuatan model.
- Mempunyai Identifikasi Tinggi: Artinya dengan data yang ada, parameter-parameter yang diestimasi harus mempunyai nilai-nilai yang unik atau dengan kata lain, hanya akan ada satu parameter saja.
- Keselarasan (*Goodness of Fit*): Tujuan analisis regresi ialah menerangkan sebanyak mungkin variasi dalam variabel tergantung dengan menggunakan variabel bebas dalam model. Oleh karena itu, suatu model dikatakan baik jika eksplanasi diukur dengan menggunakan nilai *adjusted r*<sup>2</sup> yang setinggi mungkin (mendekati 1).
- Konsistensi Dalam Teori: Model sebaiknya sejaris dengan teori. Pengukuran tanpa teori akan dapat menyesatkan hasilnya.

- Kekuatan Prediksi: Validitas suatu model berbanding lurus dengan kemampuan prediksi model tersebut. Oleh karena itu, pilihlah suatu model yang prediksi teoritisnya berasal dari pengalaman empiris.

## 1.8 Istilah-Istilah Dasar

- **Koefisien regresi baku (*Standardized regression coefficient*):** nilai statistik dalam regresi linier yang menggambarkan kekuatan dan arah hubungan linier (*linear association*) antara variabel tergantung (kriteria) dan variabel bebas (prediktor). Nilai ini disebut baku karena kisaran nilainya antara -1 sampai dengan 1 (Cramer & Howitt, 2006). Jika nilainya semakin mendekati 1 maka hal tersebut menunjukkan hubungan antara kedua variabel semakin kuat dengan mengabaikan apakah positif atau negatif. Dengan demikian variabel bebas akan semakin kuat untuk digunakan memprediksi variabel tergantung. Karena prediktor – prediktor sudah distandarisasi maka hal tersebut memungkinkan untuk membandingkan kekuatan relatif hubungan atau bobot mereka dengan variabel tergantungnya. Jika tidak terdapat tanda (positif atau negatif), maka diinterpretasikan hubungan kedua variabel positif. Hubungan positif bermakna jika nilai pada prediktor tinggi, maka nilai pada variabel tergantung juga tinggi. Sebaliknya jika terdapat tanda negatif, maka hubungan kedua variabel tersebut negatif. Hubungan negatif mempunyai makna jika jika nilai pada prediktor tinggi, maka nilai pada variabel tergantung rendah. Koefesien dengan nilai sebesar 0,50 mempunyai makna bahwa untuk setiap kenaikan simpangan baku pada nilai prediktor, maka simpangan baku pada variabel tergantungnya naik sebesar 0,5.

Koefesien regresi baku dapat diubah menjadi koefesien regresi yang tidak baku dengan cara mengkalikan koefesien regresi baku dengan deviasi standar (SD) variabel tergantung dan membaginya dengan deviasi standar prediktornya. Jika dirumuskan akan menjadi:

$$\text{Koefesien regresi tidak baku} = \text{Koefesien regresi baku} \times \frac{\text{SD variabel tergantung}}{\text{SD variabel bebas}}$$

- **Koefisien regresi tidak baku (*Unstandardized regression coefficient*):** koefesien yang belum distandarisasi yang mempunyai nilai acak dan tidak terbatas lawan dari koefisien yang distandarisasi dengan nilai  $\pm 1$ . Prediktor tertentu yang diukur dalam unit dengan nilai – nilai yang lebih besar akan terdapat kemungkinan mempunyai koefesien regresi parsial yang tidak baku yang lebih besar daripada prediktor tertentu yang diukur dalam unit nilai yang lebih kecil. Oleh karena itu, kita akan kesulitan membandingkan bobot relatif prediktor – prediktor tersebut saat mereka tidak diukur dengan menggunakan skala atau ukuran yang sama. **Koefesien regresi tidak baku dapat diubah menjadi koefesien baku dengan cara mengkalikan koefesien regresi tidak baku dengan deviasi standar variabel bebas dengan dibagi oleh deviasi standar variabel tergantungnya.**

$$\text{Koefesien regresi baku} = \text{Koefesien regresi tidak baku} \times \frac{\text{SD variabel bebas}}{\text{SD variabel tergantung}}$$

- **Hubungan linier:** hubungan linier diartikan bahwa hubungan antara variabel bebas / prediktor dengan variabel tergantung / kriteria membentuk garis lurus. Sekalipun demikian tidak semua hubungan selalu membentuk garis lurus, ada hubungan yang membentuk garis lengkung. Artinya garis kecocokan yang terbaik bukan dalam bentuk garis lurus tetapi membentuk kurva seperti huruf U. Sebagaimana dalam hubungan linier yang berbentuk garis lurus, maka kedekatan titik – titik kearah garis berbentuk kurva lengkung merupakan faktor penting dalam menentukan ukuran (kekuatan) hubungan. Jika bentuknya semakin menyebar, maka kekuatan hubungan semakin melemah. Jika digambarkan akan menjadi seperti di bawah ini:



**Gambar 1.8 Hubungan Linier**

- **Model yang Baik:** model dalam statistik mempunyai pengertian membentuk keseluruhan data dalam pertanyaan. Dengan demikian, model statistik adalah deskripsi matematik bagaimana beberapa variabel menjelaskan data tersebut dalam pertanyaan. Oleh karena itu model yang baik mampu menjelaskan data tersebut sedekat mungkin tanpa menjadikannya rumit yang menyebabkan pemahaman terhadap data tersebut menjadi kabur atau tidak jelas.
- **Asumsi klasik dalam regresi:** asumsi klasik adalah beberapa pesyaratan yang harus ditaati saat kita menggunakan prosedur regresi linier, diantaranya ialah: otokorelasi, multikolinieritas, normalitas data, heteroskedastisitas.
- **Otokorelasi:** terjadi korelasi dalam variabel bebas yang mengganggu hubungan variabel bebas tersebut dengan variabel tergantung. Untuk pengujian otokorelasi kita menggunakan nilai dari Durbin – Watson (DW). Kisaran nilai DW mulai dari 0 – 4. Tidak terjadi otokorelasi jika:  $-2 \leq DW \leq 2$  (Anderson, 2001:733)
- **Multikolinieritas:** terjadi korelasi antar variabel bebas dalam regresi linier berganda dengan nilai yang sangat tinggi atau sangat rendah. Nilai-nilai yang digunakan untuk menguji otokorelasi diantaranya: nilai *variance inflation factor* (VIF) dengan ketentuan jika nilai VIF > 5, maka terjadi multikolinieritas; nilai *condition index* dengan ketentuan jika nilai *condition index* > 5, maka terjadi multikolinieritas; nilai korelasi antar variable bebas dengan ketentuan jika nilai koefesien korelasi antara variable bebas > 0,7 atau < - 7 (Anderson, 2001:644) maka terjadi multikolinieritas.
- **Normalitas data:** Normalitas data berkaitan dengan distribusi suatu data. Data yang mempunyai distribusi normal artinya data yang distribusinya simetris sempurna. Jika

digunakan bahasa umum disebut berbentuk kurva bel. Menurut Johnston (2004) ciri-ciri data yang mempunyai distribusi normal ialah sebagai berikut: a) Kurva frekuensi normal menunjukkan frekuensi tertinggi berada di tengah-tengah, yaitu berada pada rata-rata (*mean*) nilai distribusi dengan kurva sejajar dan tepat sama pada bagian sisi kiri dan kanannya. Kesimpulannya, nilai yang paling sering muncul dalam distribusi normal ialah rata-rata (*average*), dengan setengahnya berada dibawah rata-rata dan setengahnya yang lain berada di atas rata-rata; b) Kurva normal, sering juga disebut sebagai kurva bel, berbentuk simetris sempurna; c) Karena dua bagian sisi dari tengah-tengah benar-benar simetris, maka frekuensi nilai-nilai diatas rata-rata (*mean*) akan benar-benar cocok dengan frekuensi nilai-nilai di bawah rata-rata; d) Frekuensi total semua nilai dalam populasi akan berada dalam area dibawah kurva. Perlu diketahui bahwa area total dibawah kurva mewakili kemungkinan munculnya karakteristik tersebut; e) Kurva normal dapat mempunyai bentuk yang berbeda-beda. Yang menentukan bentuk-bentuk tersebut adalah nilai rata-rata dan simpangan baku (*standard deviation*) populasi.

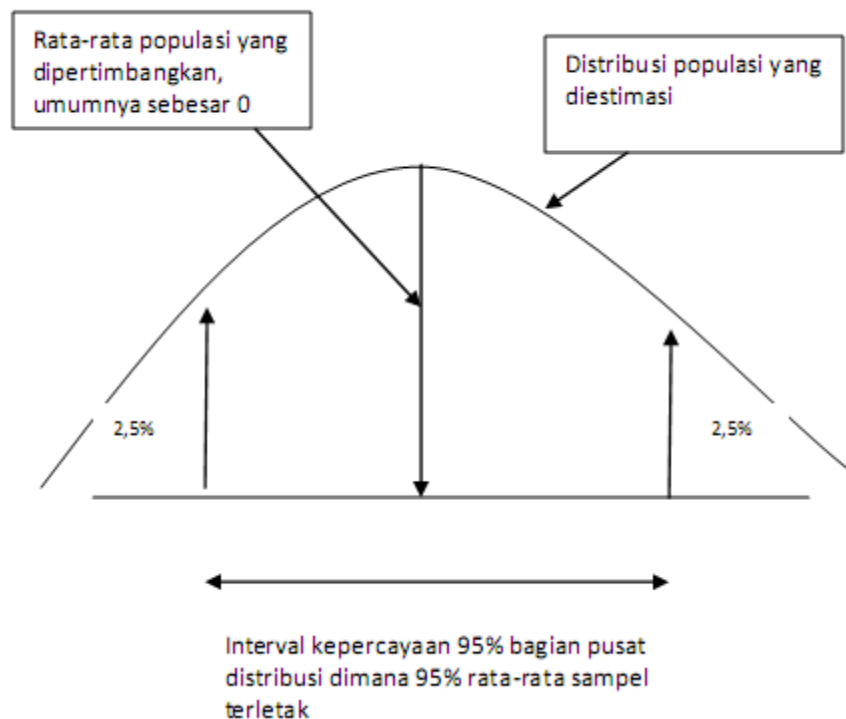
- **Heteroskedastisitas:** untuk memahami pengertian heteroskedastisitas diperlukan memahami terlebih dahulu pengertian homoskedastisitas. Homoskedastisitas adalah deskripsi data dimana varian batas kesalahan (*error terms / e*) kelihatan konstan diluar jangkauan dari nilai – nilai variabel bebas tertentu. Asumsi kesamaan varian kesalahan populasi  $\epsilon$  (dimana  $\epsilon$  diestimasi dari nilai sampel  $e$ ) kritis jika diaplikasikan pada regresi linier yang benar. Saat batas kesalahan mempunyai varian yang semakin besar, maka data disebut bersifat heteroskedastisitas. Dengan kata lain, homoskedastisitas merupakan asumsi dimana variable tergantung menunjukkan tingkatan varian yang sama untuk semua variable bebasnya. Jika penyebaran nilai varian pada semua variable bebas tidak sama maka hubungan tersebut dikatakan sebagai heteroskedastisitas. Untuk menguji homoskedastisitas (terjadinya kesamaan varian pada semua variable bebas) digunakan pengujian Levene pada data variable berskala non-metrik. Terjadi kesamaan varian jika nilai signifikansi (*sig*) pada Levene test  $> 0,05$ . Pengujian Levene dapat dijelaskan dengan membuat hipotesis awal ( $H_0$ ) yang berbunyi “varian pada semua variable bebas sama” dan hipotesis alternatif ( $H_1$ ) yang berbunyi “varian pada semua variable bebas tidak sama”. Ketentuan pengujian hipotesis didasarkan pada nilai signifikansi: jika nilai *sig*  $> 0,05$   $H_0$  diterima; jika nilai *sig*  $< 0,05$   $H_0$  ditolak. Jika variable-variabel berskala metrik kita dapat menggunakan pengujian Box’s M. Ketentuan pengujiannya sama dengan cara pengujian menggunakan Levene test. Untuk mengetahui apakah terjadi heteroskedastisitas dalam pengujian diatas dapat diketahui dari nilai signifikansinya. Jika nilai signifikansi (*sig*)  $< 0,05$ , maka dalam model tersebut terjadi heteroskedastisitas.

Terjadi heteroskedastisitas dalam model regresi jika titik – titik dalam scatterplot membentuk pola –pola tertentu atau berkumpul disatu sisi atau dekat nilai 0 pada sumbu Y pada kurva yang dihasilkan saat kita menggambar kurva dengan menggunakan SPSS. Jika titik – titik data menyebar tidak secara beraturan, maka tidak terjadi heteroskedastisitas.

- **Residual (e atau  $\epsilon$ ):** kesalahan dalam memprediksi data sampel. Kita berasumsi bahwa kesalahan random akan terjadi, tetapi kita berasumsi bahwa kesalahan ini merupakan estimasi kesalahan random yang sebenarnya pada populasi ( $\epsilon$ ), bukan semata hanya

kesalahan prediksi dalam sampel (  $e$  ). Pengertian lain ialah perbedaan antara nilai variabel tergantung yang sebenarnya dan nilai prediksi.

- **Least Square:** prosedur estimasi yang digunakan dalam regresi linier dimana koefisien regresi diestimasi untuk meminimalkan jumlah residual kuadrat keseluruhan.
- **Interval Kepercayaan:** suatu pendekatan untuk menilai informasi dalam sampel yang memberikan interval untuk suatu nilai statistik yang berkaitan dengan bagaimana sebaiknya sampel ditarik dari populasi. Sebagai contoh interval kepercayaan 95% mengacu pada interval yang menjangkau nilai statistik, misalnya rata-rata untuk sampel yang sebesar 95% ditarik dari populasinya.



**Gambar 1.9 Interval Kepercayaan sebesar 95%**

- **Signifikansi / probabilitas /  $\alpha$ :** Signifikansi atau disebut juga probabilitas merupakan tingkat ketepatan (presisi) dalam kaitannya dengan kesalahan pengambilan sampel (*sampling error*), merupakan jangkauan dimana nilai populasi yang tepat diperkirakan. Jangkauan ini sering diekspresikan dengan menggunakan poin-poin persentase, misalnya 1% atau 5%. Oleh karena itu jika seorang peneliti menemukan bahwa 60% pegawai perusahaan tertentu yang digunakan sebagai sampel sudah mengadopsi suatu metode bekerja yang direkomendasikan dengan tingkat ketepatan sebesar  $\pm 1\%$ , maka

peneliti tersebut dapat menyimpulkan bahwa antara 59% dan 61% dari pegawai perusahaan tersebut yang menjadi populasi sudah mengadopsi metode tersebut. Dalam IBM SPSS signifikansi ditulis secara default sebagai 0,05 (5%).

- **Degree of Freedom (DF) / Derajat Kebebasan (DK):** estimasi statistik yang didasarkan pada sampel yang sering berisi informasi yang redundan yang jika disertakan akan menghasilkan bias dalam estimasi tersebut. Metode penghitungan DF bervariasi tergantung dari apa yang diestimasi (dengan menggunakan prosedur apa kita membuat estimasi) dengan demikian tidak aturan umum yang dapat diterapkan. Sebagai contoh untuk menghitung DF dalam Chi Square, kita menggunakan aturan  $(\text{Kolom} - 1) \times (\text{Baris} - 1)$  sedang dalam uji t sampel bebas digunakan aturan  $(n - 2)$ .
- **Koefisien Determinasi ( $R^2$ ):** pengukuran proporsi varian variabel tergantung tentang rata – ratanya yang dapat dijelaskan oleh variabel bebas / prediktornya. Jika nilai ini semakin besar (mendekati 1), maka prediksi yang dibuat semakin akurat.
- **Koefisien Determinasi yang Disesuaikan (*adjusted*  $R^2$ ):** nilai  $R^2$  yang disesuaikan dengan mempertimbangkan jumlah variabel bebas / predictor yang dimasukkan dalam persamaan regresi dan ukuran sampel. Asumsinya jika variabel bebas ditambahkan nilai ini cenderung naik. Nilai ini sering digunakan sebagai nilai kecocokan model (*goodness of fit*) dimana jika nilainya semakin tinggi (mendekati 1), model semakin benar / akurat.
- **Linieritas:** istilah yang digunakan untuk mengekspresikan konsep yang menjelaskan bahwa model yang dibuat memiliki properti aditivitas dan homogenitas. Linier model mempunyai arti bahwa model tersebut memprediksi nilai – nilai yang berada pada garis lurus dengan memiliki perubahan unit konstan pada variabel tergantung untuk perubahan unit konstan pada variabel bebasnya. Dengan kalimat yang lebih sederhana linieritas mempunyai makna bahwa model tersebut dapat menjelaskan terjadinya perubahan pada nilai variabel tergantung dikarenakan terjadinya perubahan nilai pada variabel bebasnya.
- **Kolinieritas:** ekspresi tentang hubungan antara dua variabel bebas atau lebih dari dua variabel bebas (multikolinieritas)
- **Variance Inflation Vector:** indikator pengaruh yang menunjukkan bahwa variabel – variabel bebas lainnya mempunyai kesalahan baku dari koefisien regresi. Nilai VIF yang tinggi menunjukkan terjadinya kolinieritas ataupun multi kolinieritas diantara variabel – variabel bebas. Jika nilai  $VIF > 5$ , maka terjadi multikolinieritas.

(Tulisan ini diambil dari buku saya dengan judul “12 Jurus Ampuh SPSS untuk Riset Skripsi” yang diterbitkan oleh Penerbit Elexmedia Komputindo Kompas Gramedia Jakarta, tahun 2013)