H. MIFTACHUL 'ULUM, ST.,MM

BUKU STATISTIK

Jl. Jendral Sudirman (Sidotopo) No.11 Kepanjen - Malang
Telp. 0341 - 395 996, Fax. (0341) 395 999
www.stikeswch-malang.ac.id

BAB I
DEFINISI DAN RUANG LINGKUP, VARIABEL SAMPLING DAN DISTRIBUSI

Dalam bab ini akan diterangkan mengenai pengertian statistika, pengertian populasi dan sampel, jenis-jenis data, variabel serta teknik-teknik yang dapat digunakan dalam penelitian, selain itu akan diterangkan pula mengenai variabel sampling dan distribusi.

A. PENDAHULUAN

Statistika adalah pengetahuan yang berhubungan dengan statistik, yakni berhubungan dengan:

- cara pengumpulan data
- pengolahan dan analisis data, serta
- penarikan kesimpulan mengenai populasi

Statistik dalam pengertian awam adalah tabel/ daftar angka-angka tentang sesuatu hal/ kegiatan, sering disertai gambar diagram, grafik dan dilengkapi dengan ukuran2 pemusatan, letak, penyebaran dan ratio prosentase.

Dua Pengertian Statistik

1. Menyatakan kumpulan angka2 yang melukiskan suatu persoalan, misal: statistik penduduk, statistik kelahiran, kematian, statistik perekonomian, statistik produksi, pendapatan, harga, perdagangan, perbankan, dll.
2. Menyatakan ukuran, misal: ukuran pemusatan, letak, prosentase, angka indeks, angka perbandingan.

Statistika Hendaknya Bersifat Tak Bias artinya kesimpulan yang diperoleh sesuai dengan keadaan sebenarnya, jadi $\overline{X} = \mu$ (pada sampel $\overline{X} \neq \mu$, hanya dengan sensus $\overline{X} = \mu$)

Populasi <--------- generalisasi ------------- Sampel

parameter ---------- sampling ----------------> statistik

μ $\overline{X}$

σ S

$$Y = a + bX + e$$

Karena itu e (error) harus diminimumkan, dengan cara sample representatif, jika populasi heterogen sample diperbesar, dan dengan penerapan metode sampling yang sesuai.

Statistik adalah ukuran karakteristik sampel sedangkan **Parameter** adalah ukuran karakteristik populasi

Populasi:

- adalah kesatuan persoalan secara menyeluruh yang sudah ditentukan definisi karakteristiknya dan batas2 unit elementernya secara jelas sebagai ruang kesimpulan.
- jadi keseluruhan himpunan obyek dengan ciri yang sama
- atau kumpulan lengkap dari unit2 elementer

Sampel:

- adalah sebagian dari populasi
- merupakan himpunan bagian

Statistika Deskriptif

Statistik deskriptif adalah bagian statistika yang berhubungan dengan:

- Pengumpulan data, pengolahan dan penyajian data sebagai informasi dalam bentuk daftar/ tabel, gambar diagram, grafik dan perhitungan2 untuk menentukan statistik
- Data ini diperoleh dari penelitian nonprobabilitas
- Data ini digunakan untuk uji/ analisis2 sesuai dengan teori masing2 disiplin ilmu (uji non statistika); dan untuk menghitung ukuran2 pemusatan/ letak, penyebaran, penyimpangan, prosentase, angka indeks, dll.

Statistika Induktif/ Inferensial

Statistik induktif adalah bagian statistika yang berhubungan dengan pembuatan kesimpulan mengenai populasi, misalnya tentang:

- penaksiran karakteristik populasi
- pembuatan prediksi
- menentukan ada/ tidaknya asosiasi antara karakteristik populasi
- pembuatan generalisasi/ kesimpulan umum mengenai populasi

Statistika inferensial merupakan penerapan metode analisis dalam menginterpretasikan data statistik sampel probabilitas guna men-jelaskan populasi.

Data (Data Statistik) adalah keterangan (kuantitatif/ kualitatif) yang merupakan karakteristik unit elementer yang diselidiki, dimana kebenarannya dapat diandalkan.

Data Interen adalah data yang dikumpulkan oleh suatu badan mengenai aktivitas badan itu sendiri untuk keperluan badan tersebut.

Data Eksteren adalah data di luar aktivitas badan tersebut.

Data Primer adalah data yang dikumpulkan langsung oleh orang/ badan tertentu sebagai tangan pertama, dimana pada saat observasi data tersebut belum tersedia.
Data Sekunder adalah data yang dikumpulkan dari pihak lain, dimana pada saat observasi data tersebut telah tersedia dalam bentuk laporan atau dokumentasi.

Data Eksteren Primer adalah data eksteren dari sumber pertama

Data Eksteren Sekunder adalah data eksteren dari sumber lain (bukan sumber pertama)

Data yang merupakan karakteristik unit elementer (sampel/ populasi) dapat diukur dalam bentuk bilangan kuantitatif atau kategori kualitatif memiliki **Sifat Variabel**.

B. VARIABEL

Variabel adalah suatu konsep yang mempunyai variasi nilai (jadi lebih dari satu nilai) yg diukur dan diuji untuk menjelaskan hubungan dalam memprediksi fenomena teori.

Gambaran yang sistematis dalam teori dijabarkan dengan menghubungkan antar variabel.

1. Hubungan Variabel

Inti penelitian ilmiah adalah mencari hubungan dan kaitan pengaruh antar variabel. Pada dasarnya terjadi tiga jenis hubungan antar variabel:

a. Hubungan simetris, apabila variabel yg satu tidak disebabkan/ tidak dipengaruhi oleh variabel lainnya.
   contoh: hubungan simetris antara variabel independent

b. Hubungan resiprokal/ timbal balik, apabila pada suatu waktu variabvel X mempengaruhi variabel Y dan diwaktu lain variabel Y mempengaruhi variabel X. Jadi dapat berupa variabel independent dan dependent pada waktu yang berbeda.
c. Hubungan asimetris, apabila suatu variabel mempengaruhi variabel lainnya. Jadi variabel independent tidak pernah menjadi dependent dan sebaliknya.

2. Beberapa Tipe Hubungan Asimetris

a. Hubungan stimulus-respons yakni hubungan kausal yang mempengaruhi faktor2 luar (eksternal). Diperlukan kepekaan selektif dalam memilih faktor2 tertentu; penguasaan ilmu pengetahuan sangat membantu dalam memilih dan menempatkan faktor2 sebagai variabel yg proporsional.
b. Hubungan disposisi-respons

   Disposisi adalah kecenderungan untuk menunjukkan respons tertentu dalam situasi tertentu karena pengaruh faktor internal.

   Stimulus datang dari luar sedangkan disposisi dalam ilmu sosial ada dalam diri seseorang (seperti sikap, kemampuan dan lain-lain)
c. Hubungan prakondisi dengan akibat.

   Prakondisi adalah semacam *treatment* yang akan memberi dampak tertentu.
d. Hubungan imanen antara dua variabel.

   Kedua variabel terjalin satu sama lain; jika variabel satu berubah otomatis variabel lainnya ikut berubah.
e. Hubungan tujuan dengan cara.

   Cara mempengaruhi tujuan yang dicapai. Tujuan yang sama efektif dapat dicapai dengan cara yang berbeda efisien.
f. Hubungan *bivariat* dan *multivariat*.

Bivariat yakni hubunbgan antara dua variabel asimetris **(regresi sederhana)**

Multivariat yakni hubungan asimetris antara variabel dependent dengan beberapa variabel independent **(regresi berganda)**

3. Jenis - Jenis Variabel

Penentuan klasifikasi variabel yang benar memerlukan penguasaan dasar teoritis yang mendalam. Tinjauan teori membantu menyusun kerangka teoritis atau model yang mantap.

a. Penggolongan variabel berdasarkan fungsinya:

1) Variabel independent merupakan variabel sebab yang menjadi pokok permasalahan yg ingin diteliti.

2) Variabel dependent merupakan variabel akibat yang besarnya tergantung dari variabel independent

$$\boxed{Y = f\,(X_1, X_2, X_3, \ldots\ldots, Xn)}$$

Keterangan:

Y = variabel dependent

X = variabel independent

b. Penggolongan variabel berdasarkan keberadaan variabel dalam model

1) Variabel endogen

2) Variabel eksogen

$$\boxed{Y = a + bX + e}$$

Keterangan:

X, Y = variabel endogen

e yang dijelaskan oleh a = faktor error karena pengaruh variabel eksogen

c. Penggolongan variabel berdasarkan nilai pengukuran
   1) Variabel Kuantitatif/ Numerik, meliputi:
      a) Variabel kontinyu, dimana datanya diukur dengan nilai interval
      b) Variabel diskrit, dimana datanya diukur dengan bilangan cacah/ bukan pecahan
   2) Variabel Kualitatif/ Anumerik/ kategori, meliputi:
      a) Variabel Strata (ukuran perbedaan derajad)
      b) Variabel klaster (ukuran perbedaan jenis)

   Variabel kualitatif perbedaan derajad (strata) dapat dikuantitatifkan menjadi variabel diskrit dengan cara diberi angka skor.

C. TEKNIK SAMPLING

Teknik Sampling adalah teknik penarikan sampel dari suatu populasi.

Jenis populasi:

- populasi tak terhingga dimana banyaknya anggota tak terhingga
- populasi terhingga yang diketahui jumlah anggotanya

Sensus apabila setiap anggota populasi diteliti.

Sampling apabila hanya sebagian anggota populasi yg diteliti dengan syarat dapat mewakili populasi.

1. Alasan Dilakukan Sampling:

a. Keterbatasan biaya, waktu dan tenaga

b. Ketelitian penelitian sampel biasanya lebih tinggi jika dibandingkan sensus dengan populasi yang besar
c. Menghindari percobaan yang sifatnya merusak sebaiknya dilakukan sampling
d. Anggota populasi tak terhingga.

2. Jenis2 Teknik Sampling

Secara garis besar ada dua cara pengambilan sampel, yakni **non-probabilitas sampling** dan **probabilitas sampling**.

a. Non-probabilitas sampling (non-random sampling), meliputi:

1) Sampling seadanya *(acsidental sampling)*

Dilakukan karena populasi sulit ditentukan sejak awal. Misal penelitian karakteristik konsumen pada produksi masa dimana pembeli diwawancarai saat membeli produk tersebut.

Sampling ini hanya menunjukkan gambaran kasar, dan dalam beberapa hal sampling ini mungkin berfaedah namun dalam hal lain mungkin tidak berfaedah.

2) Sampling pertimbangan/ pilih kasih *(purposif sampling)*

Pertimbangan individu menentukan pengambilan sampel. Individu disini bisa sipeneliti atau saran para ahli, dll. Jadi ada karakteristik tertentu yang dipertimbangkan.

Misal penelitian pasar kebutuhan sandang dalam hubungan dengan masyarakat ekonomi menengah ke bawah di Surabaya yang dipilih adalah obyek di Pasar Turi; sedangkan untuk kelas menengah ke atas di Pasar Atom dan Tunjungan Plaza.

Sampling kuota tergolong kelompok sampling purpo-sif karena didasarkan pertimbangan2 tertentu yang subyektif. Berbeda dari proportional sampling yang didasarkan pada jumlah anggota unit populasi.

b. Probabilitas sampling *(random sampling)*

Asumsi dasar pemakaian statistika inferensial/ induktif adalah random sampling dimana tiap unit/ individu populasi memiliki probabilitas yang sama untuk dijadikan sampel. Jika pengambilan sampel dilakukan dengan cara non random maka pemakaian statistika inferensial perlu dipertanyakan ke-absahannya.

Random sampling dibedakan atas:

- *simple random sampling*
- *systematic random sampling*
- *stratified random sampling*
- *cluster/ area random sampling*
- *multistage random sanmpling*

1) *Simple Random Sampling*

Cara ini digunakan jika populasi dianggap homogen. Tersedia daftar dari seluruh unit populasi. Pengambilan unit sampel melalui lotre atau daftar bilangan random.

Keuntungan:

- pelaksanaannya mudah dan
- unbias karena $\overline{X}$ = u jika benar2 homogen

Kelemahan :

- sampel bisa menyebar jauh/ atau terkumpul dalam satu area
- Diperlukan daftar lengkap dari seluruh unit populasi

2) *Systematic Sampling*

Cara ini digunakan jika populasi dianggap homogen. Tersedia daftar dari seluruh unit populasi. Dibuat urutan tertentu (sistematis) untuk penentuan sampel. Atau untuk pengambilan sampel I = *simple random sampling*, sedangkan untuk II dan seterusnya ditentukan secara sistematis yakni meloncat ke nomor berikutnya dengan jarak interval tertentu.

Contoh, N = 90 , n = 30 jadi jarak sistematis 90/30 = interval 3. Hasil random sampel I = no 10 maka sampel II = no 13 dst.

Cara ini biasa disebut juga sebagai *Systematic Random Sampling*.

Keuntungan dan kelemahannya identik dengan simple random sampling.

3) *Stratified Random Sampling* (Sampling acak berstrata)

Digunakan jika populasi heterogen dan ternyata populasi tersebut terdiri dari lapisan2 (strata/ karakteristik perbedaan derajad) yang homogen.

Agar sampel lebih mewakili populasi maka stratified random sampling dibagi lagi atas:

a) *Simple stratified random sampling* jika jumlah unit populasi dalam tiap strata sama maka jumlah unit sampel dalam tiap strata juga sama.
b) *Proportional stratified random sampling* jika jumlah unit populasi dalam tiap strata tidak sama maka strata dengan unit yang besar juga diwakili unit sampel yang besar dan sebaliknya.

Cara mengambil sampel pada stratified random sampling dapat dilakukan dengan **lotre** atau **sistematik**.

4) *Cluster Random Sampling* (Sampling Klaster)

Dilakukan jika populasi heterogen dan ternyata populasi tersebut terdiri dari kelompok2 (cluster/ karakteristik perbedaan jenis) yang memiliki ciri homogen. Disebut juga *Area Random Sampling* (Sampling Area) jika kelompok adalah pembagian daerah geografis. Misal area administratif seperti: wilayah RT, Desa, Kecamatan, Kabupaten dsb; dan area geografis seperti: dataran tinggi, dataran rendah, pantai, daerah aliran sungai, dsb.

Cluster bisa juga untuk kelompok kelamin: wanita, pria, waria; kelompok warna: merah, kuning, hijau, dsb.

Jika jumlah cluster besar maka pemilihan kluster secara random, dari cluster2 tersebut kemudian diambil sampel secara random.

5) *Multistage Random Sampling* (Sampling Ganda)

Jenis2 sampling di atas adalah sampling tunggal dimana ukurannya telah ditentukan lebih dahulu, kemudian dilakukan sampling untuk memperoleh ukuran (sampling zise) tersebut. Sering kali ukuran ini berlebihan sehingga terjadi pemborosan waktu, tenaga, dan biaya. Sampling ganda memungkinkan ukuran sampel lebih kecil.

Dalam sampling ganda penelitian dimulai dengan sampel yang kecil, jika hasilnya tidak memberikan kepastian dilakukan sampling ke dua. Kesimpulannya merupakan penggabungan dari kedua sample tersebut.

6) *Sampling Sekuensial*

Cara ini berdasarkan sampling ganda, perbedaannya individu dipilih dan diteliti satu demi satu dan berdasarkan ini dibuat kesimpulan atau sampling dilanjutkan hingga tercapai tingkat yang meyakinkan dalam penelitian.

Berdasarkan sampel yang diambil dari populasi akan dipelajari karakteristik populasi (parameter). Parameter yg dimaksud ditaksir dari nilai statistik sampel yang antara lain berupa: ukuran rata2, ukuran perbandingan, simpangan baku, dan koefisien korelasi.

D. SAMPLING PROBABILITAS (SAMPLING BERPELUANG)

Dari sebuah populasi dapat diambil lebih dari sebuah sampel. Jika populasi berukuran N dan sampel berukuran n (sample size) serta pengambilan sampel tanpa pengembalian maka banyaknya sampel yang mungkin diambil (sampel probabilitas) adalah:

$$C\left\langle \begin{matrix} N \\ n \end{matrix} \right\rangle = \frac{N\,!}{n\,!\,(N-n)\,!}$$

Populasi (N) = 10

Sampel (n) = 2

$$\text{Sampel Probabilitas} = \frac{10\,!}{2\,!\,(10-2)\,!} = \frac{10.9.8\,!}{2\,!\,8\,!} = 45$$

n = 20 % dari N kombinasinya sama dengan n = 80 % dari N

n = 40 % dari N kombinasinya sama dengan n = 60 % dari N

n = 50 % dari populasi kombinasinnya paling besar

Ini adalah jumlah kombinasi atau jumlah sampel yang mungkin terjadi (sampel probabilitas) bukan ukuran besarnya sampel *(sample size)*. Berapa buah sampel probabilitas yang diambil dari suatu penelitian tergantung keadaan, sampling ganda atau sampling tunggal. Pada umumnya kesimpulan diambil hanya berdasarkan sebuah sampel (sampel tunggal).

E. DISTRIBUSI PROBABILITAS/ DISTRIBUSI PELUANG

Distribusi peluang melukiskan pengelompokan peristiwa2 dimana pada tiap kelompok telah dihitung banyaknya peristiwa yang terjadi yang dinyatakan dalam prosen.

Distribusi peluang merupakan distribusi yang diharapkan berdasarkan pada pengalaman empiris dari nilai-nilai variabel. Terdapat dua jenis distribusi peluang yakni distribusi peluang diskrit dan distribusi peluang kontinyu

1. Distribusi Peluang Diskrit

Adalah distribusi peluang dgn nilai variabel acak diskrit meliputi: distribusi Binomial dan distribusi Poisson.

Apabila untuk nilai2 diskrit $X = X_1, X_2, .., X_n$ didapat harga peluang $P_{(X1)}, P_{(X2)}, .., P_{(Xn)}$ maka jumlah peluang tersebut = 1 atau $\Sigma \; P_{(Xi)} = 1$

a. Distribusi Binomial

Peluang terjadinya suatu peristiwa tepat sebanyak X kali diantara percobaan sebanyak N, dapat ditentukan dengan rumus:

$$P_{(x)} = \binom{N}{X} \pi^{X} (1 - \pi)^{N \;\; x}$$

$$\binom{N}{X} = \frac{N\,!}{X!\,(N - x)\,!}$$

Parameter untuk distribusi binomial: N dan π, dengan rata2 dan simpangan baku adalah:

$$u = N\,\pi$$

Simpangan baku menyatakan berapa besar pencarannya yang diharapkan dihitung mulai dari u.

Soal:

Diketahui produksi 15 % rusak. Jika diteliti 30 unit secara acak, hitung peluang: a) bagus semua, b) 1 rusak, c) paling sedikit 1 rusak.

b. Distribusi Poisson

Digunakan jika N cukup besar sedangkan peluang π sangat kecil. Pendekatan ini sangat baik jika $N\,\pi \le 5$ dan $\pi \le 0{,}1$ dengan rumus:

$$P_{(x)} = \frac{e^{-\alpha}\alpha^{x}}{x!}$$

Parameter untuk distribusi Poisson adalah $\alpha = N\,\pi$ dengan rata-rata dan simpangan baku.

$u = \alpha$

$\sigma = \sqrt{\alpha}$

Soal:

Produk A diiklankan di koran "X" dengan 100 ribu pembaca. Jika peluang pembaca akan membalas iklan = 0,00002, hitung peluang hanya seorang yang membalas iklan. ($\alpha = N\,\pi = 100.000 * 0{,}00002 = 2$).

Distribusi Binomial dan Poisson tidak dibicarakan. Karena materi kita menyangkut Regresi dan Korelasi maka yang dibicarakan adalah distribusi normal dimana uji t dan uji F dalam distribusi tersebut berdistribusi normal.

2. Distribusi Peluang Kontinyu

Adalah distribusi peluang dgn nilai variabel acak kontinyu meliputi: distribusi Normal, distribusi t, dan distribusi Chi Kuadrat.

3. Distribusi Normal

Distribusi peluang normal atau disingkat distribusi normal disebut juga distribusi **Gauss** karena jasa Carl Gauss yang banyak mengungkapkan distribusi normal pada akhir abad ke 18. Ini merupakan distribusi terpenting yang banyak digunakan dalam statistika.

Tinggi ordinat kurva normal diukur dengan rumus

$$\boxed{Y = \frac{1}{\sigma\sqrt{2\pi}} e^{1/2} \; \frac{(\overline{X} - u)}{\sigma}^{2}}$$

Dimana:

π = nilai konstanta 3,1416

e = logaritma Napier 2,7183

u = parameter harga rata2 distribusi normal

σ = parameter simpangan baku distribusi normal

Nilai Y merupakan tinggi kurva dihitung mulai dari sumbu datar untuk harga X variabel acak kontinyu yang harganya $-\infty < X < +\infty$. Dalam aplikasinya tidak banyak tertarik pada nilai Y (tinggi kurva normal) melainkan pada luas daerah di bawah kurva normal.

Sifat-2 Distribusi Normal

1) Grafiknya selalu ada di atas sumbu datar X
2) Simetris terhadap X = u
3) Mempunyai satu modus yakni nilai terbesar untuk Y yg dicapai saat X = u yg besarnya = 0,3989/ σ
4) Grafiknya berasimtutkan (mendekati) sumbu datar X mulai dari X = u + 3 σ ke kanan dan X = u - 3 σ ke kiri
5) Luas daerah di bawah kurva normal selalu sama dengan satu unit persegi

Bagi tiap pasang u dan σ yang diketahui, grafiknya akan selalu memenuhi sifat-2 di atas hanya bentuknya saja yang berlainan (yakni lebar sempitnya dan tinggi rendahnya grafik).

Makin besar σ makin lebar dan makin rendah grafik → kurva Z, F

Makin kecil σ makin sempit dan makin tinggi grafik → kurva t

Agar mempermudah penggunaannya maka distribusi normal dengan rata2 u dan simpangan baku σ ditransformasikan menjadi distribusi normal standar yang mempunyai rata-2 u = 0 dan simpangan baku σ = 1 dimana variabel acak X diubah menjadi variabel acak Z (sumbu datar distribusi normal) dengan rumus.

$$\boxed{Z = \frac{X - u}{\sigma}}$$

Luas daerah distribusi normal standar menjadi

$$\boxed{Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}(Z)^2}$$

yang telah dihitung dalam 4 desimal dam disusun dalam daftar distribusi normal standar. Daftar ini berisi luas bagian daerah dibawah kurva normal dihitung mulai dari Z = 0 sampai dengan Z berharga + dimana $Z_{hitung} = (X - u)/\ \sigma$. Untuk Z berharga - identik dengan yang + karena simetris.

Contoh soal:

Upah sejumlah karyawan suatu perusahaan berdistribusi normal. Jika diketahui upah rata2 per bulan (u) = Rp 5.675,- dan simpangan bakunya (σ) = Rp 1.528,- Hitung:

a) Berapa % karyawan yang upahnya antara Rp 3.500,- s/d Rp 7.500,-

Batas bawah Z = (X - u)/ σ = (3.500 - 5.675)/ 1528 = - 1,42 --> = 0,4222

Batas atas Z = (7.500 - 5.675)/ 1528 = 1,19 --> = 0,3830 Jadi % karyawan = 42,22 % + 38,30 % = 80,52 %

b) Berapa % karyawan yang jumlah upahnya paling sedikit Rp. 2.000,-

Z = (2.000 - 5.675)/ 1.528 = - 2,41 ---> = 0,4920

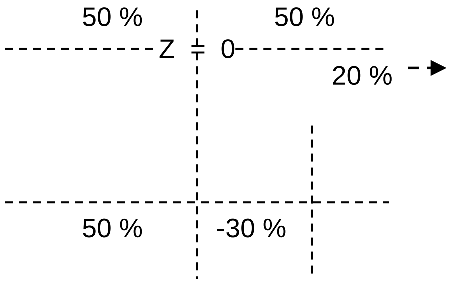
Jadi % karyawan = 49,20% + 50 % = 99,20 %

c) Berapa % karyawan yang jumlah upahnya paling besar Rp. 10.000,-

Z = (10.000 - 5.675)/ 1.528 = 2,83 ---> = 0,4977

Jadi % karyawan = 50% + 49,77% = 99,77 %

d) Jika 20 % karyawan memiliki upah tergolong tinggi, hitung jumlah upah minimum untuk golongan tersebut.



Upah tinggi , Z = 0,84 → 30%

50% - 30% = 20% upah tinggi

0,84 = (x – 5.675)/ 1.528

1.283,52 = X – 5.675

X = 6.958,52

Jadi jumlah upah minimumnya = Rp 6.958,52

F. DISTRIBUSI SAMPLING

Dalam distribusi sampel dipelajari karakteristik populasi (parameter) berdasarkan statistik sampel antara lain tentang rata2, perbandingan, dan simpangan baku.

Jika masing2 kombinasi/ masing2 sampel probabilitas dihitung nilai statistiknya (rata-2, perbandingan, simpangan baku) maka nilai-2 tersebut akan berbeda untuk tiap sampel.

Jika nilai-2 statistik tersebut dikumpulkan dan disajikan dalam suatu daftar atau grafik maka akan diperoleh Distribusi Sampling.

Jika yg disajikan nilai rata2 akan diperoleh distribusi sampling rata-2, jika nilai perbandingan diperoleh distribusi sampling perbandingan, jika selisih rata2 diperoleh distribusi sampling selisih rata-2 dst untuk distribusi sampling selisih perbandingan.

$\overline{X}$ = rata2 hitung sampel ($x = \Sigma X/n$)

u = rata2 hitung populasi ($u = \Sigma X/N$)

u_x = rata2 hitung untuk distribusi sampling rata2 ($u_x = u$)

Simpangan baku sampel = ukuran dispersi/ kekeliruan/ kesa-lahan standar dari nilai data terhadap nilai statistiknya rata2 atau perbandingan dll).

$$\boxed{\sigma = \sqrt{\frac{\Sigma(\overline{X} - u)^2}{N}}}$$ Simpangan baku populasi

$$\boxed{\sigma = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}}$$ Simpangan baku sampel

u = rata-rata hitung populasi ($u = \Sigma X/ N$)

u_x = rata-rata hitung untuk distribusi sampling rata-rata ($u_x = u$).

Simpangan baku sampel = ukuran dispersi/ kekeliruan/ kesalahan standar dari nilai data terhadap nilai statistiknya (rata-rata atau perbandingan dll)

$$\sigma = \sqrt{\frac{\Sigma(\overline{X} - u)^2}{N}}$$ Simpangan baku populasi

$$\sigma = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}$$ Simpangan baku sampel

1. Distribusi Sampling Rata2

Distribusi sampling rata2 memiliki rata2 $u_x = u$ (Rata2 dari semua sampel probabilitas = rata2 populasi) dan simpangan baku.

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$ Simpangan baku rata-2 untuk $n/N \leq 5\%$ (sampel kecil)

$$\sigma_x = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$ Simpangan baku rata-2 untuk $n/N > 5\%$ (sampel besar)

Dalil Limit Pusat:

Jika ukuran sampel n cukup besar maka distribusi sampling rata2 ternyata mendekati distribusi normal dengan.

$$Z = \frac{X - u_x}{\sigma_x}$$ jadi

$$Z = \frac{\overline{X} - u_x}{\sigma_x} = \frac{\overline{X} - u_x}{\sigma / \sqrt{n}}$$ untuk sampel kecil

$$Z = \frac{\overline{X} - u}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}}}$$ untuk sampel besar

Jika simpangan baku populasi (σ) diketahui dan selisih rata2 yang dikehendaki dari dua sampel probabilitas (d) diketahui maka ukuran sampel (sample zise) dapat dihitung dengan rumus:

$$\frac{\sigma}{\sqrt{n}} \leq d$$

Contoh soal:

Dari populasi 40.000 karyawan telah diambil sampel secara acak 100 orang untuk diteliti tingkat upahnya. Jika diketahui rata2 tingkat upah seluruh anggota populasi Rp 27.500,- per bulan dengan simpangan baku = Rp 10.000,-

a) Hitung probabilitas sampel tersebut dengan upah antara Rp 25.000,- s/d Rp 30.000 (Hitung peluang karyawan dari sampel tersebut dengan upah antara 25.000 s/d 30.000)P ($25000 \leq X \leq 30000$) = ?

b) Hitung probabilitas sampel tersebut dengan upah paling rendah Rp 20.000,- (Hitung peluang karyawan dari sampel tersebut dengan upah paling rendah 20.000) P ($X \geq 20000$) = ?

c) Tentukan jumlah ukuran sampel (sample zise) apabila dikehendaki perbedaan rata-2 upah untuk tiap dua sampel probabilitas paling besar Rp 500,-

Jawab:

N = 40.000; n = 100; u = 27.500; σ = 10.000

n/N = 100/40.000 = 0,0025 = 0,25 % < 5 % (termasuk sampel kecil)

a) $Z = \dfrac{\overline{X} - u}{\sigma/\sqrt{n}}$

$$\text{Batas bawah} = \frac{25.000 - 27.500}{10.000/\sqrt{100}} = -2,5 \rightarrow = 0,4938$$

$$\text{Batas atas} = \frac{30.000 - 27.500}{10.000/\sqrt{100}} = +2,5 \rightarrow = 0,4938$$

Jadi karyawan dengan upah antara Rp 25.000,- s/d Rp 30.000,- mempunyai peluang 49,38 % + 49,38 % = 98,76 %

b) $$\text{Batas bawah} = \frac{20.000 - 27.500}{10.000/\sqrt{100}} = -7,5 \dashrightarrow = 0,5000$$

Jadi karyawan dengan rata2 upah paling rendah Rp 2.000,- mempunyai peluang 100 %

c) Ukuran sampel (sample zise) dapat dihitung dengan rumus

$$\frac{\sigma}{\sqrt{n}} \leq d \; ; \; \frac{10.000}{\sqrt{n}} \leq 500 \; ; \; \sqrt{n} \geq \frac{10.000}{500} \; \sqrt{n} \geq 20 \; ; \; n \geq 400$$

2. Distribusi Sampling Perbandingan

Distribusi sampling perbandingan p = X/n mempunyai rata2 perbandingan up = π dan simpangan baku perbandingan sbb:

$$\boxed{\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}}$$

Simpangan baku rata-2 untuk n/N $\leq$ 5% (sampel kecil)

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Simpangan baku rata-2 untuk n/N > 5% (sampel besar)

Dalil Limit Pusat:

Jika ukuran sampel n cukup besar maka distribusi sampling perbandingan p = X/n ternyata mendekati distribusi normal dengan

$$Z = \frac{x/n - \pi}{\sigma_p}$$

Jadi

$$Z = \frac{x/n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

untuk sampel kecil

$$Z = \frac{x/n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}}$$

untuk sampel besar

Dari standar baku perbandingan σ_p dapat ditentukan ukuran sampel sample zise) minimum bila perbandingan maksimum yang dikehendaki untuk dua sampel probabilitas diketahui, dimana nilai n dihitung dari:

$$\sqrt{\frac{\pi(1-\pi)}{n}} \leq d$$

Jika π dari populasi tidak diketahui maka digunakan nilai $\pi(1-\pi)$ yang maksimum yakni $\pi(1 - \pi)$= 0,50 * 0,50 = 0,25

Contoh soal:

Dalam setiap pengiriman barang ternyata rata 10 % rusak. Jika pada setiap pengiriman barang diambil sebuah sampel acak terdiri dari 100 unit barang, hitung:

a. Peluang barang rusak dari sampel tersebut paling kecil 15 % Hitung probabilitas sampling tersebut dengan barang rusak paling kecil 15 %) .
b. Berapa ukuran sampel (sample zise) minimal agar prosentase kerusakan yang diharapkan akan berbeda antara tiap dua sampel probabilitas, tidak lebih dari 2 %.

Jawab:

up = π = 0,10 ; N tak terhingga (tidak dibatasi);

n = 100; n/N akan kecil < 5 %

a) P(x $\geq$ 0,15) = ?

$$Z = \frac{x/n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$ untuk sampel kecil

$$Z = \frac{0{,}15 - 0{,}10}{\sqrt{\frac{0{,}10(1-0{,}10)}{100}}} = \frac{0{,}05}{0{,}03} = 1{,}67$$

Z1,67 $\geq$ 0,50 – 0,4525 = 0,0475 = 4,75 %

b) Sample zise dengan d = 0,02

$$\sqrt{\frac{\pi(1 \quad \pi)}{n}} \leq d \; ; \; \sqrt{\frac{0{,}10 \; (1 \; - \; 0{,}10)}{n}} \leq 0{,}02$$

$$\sqrt{\frac{0{,}09}{n}} \leq 0{,}02 \quad ; \quad \frac{0{,}09}{n} \leq 0{,}0004 \quad ; \quad n \geq \frac{0{,}09}{0{,}0004}$$

n $\geq$ 225

3. Distribusi Sampling Selisih Rata-Rata

Untuk mengetahui apakah antara du (2) sampel terdapat perbedaan nilai rata-2 atau tidak.

Dua populasi masing-masing:

N_1 dengan rata-2 populasi u_1 N_2 dengan rata-2 populasi u_2

Dan simpangan baku σ_1 σ_2

Sampel n_1 n_2

Rata-2 sampel $\overline{X}_{1i}$

$\overline{X}_{2j}$

Selisih rata-2 sr $= (\overline{X}_{1j} - \overline{X}_{2j}) = (\overline{X}_1 - \overline{X}_2)$

Rata-2 dari selisih rata-2 u_{sr} $= u_1 - u_2$ atau $= u_2 - u_1$

Simpangan baku selisih rata-rata:

$$\sigma_{sr} = \sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

Dalil limit pusat:

Jika ukuran sampel n_1 dan n_2 cukup besar maka distribusi sampling selisih rata-rata ternyata mendekati distribusi normal dengan:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - u_{sr}}{\sigma_{sr}}$$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (u_1 - u_2)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}}$$

Contoh Soal:

Dari dua populasi lampu dop jenis A dan jenis B akan diteliti rata-rata daya tahan pakai masing-2 produk tersebut. Jika diambil sampel acak dari masing-2 produk sebanyak $n_A = n_B = 125$ sedangkan daya tahan pakai produk A rata-2 1400 jam dengan simpangan baku = 200 jam dan produk B rata-2 daya tahan pakai = 1200 jam dengan simpangan baku = 100.

a. Hitung peluang produk A paling sedikit 300 jam lebih dari B.

   $P(X_A - X_B \geq 300 \text{ jam})$

Jawab:

$n_A = n_B = 125$

$u_A = 1400$ jam $\qquad u_B = 1200$ jam

$\sigma_A = 200$ jam $\qquad \sigma_B = 100$ jam

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (u_1 - u_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$$

$$Z = \frac{(300) - (1400 - 1200)}{\sqrt{\frac{200^2}{125} + \frac{100^2}{125}}} = 5 = 50\%$$

Jadi 0,50 – 0,50 = 0% atau praktis tidak terdapat selisih rata-2 antara kedua lampu tersebut akan lebih dari 300 jam.

4. Distribusi Sampling Selisih Perbandingan

Untuk mengetahui apakah antara dua (2) sampel terdapat perbedaan nilai perbandingan atau tidak.

Dua populasi masing-masing:

N_1 dengan rata-2 populasi	π_1	N_2 dengan rata-2 populasi π_2
Dan simpangan baku	$\sqrt{\pi_1(1-\pi_1)}$	$\sqrt{\pi_2(1-\pi_2)}$
Sampel	n_1	n_2

Selisih perbandingan $sp = \frac{X_1}{n_1} - \frac{X_2}{n_2}$

Rata-2 dari selisih perbandingan $u_{sp} = \pi_1 - \pi_2$

Simpangan baku selisih rata-rata:

$$\sigma_{sp} = \sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}}$$

Jika perbandingan kedua populasi tidak diketahui maka dianggap

$\pi_1 = \pi_2 - \pi$

Dalil Limit Pusat:

Jika ukuran sampel n_1 dan n_2 cukup besar maka distribusi sampling selisih perbandingan ternyata mendekati distribusi normal dengan.

$$Z = \frac{[\frac{X_1}{n_1} - \frac{X_2}{n_2}] - u_{sp}}{\sigma_{sr}}$$

$$Z = \frac{[\frac{X_1}{n_1} - \frac{X_2}{n_2}] - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

Contoh:

Produk A dihasilkan oleh perusahaan 1 dan 2

Tingkat kerusakan perusahaan 1 $\pi_1 = 5\%$

Tingkat kerusakan perusahaan 2 $\pi_2 = 4\%$

Jika diambil sampel acak $n_1 = n_2 = 100$ unit barang

Hitung: peluang kerusakan barang yang dihasilkan oleh perusahaan 1 akan berbeda tidak lebih dari 0,5% bila dibandingkan kerusakan barang yang dihasilkan pada perusahaan 2.

$$P\,[(\frac{x_1}{n_1} - \frac{x_2}{n_{21}}) \leq 0{,}005] = ?$$

Jawab:

$$Z = \frac{[\frac{X_1}{n_1} - \frac{X_2}{n_2}] - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

$$Z = \frac{[0{,}005] - (0{,}05 - 0{,}04)}{\sqrt{\frac{0{,}05(1-0{,}05)}{100} + \frac{0{,}04(1-0{,}04)}{100}}}$$

$$Z = \frac{[0{,}005] - (0{,}05 - 0{,}04)}{\sqrt{\frac{0{,}05(0{,}95)}{100} + \frac{0{,}04(0{,}96)}{100}}} = -0{,}17 = 6{,}75\%$$

Jadi:

$$P\,[(\frac{X_1}{n_1} - \frac{X_2}{n_{21}}) \leq 0{,}005] = 0{,}50 - 6{,}75\ \% = 43{,}25\ \%$$

BAB II

ANALISA REGRESI DAN KORELASI SEDERHANA

A. PENDAHULUAN

Analisa Regresi menyatakan bentuk hubungan dan pengaruh variabel bebas terhadap variabel tak bebas. Bentuk hubungan dinyatakan dalam model persamaan regresi yang signifikan dimana variabel tak bebas (Y) merupakan fungsi dari variabel bebas (X). Jadi $Y = f(X_1, X_2, X_3, \ldots X_n)$.

Sedangkan pengaruh ditunjukkan oleh tanda (+/-) dan besarnya koefisien arah regresi. Tanda + menyatakan pengaruh searah, sedangkan tanda - menyatakan pengaruh berlawanan arah.

Interpretasi koefisien arah regresi tergantung pada bentuk persamaan regresi itu sendiri, misalnya untuk persamaan linear maka koefisien arah menyatakan pengaruh marginal $= \delta Y / \delta X$ sedangkan untuk persamaan **Cobb-Douglass** menyatakan pengaruh elastisitas = marginal/ rata2 $= \delta Y / \delta X : Y / X$

Diperlukan dasar-dasar teoritis dan pengetahuan tentang hubungan kausal antar variabel sesuai masalah yang dipelajari guna mengklasifikasi variabel ke dalam bentuk bebas dan tidak bebas. Jadi telah diketahui variabel mana yang variasinya dipengaruhi/ bergantung pada variabel lainnya (dependent variable) dan variabel mana yang mempengaruhinya (independent variable).

Analisia Regresi berbeda dengan analisa Varians karena tujuan analisa tersebut berbeda. Dalam analisa varians kita tidak mencari bentuk hubungan antar variabel, melainkan membandingkan efek dari variabel-2 tersebut. Walaupun demikian terdapat hubungan antara analisa regresi dengan analisa varian, bahkan analisa varian (ANAVA) digunakan untuk menguji signifikansi dari suatu model regresi. Disamping itu digunakan juga uji t untuk menguji koefisien regresi parsial.

Analisa Korelasi menyatakan derajad keeratan hubungan antar variabel yang dikemukakan dalam %, disamping itu menyatakan juga arah hubungan antar variabel yang dikemukakan dalam tanda +/-. Tanda (+) menyatakan hubungan searah sedangkan tanda (-) menyatakan hubungan berlawanan arah (hubungan terbalik). Nilai korelasi (r) juga diuji dengan uji t.

Dalam analisa korelasi tidak terdapat perbedaan yang tegas antara variabel bebas maupun tak bebas.

Analisis regresi dan korelasi memiliki banyak kesamaan terutama dalam teknik-2 perhitungannya.

Perlu diingat bahwa korelasi berhubungan langsung dengan bentuk persamaan regresi atau bentuk regresi menentukan nilai koefisien korelasi.

Analisa regresi dapat diklasifikasikan atas dasar:

1) **Jumlah variabel bebas, meliputi:**
   a) Regresi sederhana bila hanya menganalisis satu variabel bebas
   b) Regresi berganda bila menganalisis lebih dari satu variabel bebas
2) **Bentuk persamaan regresi, meliputi:**
   a) Regresi linear bila pengaruh variabel bebas terhadap variabel tidak bebas bersifat konstan *(constant rate)*
   b) Regresi non-linear bila pengaruh variabel bebas terhadap variabel tidak bebas tidak bersifat konstan (misal increasing rate atau decreasing rate).

Secara garis besar ada 4 macam analisa regresi, yaitu:

1) Regresi linear sederhana
2) Regresi linear berganda
3) Regresi non linear sederhana
4) Regresi non linear berganda

B. REGRESI LINEAR SEDERHANA

Regresi linear sederhana mempelajari bentuk hubungan dan pengaruh yang diduga bersifat konstan antara satu variabel bebas (X) terhadap variabel tak bebas (Y). Misal, analisis regresi linear sederhana antara variabel bebas/ independent jumlah pendapatan mingguan X_i terhadap belanja konsumsi keluarga sebagai variabel terikat/ dependent Y_i dari 10 keluarga sampel di desa A dengan data sebagai berikut:

Tabel : 1

Regresi luas lahan (X) terhadap biaya produksi (Y)

N	Y_i	X_i	$\hat{Y} = Y_i - e$ $\hat{Y}_i = b_0 + b_1 X$ $Y = 8{,}58 + 62{,}08\, X$	$e = Y_i - \hat{Y}$
1	59,2	0,7	52,0	7,2
2	97,8	1,5	101,7	- 3,9
3	98,6	1,9	126,5	-27,9
4	38,2	0,5	39,6	- 1,4
5	14,4	0,2	21,0	- 6,6
6	159,6	2,1	138,9	20,7
7	37,0	0,5	39,6	- 2,6
8	17,7	0,2	21,0	- 3,3
9	26,1	0,4	33,4	- 7,3
10	5,4	0,1	14,8	- 9,4
11	37,0	0,4	33,4	3,6
12	38,6	0,4	33,4	5,2
13	19,4	0,3	27,2	- 7,8
14	8,1	0,1	14,8	- 6,7
15	22,3	0,2	21,0	1,3
16	8,3	0,1	14,8	- 6,5
17	43,8	0,3	27,2	16,6
18	51,6	0,4	33,4	18,2
19	50.1	0,4	33,4	16,7
20	8,9	0,1	14,8	- 5,9

Untuk memperkirakan model regresi, yang dilakukan pertama kali adalah melihat distribusi data dari diagram pencar (scatter diagram) dengan cara ploting titik-titik yang menghubungkan antara total biaya produksi (sumbu Y) dengan luas lahan (sumbu X). Dari diagram pencar tampak tendensi model penyebaran data apakah linier atau non-linier.

Titik-titik tersebut bisa terletak dalam satu garis/ kurva, namun dalam prakteknya terdapat berbagai kemungkinan bentuk/ model kurva yang dapat dibuat diantara titik-titik tersebut dan titik diagram pencar tidak terletak pada satu garis.

1. Metode Least Square

Menurut teori regresi bahwa garis yang paling mewakili ialah garis yang dibuat sedemikian rupa sehingga total errornya yakni: $e = \sum (Y_i - \hat{Y})$ yang terjadi dapat ditekan sekecil mungkin.

Terdapat 2 teori yakni *Least Square Method* dan *Maximum Likelihood Estimation* yang membuktikan bahwa minimisasi jumlah kuadrat dari error merupakan teknik estimasi yang terbaik. Disini kita hanya membicarakan Metode Jumlah Kuadrat Terkecil *(Least Square Method)* karena perhitungannya lebih sederhana.

Metode Least Square digunakan untuk meminimumkan jumlah kuadrat dari error yakni:

$$\boxed{\sum (Y_i - \hat{Y})^2 \text{ ----> minimum}}$$

Beberapa Keunggulan Metode Least Square:

a. Dengan cara mengkuadratkan maka semua error akan positip

b. Dengan mengkuadratkan maka nilai error yang kecil akan diperbesar dan bila nilai ini diminimumkan maka garis regresi yg dihasilkan akan mendekati ketepatan sebagai penduga.

c. Perhitungan aljabarnya cukup sederhana

Jika diagram pencar dari data luas lahan (X) dan total biaya produksi (Y) di atas bertendensi linear maka model regresi yang digunakan adalah regresi linear sederhana, dengan formula umum:

$$Y = \beta_o + \beta_1 X + e$$

β_o dan β_1 adalah koefisien dari persamaan regresi yang merupakan bilangan tetap yang nilainya akan diestimasi.

β_o disebut koefisien intersep regresi

β_1 disebut koefisien arah regresi

Estimasi dengan metode least square melalui perhitungan sbb:

Karena $\hat{Y} = \beta_o + \beta_1 X$

Sehingga besarnya jumlah kuadrat error e adalah:

$$S = \sum e_i^2 = \sum (Y_i - \beta_o - \beta_1 X)^2$$

Agar persamaan S minimum maka turunan pertamanya terhadap β_o dan β_1 harus = 0

$S \quad = \sum (Y - \beta_o - \beta_1 X)^2$

$S \quad = \sum (Y^2 - 2\beta_o Y - 2\beta_1 XY + \beta_o^2 + 2\beta_o \beta_1 X + \beta_1^2 X^2)$

$S \quad = \sum Y^2 - 2\beta_o \sum Y - 2\beta_1 \sum XY + n\beta_o^2 + 2\beta_o \beta_1 \sum X + \beta_1^2 \sum X^2$

Agar S minimum maka

$\delta S / \delta\beta_o = 0$ jadi $-2 \sum Y + 2 n\beta_o + 2 \beta_1 \sum X = 0 \,.\, (-½)$

$$- \sum Y + n \beta_o + \beta_1 \sum X = 0$$

$$n \beta_o + \beta_1 \sum X = \sum Y$$

$$\beta_o = \sum Y/n - \beta_1 \sum X/n$$

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}$$

$\delta S / \delta\beta_1 = 0$ jadi $-2\sum XY + 2\beta_o \sum X + 2\beta_1 \sum X^2 = 0 \;.\; (-½)$

$$\sum XY - \beta_o \sum X - \beta_1 \sum X^2 = 0$$

$$\sum XY - (\sum Y/n - \beta_1 \sum X/n)\sum X - \beta_1 \sum X^2 = 0$$

$$\sum XY - \sum X \sum Y/n + \beta_1 (\sum X)^2/n - \beta_1 \sum X^2 = 0$$

$$\beta_1 \sum X^2 - \beta_1 (\sum X)^2/n = \sum XY - \sum X \sum Y/n$$

$$\beta_1 [\sum X^2 - (\sum X)^2/n] = \sum XY - (\sum X \sum Y)/n$$

$$\beta_1 = \frac{\sum XY - (\sum X \sum Y)/n}{\sum X^2 - (\sum X)^2/n}$$

Atau, jika notasi diganti dengan b_o dan b_1 maka dapat dicari secara simultan sebagai berikut

$\sum (Y - b_o - b_1 X) = 0$

$\sum Y - nb_o - b_1 \sum X = 0$

$nb_o + b_1 \sum X = \sum Y$ ....(1)

$\sum X (Y - b_o - b_1 X) = 0$

$\sum XY - b_o \sum X - b_1 \sum X^2 = 0$

$b_o \sum X + b_1 \sum X^2 = \sum XY$ ....(2)

Dari persamaan (1) $\quad nb_o + b_1 \sum X = \sum Y$

dan persamaan (2) $\quad b_o \sum X + b_1 \sum X^2 = \sum XY$

---

dapat dicari b_o dan b_1 sehingga diperoleh hasil sbb:

$$b_1 = \frac{\Sigma X\,Y - [(\Sigma X)(\Sigma Y)]/n}{\Sigma X^2 - (\Sigma X)^2/n} \bullet \frac{n}{n}$$

atau $$b_1 = \frac{n\,\Sigma X\,Y - \Sigma X\,\Sigma Y}{n\,\Sigma X^2 - (\Sigma X)^2}$$

atau $$b_1 = \frac{\Sigma (X - \overline{X})(Y - \overline{Y})}{\Sigma (X - \overline{X})^2}$$

atau $$b_1 = \frac{\Sigma\, x_i\, y_i}{\Sigma\, x_i^{\,2}}$$

dimana $x = X - \overline{X}$ dan $y = Y - \overline{Y}$

Sedangkan b_0 diperoleh dari persamaan (1), yakni:

$$nb_0 + b_1 \sum X = \sum Y \text{ ....(1)}$$

$$b_0 = \frac{\Sigma Y - b_1\, \Sigma X}{n} = \overline{Y} - b_1 \overline{X}$$

Jadi model persamaan regresi linear sederhana yang dicari adalah:

$$\boxed{Y = b_o + b_1 X}$$

atau $Y = \overline{Y} - b_1 \overline{X} + b_1 X$

$Y = \overline{Y} + b_1 (X - \overline{X})$

Misal analisis regresi linear sederhana antara variabel independent luas lahan X (ha) dengan biaya produksi (real cost) Y (Rp.000,-) dari 20 petani sampel di desa A dengan data sbb:

Tabel 1. Regresi luas lahan X terhadap biaya produksi Y.

N	Y	X	XY	X^2	$X - \bar{X}$ = x	$Y - \bar{Y}$ = y	$(X - \bar{X})(Y - \bar{Y})$ = (xy)	$(X - \bar{X})^2$ = x^2
1	59,2	0,7	41,44	0,49	0,16	17,095	2,7352	0,0256
2	97,8	1,5	146,70	2,25	0,96	55,695	53,4672	0,9216
3	98,6	1,9	187,34	3,61	1,36	56,495	76,8332	1.8496
4	38,2	0,5	19,10	0,25	-0,04	-3,905	0,1562	0,0016
5	14,4	0,2	2,88	0,04	-0,34	-27,705	9,4197	0,1156
6	159,6	2,1	335,16	4,41	1,56	117,495	183,2922	2,4336
7	37,0	0,5	18,50	0,25	-0,04	-5,105	0,2042	0,0016
8	17,7	0,2	3,54	0,04	-0,34	-24,405	8,2977	0,1156
9	26,1	0,4	10,44	0,16	-0,14	-16,005	2,2407	0,0196
10	5,4	0,1	0,54	0,01	-0,44	-36,705	16,1502	0,1936
11	37,0	0,4	14,80	0,16	-0,14	-5,105	0,7147	0,0196
12	38,6	0,4	15,44	0,16	-0,14	3,505	0,4907	0,0196
13	19,4	0,3	5,82	0,09	-0,24	-22,705	5,4492	0,0576
14	8,1	0,1	0,81	0,01	-0,44	-34,005	14,9622	0,1936
15	22,3	0,2	4,46	0,04	-0,34	-19,805	6,7337	0,1156
16	8,3	0,1	0,83	0,01	-0,44	-33,805	14,8742	0,1936
17	43,8	0,3	13,14	0,09	-0,24	1,695	0,4068	0,0576
18	51,6	0,4	20,64	0,16	-0,14	9,495	1,3293	0,0196
19	50.1	0,4	20,04	0,16	-0,14	7,995	1,1193	0,0196
20	8,9	0,1	0,89	0,01	-0,44	-33,205	14,6102	0,1936
JMH	842,1	10,8	862,51	12,40	0	0	413,4868	6,568
MEAN	42,105	0,54						

$\sum X = 10,8 \qquad \bar{X} = 0,54 \qquad \sum Y = 842,1 \qquad \bar{Y} = 42,105$

$\sum X^2 = 12,4 \qquad \sum XY = 862,51$

$\sum(X)^2 = (10,8)^2 = 116,64 \qquad \sum X \sum Y = (10,8)(842,1) = 9094.68$

Koefisien b_1 dapat dicari dengan menggunakan rumus:

$$b1 = \frac{\sum XY - [(\sum X)(\sum Y)]/n}{\sum X^2 - (\sum X)^2/n}$$

$$= \frac{862,51 - 9094,68/20}{12,4 - 116,64/20} = \frac{407,776}{6,568} = 62,08$$

atau $$b_1 = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{20\,(862{,}51) - 9094{,}68}{20\,(12{,}4) - 116{,}64} = \frac{8155{,}52}{131{,}36} = 62{,}08$$

atau $\quad b1 = \dfrac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \dfrac{413{,}4868}{6{,}568} = 62{,}95$

atau $\quad b1 = \dfrac{\sum xy}{\sum x^2}$

Sedangkan bo diperoleh dari persamaan (1), yakni:

$$nb_o + b_1 \sum X = \sum Y$$

$$bo = \frac{\sum Y - b_1 \sum X}{n} = \quad b_0 = \frac{\sum Y - b_1 \sum X}{n} = \bar{Y} - b_1 \bar{X}$$

$$= 42{,}105 - 62{,}08\,(0{,}54) = 8{,}58$$

Jadi model persamaan regresi linear sederhana yang dicari adalah: $Y = b_o + b_1 X$

$Y = 8{,}58 + 62{,}08\,X$

atau $\quad Y = b_o + b_1 X \qquad$ karena $bo = \bar{Y} - b_1 \bar{X}$

maka $\quad Y = \bar{Y} - b_1 \bar{X} + b_1 X$

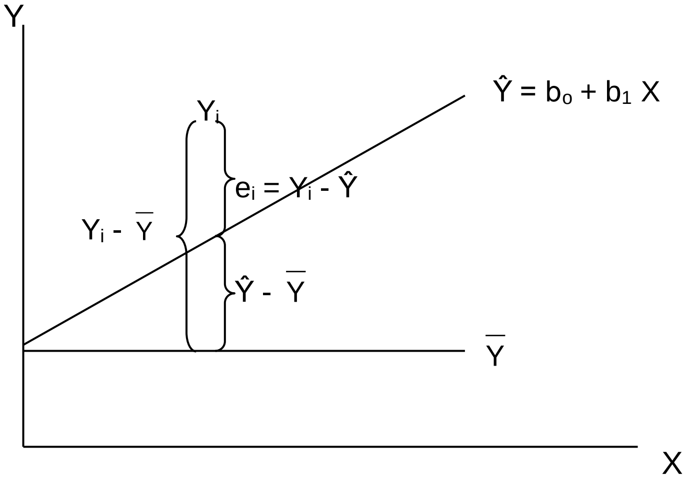
$Y = \bar{Y} + b_1 (X - \bar{X})$

$Y = 42{,}105 + 62{,}08\,(X - 0{,}54)$

$Y = 42{,}105 + 62{,}08\,X - 33{,}5232$

$Y = 8{,}58 + 62{,}08\,X$

2. Presisi Persamaan Regresi



Tampak terjadi hubungan bahwa:

Error = Total - Regresi

$(Y - \hat{Y}) = (Y - \bar{Y}) - (\hat{Y} - \bar{Y})$ Jumlah kuadratnya adalah:

$$\sum (Y - \hat{Y})^2 = \sum [(Y - \bar{Y}) - (\hat{Y} - \bar{Y})]^2$$

$$\sum (Y - \hat{Y})^2 = \sum [(Y - \bar{Y})^2 - 2(Y - \bar{Y})(\hat{Y} - \bar{Y}) + (\hat{Y} - \bar{Y})^2]$$

$$\sum (Y - \hat{Y})^2 = \sum (Y - \bar{Y})^2 - 2\sum(Y - \bar{Y})(\hat{Y} - \bar{Y}) + \sum(\hat{Y} - \bar{Y})^2$$

Karena $\hat{Y} = \bar{Y} + b_1 (X - \bar{X})$

$\hat{Y} - \bar{Y} = b_1 (X - \bar{X})$ maka

$$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - 2\sum(Y - \bar{Y})\, b_1(X - \bar{X}) + \sum(Y - \bar{Y})^2$$

$$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - 2b_1 \sum(Y - \bar{Y})(X - \bar{X}) + \sum(\hat{Y} - \bar{Y})^2$$

Karena $b_1 = \dfrac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$ atau

$$\sum(Y - \bar{Y})(X - \bar{X}) = b_1 \sum(X - \bar{X})^2$$

$= \sum b_1 (X - \bar{X})^2$ maka

$$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - 2\sum b_1^2 (X - \bar{X})^2 + \sum(\hat{Y} - \bar{Y})^2$$

Karena $(\hat{Y} - \bar{Y})^2 = b_1^2 (X - \bar{X})^2$ maka

$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - 2\sum(\hat{Y} - \bar{Y})^2 + \sum(\hat{Y} - \bar{Y})^2$

$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - \sum(\hat{Y} - \bar{Y})^2$ atau

$\sum(Y - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y - \hat{Y})^2$ yakni

SS Total = SS Regresi + SS Error

Suatu garis regresi dikatakan sebagai penduga yang baik jika jumlah Kuadrat Regresinya cukup besar atau

$$R^2 = \frac{\text{SS Regres}}{\text{SS Total}} \text{ mendekati } 1$$

Besarnya derajad bebas (df) dari setiap Jumlah Kuadrat di atas sbb:

SS Total = SS Regresi + SS Error

(n-1) = (k) + (n-k-1)

dimana:

n = jumlah pengamatan/ sampel

k = jumlah variabel bebas

Tabel Analisa Varians (Anava) dari analisis regresi linear sederhana

Sumber Variasi	db	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Total} 0,05 0,01
Regresi Error Total	1 n-2 n-1	$b_1 \sum xy$ SST – SSE $\sum y^2$	MSR= SSR/1 S^2=SSE/(n-2)	MSR/MSE	

SS Total = SS Regresi + SS Error

$\sum(Y - Y)^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y - \hat{Y})^2$ yakni

SS Total $= \sum(Y - \bar{Y})^2 = \sum y^2$

SS Regresi $= \sum (\hat{Y} - \bar{Y})^2 = \sum b_1^2 (X - \bar{X})^2 = b_0 \sum b_1(X - \bar{X})^2$

Karena $b_1 = \dfrac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$ maka

$$\text{SS Regresi} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \sum b_1 (X - \bar{X})^2$$

$$= \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} b_1 \sum(X - \bar{X})^2$$

$$= b_1 \sum(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= b_1 \sum x y$$

Tabel analisa varians (Anava) dari analisis regresi linier sederhana untuk data luas lahan dan biaya produksi.

Sumber Variasi	db	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Total} 0,05 0,01
Regresi	1	25316,8797	25316,8797	176,17**	
Error	18	2586,6898	143,7050		
Total	19	27903,5695			

3. Asumsi Analisa Regresi

a. $E(e_i) = 0$ dan $V(e_i) = \sigma^2$

Artinya e_i adalah variabel random dengan rata-2 = 0 dan varians $= \sigma^2$

b. $Cov(e_i, e_j) = 0$

Artinya tidak ada korelasi antara e_i dan e_j untuk $i \neq j$. Jadi $E(Y_i) = \beta_0 + \beta_1 X_1$ dengan $V(Y_i)$ juga $= \sigma^2$ dan tidak ada korelasi antara Y_i dan Y_j untuk $i \neq j$.

c. $e_i \infty N(0, \sigma^2)$

Artinya ei berdistribusi normal dengan rata-2 = 0 dan varians σ^2 akibatnya e_i dan e_j bukan saja tidak berkorelasi tetapi juga independent (tidak saling tergantung)

4. Contoh Regresi Linier Sederhana

Dari hasil penelitian pengaruh pendapatan mingguan (X) terhadap belanja konsumsi mingguan (Y) 10 sampel keluarga sbb:

Tabel 2. Regresi pendapatan X terhadap belanja konsumsi Y.

N	Y	X	XY	X^2	$X-\bar{X}$ $= x$	$Y-\bar{Y}$ $= y$	$(X-\bar{X})(Y-\bar{Y})$ $= xy$	$(X-\bar{X})^2$ $= x^2$	Y^2
1	70	80	5600	6400	-90	-41	3690	8100	1681
2	65	100	6500	10000	-70	-46	3220	4900	2116
3	90	120	10800	14400	-50	-21	1050	2500	441
4	95	140	13300	19600	-30	-16	480	900	256
5	110	160	17600	25600	-10	- 1	10	100	1
6	115	180	20700	32400	10	4	40	100	16
7	120	200	24000	40000	30	9	270	900	81
8	140	220	30800	48400	50	29	1450	2500	841
9	155	240	37200	57600	70	44	3080	4900	1936
10	150	260	39000	67600	90	39	3510	8100	1521
JML	1110	1700	205500	322000	0	0	16800	33000	8890
MEAN	111	170							

$\Sigma X = 1700$ $\quad \bar{X} = 170$ $\quad \Sigma Y = 1110$ $\quad \bar{Y} = 111$

$\mathbf{\Sigma X^2 = 322000}$ $\quad \mathbf{\Sigma XY = 205500}$

$(\Sigma X)^2 = (1700)^2 = 2890000$ $\quad \Sigma X \, \Sigma Y = (1700)(1110) = 1887000$

$\Sigma (X-\bar{X})(Y-\bar{Y}) = 16800$ $\quad \Sigma (X_i - \bar{X})^2 = 33000$

Koefisien b_1 dapat dicari dengan menggunakan rumus:

$$b_1 = \frac{\Sigma XY - [(\Sigma X)(\Sigma Y)]/n}{\Sigma X_i^{\,2} - (\Sigma X)^2/n}$$

$$= \frac{205500 - 1887000/10}{322000 - 2890000/10} = \frac{16800}{33000} = 0{,}50$$

atau

$$b_1 = \frac{n\ \Sigma XY - \Sigma X\ \Sigma Y}{n\ \Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{10(205500) - 1887000}{10(322000) - 2890000} = \frac{16800}{33000} = 0{,}509$$

atau

$$b_1 = \frac{\Sigma\,(X - \overline{X})\,(Y - \overline{Y})}{\Sigma(X - \overline{X})^2} = \frac{16800}{33000} = 0{,}509$$

atau

$$b_1 = \frac{\Sigma\, xy}{\Sigma\, x^2} = \frac{16800}{33000} = 0{,}509$$

Sedangkan b_0 diperoleh dari persamaan (1), yakni:

$nb_0 + b_1\ \Sigma X = \Sigma Y$

$$b_0 = \frac{\Sigma Y - b_1\ \Sigma X}{n} = \overline{Y} - b_1 \overline{X}$$

$$= 111 - 0{,}509\ (170) = 24{,}47$$

Jadi model persamaan regresi linear sederhana yang dicari adalah:

$Y = b_o + b_1 X$

$Y = 24{,}47 + 0{,}509\ X$

atau $Y = b_o + b_1 X$ karena $b_o = \overline{Y} - b_1 \overline{X}$

maka $Y = \overline{Y} - b_1 \overline{X} + b_1 X$

$Y = Y + b_1 (X - \overline{X})$

$Y = 111 + 0{,}509\ (X - 170)$

$Y = 111 + 0{,}509\ X - 86{,}53$

$Y = 24{,}47 + 0{,}509\ X$

5. Analisa Varians untuk Uji F

Anava untuk regresi linear sederhana dari data pendapatan (X) dengan pengeluaran konsumsi (Y)

Sumber Variasi S.V	Db Df	Jumlah Kuadrat SS	Rata-rata Kuadrat MS	F_{hitung}	F_{tabel} 0,05 0,01
Regresi Error Total	1 8 9	8551,20 338,80 8890,00	8551,20 42,35	201,92	

JK Total = JK Regresi + JK Error

$\sum (Y - Y)^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2$ yakni

JK Total $= \sum (Y - \bar{Y})^2 = \sum y_i^2 = 8890$

JK Regresi $= \sum (\hat{Y} - \bar{Y})^2 = b_1 \sum(X - \bar{X})(Y - \bar{Y})$

$= b_1 \sum x\, y = 0{,}509\ (16800) = 8551{,}2$

Hipotesis untuk uji F overall

Ho : $\beta_1 = 0$

Ha : $\beta_1 \neq 0$

F hitung = MSR/MSE dengan db = (1; 8)

F hitung = MSR/MSE = 8551,20/ 42,35 = 201,92

Ftabel 0,95 (1; 8) = 5,32

Ftabel 0,99 (1; 8) = 11,26

C.

D. Karena F_{hitung} = 201,92* > F_{tabel} 0,99 (1; 8) = 11,26 maka disimpulkan bahwa regresi tersebut sangat berbeda nyata sekali pada tingkat kepercayaan 99 % sehingga dapat digunakan**

sebagai model untuk memprediksi pengaruh variabel X (pendapatan) dengan variabel Y (pengeluaran konsumsi)

6. Varians dan Standar Error untuk uji t

$$b1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

$\sum (X\,Y - \bar{X}\,Y - \bar{Y}\,X + \bar{X}\,\bar{Y})$

$= \sum (X_i\,Y_i - X\,Y_i - X_i\,Y - X\,Y)$

$= \sum [(X_i\,Y_i - X\,Y_i) - (X_i\,Y - X\,Y)]$

$= \sum [(X_i - X)\,Y_i - (X_i - X)\,Y]$

$\sum (X_i - X)\,Y$

$= Y \sum (X_i - X)$

$= Y (\sum X_i - n\,X) \quad \text{--->} \sum X_i = n \dfrac{\sum X_i}{n} = n\,X$

$= Y (\,n\,X - n\,X) = 0$

$$b1 = \frac{\sum (X_i - X)Y_i}{\sum (X_i - X)^2} = \frac{\sum (X_i - X)}{\sum (X_i - X)^2}$$

$$= \frac{1}{\sum (X_i - X)}\;Y_i$$

Jika fungsi F $= a_1\,Y_1 + a_2\,Y_2 + ... + a_n\,Y_n$

maka $V(F) = a_1^2\,V(Y_1) + a_2^2\,V(Y_2) + ... + a_n^2\,V(Y_n)$

$= a_1^2\,V(Y_1) + a_2^2\,V(Y_2) + ... + a_n^2\,V(Y_n)$

$= \sum a_i^2\,V(Y_i)$

$V(Y_i) = \sigma^2 \qquad = (\sum a_i^2)\,\sigma^2$

a. Varians b_1 atau $V(b_1)$

$$V(b_1) = \frac{1}{\sum (X_i - X)^2} \sigma y.x^2 = \frac{\sigma y.x^2}{\sum (X_i - X)^2} = \frac{s_y.x^2}{\sum x_i^2}$$

$$V(b_1) = \frac{S_y.x^2}{\sum x_i^2} = \frac{MSE}{\sum X_i^2 - (\sum x_i)^2/n} = \frac{42{,}35}{33000} = 0{,}0013$$

b. Standar error b_1 atau $S.e(b_1)$ atau $s(b_1)$ = akar dari $V(b_1)$

$$s.e(b_1) = \sqrt{\frac{S_y.x^2}{\sum (X_i - X)^2}} = \frac{S_y.x}{\sqrt{\sum (X_i - X)^2}}$$

$$s.e(b_1) = \sqrt{V(b_1)} = \sqrt{0{,}0013} = 0{,}036$$

c. Confidence limit (batas kepercayaan) untuk α_1

$$= b_1 \pm \frac{t\,(n - 2;\ \tfrac{1}{2}\,\alpha)\ Sy.x}{\sqrt{\sum (Xi - X)^2}}$$

$$= b_1 \pm t\,(n\text{-}2;\ 1/2\alpha)\ S.e\,(b_1)$$

Jika $\alpha = 0{,}05$

$= b_1 \pm t\,(8;\ 1/2 * 0{,}05)\ S.e(b_1)$

$= 0{,}509 \pm t\,(8;\ 1/2 * 0{,}05)\ 0{,}036$

$t_{tabel}\ (8;\ 0{,}025) = 2{,}306$

$= 0{,}509 \pm t\,(8;\ 0{,}025)\ 0{,}036$

$= 0{,}509 \pm 2{,}306\ (0{,}036)$

$= 0{,}509 \pm 0{,}0830$

Jadi confidence limit untuk β_1 adalah:

$0,426 \leq \beta_1 \leq 0,592$

d. Hipotesis untuk uji t parsial terhadap β_1

Ho : $\beta_1 = 0$

Ha : $\beta_1 \neq 0$

Uji t ;

$$t_{hitung} = \frac{(b_1 - \beta)}{S.e\,(b_1)} = \frac{(b_1 - \beta)}{\dfrac{S_{y.x}}{\sqrt{\Sigma(X_{1i} - X_1)}}} = \frac{(b_1 - \beta)\,S_x}{S_{y.x}}\sqrt{n-1}$$

$t_{hitung} = b_1 / S.e\,(b_1) = 0,509/\ 0,036 = 14,1388.$

Hasil t_{hitung} dibandingkan dgn t_{tabel} untuk d.f = n-k-1 = n-2 dengan taraf nyata *(level of significance)* p = 100(1-α) %. Karena Hipotesis menyatakan sama dengan atau Ho = 0 maka digunakan uji dua pihak p = 1 - 1/2α dimana α simetris 1/2 α dipihak kiri dan 1/2 α dipihak kanan.

Jika α = 0,05 taraf nyata (level of significance) = 100(1-0,05) = 100 (0,95) % = 95 %

t_{tabel} (n-2; 1/2 * 0,05) yakni t_{tabel} (8; 0,025) = 2,306

Jika α = 0,01 atau taraf nyata = 100(1-0,01) = 100 (0,99) = 99%.

t_{tabel} (n-2; 1/2 * 0,01) yakni t_{tabel} (8; 0,005) = 3,355

Karena t_{hitung} = 14,1388*** > t_{tabel} (8; 0,005) = 3,355 maka disimpulkan bahwa koefisien regresi b_1 secara parsial sangat berbeda nyata sekali pada tingkat kepercayaan 99 %.

Artinya Ho: $\alpha_1 = 0$

Ho diterima atau Ha ditolak

Ha: $\alpha_1 \neq 0$

Ho ditolak atau Ha diterima

Nilai t $_{hitung}$ bisa + atau - tergantung nilai t = 14,208 terletak dalam ½ α -- ½ α Jadi terdapat pengaruh antara X (pendapatan) dgn belanja konsumsi (Y).

Uji parsial ini akan berguna untuk analisis regresi berganda guna melihat variabel manakah yang secara parsial lebih berpengaruh dibandingkan variabel lainnya.

Walaupun uji F overall non-significance masih ada kemungkinan diantara variabel regresi berganda yang significance dalam uji t partial.

Jika uji F significance dalam regresi linear sederhana maka secara otomatis uji t nya juga significance, dan sebaliknya jika nonsignificance.

C. KOLEKSI RUMUS DAN PERHITUNGAN UNTUK MENENTUKAN STANDAR DEVIASI, VARIANS DAN STANDAR ERROR UNTUK REGRESI LINIER SEDERHANA

Tabel 2. Regresi pendapatan X_1 terhadap belanja konsumsi Y.

N	Y_i	X_{1i}	$X_{1i}Y_i$	X_{1i}^2	$\bar{Y}$	$Yi-\bar{Y}$	$(Yi-\bar{Y})^2$ e	$(Yi-\bar{Y})^2$ e^2
1	70	80	5600	6400	4900	65,182	4,818	23,213124
2	65	100	6500	10000	4225	75,364	-10,364	107,412496
3	90	120	10800	14400	8100	85,545	4,455	19,847025
4	95	140	13300	19600	9025	95,727	- 0,727	0,528529
5	110	160	17600	25600	12100	105,909	4,091	16,736281
6	115	180	20700	32400	13225	116,091	- 1,091	1,190281
7	120	200	24000	40000	14400	126,276	- 6,273	39,350529
8	140	220	30800	48400	19600	136,455	3,545	12,567025
9	155	240	37200	57600	24025	146,636	8,364	69,956496
10	150	260	39000	67600	22500	156,818	- 6,818	46,485124
JML	1110	1700	205500	322000	132100			337,286910
Mean	111	170						

$\sum X_{1i} = 1700$ $\quad X_1 = 170$ $\quad \sum Y_i = 1110$ $\quad Y = 111$

$\sum X_{1i}^2 = 322000$ $\quad \sum X_{1i} Y_i = 205500$

$(\sum X_{1i})^2 = (1700)^2 = 2890000$ $\quad \sum X_{1i} \sum Y_i = (1700)(1110) = 1887000$

$\sum (X_{1i} - X1)(Y_i - Y) = 16800$ $\quad \sum (X_{1i} - X_1)^2 = 33000$

$\sum Y_i^2 = 132100$

Koefisien regresi b_o dan b_1 adalah $Y = 24,47 + 0,509\ X_1$

Varians dan standar error s.e b_1 dicari dari titik taksiran yakni dari

Standar deviasi (Simpangan Baku Taksiran) $s_{y.x1}$

Standar Deviasi untuk Estimator (Taksiran) $s_{y.x1}$

$$s_{y.x1} = \sqrt{\frac{\sum (Y_i - Y)^2}{n - 2}} = \frac{337,286910}{8} = 6,50$$

$$\text{atau } s_{y.x1} = \sqrt{\frac{\sum Yi^2 - a \sum Yi - b \sum X_{1i} Y_i}{n - 2}}$$

$$= \frac{132100 - 24,47\ (1110) - 0,509\ (205500)}{8}$$

$$= 6,507687761$$

$$s_{y.x}^2 = 42,35$$

$$\text{atau } sy.x = \sqrt{MSE} = \sqrt{s_{y.x}^2}$$

$$= \sqrt{SSE/\ dfE}$$

$$= \frac{SST - SSR}{dfE}$$

$$= \sqrt{\frac{\sum y_i^2 - b1 \sum X1i\ yi}{n - 2}} = \frac{8890 - 0,509\ (16800)}{8}$$

$$= \sqrt{42,35} = 6,507687761$$

MSE = SSE/ dfE

SSE = SST – SSR

$$MSE = \frac{\sum Y_i^2 - b_1 \sum X_{1i} Y_i}{n - 2} = \frac{8890 - 0{,}509\ (16800)}{8} = 42{,}35$$

$sy.x^2$ sebagai penduga terhadap $\sigma y.x^2$ dapat juga dicari dengan rumus:

$$sy.x^2 = \frac{1}{n-2} \left[\sum Yi^2 - \frac{(\sum Yi)^2}{n} - b1 \left\{ \sum X1i\ Yi - \frac{\sum X1i \sum Yi}{n} \right\}\right]$$

$$= \frac{1}{8} \left[132100 - \frac{(1110)^2}{10} - 0{,}509 \left\{205500 - \frac{(1700)(1110)}{10}\right\}\right]$$

$$= \frac{1}{8} [132100 - 123210 - 0{,}509 \{205500 - 188700\}] = 42{,}35$$

$$\text{Rumus umum } s_{y.x}^2 = \frac{\sum (Y_i - Y)^2}{n - k - 1}\text{ ; } k = \text{jumlah variabel bebas}$$

Varians b_1

$$V(b1) = \frac{s_{y.x}^2}{\sum (X_{1i} - X_1)^2} = \frac{42{,}35}{33000} = 0{,}001283 = 0{,}0013$$

$$\text{atau } = \frac{MSE}{\sum X_{1i}^2 - (\sum X_{1i})^{2/n}} = \frac{42{,}35}{322000 - 2890000/10} = \frac{42{,}35}{33000} = 0{,}0013$$

$$\text{atau } = \frac{MSE}{\Sigma X_{1i}^{2} - n \frac{\Sigma X_{1i}}{n} \frac{\Sigma X_{1i}}{n}} = \frac{MSE}{\Sigma X_{1i}^{2} - n\,(X_1)^2}$$

$$\text{atau } = \frac{s_{y.x}^2}{\sum X_{1i}^2 - n\,(X_1)^2} = \frac{42,35}{322000 - 10\,(170)^2} = \frac{42,35}{33000} = 0,0013$$

$$\text{Rumus umum } \quad V(b_1) = \frac{s_{y.x}^2}{\sum (X_{1i} - X_1)^2}$$

Standar Error b_1

$$se\,(b_1) = \frac{S_{y.x}}{\sqrt{\sum (X_{1i} - X_1)^2}} = \frac{6,507687761}{\sqrt{33000}} = 0,035823642 = 0,036$$

$$\text{atau } \quad se\,(b_1) = \frac{S_{y.x}}{\sqrt{\sum X_{1i}^2 - n\,(X_1)^2}}$$

$$= \frac{6,507687761}{\sqrt{322000 - 10\,(170)^2}} = \frac{6,507687761}{\sqrt{33000}} = 0,036$$

atau $se\,(b_1) = \sqrt{V(b_1)} = \sqrt{0,001283} = 0,0358 = 0,036$

Rumus umum se(bi) $= \sqrt{V(b_i)}$

Menguji Koefisien Arah (b_1) Regresi Linear Sederhana

Untuk menguji hipotesis mengenai koefisien arah b1 diperlukan

- perumusan Hipotesis (H) atau disebut Hipotesis nol (Ho) dan
- perumusan Alternatif (A) disebut Hipotesis alternatif (Ha)

Jika Ho: $\beta_1 = 0$ maka Ha: $\beta_1 \neq 0$ Hipotesis menyatakan sama

Jika Ho: $\beta_1 \geq 0,75$ maka Ha: $\beta_1 < 0,75$ Hipotesis minimum

Jika Ho: $\beta_1 \leq 0,75$ maka Ha: $\beta_1 > 0,75$ Hipotesis maksimum

Uji t - parsial untuk b_1

$$t_{hitung} = \frac{(b_1 - \beta o)\, s_x}{S_{y.x}} \sqrt{n - 1}$$

Jika $\beta_o = 0$ (diketahui melawan alternatif, bukan intersept)

Ho: $\beta_1 = 0$ tidak terdapat pengaruh antara X dengan Y

Ha: $\beta_1 \neq \beta o$ terdapat pengaruh antara X dengan Y

Rumus t_{hitung} dengan menggunakan Standar Deviasi:

$$t_{hitung} = \frac{(b_1)\, s_{x1}}{s_{y.x1}} \sqrt{n - 1} = \frac{(0{,}509)\ 60{,}553}{6{,}507687761} \sqrt{9} = 14{,}208$$

s_{x1} = standar deviasi untuk variabel X1

$$= \sqrt{\frac{\sum (X_{1i} - X_1)^2}{n}} \quad \text{untuk sampel besar } n > 30$$

$$= \sqrt{\frac{\sum (X_{1i} - X_1)^2}{n - 1}} \quad \text{untuk sampel kecil } n \leq 30$$

$$= \sqrt{33000/ 9} = 60{,}553$$

$s_{y.x1}$ = standar deviasi estimator (taksiran)

$$s_{y.x1} = \sqrt{\frac{\sum(Yi - Y)^2}{n - 2}} = \sqrt{MSE} = \sqrt{42{,}35} = 6{,}507687761$$

atau

$$t_{\text{hitung}} = \frac{b_1 \sqrt{\sum(X_{1i} - X_1)^2}}{s_{y.x}} = \frac{0,509 \sqrt{33000}}{6,507687761} = 14,208$$

Rumus t_{hitung} dengan menggunakan Standar Error:

$$t_{\text{hitung}} = \frac{b_1}{se(b_1)} = \frac{0,509}{0,035823642} = 14,208$$

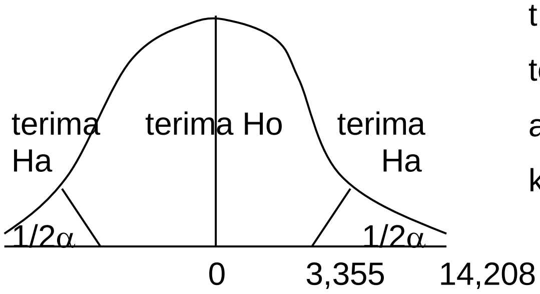
$$\text{Rumus umum } t_{\text{hitung}} = \frac{bi}{se\ (bi)}$$

Hasil t_{hitung} dibandingkan dengan t-tabel dimana t berdistribusi Student t dengan db = (n-k-1); p = 1 - ½ α dimana k adalah jumlah variabel bebas.

Karena Hipotesis menyatakan sama Ho = 0 maka digunakan uji dua (2) pihak karena itu p = 1 – ½ α dimana α simetris ½ α di pihak kanan dan ½ α di pihak kiri.

α = 0,05 maka taraf nyata (level of significance) = 100 (1-α)
% = 100(1-0,05) % = 95 %
Jika α = 0,01 maka taraf nyata = 99 %
t-tabel (8; 0,025) = 2,306 dan t-tabel (8; 0,005) = 3,355

Karena t_{hitung} = 14,208 > t_{tabel} (8; 0,005) = 3,355 maka b1 significance pada taraf nyata 99 % (very highly significance) artinya Ho: $\alpha 1$ = 0 ditolak atau Ha: $\alpha 1 \neq 0$ diterima. (nilai t_{hitung} bisa + atau - tergantung nilai b)



t = 14,208 terletak dalam daerah terima Ha Jadi terdapat pengaruh antara X (pendapatan) dgn belanja konsumsi(Y)

D. ANALISA KORELASI SEDERHANA

Regresi dinyatakan dalam bentuk persamaan matematis atau kurva bentuk hubungan dan pengaruh antara variabel bebas dengan variabel tergantung, sedangkan Korelasi dinyatakan dalam persentase keeratan hubungan antar variabel.

Dalam analisa korelasi tidak terlalu dipertimbangkan kedudukan variabel dependent dan independent, artinya korelasi X terhadap Y akan sama dengan korelasi Y terhadap X karena X dan Y keduanya adalah variabel random sedangkan X dalam regresi bersifat fixed dan Y nya random. Jadi:

$$\boxed{r_{xy} = r_{yx}}$$

Koefisien korelasi untuk statistik sampel diberi notasi r, sedangkan untuk parameter populasi diberi notasi ζ (baca rho).

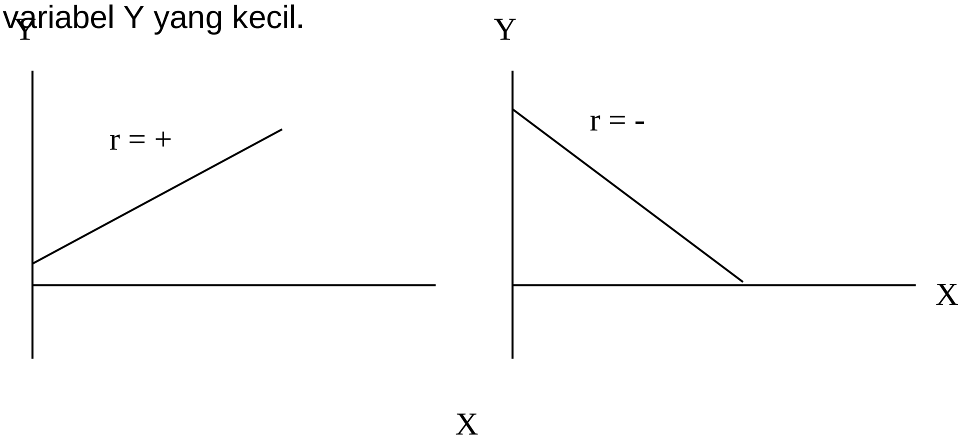
Koefisien korelasi r_{xy} menunjukkan derajad keeratan hubungan regresi antara variabel X dan Y dan bagaimana arah hubungannya (+/-). Sebaiknya terlebih dahulu menentukan bentuk persamaan regresi yang relevan (yang terbaik sebagai estimator) sebelum menentukan korelasinya.

1. Batas-Batas Koefisien Korelasi

Koefisen korelasi dinyatakan dalam persen dan memiliki nilai antara -1 dan +1 atau $-1 < r < +1$

Korelasi + atau hubungan searah artinya nilai variabel X yang kecil berpasangan dengan nilai variabel Y yang kecil dan nilai variabel X yang besar juga berpasangan dengan nilai variabel Y yang besar.

Korelasi - atau hubungan terbalik artinya nilai variabel X yang kecil berpasangan dengan nilai variabel Y yang besar dan sebaliknya nilai variabel X yang besar berpasangan dengan nilai variabel Y yang kecil.



2. Menghitung Koefisien Korelasi

a. Koefisien korelasi Produk Momen Pearson

$$r_{xy} = \sqrt{1 - \frac{SSE}{SST}} = \sqrt{1 - \frac{\sum (Y_i - Y)^2}{\sum (Y_i - Y)^2}}$$

Jika regresi cocok dengan letak titik2 pada diagram pencar, maka hasil bagi

$$\frac{\sum (Y_i - Y)^2}{\sum (Y_i - Y)^2} = \text{mendekati 0, sehingga } r \text{ mendekati} = 1$$

Jika $r_{xy} = 1$ artinya letak titik2 dalam diagram pencar berada persis pada regresi yang searah.

Jika $r_{xy} = -1$ artinya letak titik2 dalam diagram pencar berada persis pada regresi yang berlawanan.

Makin terpencar letak titik2 itu dari sebuah regresi nilai r korelasinya makin mendekati $= 0$.

Jika $r = 0$ bukan berarti antara variabel X dan Y tidak terdapat hubungan, tetapi tidak terdapat hubungan seperti regresi yang digunakan sehingga perlu dirobah dengan model regresi yang sesuai untuk menemukan nilai korelasi tertentu.

b. Korelasi sederhana yang dihitung dari standar deviasi s_x dan s_y

$$r_{xy} = \frac{\sum (X_i - X)(Y_i - Y)}{(n-1)\, s_x \,.\, s_y} = \frac{\sum X_i\, Y_i}{(n-1)\, s_x \,.\, s_y}$$

c. Rumus-rumus lainnya untuk menghitung koefisien korelasi sederhana

$$r_{xy} = \frac{n \sum X_i\, Y_i - \sum X_i \sum Y_i}{\sqrt{\{n \sum X_i^2 - (\sum X_i)^2\}\{n \sum Y_i^2 - (\sum Y_i)^2\}}} \qquad \text{atau}$$

$$r_{xy} = \frac{\sum (X_i - X)\,(Y_i - Y)}{\sqrt{\{\sum (X_i - X)^2\}\{\sum (Y_i - Y)^2\}}} \qquad \text{atau}$$

$$r_{xy} = \frac{\sum X_i Y_i - \dfrac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{\left\{\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}\right\}\left\{n \sum Y_i^2 - \dfrac{(\sum Y_i)^2}{n}\right\}}}$$

3. Hubungan antara Korelasi dengan Regresi

$$b1 = \frac{\sum(Y_i - Y)^2}{\sum(X_i - X)^2} * r_{xy}$$

$$sy^2 = \frac{\sum(Y_i - Y)^2}{n - 1} \text{ atau n - 1)} sy^2 = \sum(Y_i - Y)^2$$

$$sx^2 = \frac{\sum(X_i - X)^2}{n - 1} \text{ atau } (n - 1)\, sx^2 = \sum (X_i - X)^2$$

Jadi hubungannya $b_1 = (s_y / s_x)\, r_{xy}$ atau $r_{xy} = \dfrac{b_1}{s_y / s_x}$

Walaupun terdapat hubungan yang sangat erat antara b_1 regresi dengan r_{xy} korelasi namun interpretasi b_1 sangat berlainan dengan r_{xy} dimana:

r_{xy} = mengukur eratnya hubungan antara X dan Y, sedangkan

b_1 = mengukur besarnya perobahan pada Y yang diakibatkan oleh perobahan setiap unit X

4. Koefisien Determinasi

Koefisien determinasi adalah kuadrat koefisien korelasi (r^2). Kalau koefisien korelasi $-1 < r < +1$ maka koefisien determinasi tidak pernah negatif atau $0 < r^2 < 1$

Koefisien determinasi juga dinyatakan dalam persen yang menginterpretasikan bahwa variasi variabel Y disebabkan r^2 % oleh perubahan (variasi) variabel X.

$$r^2 = \frac{SSR}{SST}$$

Koefisien determinasi untuk regresi linear sederhana

$$r^2 = \frac{SSR}{SST} = \frac{b_0 \sum x_i y_i}{\sum y_i^2}$$

5. Contoh Menghitung Koefisien Korelasi Sederhana

Contoh, dari hasil penelitian pengaruh pendapatan mingguan (X) terhadap belanja konsumsi mingguan (Y) dari 10 sampel keluarga diperoleh hasil sbb:

$\sum(Xi)^2$ = $(1700)^2$ = 2890000

$\sum Xi \sum Yi$ = (1700)(1110) = 1887000

$(\sum Yi)^2$ = $(1110)^2$ = 1232100

Y = 24,47 + 0,509 X

Tabel 2. Regresi pendapatan X terhadap belanja konsumsi Y.

n	Yi	Xi	XiYi	Xi^2	Xi-X (xi)	Yi-Y (yi)	(Xi-X)(Yi-Y) (xi yi)	$(Xi-X)^2$ (xi^2)	$(Yi-Y)^2$ (yi^2)
1	70	80	5600	6400	-90	-41	3690	8100	1681
2	65	100	6500	10000	-70	-46	3220	4900	2116
3	90	120	10800	14400	-50	-21	1050	2500	441
4	95	140	13300	19600	-30	-16	480	900	256
5	110	160	17600	25600	-10	-1	10	100	1
6	115	180	20700	32400	10	4	40	100	16
7	120	200	24000	40000	30	9	270	900	81
8	140	220	30800	48400	50	29	1450	2500	841
9	155	240	37200	57600	70	44	3080	4900	1936
10	150	260	39000	67600	90	39	3510	8100	1521
JML	1110	1700	205500	322000	0	0	16800	33000	8890
MEAN	111	170							

Lanjutan Tabel 2

n	Yi	Xi	XiYi	Xi^2	Yi^2	Y	Yi − Y e	$(Yi - Y)^2$ e^2
1	70	80	5600	6400	4900	65,182	4,818	23,213124
2	65	100	6500	10000	4225	75,364	-10,364	107,412496
3	90	120	10800	14400	8100	85,545	4,455	19,847025
4	95	140	13300	19600	9025	95,727	-0,727	0,528529
5	110	160	17600	25600	12100	105,909	4,091	16,736281
6	115	180	20700	32400	13225	116,091	-1,091	1,190281
7	120	200	24000	40000	14400	126,276	-6,273	39,350529
8	140	220	30800	48400	19600	136,455	3,545	12,567025
9	155	240	37200	57600	24025	146,636	8,364	69,956496
10	150	260	39000	67600	22500	156,818	-6,818	46,485124
JML	1110	1700	205500	322000	132100			337,286910
MEAN	111	170						

Analisa Varians (ANAVA) untuk regresi linear sederhana dari data pendapatan (X) dengan pengeluaran konsumsi (Y).

Sumber Variasi S.V	db df	Jumlah Kuadrat SS	Rata-rata Kuadrat MS	F-hitung	F-tabel 0,05 0,01
Regresi	1	8551,20	8551,20	201,92	
Error	8	338,80	42,35		
Total	9	8890,00			

SS Total = SS Regresi + SS Error

$\sum(Y_i - Y)^2 = \sum(Y - Y)^2 + \sum(Y_i - Y)^2$ yakni

SS Total $= \sum(Y_i - Y)^2 = \sum y_i^2 = 8890$

SS Regresi $= \sum(Y - Y)^2 = b_1 \sum(X_i - X)(Y_i - Y)$

$= b_1 \sum X_i y_i = 0{,}509 (16800) = 8551{,}2$

SS Error = 338,80 $\neq$ 337,38 karena koefisien b_1 mengalami pembulatan dari 0,509090909 (16800) = 8552,727273

Jadi 8890 - 8552,72 = 337,28

a. Koefisien Korelasi Produk Momen Pearson

$$r_{xy} = \sqrt{1 - \frac{SSE}{SST}} = \sqrt{1 - \frac{\sum (Yi - Y)^2}{\sum (Yi - Y)^2}}$$

$$= \sqrt{1 - \frac{338{,}80}{8890}} = 0{,}98$$

b. Korelasi dihitung dengan Standar Deviasi s_x dan s_y

$$s_x = \sqrt{\frac{\sum (Xi - X)^2}{n - 1}} = \sqrt{\frac{33000}{9}} = 60{,}55300708$$

$$s_y = \sqrt{\frac{\sum (Yi - Y)^2}{n - 1}} = \sqrt{\frac{8890}{9}} = 31{,}42893218$$

$$r_{xy} = \frac{\sum (Xi - X)(Yi - Y)}{(n - 1)\ sx\ .\ sy} = \frac{16800}{(9)\ (60{,}55)\ (31{,}43)} = 0{,}98$$

$$= \frac{\sum Xi\ yi}{(n-1)\ sx\ .\ sy} = 0{,}98$$

c. Rumus2 lain utk menghitung koefisien korelasi sederhana

1) $$r_{xy} = \frac{n \sum Xi\ Yi - \sum Xi \sum Yi}{\sqrt{\{n \sum Xi^2 - (\sum Xi)^2\}\{n \sum Yi^2 - (\sum Yi)^2\}}}$$

$$= \frac{10\ (205500) - (1887000)}{\sqrt{\{10\ (322000) - 2890000\}\ \{10\ (132100) - (1232100)}}$$

$$= \frac{168000}{\sqrt{\{330000\}\ \{88900\}}} = 0{,}98$$

2) $$r_{xy} = \frac{\sum (Xi - X)(Yi - Y)}{\sqrt{\{\sum (Xi - X)^2\}\{\sum (Yi - Y)^2\}}}$$

$$= \frac{\sum Xi\ yi}{\sqrt{\sum Xi^2 \sum yi^2}} = \frac{16800}{\sqrt{(33000)(8890)}} = 0{,}98$$

3) $$r_{xy} = \frac{\sum Xi\ Yi - \dfrac{(\sum Xi)(\sum Yi)}{n}}{\sqrt{\left\{\sum Xi^2 - \dfrac{(\sum Xi)^2}{n}\right\}\left\{\sum Yi^2 - \dfrac{(\sum Yi)^2}{n}\right\}}}$$

$$= \frac{205500 - 1887000/10}{\sqrt{(322000 - 2890000/10)(132100 - 1232100/10)}}$$

$$= \frac{16800}{\sqrt{(33000)(8890)}} = 0{,}98$$

4) $b1 = (sy/\ sx)\ r_{xy}$ atau $r_{xy} = \dfrac{b1}{sy/\ sx}$

$$r_{xy} = \frac{0{,}509}{31{,}4289/\ 60{,}553} = 0{,}98$$

6. Menguji Koefisien Korelasi Sederhana

Mirip dengan uji t untuk regresi linear sederhana yaitu:

Rumus umum $t_{\text{hitung}} = \dfrac{bi - \beta}{se\ (b_i)}$

Pengujian koefisien korelasi dengan uji t

Untuk $\tau_{xy} = 0$ statistik sampel r_{xy} bersifat tak bias dengan varians = $(1 - r_{xy}^2)/\ (n - 2)$

$se(r_{xy}) = \sqrt{\dfrac{1 - r_{xy}^2}{n - 2}}$ karena itu rumus t_{hitung} yakni:

$t_{\text{hitung}} = \dfrac{r_{xy} - \tau_{xy}}{se(r_{xy})} = \dfrac{r_{xy} - \tau_{xy}}{\sqrt{\dfrac{1 - r_{xy}^2}{n - 2}}}$ karena Ho: $\tau = 0$ maka

$$t_{\text{hitung}} = \frac{r_{xy}}{\sqrt{\dfrac{1 - r_{xy}^2}{n - 2}}} = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$$

Hipotesa Ho : $\tau = 0$

Alternatif Ha : $\tau \neq 0$

Bandingkan t_{hitung} dengan $t_{\text{tabel}}\ (n - 2;\ \alpha/2)$

Untuk t_{hitung} positif, apabila $t_{hitung} \geq t_{tabel}$ 0,05 maka kesimpulannya Ho ditolak yang berarti ada korelasi antara X dan Y.

$$t_{hitung} = \frac{0{,}98}{\sqrt{\frac{1 - 0{,}96}{8}}} = 13{,}85$$

t_{tabel} 0,05 (8; 0,025) = 2,3060

t_{tabel} 0,01 (8; 0,005) = 3,3554

Karena t_{hitung} positif > t_{tabel} 0,01 atau berbeda nyata pada koefisien kepercayaan 99 %, jadi kesimpulannya Ho ditolak atau Ha diterima yang berarti ada korelasi antara Pendapatan (X) dengan pengeluaran konsumsi mingguan rumah tangga (Y).

BAB III
ANALISA REGRESI DAN KORELASI BERGANDA

A. LEAST SQUARE METHODE UNTUK REGRESI LINEAR BERGANDA DENGAN 2 VARIABEL INDEPENDENT

$$Y = b_o + b_1 X_1 + b_2 X_2 + e_i$$

Estimasi dengan metode least square melalui perhitungan sbb:

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

Sehingga besarnya jumlah kuadrat dari error e adalah:

$$\sum e_i^2 = \sum (Y_i - \beta_o - \beta_1 X_{1i} - \beta_2 X_{2i})^2$$

Agar persamaan $\sum e_i^{\,2}$ minimum maka f.o.c (turunan pertamanya) terhadap β_o , β_1 dan β_2 harus = 0

$$\delta\sum e_i^{\,2} / \delta\beta_o \rightarrow 2 \sum (Y_i - \beta_o - \beta_1 X_{1i} - \beta_2 X_{2i})(-1) \quad = 0 * -1/2$$

$$\delta\sum e_i^{\,2} / \delta\beta_1 \rightarrow 2 \sum (Y_i - \beta_o - \beta_1 X_{1i} - \beta_2 X_{2i})(-X_{1i}) = 0 * -1/2$$

$$\delta\sum e_i^{\,2} / \delta\beta_2 \rightarrow 2 \sum (Y_i - \beta_o - \beta_1 X_{1i} - \beta_2 X_{2i})(-X_{2i}) = 0 * -1/2$$

$$b_o n + b_1 \sum X_{1i} + b_2 \sum X_{2i} = \sum Y_i$$

$$b_o \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i} = \sum X_{1i} Y_i$$

$$b_o \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2 = \sum X_{2i} Yi$$

$$b1 = \frac{(\sum y_i x_{1i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}$$

$$b2 = \frac{(\sum y_i x_{2i})(\sum x_{1i}^2) - (\sum y_i x_{1i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}$$

$$bo = Y - b_1 X_1 - b_2 X_2$$

B. VARIANS DAN STANDAR ERROR b1, b2

$$V(b_1) = \frac{\sum x2_i^2}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_2)^2}$$

$$se(b_1) = + \sqrt{V(b_1)}$$

$$V(b_2) = \frac{\sum x_{1i}^2}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_2)^2}$$

$$se(b_2) = + \sqrt{V(b_2)}$$

$$\sigma^2 = \frac{\sum e_i^2}{n - k - 1} = \frac{SSE}{n - k - 1} = MSE$$

C. KOEFISIEN DETERMINASI DAN KORELASI BERGANDA

1. Rumus Koefisien Determinasi (R^2)

$$R^2 \text{ adjusted} = 1 - \frac{\sum e_i^2 / n\text{-}k\text{-}1}{\sum y_i^2 / n - 1} = 1 - \frac{MSE}{sy^2}$$

$$\text{Varians Y yakni } sy^2 = \frac{\sum(Y_i - Y)^2}{n - 1}$$

2. Sifat R^2 adjusted

a) k = 1 maka R^2 adjusted = R^2

b) k bertambah maka R^2 adjusted bertambah besar namun relatif lebih kecil dari R^2 (R^2 adjusted < R^2)

$$R^2 = \frac{SSR}{SST} = \frac{b_1 \sum y_i x_{1i} + b_2 \sum y_i x_{2i}}{\sum y_i^2}$$

$$= \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

3. Koefisien korelasi merupakan akar dari koefisien determinasi

$R = \pm \sqrt{R^2}$

Batas2 nilai R antara -1 s/d + 1 atau $-1 < R < +1$

Batas2 nilai R^2 antara 0 s/d 1 atau $0 < R^2 < 1$

D. KOEFISIEN KORELASI PARSIAL

Secara langsung koefisien korelasi sederhana dapat diukur dengan rumus sum product (SP) dan sum square (SS) sbb:

$$\boxed{r^2 = b_1^2 \left\langle \frac{\Sigma x_i^2}{\Sigma y_i^2} \right\rangle}$$

$$\text{karena } b_1 = \frac{(\Sigma x_i y_i)}{(\Sigma x_i^2)}$$

$$\boxed{r^2 = \frac{(\Sigma x_i y_i)^2}{(\Sigma x_i^2)^2} \left\langle \frac{\Sigma x_i^2}{\Sigma y_i^2} \right\rangle = \frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2 \Sigma y_i^2} \left\langle \frac{\Sigma x_i^2}{\Sigma y_i^2} \right\rangle}$$

$$r^2 = \frac{(\sum x_i\, y_i)^2}{\sum x_1^2 \sum y_i^2}$$ jadi koefisien korelasinya

$$r = \frac{\sum x_i\, y_i}{\sqrt{\sum x_1^2 \sum y_i^2}} = \frac{Spx_i\, y_i}{\sqrt{SSX\ SSY}}$$

Koefisien korelasi tersebut mengukur besarnya derajad hubungan linear antara dua variabel yakni X dan Y.

Dalam membahas Regresi berganda dengan 2 variabel bebas terdapat 3 (tiga) nilai koefisien korelasi sederhana, masing2:

1. Hubungan Y dengan X_1 yakni r_{yx1} atau r_{y1}

$$r_{y1} = \frac{SPx_1\, y_i}{\sqrt{SSX_1\ SSY}} = \frac{\sum x_{1i}\, y_i}{\sqrt{\sum x_1^2 \sum y_i^2}}$$

2. Hubungan Y dengan X_2 yakni $r_{yx}{}^2$ atau $r_y{}^2$

$$ry^2 = \frac{SPx_2\, yi}{\sqrt{SSX_2\ SSY}} = \frac{\sum x_{2i}\, y_i}{\sqrt{\sum x_2^2 \ \sum y_i^2}}$$

3. Hubungan X_1 dengan X_2 yakni $rx_1\, x_2$ atau r_{12}

$$r_{12} = \frac{SPx1\ x2}{\sqrt{SSX_1\ SSX_2}} = \frac{\sum x_{1i}\, x_{2i}}{\sqrt{\sum x1^2 \ \sum x_2^2}}$$

Koefisien korelasi sederhana di atas bukan merupakan derajad keeratan hubungan yang sebenarnya antara dua variabel yang dikorelasi karena munculnya variabel ke tiga. Koefisien korelasi parsial perlu mengeliminir faktor koreksinya dengan rumus sbb:

1. Hubungan Y dengan X_1 dengan anggapan X_2 konstan ($r_{y1.2}$)

$$r_{y1.2} = \frac{r_{y1} - r_{y2}\ r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}$$

2. Hubungan Y dengan X_2 dengan anggapan X_1 konstan ($r_{y2.1}$)

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\ r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}$$

3. Hubungan X_1 dengan X_2 dengan anggapan Y konstan ($r_{12.y}$)

$$r_{12.y} = \frac{r_{12} - r_{y1}\ r_{y2}}{\sqrt{(1-r_{y1}^2)(1-r_{y2}^2)}}$$

E. PENGUJIAN KOEFISIEN REGRESI b1, b2

1. Pengujian Koefisien Arah b_1, b_2 Secara Serempak

   Dengan Uji F sesuai metoda Analisis Varians (ANAVA)

   Hipotesis Ho: $\beta_1 = \beta_2 = 0$

   Alternatif Ha: $\beta_1 \neq \beta_2 \neq 0$

   Tabel ANAVA pengaruh serempak variabel independent

Sumber Variasi S.V	Db Df	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Tabel} 0,05 0,01
Regresi Error Total	K n-k-1 n-1	SSR SSE SST	MSR MSE	MSR/MSE	

$SST = \sum(Y_i - Y)^2 = \sum y_i^2$

$SSR = \sum(\hat{Y} - \overline{Y})^2 = b_1 \sum y_i x_{1i} + b_2 \sum y_i x_{2i}$

$SSE = \sum e_i^2 = \sum (Y_i - Y)^2 = SST - SSR$

$= \sum y_i^2 - b_1 \sum y_i x_{1i} - b_2 \sum y_i x_{2i}$

$MS = SS/ df$

Berdasarkan asumsi normal untuk disturbans e_i maka nilai F_{hitung} adalah:

$$F_{hitung} = \frac{MSR}{MSE}$$

yang mengikuti distribusi F dengan derajad bebas = (k; n-k-1)

Kaidah keputusan uji F ini adalah:

Jika $F_{hitung} \leq F_{tabel}$ (α; df = k; n-k-1) ..... maka Ho diterima (non-significance)

Jika $F_{hitung} > F_{tabel}$ (α ; df = k; n-k-1) ..... maka Ho ditolak (significance)

Kaedah keputusan tolak Ho (terima Ha) berarti koefisien arah regresi secara serempak dapat digunakan sebagai penduga (estimator) yang dipercaya untuk memprediksi pengaruh semua variabel bebas X secara serempak terhadap Y.

2. Pengujian Koefisien Arah b_1, b_2 Secara Parsial

   Model pengujian koefisien regresi partial dengan uji t

   Pengujian koefisien b_1

   Hipotesis Ho: $\beta_1 = 0$

   Alternatif Ha: $\beta_1 \neq 0$

$$t_{hitung} = \frac{b_1}{s.e(b_1)}$$

   Pengujian koefisien b_2

   Hipotesis Ho: $\beta_2 = 0$

   Alternatif Ha: $\beta_2 \neq 0$

$$t_{hitung} = \frac{b_2}{s.e(b_2)}$$

yang mengikuti distribusi t dengan derajad bebas df = n-k-1

Kaidah keputusan uji t ini adalah:

Jika $t_{hitung} < t_{tabel}$ (1/2 α ; df = n-k-1) ..... maka Ho diterima (nonsignificance)

Jika $t_{hitung} \geq t_{tabel}$ (1/2 α ; df = n-k-1) ..... maka Ho ditolak (significance)

Kaedah keputusan tolak Ho (terima Ha) berarti koefisien arah tersebut secara partial dapat digunakan sebagai penduga (estimator) yang dipercaya untuk memprediksi pengaruh variabel bebas X secara individu terhadap Y.

Koefisien korelasi dan determinasi berhubungan dengan koefisien regresi dimana hasil uji koefisien regresi akan identik dgn hasil uji koefisien korelasi. Karena itu jika koefisien regresi telah diuji tidak perlu lagi menguji koefisien korelasinya.

Artinya jika uji koefisien regresi secara serempak dari suatu model regresi hasilnya nonsignificance maka hasil uji koefisien korelasi bergandanya juga akan non-significance, dan sebaliknya.

Pengujian parsial diperlukan karena walaupun uji serempak menyatakan nonsignificance kemungkinan hasil uji parsialnya ada yang significance.

Sebaliknya jika uji serempak menyatakan significance tetapi uji parsial dari masing2 variabel independent menyatakan nonsignificance menunjukkan terjadinya kasus “multikolinearitas”.

F. MODIFIKASI ANAVA

Modifikasi Anava kurang populer, dan terutama hanya digunakan oleh analis yang benar2 ahli statistik untuk mempelajari pengaruh variabel X_2 dengan syarat bahwa X_2 ini merupakan tambahan kepada X_1 yang berpengaruh terhadap Y atau (X_2/ X_1).

Tabel ANAVA pengaruh parsial variabel independent

Sumber Variasi S.V	Db Df	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Total} 0,05 0,01
Regresi b_1,b_2 (Pengaruh X_1, X_2)	K = 2	SSR	SSR/k	F_s	
Regresi b_1 (Pengaruh X_1)	1	SSR1	SSR1/k	F_1	
Regresi b_2/ b_1 (Pengaruh X_2/ X_1)	1	SSR2	SSR2/k	F_2	
Error 2	n-k-1	SSE	SSE/ n-k-1		
Total	n-1	SST			

$SSR = b_1 \sum y_i x_{1i} + b_2 \sum y_i x_{2i}$

$SSR_1 = b_1 \sum y_i x_{1i}$

$SSR_2 = SSR - SSR_1$

$SSE = SST - SSR$

$SST = \sum y_i^2$

1. Pengujian pengaruh koefisien regresi secara serempak dengan Fs

Ho: $\beta_1 = \beta_2 = 0$

Ha: $\beta_1 \neq \beta_2 \neq 0$

$$Fs = \frac{MSR}{MSE}$$

Fs dibandingkan terhadap F_{tabel} dengan df = {k/ (n-k-1)}

2. Pengujian pengaruh b_1 individu dengan F_1

Ho: $\beta_1 = 0$

$$F1 = \frac{MSR_1}{MSE}$$

Ha: $\beta_1 \neq 0$

F_1 dibandingkan terhadap F $_{tabel}$ dengan df = {1/(n-k-1)}

3. Pengujian pengaruh b_2 parsial setelah b_1 dengan F_2

Ho: $\beta_2 = 0$

$$F_2 = \frac{MSR_2}{MSE}$$

Ha: $\beta_2 \neq 0$

F_2 dibandingkan terhadap F $_{tabel}$ dengan df = {1/ (n-k-1)}

G. CONTOH REGRESI LINEAR BERGANDA DENGAN 2 VARIABEL INDEPENDENT

Contoh 1.

Dari hasil penelitian pengaruh disposible income (X_1) dari waktu ke waktu th 1956 s/d 1970 (X_2) terhadap belanja konsumsi personil di USA (Y) diperoleh hasil sbb (dalam milyar dolar):

n	Y_i	X_1	X_2	$X_{1i}\, Y_i$	$X_{2i}\, Y_i$	$X_{1i}\, X_{2i}$	X_{1i}^2	X_{2i}^2
1	281,3	309,3	1	87006,09	281,3	309,3	95666,49	1
2	288,1	316,1	2	91068,41	576,2	632,2	99919,21	4
3	290	318,8	3	92452	870	956,4	101633,44	9
4	307,3	333	4	102330,9	1229,2	1332	110889	16
5	316,1	340,3	5	107568,83	1580,5	1701,5	115804,09	25
6	322,5	350,5	6	113036,25	1935	2103	122850,25	36
7	338,4	367,2	7	124260,48	2368,8	2570,4	134835,84	49
8	353,3	381,2	8	134677,96	2826,4	3049,6	145313,44	64
9	373,7	408,1	9	152506,97	3363,3	3672,9	166545,61	81
10	397,7	434,8	10	172919,96	3977	4348	189051,04	100
11	418,1	458,9	11	191866,09	4599,1	5047,9	210589,21	121
12	430,1	477,5	12	205372,75	5161,2	5730	228006,25	144
13	452,7	499	13	225897,3	5885,1	6487	249001	169
14	469,1	513,5	14	240882,85	6567,4	7189	263682,25	196
15	476,9	533,2	15	254283,08	7153,5	7998	284302,24	225
JML	5515,3	6041,4	120	2296129,92	48374	53127,2	2518089,36	1240
ME	367,69	402,76	8					

$n = 15$

$\sum Y_i = 5515,3$ $\sum X_{1i}Y_i = 2296129,92$ $\sum X_{2i}Y_i = 48374$

$\sum X_{1i} = 6041,4$ $\sum X_{1i}^2 = 2518089,36$

$\sum X_{2i} = 120$ $\sum X_{2i}^2 = 1240$ $\sum X_{1i}X_{2i} = 53127,2$

$\sum y_i^2 = 66059,5375$ $\sum y_i x_{1i} = 74787,6920$

$\sum x_{1i}^2 = 84855,096$ $\sum y_i x_{2i} = 4251,60$

$\sum x_{2i}^2 = 280$ $\sum x_{1i} x_{2i} = 4786$

1. Koefisien Regresi

$$b_o n + b_1 \sum X_{1i} + b_2 \sum X_{2i} = \sum Y_i$$

$$b_o \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i} = \sum X_{1i} Y_i$$

$$b_o \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2 = \sum X_{2i} Y_i$$

$$15\ b_o + 6041,4\ b_1 + 120\ b_2 = 5515,3 \quad (1)$$

$$6041,4\ b_o + 2518089,36\ b_1 + 53127,2\ b_2 = 2296129,92 \quad (2)$$

$$120\ b_o + 53127,2\ b_1 + 1240\ b_2 = 48374 \quad (3)$$

$$15\ b_o + 6041,4\ b_1 + 120\ b_2 = 5515,3 \quad * 402,76$$

$$6041,4\ b_o + 2433234,264\ b_1 + 48331,2\ b_2 = 2221342,228$$

$$6041,4\ b_o + 2518089,36\ b_1 + 53127,2\ b_2 = 2296129,92$$

$$-\ 84855,096\ b_1 - 4796\ b_2 = -\ 74787,692 \quad (4)$$

$$15\ b_o + 6041,4\ b_1 + 120\ b_2 = 5515,3 \quad * 8$$

$$120\ b_o + 48331,2\ b_1 + 960\ b_2 = 44122,4$$

$$120\ b_o + 53127,2\ b_1 + 1240\ b_2 = 48374 \quad (3) \quad *17,12857143$$

$$-\ 4796\ b_1 - 280\ b_2 = -\ 4251,6 \quad (5)$$

$$-\ 82148{,}62857\ b_1 - 4796\ b_2 = -\ 72823{,}83428$$
$$-\ 84855{,}096\ b_1 - 4796\ b_2 = -\ 74787{,}692 \qquad (4)$$

$$2706{,}46743\ b_1 = 1963{,}85772$$
$$\mathbf{b_1 = 0{,}725616609 = 0{,}73}$$

$$-\ 4796\ b_1 - 280\ b_2 = -\ 4251{,}6 \qquad (5)$$
$$-\ 4796\ (0{,}7256) - 280\ b_2 = -\ 4251{,}6$$
$$-\ 3480{,}057259 - 280\ b_2 = -\ 4251{,}6$$
$$280\ b_2 = 771{,}542741$$
$$\mathbf{b_2 = 2{,}755509789 = 2{,}76}$$

$$15\ b_o + 6041{,}4\ b_1 + 120\ b_2 = 5515{,}3 \qquad (1)$$
$$15\ b_o + 6041{,}4\ (0{,}725616609) + 120\ (2{,}755509789) = 5515{,}3$$
$$15\ b_o = 800.8986438$$
$$\mathbf{b_o = 53{,}39}$$

Persamaan Regresi Linear Berganda

$Y = 53{,}39 + 0{,}73\ X_1 + 2{,}76\ X_2$

ANALISA VARIANS (ANAVA)

Sumber Variasi S.V	db df	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Total} 0,05 0,01
Regresi Error Total	2 12 14	65982,52 77,02 66059,54	32991,26 6,42	5140,132	

SS Total $= \sum y_i^2 = \sum(Y_i - Y)^2 = 66059{,}54$

SS Regresi $= \sum y_i^2 = \sum (Y - Y)^2 = b_1 \sum x_{1i}\, y_i + b_2 \sum x_{2i}\, y_i$

$0{,}725616609(74787{,}692) + 2{,}755509789(4251{,}6) = 65982{,}52$

SS Error $= \sum e_i^2 = SST - SSR = 66059{,}54 - 65982{,}52 = 77{,}02$

Hipotesis untuk uji F overall

Ho : $\beta_1 = \beta_2 = 0$

Ha : $\beta_1 \neq \beta_2 \neq 0$

F_{hitung} = MSR/ MSE = 32991,26/ 6,42 = 5140,132

F_{tabel} 0,95 (2; 12) =

F_{tabel} 0,99 (2; 12) =

Uji t parsial untuk koefisien regresi b_1 dan b_2

Ho : $\beta_1 = 0$

Ha : $\beta_1 \neq 0$

Koefisien korelasi parsial

Modifikasi lain dari ANAVA

Tabel ANAVA pengaruh parsial variabel independent

<table>
<tr><td>Sumber Variasi
S.V</td><td>db

df</td><td>Jumlah
Kuadrat
(SS)</td><td>Rata2
Kuadrat
(MS)</td><td>F_{Hitung}</td><td>F_{Total}
0,05 0,01</td></tr>
<tr><td>Regresi b1/bo
Error 1
Regresi b1,b2/bo
Regresi b2/bo,b1
Error 2
Total</td><td>1

k

n-k-1
n-1</td><td>

SSR

SST</td><td>

MSR

MSE
</td><td>

MSR/MSE

</td><td></td></tr>
</table>

CONTOH 2

Dari hasil penelitian terhadap 10 rumah tangga mengenai pengaruh pendapatan dalam ribuan rupiah (X1) dan jumlah anggota keluarga (X2) terhadap belanja konsumsi harian dalam ratusan rupiah (Y) diperoleh hasil sbb (data hipotetik):

n	Yi	X1	X2	X1i Yi	X2i Yi	X1i X2i	$X1i^2$	$X2i^2$	Yi^2
1	23	10	7	230	161	70	100	49	529
2	7	2	3	14	21	6	4	9	49
3	15	4	2	60	30	8	16	4	225
4	17	6	4	102	68	24	36	16	289
5	23	8	6	184	138	48	64	36	529
6	22	7	5	154	110	35	49	25	484
7	10	4	3	40	30	12	16	9	100
8	14	6	3	84	42	18	36	9	196
9	20	7	4	140	80	28	49	16	400
10	19	6	3	114	57	18	36	9	361
JML	170	60	40	1122	737	267	406	182	3162
MEAN	17	6	4						

$n = 10$

$\sum Y_i = 170 \quad Y = 17 \quad X_1 = 6 \quad X_2 = 4$

$\sum X_{1i} = 60 \quad \sum X_{1i}Y_i = 1122 \quad \sum X_{1i}^2 = 406 \quad \sum Y_i^2 = 3162$

$\sum X_{2i} = 40 \quad \sum X_{2i}Y_i = 737 \quad \sum X_{2i}^2 = 182 \quad \sum X_{1i}X_{2i} = 267$

1. Menghitung Koefisien Regresi

$\hat{Y} = b_o + b_1 X_1 + b_2 X_2 + e_i$

$b_o\, n + b_1 \sum X_{1i} + b_2 \sum X_{2i} = \sum Y_i$

$b_o \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i} = \sum X_{1i} Y_i$

$b_o \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2 = \sum X_{2i} Y_i$

$10\, b_o + 60\, b_1 + 40\, b_2 = 170 \qquad (1)$

$60\, b_o + 406\, b_1 + 267\, b_2 = 1122 \qquad (2)$

$40\, b_o + 267\, b_1 + 182\, b_2 = 737 \qquad (3)$

$10\, b_o + 60\, b_1 + 40\, b_2 = 170 \qquad (1)$

$60\, b_o + 360\, b_1 + 240\, b_2 = 1020 \qquad (1)$

$60\, b_o + 406\, b_1 + 267\, b_2 = 1122 \qquad (2)$

$-\ 46\, b_1 -\ 27\, b_2 = -\ 102 \qquad (4)$

$10\ b_o\ +\ 60\ b_1 +\ 40\ b_2 \quad = \ 170 \qquad (1)$

$40\ b_o\ + 240\ b_1\ + 160\ b_2 \quad = \ 680 \qquad (1)$

$40\ b_o\ + 267\ b_1\ + 182\ b_2 \quad = \ 737 \qquad (3)$

$(5) \quad -\ 27\ b_1 \ -\ 22\ b_2 \quad = -\ 57\ \ *\ 27$

$(4) \quad -\ 46\ b_1 \ -\ 27\ b_2 \quad = -\ 102\ \ *\ 22$

$-\ 729\ b_1\ -\ 594\ b_2 \quad = -1539$

$-1012\ b_1\ -\ 594\ b_2 \quad = -2244$

$283\ b_1 \quad = \ 705$

$\mathbf{b_1 \quad = \ 2{,}49116607}$

$(5) \quad -\ 27\ b_1 \ -\ 22\ b_2 \quad = \ -\ 57$

$-27(2{,}49116607)\ -\ 22\ b_2 \ = \ -\ 57$

$22\ b_2 \quad = \ -\ 10{,}2614838$

$\mathbf{b_2 \quad = \ -\ 0{,}46643108}$

$10\ b_o\ +\ 60\ b_1\ +\ 40\ b_2 \quad = \ 170 \qquad (1)$

$10\ b_o + 60\ (2{,}49116607) +\ 40\ (-\ 0{,}46643108)\ =\ 170$

$10\ b_o\ =\ 39{,}187279$

$\mathbf{b_o\ =\ 3{,}9187279}$

Persamaan regresi linear berganda

$Y = 3{,}92 + 2{,}49\ X_1 - 0{,}47\ X_2$

Regresi ini dapat digunakan untuk menaksir (mengestimasi) jumlah belanja konsumsi = Rp 1.698,- jika rata2 jumlah penghasilan Rp 6.000,- dan jumlah anggota keluarga 4 orang.

Yang diperoleh dari Y = 3,92 + 2,49 (6) - 0,47 (4) = 16,98

Karena belanja konsumsi Y dalam ratusan rupiah maka Y = 6,98 (Rp 100) = Rp 1.698,-

Model regresi tersebut baru dapat diterima sebagai estimator jika hasil uji serempak dengan ANAVA menunjukkan signifikan.

2. Pengujian Koefisien Regresi b_1, b_2 Secara Serempak

Dengan Uji F sesuai metoda Analisis Varians (ANAVA)

Hipotesis Ho: $\beta_1 = \beta_2 = 0$

Alternatif Ha: $\beta_1 \neq \beta_2 \neq 0$

Tabel ANAVA pengaruh serempak variabel independent

Sumber Variasi S.V	Db Df	Jumlah Kuadrat (SS)	Rata2 Kuadrat (MS)	F_{Hitung}	F_{Tabel} 0,05 0,01
Regresi	2	227,51	113,76	17,89	
Error	7	44,49	6,36		
Total	9	272,00			

$$\text{SS Total} = \sum(Yi - Y)^2 = \sum yi^2 = \sum Yi^2 - \frac{(\sum Yi)^2}{n}$$

$$= 3162 - \frac{(170)^2}{10} = 272$$

$$\text{SS Regresi} = \sum (Y - Y)^2 = b1 \sum x1i\, yi + b2 \sum x2i\, yi$$

$$= b_1\left\langle \Sigma X_1 Y_1 - \frac{\Sigma X_1 \Sigma Y_1}{n} \right\rangle + b_2\left\langle \Sigma X_2 Y_2 - \frac{\Sigma X_2 \Sigma Y_i}{n} \right\rangle$$

$$= 2{,}49116607 \{1122 - \frac{60\ (170)}{10}\} - 0{,}46643108 \{737 - \frac{40\ (170)}{10}\}$$

$$= 254{,}0989391 - 26{,}58657156 = 227{,}5123675$$

SS Error = $\sum ei^2$ = SST - SSR

= 272 - 227,5123675 = 44,4876325

Berdasarkan asumsi normal untuk disturbans ui maka nilai F_{hitung} adalah:

$$F_{hitung} = \frac{MSR}{MSE} = \frac{113{,}76}{6{,}36} = 17{,}89$$

yang mengikuti distribusi F dengan derajad bebas = (2; 7)

Kaidah keputusan uji F ini adalah:

F_{hitung} = 17,89 > F_{tabel} (α= 0,01 ; df= 2; 7) = 9,55 maka tolak Ho (**highly significance** atau berbeda nyata pada koefisien kepercayaan 99 %).

Kaedah keputusan tolak Ho (terima Ha) berarti koefisien arah regresi secara serempak dapat digunakan sebagai penduga (estimator) yang dipercaya untuk memprediksi pengaruh semua variabel bebas X secara serempak terhadap Y.

3. Pengujian Koefisien Arah b_1, b_2 Secara Parsial

Dengan uji t, untuk itu diperlukan V (b1), se (b1) dan V (b2) , se (b2) sbb:

$$V(b1) = \frac{\sum x2i^2}{(\sum x1i^2)(\sum x2i^2) - (\sum x1i\ x2)^2}$$

$$\sigma^2 = \frac{\sum ei^2}{n - k - 1} = \frac{SSE}{n - k - 1} = MSE$$

$$V(b_1) = \frac{\Sigma X_{2i}^{\ 2} - \dfrac{(\Sigma X_2)^2}{n}}{\{\Sigma X_{1i}^{\ 2} - \dfrac{(\Sigma X_{1i})^2}{n}\}\{\Sigma X_{2i}^{\ 2} - \dfrac{(\Sigma X_{2i})^2}{n}\} - \{\Sigma X_{1i}\ \Sigma X_{2i} - \dfrac{\Sigma X_{1i}\Sigma X_{2i}}{n}\}^2} MSE$$

$$= \frac{182 - (40)^2/10}{\{406 - (60)^2/10\}\ \{182 - (40)^2/10\} - \{267 - (60)(40)/10\}^2} * 6{,}36$$

$$= \frac{22}{\{46\}\{22\} - \{27\}^2} * 6{,}36 = \frac{22}{283} * 6{,}36 = 0{,}494416961$$

se(b1)= + $\sqrt{}$ V(b1) = 0,703147894

$$V(b2) = \frac{\sum x1i^2}{(\sum x1i^2)(\sum x2i^2) - (\sum x1i\ x2)^2}\ \sigma^2$$

$$V(b_1) = \frac{\Sigma X_{2i}^{\ 2} - \dfrac{(\Sigma X_2)^2}{n}}{\{\Sigma X_{1i}^{\ 2} - \dfrac{(\Sigma X_{1i})^2}{n}\}\{\Sigma X_{2i}^{\ 2} - \dfrac{(\Sigma X_{2i})^2}{n}\} - \{\Sigma X_{1i}\ \Sigma X_{2i} - \dfrac{\Sigma X_{1i}\Sigma X_{2i}}{n}\}^2} MSE$$

$$= \frac{46}{283} * 6{,}36 = 1{,}033780919$$

se(b_2) = $\sqrt{}$ V(b2) = 1,016750175

a. Pengujian koefisien b_1

Hipotesis Ho: $\beta_1 = 0$

Alternatif Ha: $\beta_1 \neq 0$

$$t_{\text{hitung}} = \frac{b_1}{s.e\ (b_1)} = \frac{2{,}49116607}{0{,}703147894} = 3{,}543$$

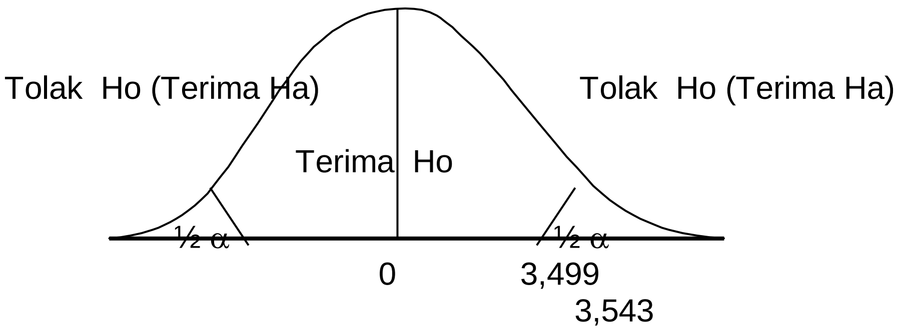
t $_{tabel}$ (1/2 α ; df = n-k-1) dgn koefisien kepercayaan α = 99 %

atau p = 1 - α = 1 - 0,99 = 0,01 ; v atau df = 10 - 2 - 1 = 7

t $_{tabel}$ (1/2 α ; df = n-k-1) =

t $_{tabel}$ (p = 0,005; v = 7) = 3,4995 untuk tabel 1 arah

t $_{tabel}$ (p = 0,01 ; v = 7) = 3,4995 untuk tabel 2 arah (artinya nilai t pada 0,01 adalah nilai pada kedua arahnya yang 0,005)



Uji pihak kanan t $_{hit}$ = 3,543 > t $_{tabel}$ (p = 0,005 ; v = 7) = 3,4995 maka tolak Ho atau terima Ha (highly significance) artinya koefisien arah b_1 secara partial (tanpa pengaruh variabel X_2) dapat digunakan sebagai penduga (estimator) yang dipercaya untuk memprediksi pengaruh variabel bebas X_1 terhadap Y.

b. Pengujian koefisien b2

Hipotesis Ho: $\beta_2 = 0$

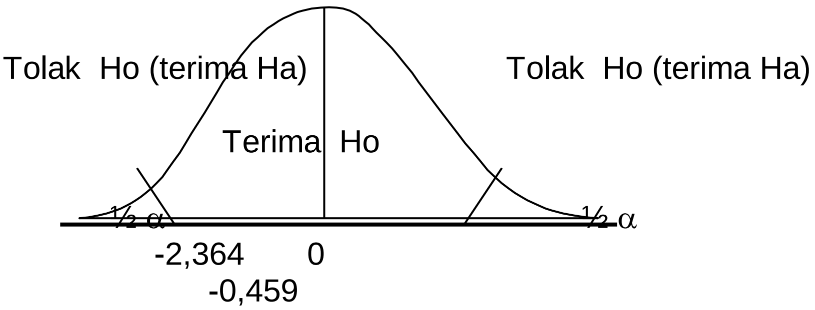
Alternatif Ha: $\beta_2 \neq 0$

$$t_{hitung} = \frac{b_2}{s.e\,(b_2)} = \frac{-\,0{,}46643108}{1{,}016750175} = -\,0{,}459$$

t $_{tabel}$ (½ α ; df = n-k-1) =

t $_{tabel}$ (p = 0,025; v = 7) = 2,3646 untuk tabel 1 arah

t $_{tabel}$ (p = 0,05 ; v = 7) = 2,3646 untuk tabel 2 arah (artinya nilai t pada 0,05 adalah nilai pada kedua arahnya yang 0,025)



Karena t $_{hitung}$ negatif = - 0,459 maka uji t adalah uji pihak kiri.

t_{hit} = - 0,459 > t_{tabel} (p = 0,025; v = 7) = - 2,3646 atau terletak dalam daerah terima Ho maka terima Ho atau tolak Ha (nonsignificance) artinya koefisien arah b_2 secara partial (tanpa pengaruh variabel X_1) tidak dapat digunakan sebagai penduga (estimator) yang dipercaya untuk memprediksi pengaruh variabel bebas X_2 terhadap Y.

4. KOEFISIEN DETERMINASI DAN KORELASI BERGANDA

$$R^2 \text{ adjusted} = 1 - \frac{\sum ei^2 / n\text{-}k\text{-}1}{\sum yi^2 / n - 1} = 1 - \frac{MSE}{MST} = 1 - \frac{6,36}{272/9} = 0,79$$

$$R^2 = \frac{SSR}{SST} = \frac{227,51}{272} = 0,836 = 0,84$$

$$R = \pm \sqrt{R^2} = 0,91$$

5. KOEFISIEN KORELASI PARSIAL

a. Hubungan Y dengan X_1 yakni r_{yx1} atau r_{y1}

$$r_{y1} = \frac{SPx_1\, y_i}{\sqrt{SSX_1\, SSY}} = \frac{\sum x_{1i}\, y_i}{\sqrt{\sum x_1^2 \sum y_i^2}}$$

$$= \frac{\Sigma X_1 Y - \dfrac{\Sigma X_{1i} Y_i}{n}}{\sqrt{\Sigma X_1^{\ 2} - \dfrac{(\Sigma X_1)^2}{n} * \Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}}}$$

$$= \frac{1122 - (60)(170)/10}{\sqrt{\{406 - (60)^2/10\}\{3162 - (170)^2/10\}}} = 0{,}911878138$$

b. Hubungan Y dengan X_2 yakni $r_{yx}{}^2$ atau $r_y{}^2$

$$r_y{}^2 = \frac{SPx_2\, y_i}{\sqrt{SSX_2\ SSY}} = \frac{\sum x_{2i}\, y_i}{\sqrt{\sum x_2{}^2 \sum y_i{}^2}}$$

$$= \frac{\Sigma X_2 Y - \dfrac{\Sigma X_2\ \ \Sigma Y_i}{n}}{\sqrt{\Sigma X_2^{\ 2} - \dfrac{(\Sigma X_2)^2}{n} * \Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}}}$$

$$= \frac{737 - (40)(170)/10}{\sqrt{\{182 - (40)^2/10\}\{3162 - (170)^2/10}} = 0{,}736849958$$

c. Hubungan X_1 dengan X_2 yakni r_{x1x2} atau r_{12}

$$r_{12} = \frac{SPx1\ x2}{\sqrt{SSX1\ \ SSX2}} = \frac{\sum x1i\ x2i}{\sqrt{\sum x_1{}^2\ \ \sum x_2{}^2}}$$

$$= \frac{\Sigma X_1 X_2 - \dfrac{\Sigma X_1\ \ \Sigma X_2}{n}}{\sqrt{\Sigma X_1^{\ 2} - \dfrac{(\Sigma X_1)^2}{n} * \Sigma X_2^{\ 2} - \dfrac{(\Sigma X_2)^2}{n}}}$$

$$= \frac{267 - (60)(40)/10}{\sqrt{\{406 - (60)^2/10\}\{182 - (40)^2/10\}}} = 0{,}848737728$$

Koefisien korelasi sederhana di atas bukan merupakan derajad keeratan hubungan yang sebenarnya antara dua variabel yang dikorelasi karena munculnya variabel ke tiga.

Karena itu **Koefisien korelasi parsial** perlu mengeliminir faktor koreksinya dengan rumus sbb:

a. Korelasi Y dengan X_1 dengan asumsi X_2 konstan ($r_{y1.2}$)

$$r_{y1.2} = \frac{r_{y1} - r_{y2}\, r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}} = \frac{0{,}911878138 - 0{,}625392359}{\sqrt{(1- 0{,}54294786)(1-0{,}72035573)}}$$

$$= \frac{0{,}286485778}{0{,}357508058} = 0{,}8013 = 0{,}80$$

b. Korelasi Y dengan X_2 dengan asumsi X_1 konstan ($r_{y2.1}$)

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\, r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}} = \frac{0{,}736849958 - 0{,}773945379}{\sqrt{(1-0{,}831521738)(1-0{,}72035573)}}$$

$$= \frac{- 0{,}03709542}{0{,}217057551} = - 0{,}1709 = - 0{,}17$$

c. Korelasi X1 dengan X2 dengan asumsi Y konstan ($r_{12.y}$)

$$r_{12.y} = \frac{r_{12} - r_{y1}\, r_{y2}}{\sqrt{(1-r_{y1}^2)(1-r_{y2}^2)}} = \frac{0{,}848737728 - 0{,}671917367}{\sqrt{(1-0{,}831521738)\,(1-0{,}54294786)}}$$

$$= \frac{0{,}17682036}{0{,}277494774} = 0{,}6372 = 0{,}64$$

MATRIX KORELASI PARSIAL

R	X1	X2	Y
X1	1	0,64	0,80
X2	0,64	1	-0,17
Y	0,80	-0,17	1

BAB IV

ANALISA REGRESI NONLINIER

A. PENDAHULUAN

Asumsi linear tidak selamanya dapat digunakan untuk semua variabel karena model populasi dari variasi data tertentu tidak linear sehingga regresi non-linear dibutuhkan. Terdapat banyak bentuk kurva non-linear yang dapat digunakan untuk menyatakan hubungan antara dua atau lebih variabel, karena itu dalam analisis hasil penelitian biasanya ditentukan terlebih dahulu bentuk kurva yang paling mendekati dalam mengekspresikan variasi data melalui scatterplot diagram. Seringkali diperlukan juga pengalaman maupun informasi literatur dalam memipilih tipe kurva regresi yang lebih logis untuk diuji signifikansinya. Dalam tulisan ini akan diuraikan beberapa bentuk persamaan fungsi dan gambar kurva regresi non-linear, antara lain:

- Regresi polinomial
- Fungsi perpangkatan
- Double log transformation
- Fungsi exponential
- Fungsi logaritmik
- Semi-log transformation
- Fungsi reciprocal

Terdapat banyak bentuk kurva non-linear yang dapat digunakan untuk menyatakan hubungan antara dua atau lebih variabel, karena itu dalam analisis hasil penelitian biasanya ditentukan terlebih dahulu bentuk kurva yang paling mendekati mengekspresikan data melalui scatterplot diagram. Pekerjaan ini tidak mudah dan kadang2 tidak mungkin dilakukan. Seringkali melalui pengalaman maupun informasi literatur dapat dipilih tipe kurva regresi yang lebih logis untuk diuji signifikansinya.

B. REGRESI POLINOMIAL

$$Y = bo + b1\ X + b2\ X2 + \ldots + br\ Xr$$

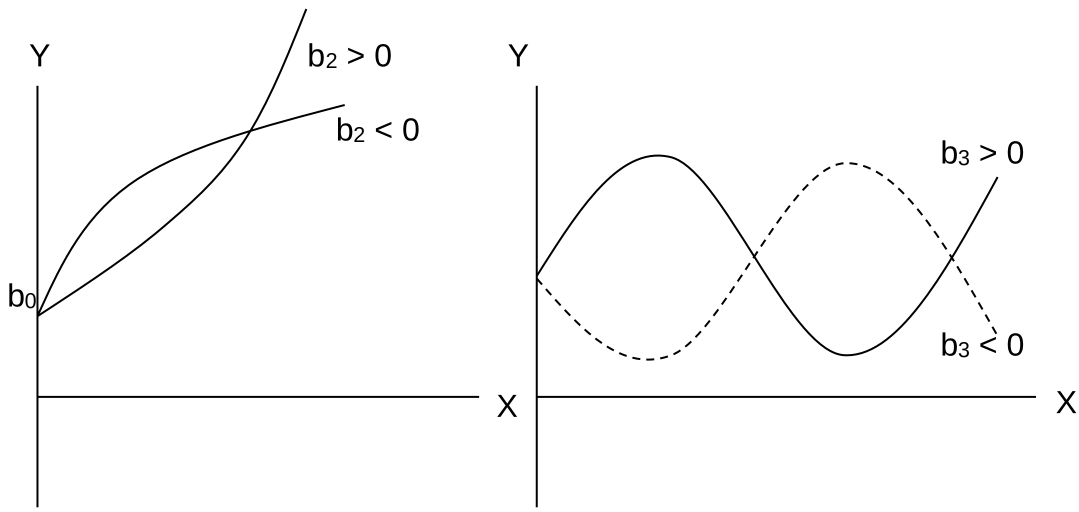
Koefisien regresinya bersifat linear. Disebut fungsi polinomial berderajad r *(rth degree polinomial)* karena pangkat tertinggi dari X adalah r. Untuk menghitung koefisien regresi polinomial berderajad r diperlukan $n > r + 1$ pasang data. Regresi polinomial yang banyak digunakan adalah:

$Y = bo + b1\ X + b2\ X2$ atau

$Y = bo + b1\ X1 + b2\ X2$

dimana $X_2 = X_1^2$ dan penyelesaiannya akan persis sama dengan penyelesaian regresi linear berganda.

a. **Fungsi Kwadratik** atau regresi parabolik r = 2 adalah fungsi polinomial berderajad 2



Regresi parabolik r = 2 Regresi kubik r = 3

Regresi polinomial memiliki koefisien intersep b_o dan koefisien arah b_1 dan b_2.

Jika $b_2 > 0$ maka pengaruh X terhadap Y dengan marginal yang meningkat (increassing rate) yakni pada setiap penambahan 1 (satu) unit variabel X akan menyebabkan penambahan variabel Y yang lebih besar dari sebelumnya.

Sebaliknya jika $b_2 < 0$ maka pengaruh X terhadap Y dengan marginal yang menurun (decreassing rate) yakni pada setiap penambahan 1 (satu) unit variabel X akan menyebabkan penambahan variabel Y yang lebih kecil dari sebelumnya.

b. **Regresi Kubik** $r = 3$ adalah fungsi polinomial berderajad 3.

$Y = b_o + b_1 X + b_2 X_2 + b_3 X_3$ atau

$Y = b_o + b_1 X_1 + b_2 X_2 + b_3 X_3$

dimana $X_2 = X_1^2$ dan $X_3 = X_1^3$

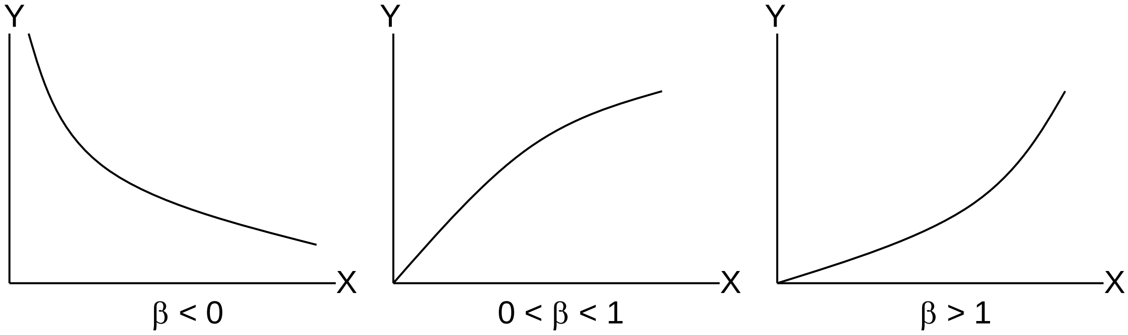
Penyelesaiannya juga akan persis sama dengan penyelesaian regresi linear berganda.

C. FUNGSI PERPANGKATAN

$$Y = \alpha X^{\beta}$$

Diasumsikan bahwa X selalu positif

Tiga buah grafik untuk berbagai harga β dengan $\alpha > 0$ adalah



$\beta < 0 \rightarrow Y = \alpha X^{-\beta} = \alpha / X^{\beta}$

Jadi Y semakin berkurang dengan bertambahnya nilai X dan pada X = 0 maka Y akan tak terhingga

$0 < \beta < 1 \rightarrow Y = \alpha X^{\beta}$

jadi Y bertambah dengan bertambahnya nilai X namun pertambahan tersebut dengan marginal yang berkurang dan pada X = 0 maka Y = 0

$\beta > 1 \rightarrow$ Y bertambah dengan marginal yang semakin besar jika X ditingkatkan, dan pada X = 0 maka Y = 0

Penyelesaiannya melalui transformasi logarithma ke dalam bentuk linear sehingga $Y = \alpha X^{\beta}$ menjadi:

$$\log Y = \log \alpha + \beta \log X \quad \text{atau}$$

$$Y = a + b X$$

dimana $Y = \log Y$

$a = \log \alpha$; jadi α = anti - log a = 10^{a}

$X = \log X$

Penyelesaian selanjutnya persis sama dengan regresi linear sederhana.

Nilai prediksi α diperoleh melalui antilog a (invers log a). Sedangkan nilai b pada persamaan regresi linear tetap merupakan nilai β pada fungsi perpangkatannya, namun nilai β disini diprediksikan sebagai koefisien elastisitas (bukan koefisien marginal). Jika $\beta < 0$ maka pengaruh X terhadap Y berkorelasi negatif dengan penurunan Y yang bertambah besar untuk setiap penambahan 1 unit X. Ini terjadi karena pengaruh faktor eksternal yang dominan sehingga jika $X = 0$ maka Y menjadi tak terhingga.

Jika $0 < b < 1$ maka pengaruh X terhadap Y berkorelasi positif dengan kenaikan Y yang bertambah kecil untuk setiap penambahan 1 unit X atau identik dengan *"the law of deminishing marginal return"* dalam teori produksi. Ini terjadi karena pengaruh faktor2 tetap *(fixed)* yang semakin langka *(scarcity)* sehingga jika faktor variabelnya ditambah menyebabkan daya dukung dari faktor tetap semakin terbatas dan suatu saat justru terjadi over capacity dimana jika faktor X masih ditambahkan menyebabkan Y justru menurun. Jika X dan Y adalah pengaruh input - output maka b antara 0 dan 1 terjadi pada tahap produksi II yang rasional.

Jika $\beta > 1$ maka pengaruh X terhadap Y berkorelasi positif dengan kenaikan Y yang bertambah besar untuk setiap penambahan 1 unit X atau identik dengan "increassing rate". Ini terjadi karena pengaruh faktor2 tetap (fixed) yang sebagian besar belum digunakan (kapasitas nganggur) sehingga jika faktor variabelnya ditambah menyebabkan daya dukung dari faktor tetap semakin bertambah. Jika X dan Y adalah pengaruh input - output maka $\beta > 1$ terjadi pada tahap produksi I yang irasional karena terdapat kapasitas nganggur.

Dalam tipe 2 dan 3, jika $X = 0$ maka Y juga akan = 0 atau intersep = 0 artinya faktor X mutlak diperlukan.

Nilai X = 0 pada tipe 1 tidak akan pernah terjadi karena jika X dikurangisampai tingkat tertentu maka fungsi produksi akan berobah bentuk menjadi tipe2 selanjutnya bisa menjadi tipe 3.

D. FUNGSI COB-DOUGLASS

Adalah salah satu bentuk fungsi perpangkatan yang banyak digunakan dalam penelitian produksi pertanian. Interpretasinya identik di atas.

$Y = \alpha \; X1^{\beta 1} \;\; X2^{\beta 2} \ldots \;\; Xn^{\beta n}$

Transformasinya ke dalam bentuk linear adalah sbb:

$\log Y = \log \alpha + \beta 1 \log X1 + \beta 2 \log X2 + \ldots + \beta n \log Xn$

Penyelesaiannya sama dengan regresi linear berganda yakni:

$Y = a + b1\ X1 + b2\ X2 + \ldots + bn\ Xn$

Dimana: Y = log Y
X1 = log X1
X2 = log X2
Xn = log Xn
a = log α ; jadi α = anti-log a (atau invers log a).
b1 = $\beta 1$; b2 = $\beta 2$; bn = βn

$\beta 1$, $\beta 2$, ..., βn = koefisien elastisitas, berbeda dengan linear berganda dimana $\beta 1$, $\beta 2$, ..., βn = koefisien marjinal.

E. DOUBLE LOG TRANSFORMATION

Merupakan variasi dari fungsi perpangkatan. Ada dua bentuk double log transformation yakni:

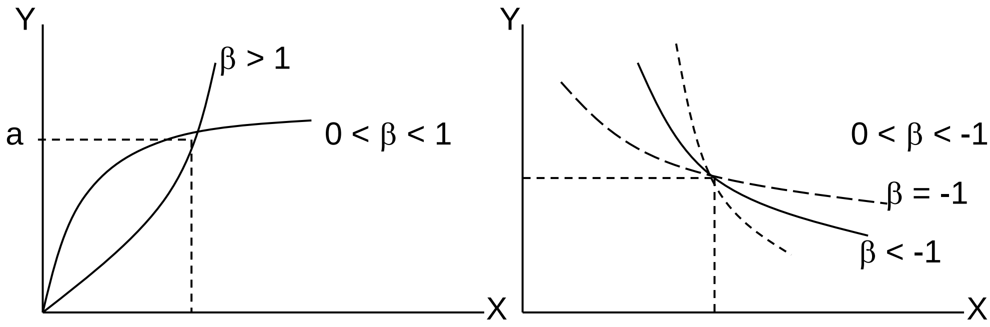
1. $\log Y = \alpha + \beta \log X$

   Identik dengan fungsi perpangkatan $Y = a\ X^{\beta}$

   namun $a = \log \alpha$ (transformasi log) dan $dY/dX = a \,.\, \beta \,.\, X^{\beta - 1}$

   Sehingga bila β positif > 1 koefisien arahnya akan semakin bertambah dengan makin bertambahnya nilai X.

   Sedangkan bila $0 < \beta < 1$ koefisien arah semakin berkurang dengan makin bertambah nilai X.



Grafik log Y = α + β log X　　　Grafik log Y = α - β log X

2. $\log Y = \alpha - \beta \log X$

   Identik dengan fungsi perpangkatan $Y = a X^{-\beta}$

   namun $a = \log \alpha$ (transformasi log)

   dan $dY/dX = -a \,.\, \beta \,.\, X^{-(\beta - 1)} = -a \,.\, \beta \,.\, X^{-(\beta + 1)}$

   $$= -\frac{a \,.\, \beta}{X^{(\beta + 1)}}$$

   Sehingga koefisien arah negatif dimana Y semakin berkurang dengan makin bertambahnya nilai X dan sebaliknya.

   Bila $\beta = 1$ menghasilkan rectangular hyperbola dimana locus dari titik-titik hasil-kali koordinat XY merupakan suatu bilangan konstant.

F. FUNGSI EXPONENTIAL

$$\boxed{Y = \alpha\, e^{\beta X}}$$

Transformasinya adalah:

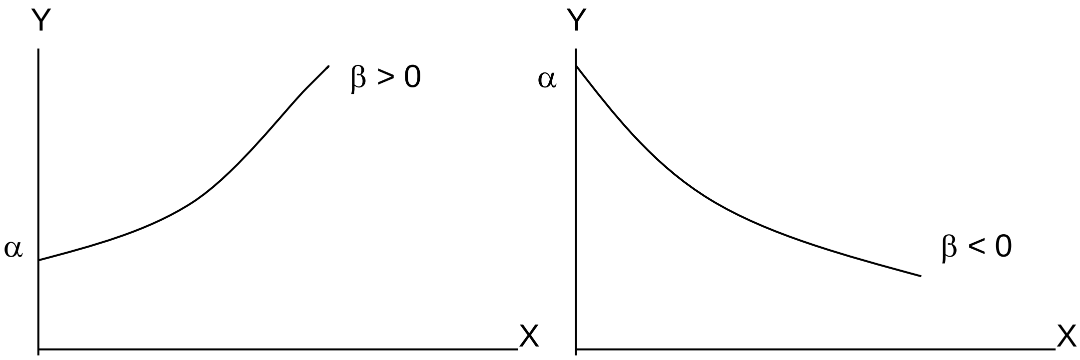
$$\boxed{\begin{array}{l} \ln\ Y = \ln \alpha + \beta X \\ Y = a + b X \end{array}}$$

Dimana: Y = ln Y

$a = \ln \alpha$; jadi α = anti - ln a = e^{a} = $2{,}7183^{a}$

$b = \beta$

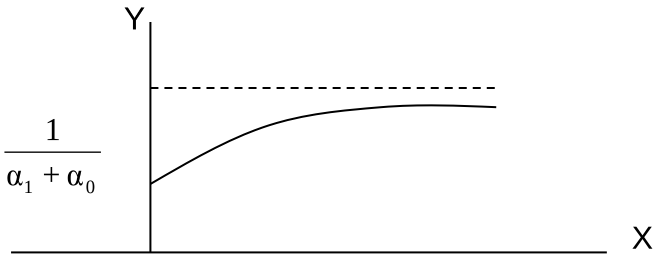
Penyelesaiannya identik dengan fungsi linear sederhana



Grafik fungsi $Y = \alpha \; e^{\beta X}$

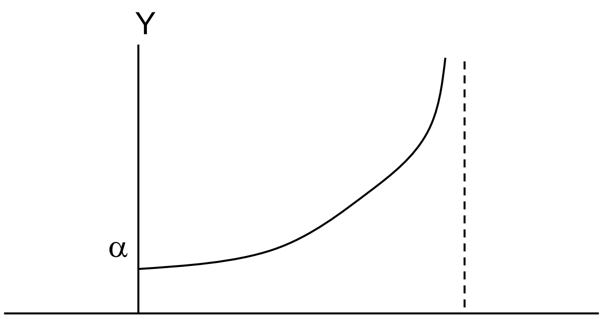
Model lainnya yang tergolong fungsi eksponential yakni:

1. Kurva Logistik



$$Y = \frac{1}{\alpha_0 + \alpha_1 \; e^{-\beta x}}$$

2. Model Pertumbuhan Populasi (Population Growth Model)



$N_t = N_0 \; e^{rt}$ identik dengan

$Y = \alpha \; e^{\beta x}$

X = t

N_t = jumlah populasi th t N_o = jumlah populasi awal

r = kecepatan pertumbuhan per tahun t = jumlah tahun

3. Logarithmic Reciprocal Transformation

Bila fungsi eksponensialnya $Y = e^{\alpha-\beta/x}$ maka dalam analisisnya dilakukan transformasi ke logarithmic reciprocal sehingga menjadi bentuk linear

$$\boxed{\ln Y = \alpha - \beta / X}$$

Untuk $X = 0$ tidak dapat ditentukan besarnya Y, namun untuk X mendekati 0 maka Y juga akan mendekati 0.

Karena itu titik (0;0) dianggap sebagai titik awal dari fungsi ini.

$$\boxed{\frac{dY}{dX} = e^{\alpha-\beta/x}\left\langle\frac{-\beta}{X^2}\right\rangle}$$

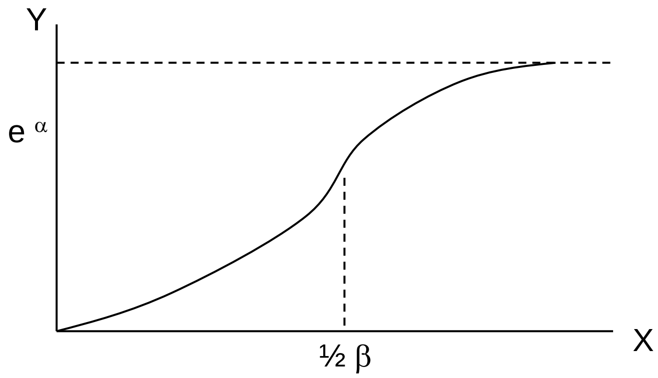
yang berarti sudut kemiringan (koefisien arah) dari fungsi ini akan positip untuk nilai X positif (X > 0).

$$\boxed{\frac{d^2Y}{dX^2} = e^{\alpha-\beta/x}\left\langle\frac{\beta^2}{X^4} - \frac{2\beta}{X^3}\right\rangle}$$

Terdapat titik balik (inflection point) pada $X = \frac{1}{2}\beta$.

Di sebelah kiri titik ini koefisien arah akan bertambah sedangkan di sebelah kanannya akan berkurang dengan semakin bertambah-nya nilai X.

Untuk $X = \infty$ maka $Y = e^{\alpha}$ sehingga grafik fungsinya adalah sebagai berikut:



$$\text{Ln } Y = \alpha - \beta / X$$

$$Y = e^{\alpha - \beta / x}$$

G. FUNGSI LOGARITHMIK

$$\boxed{e^{Y} = \alpha X^{\beta} \quad \text{(X positif)}}$$

Transformasi ke logarithma naturalnya adalah:

$$\boxed{Y = \ln \alpha + \beta \ln X}$$

Penyelesaiannya identik dengan regresi linear sederhana

$$\boxed{Y = a + b X}$$

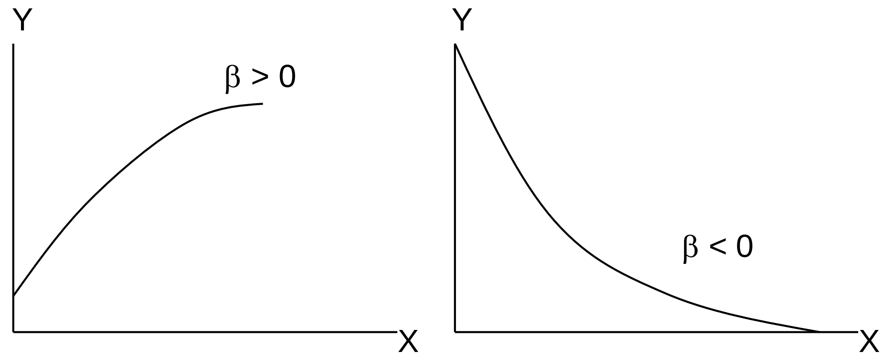
Dimana:

$Y = Y$

$X = \ln X$

$a = \ln \alpha$; jadi α = anti-ln $a = 2{,}7183^{a}$

$b = \beta$



Grafik fungsi $e^{Y} = \alpha X^{\beta}$

H. SEMI-LOG TRANSFORMATION FUNCTION

$$\boxed{Y = \alpha + \beta \ln X}$$

dimana $\frac{dY}{dX} = \frac{\beta}{X}$

karena itu besarnya koefisien arah (sudut kemiringan kurva) akan semakin berkurang dengan semakin bertambahnya nilai X.

Pada saat $Y = 0$ maka $\ln X = - \alpha / \beta$

Sehingga titik potong kurva dengan sumbu X terletak pada:

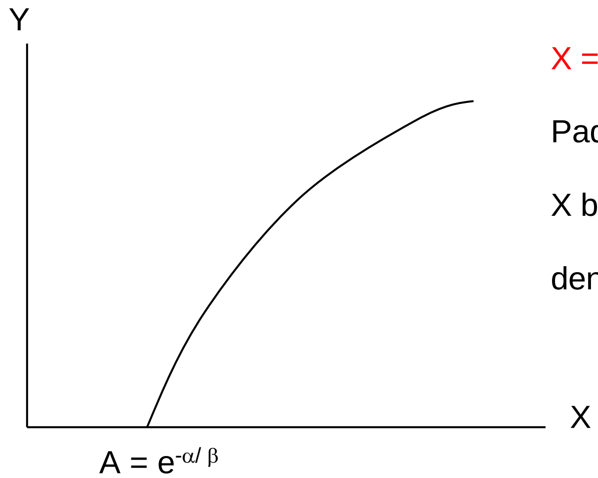
$$\boxed{X = e^{-\alpha/\beta}}$$

Invers dari fungsi ini ialah $X = e^{-\alpha/\beta}\ e^{Y/\beta}$

yang dapat ditulis sebagai berikut $X = A\,B^{Y}$

dimana $A = e^{-\alpha/\beta}$ dan $B = e^{1/\beta}$

Fungsi ini sering disebut *"Steady Growth Function"*



$X = A\,B^Y$

Pada $Y = 0$ maka $X = A$; jika

X bertambah maka Y meningkat

dengan marginal berkurang

I. FUNGSI RECIPROCAL

Ada 2 (dua) bentuk fungsi reciprocal, yakni:

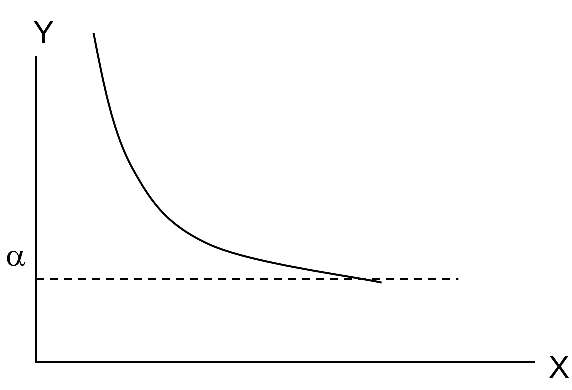
1. $\boxed{Y = \alpha + \beta / X}$

$$\frac{dY}{dX} = -\frac{\beta}{X^2}$$

artinya sudut kemiringan dari fungsi ini bersifat negatif dengan marginal yang semakin besar. Jadi nilai Y akan semakin berkurang (marginalnya turun semakin besar) jika nilai X bertambah.

Untuk $X = 0$ maka $Y = \infty$

$X = \infty$ maka Y = mendekati α

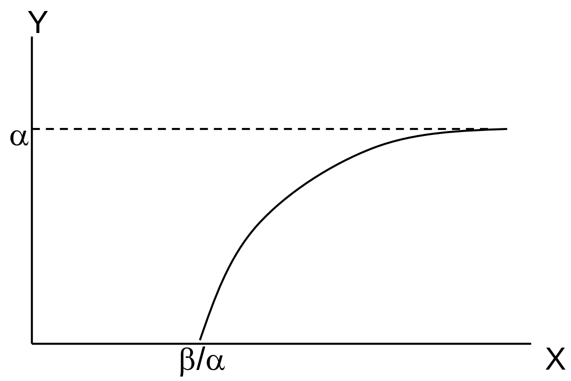


2. $\boxed{Y = \alpha - \beta / X}$

$$\frac{dY}{dX} = \frac{\beta}{X^2}$$

artinya sudut kemiringan dari fungsi ini bersifat positif dengan marginal yang semakin kecil.

Jadi nilai Y akan bertambah namun dengan pertambahan (marginal) yang berkurang jika nilai X bertambah.



Untuk X = 0 maka Y $= \infty$

X = ∞ maka Y $= \alpha$

Y = 0 maka $\beta/X = \alpha$

$\alpha . X = \beta$

$X = \beta / \alpha$

BAB V

ANALISA REGRESI VARIABEL DUMMY

A. PENDAHULUAN

Dilihat dari cara mengukur, maka variabel dibedakan atas:

1. Variabel Anumerik; disebut juga variabel nominal atau variabel kualitatif, atau variabel kategori baik yang bersifat perbedaan jenis (klaster) maupun perbedaan derajad (strata).
2. Variabel Numerik; disebut juga sebagai variabel kuantitatif baik yang bersifat kontinum maupun diskrit.

(Variabel anumerik yang strata dapat dijadikan numerik dengan cara diskore).

Disisi lain skala pengukuran dibedakan atas:

1. skala nominal
2. skala ordinal
3. skala interval, dan
4. skala ratio

Skala nominal dan ordinal bisa digunakan untuk mengukur variabel kualitatif, sedangkan skala interval dan ratio digunakan untuk mengukur variabel numerik.

Variabel kualitatif sering disebut variabel dummy. Dalam analisis regresi sering dijumpai bahwa variabel dependen tidak hanya dipengaruhi oleh variabel kuantitatif tetapi dipengaruhi pula oleh variabel kualitatif (mis: jenis kelamin, ras, warna kulit, agama, kebangsaan, perang, musim, pemogokan, kebijakan, dll).

Metode untuk membuat variabel kualitatif menjadi kuantitatif adalah dengan membentuk variabel buatan (dummy) yang mengambil nilai 1 atau 0; karena itu variabel kualitatif sering disebut variabel dummy .

D = 1 menunjukkan keberadaan kategori tertentu (mis: laki2, lulusan PT, dan lain-lain).

D = 0 artinya tidak tergolong kategori tersebut tetapi kategori lain (misal: perempuan, bukan lulusan PT, dll).

Variabel yg diberi nilai 1 dan 0 disebut variabel dummy/ variabel binary/ variabel kategori/ variabel dichotom.

Dalam analisis variabel dummy dikenal aturan umum sebagai berikut:

1. Jika suatu variabel kualitatif mempunyai m kategori, hanya dibuat m - 1 variabel dummy (menghindari multikolineariti).
2. Penetapan nilai 0 dan 1 bersifat *arbitrary* (tanpa dasar) artinya dapat dipertukarkan antar kategori.
3. Kategori yang diberi nilai 0 disebut kategori dasar/ kontrol/ perbandingan. Merupakan dasar bahwa perbandingan dibuat dalam kategori tersebut, ditetapkan bersifat apriori. Intersep α_o = intersep untuk kategori dasar.
4. Koefisien α_1 disebut koefisien intersep diferensial karena menyatakan berapa banyak nilai unsur intersep dari kategori nilai 1 berbeda dari koefisien intersep kategori dasar.

B. REGRESI ATAS 1 VARIABEL KUANTITATIF DAN 1 VARIABEL KUALITATIF DENGAN 2 KATEGORI

Misal regresi gaji karyawan pertahun (Y) terhadap jenis kelamin (D) dan masa kerja (pengalaman mengajar) (X) seperti pada lampiran 1.

$$\boxed{Y_i = \alpha_o + \alpha_1 D_1 + \alpha X + u}$$

Y = gaji karyawan per tahun (Rp)

X_1 = masa kerja (th)

D_1 = Variabel kualitatif jenis kelamin dengan 2 kategori: laki2 dan wanita

D = 1 karyawan laki-2

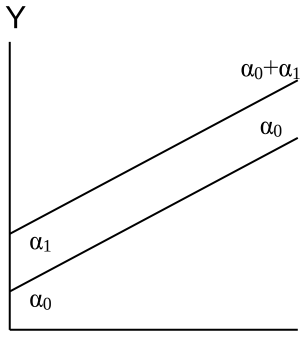
D = 0 karyawan wanita

α_o = intersep karyawan wanita à1 = intersep karyawan laki2

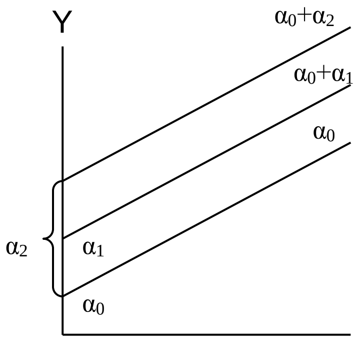
Ekspektasi gaji karyawan per tahun (lihat Gambar 1)

Gaji karyawan wanita per tahun $E(Y_i \mid X_i, D_i = 0) = \alpha_o + \alpha X_i$

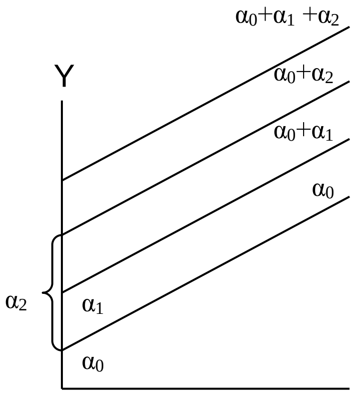
Gaji karyawan pria per tahun $E(Y_i \mid X_i, D = 1) = (\alpha_o + \alpha_1) + \alpha X_i$



Gambar 1



Gambar 2



Gambar 3

Contoh Soal:

n	Y (Juta Rp)	D_1 (? / ?)	X (jml th)
1	22	1	8
2	20	1	6
3	18	0	8
4	17	1	4
5	16	0	7
6	14	0	4

NB: Jumlah observasi untuk setiap kategori tidak perlu sama. Jumlah wanita tidak perlu harus = jumlah pria.

$Y = \alpha_0 + \alpha_1 D_1 + \beta X_1$

$Y = 9{,}16 + 4{,}03\ D_1 + 1{,}08\ X_1$

t = (7,674) (6,895) R^2 = 0,95

df = n – k - 1 = 6 - 3 = 3 t tabel 0,05 = 3,1825 t tabel 0,01 = 5,8409

Hasil analisis menunjukkan bahwa semua koefisien variabel independent signifikan pada taraf kepercayaan 99 %.

Ekspektasi gaji karyawan wanita/ tahun dengan masa kerja 10 tahun adalah:

$E(Y \mid X, D_1 = 0) = \alpha_o + \beta X_1$

$Y = 9{,}16 + 1{,}08\ (10) = \text{Rp}.19.960.000,-$

Ekspektasi gaji karyawan pria/ tahun dengan masa kerja 10 tahun adalah:

$E(Y \mid X, D_1 = 1) = (\alpha_o + \alpha_1) + \beta X_1$

$Y = 13{,}19 + 1{,}08\ (10) = \text{Rp}.23.990.000,-$

C. REGRESI ATAS 1 VARIABEL KUANTITATIF DAN 1 VARIABEL KUALITATIF DENGAN LEBIH DARI 2 KELAS/ KATEGORI

Misal regresi pengeluaran tahunan untuk perawatan kesehatan (Y) terhadap pendapatan (X) dan pendidikan (D). Pendidikan dengan 3 kategori yang mutually exclusive, yakni:

a. lebih rendah dari SLTA

b. SLTA, dan

c. Perguruan Tinggi

Persamaan regresi linearnya adalah:

$$\boxed{Y = \alpha_o + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \beta X_i + u_i}$$

Y = pengeluaran tahunan untuk perawatan kesehatan (Rp)

X = pendapatan per tahun (Rp)

D1= 1 pendidikan SLTA

= 0 untuk yang lain (< SLTA)

D_2 = 1 pendidikan Perguruan Tinggi

= 0 untuk yang lain (< SLTA)

Secara *arbitrary* pendidikan lebih rendah dari SLA ditetapkan sebagai kategori dasar.

α_o = intersep pendidikan < SLTA

α_1 = intersep pendidikan SLTA

α_2 = intersep PT

Ekspektasi pengeluaran tahunan untuk pemeliharaan kesehatan (lihat Gambar2)

- Pendidikan < SLTA $E(Y \mid X, D_1=0, D_2=0) = \alpha_o + \beta X_i$
- Pendidikan SLTA $E(Y \mid X, D_1=1, D_2=0) = (\alpha_o + \alpha_1) + \beta X_i$
- Pendidikan PT $E(Y \mid X, D_1=0, D_2=1) = (\alpha_o + \alpha_2) + \beta X_i$

Regresi di atas dapat dikembangkan untuk lebih dari satu variabel kuantitatif.

Hasil observasi dari 20 responden karyawan industri sanitair pilar diperoleh data untuk regresi tingkat upah (Y) terhadap produktivitas (X_1), umur (X_2) dan pendidikan (D_1), lihat lampiran 2. Pendidikan dengan 3 kategori mutually exclusive:

a. SLTA

b. SLTP, dan

c. SD (kategori dasar)

Persamaan regresinya adalah:

$$\boxed{Y = \alpha_o + \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 X_1 + \beta_2 X_2 + u}$$

Y = upah industri sanitair pilar (Rp)

X_1 = produktivitas karyawan (unit)

X_2 = umur karyawan (th)

D_1 = 1 pendidikan SLTA

= 0 untuk yang lain

$D_2 = 1$ pendidikan SLTP

$= 0$ untuk yang lain

Secara *arbitrary* pendidikan SD ditetapkan sebagai kategori dasar

α_o = intersep pendidikan SD

α_1 = intersep pendidikan SLTA

α_2 = intersep SLTP

Ekspektasi pendapatan per bulan

SD $E(Y \mid X_1, X_2;\ D_1=0,\ D_2=0) = \alpha_o + \beta_1 X_1 + \beta_2 X_2$

SLTA $E(Y \mid X_1, X_2;\ D_1=1,\ D_2=0) = (\alpha_o + \alpha_1) + \beta_1 X_1 + \beta_2 X_2$

SLTP $E(Y \mid X_1, X_2;\ D_1=0,\ D_2=1) = (\alpha_o + \alpha_2) + \beta_1 X_1 + \beta_2 X_2$

Sehingga akan diperoleh hasil sebagai berikut:

$Y = -\ 7641{,}3642 + 1090{,}31\ D_1 + 1630{,}6327\ D_2 + 2962{,}5897\ X_1 - 14{,}6063\ X_2$

t	=	(1,897)	(2,879)	(18,895)	(-0,254)
prob.	=	(0,077)	(0,011)	(0,0000)	(0,803)

F_{ratio} = 178,006 (P = 1,910E-12)

R^2 = 0,97

Secara simultan persamaan regresi tersebut dapat diterima sebagai estimator dengan tingkat kepercayaan di atas 99 % (F_{ratio} = 178,006). Secara parsial tampak bahwa koefisien SLTA signifikan pada $\alpha = 0{,}08$ (taraf kepercayaan 92 %), SLTP pada $\alpha = 0{,}02$ (taraf kepercayaan 98 %) dan produktivitas signifikan pada $\alpha = 0{,}01$ (taraf kepercayaan 99 %) sedangkan variabel umur non-signifikan.

Interpretasi regresi dilakukan setelah mengeleminir/ mengeluarkan variabel X_2 dari persamaan regresi di atas (dengan regresi stepwise). Hasilnya akan menunjukkan koefisien probabilitas yang lebih bagus (semakin kecil).

Jika variabel umur dikeluarkan akan diperoleh persamaan regresi sebagai berikut:

$Y = -7624{,}2127 + 1068{,}5794\ D_1 + 1660{,}7573\ D_2 + 2942{,}6172\ X_1$

t	=	(1,938)	(3,090)	(22,304)
prob.	=	(0,070)	(0,007)	(0,0000)

F_{ratio} = 252,059 (P = 1,8E-13)

R^2 = 0,97

Ekspektasi tingkat upah per bulan menurut tingkat pendidikan jika produktivitas rata-rata = 19 unit/ bulan:

SD $E(Y \mid X_1,\ D_1=0,\ D_2=0) = \alpha_o + \beta_1 X_1$

= - 7624,2127 + 2942,6172 (19) = Rp.48.285,50

SLTA $E(Y \mid X_1,\ D_1=1,\ D_2=0) = (\alpha_o + \alpha_1) + \beta_1 X_1$

= - 6555,6333 + 2942,6172 (19) = Rp.49.354,10

SLTP $E(Y \mid X1,\ D1=0,\ D2=1) = (\alpha_o + \alpha_2) + \beta_1 X_1$

= - 5963,4554 + 2942,6172 (19) = Rp.49.946,30

Ekspektasi semacam ini seolah-olah menunjukkan hanya diskriminasi upah antar tingkat pendidikan yang menyebabkan perbedaan tingkat upah. Seharusnya kesimpulan tidak demikian karena tingkat produktivitas antar kategori pendidikan juga berbeda, dimana rata-rata produkstivitas karyawan dengan pendidikan SLTP lebih tinggi (19,625) berikutnya SD (18,857) dan SLTA produktivitasnya 18,2.

Jadi kesimpulannya bahwa latar belakang pendidikan dan produktivitas, keduanya mempengaruhi tingkat upah.

Ekspektasi tingkat upah per bulan menurut tingkat pendidikan dan produktivitas adalah:

SD $E(Y \mid X_1,\ D_1=0,\ D_2=0) = \alpha_o + \beta_1 X_1$

= - 7624,2127 + 2942,6172 (18,857) = Rp.47.865,-

SLTA $E(Y \mid X_1,\ D_1=1,\ D_2=0) = (\alpha_o + \alpha_1) + \beta_1 X_1$

= - 6555,6333 + 2942,6172 (18,2) = Rp.47.000,-

SLTP $E(Y \mid X1, \quad D1=0, \quad D2=1) = (\alpha_o + \alpha_2) + \beta_1 X_1$

$= - 5963,4554 + 2942,6172 (19,625) = Rp.51.785,41$

D. REGRESI ATAS 1 VARIABEL KUANTITATIF DAN 2 VARIABEL KUALITATIF MASING-MASING DENGAN 2 KATEGORI

Misal regresi gaji dosen pertahun (Y) terhadap masa kerja (X), jenis kelamin (D_1) dan warna kulit (D_2).

Jenis kelamin dengan 2 kategori: laki-2 dan wanita

Warna kulit dengan 2 kategori: hitam dan putih

Karena terdiri dari 2 kategori, jadi memerlukan 1 variabel dummy untuk masing-2 variabel kualitatif jenis kelamin dan warna kulit.

Persamaan regresi linearnya adalah:

$$\boxed{Y_i = \alpha_o + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \beta X_i + u_i}$$

Y = Gaji dosen per tahun (Rp)

X = masa kerja (th)

D_1 = 1 laki-laki

= 0 lainnya

D_2 = 1 kulit putih

= 0 lainnya

Kategori dasar/ kategori yang diabaikan adalah dosen wanita berkulit hitam.

Ekspektasi gaji dosen (lihat gambar 3)

- Wanita kulit hitam $E(Y_i \mid X_i, \quad D_1=0, \quad D_2=0) = \alpha_o + \beta X_i$
- Laki-2 kulit hitam $E(Y_i \mid X_i, \quad D_1=1, \quad D_2=0) = (\alpha_o + \alpha_1) + \beta X_i$
- Wanita kulit putih $E(Y_i \mid X_i, \quad D_1=0, \quad D_2=1) = (\alpha_o + \alpha_2) + \beta X_i$
- Laki-2 kulit putih $E(Y_i \mid X_i, \quad D_1=0, \quad D_2=1) = (\alpha_o + \alpha_1 + \alpha_2) + \beta X_i$

Regresi-2 di atas berbeda intersep (α) namun arahnya (β) sama, karena itu garis regresinya sejajar.

Contoh Soal:

n	**Y** Juta RP	D_1 ? / ?	D_2 Kulit	X Jml / th
1	17	0	0	9
2	20	1	0	8
3	18	0	1	10
4	22	1	1	10
5	16	0	0	7
6	19	1	0	8
7	18	0	1	7
8	20	1	1	9
9	16	0	0	5
10	17	1	0	6
11	17	0	1	6
12	18	1	1	7

Persamaan umum Regressinya adalah sebagai berikut:

$$Y = \alpha_o + \alpha_1 D_1 + \alpha_2 D_2 + \alpha X + u$$

Y = Gaji karyawan per tahun (Rp)

X = masa kerja (th)

D1 = 1 laki2

= 0 lainnya

D2 = 1 kulit putih

= 0 lainnya

Kategori dasar (kategori yg diabaikan) adalah karyawan wanita berkulit hitam.

Ekspektasi gaji karyawan menurut warna kulit adalah:

- Wanita kulit hitam $E(Y \mid X, \quad D_1=0, \quad D_2=0) = \alpha_o + \beta X$
- Laki2 kulit hitam $E(Y \mid X, \quad D_1=1, \quad D_2=0) = (\alpha_o + \alpha_1) + \beta X$
- Wanita kulit putih $E(Y \mid X, \quad D_1=0, \quad D_2=1) = (\alpha o + \alpha 2) + \beta X$
- Laki2 kulit putih $E(Y \mid X, \quad D1=1, \quad D2=1) = (\alpha o + \alpha 1 + \alpha 2) + \beta X$

Keempat regresi di atas berbeda intersep (α) namun arahnya (β) sama karena itu garis regresinya sejajar (lihat gambar 3)

Persamaan regresi dari hasil analisis adalah:

$Y = 12{,}68 + 1{,}98\ D_1 + 0{,}80\ D_2 + 0{,}53\ X$

t = (3,551) (1,391) (2,806)

p = (0,0075) (0,2018) (0,02297)

F_{ratio} = 10,815 (P = 3,455E-03)

R^2 = 0,73

df = n-k-1 = 12 - 4 = 8 t tabel 0,30 = 1,108

t tabel 0,05 = 2,306

t tabel 0,01 = 3,355

Hasil analisis menunjukkan bahwa secara serempak persamaan regresi dapat diterima sebagai estimator pada taraf kepercayaan di atas 99 % (F_{ratio} = 10,815).

Secara parsial variabel kualitatif jenis kelamin signifikan pada taraf kepercayaan 99 % dan variabel masa kerja signifikan pada taraf kepercayaan 97 % sedangkan variabel kualitatif warna kulit hanya signifikan pada taraf kepercayaan 79 %. Artinya:

1. Ada faktor pada jenis kelamin dan warna kulit yang mempengaruhi besarnya gaji, namun faktor jenis kelamin berpengaruh lebih dominan (jika dikehendaki indikator dari faktor tersebut dapat diteliti lebih lanjut).
2. Koefisien masa kerja 0,53 artinya setiap tambahan 1 tahun masa kerja diharapkan gaji akan bertambah 0,53 juta rupiah

Ekspektasi pengaruh jenis kelamin, warna kulit dan masa kerja terhadap besarnya gaji per tahun (untuk masa kerja 15 tahun):

- Wanita kulit hitam $E(Y \mid X, D_1=0, D_2=0) = \alpha_o + \beta X$

  $Y = 12{,}68 + 0{,}53\ X = 12{,}68 + 0{,}53\ (15)$ = Rp 20.630.000,-

- Wanita kulit putih $E(Y \mid X, D_1=0, D_2=1) = (\alpha_o + \alpha_2) + \beta X$

  $Y = 13{,}48 + 0{,}53\ X = 13{,}48 + 0{,}53\ (15) =$ Rp 21.430.000,-
- Laki2 kulit hitam $E(Y \mid X, D_1=1, D_2=0) = (\alpha_o + \alpha_1) + \beta X$

  $Y = 14{,}66 + 0{,}53\ X = 14{,}66 + 0{,}53\ (15) =$ Rp 22.610.000,-
- Laki2 kulit putih $E(Y \mid X, D_1=1, D_2=1) = (\alpha_o + \alpha_1 + \alpha_2) + \beta X$

  $Y = 15{,}46 + 0{,}53\ X = 15{,}46 + 0{,}53\ (15) =$ Rp 23.410.000,-

Jadi laki-laki mempunyai ekspektasi lebih tinggi dari wanita dan kulit putih mempunyai ekspektasi lebih tinggi dari kulit hitam.

(Latihan: coba dikerjakan sekali lagi dengan kategori yang dibalik jadi $D_1 = 1$ untuk wanita dan $D_2 = 1$ untuk kulit hitam, disertai kesimpulan).

Jika dikerjakan sekali lagi dengan kategori yang dibalik jadi $D_1 = 1$ untuk wanita dan $D_2 = 1$ untuk kulit hitam, maka kesimpulannya akan tetap sama walaupun koefisien intersep regresinya berubah, koefisien arah regresinya tetap sama namun dengan tanda yang berlawanan (+ menjadi -). Karena itu uji t berlawanan arah menjadi negatif namun dengan koefisien yang tetap sama demikian juga probabilitas dan F_{ratio} tetap sama, lihat lampiran 4.

Persamaan umum regresinya adalah:

$$\boxed{Y = \alpha_o + \alpha_1 D_1 + \alpha_2 D_2 + \beta X + u}$$

Y = Gaji karyawan per tahun (Rp)

X = masa kerja (th)

D1 = 1 wanita

= 0 lainnya

D2 = 1 kulit hitam

= 0 lainnya

Kategori dasar adalah karyawan laki-2 berkulit putih.

Ekspektasi gaji karyawan menurut warna kulit adalah:

- Laki2 kulit putih E(Y | X, D_1=0, D_2=0) = $\alpha_o + \beta X$
- Wanita kulit putih E(Y | X, D_1=1, D_2=0) = $(\alpha_o + \alpha_1) + \beta X$
- Laki2 kulit hitam E(Y | X, D_1=0, D_2=1) = $(\alpha_o + \alpha_2) + \beta X$
- Wanita kulit hitam E(Y | X, D_1=1, D_2=1) = $(\alpha_o + \alpha_1 + \alpha 2) + \beta X$

Persamaan regresi dari hasil analisis adalah:

Y = 15,46 - 1,98 D_1 - 0,80 + 0,53 X

t = (-3,551) (-1,391) (2,806)

p = (0,0075) (0,2018) (0,02297)

F_{ratio} = 10,815 (P = 3,455E-03)

R^2 = 0,73

Ekspektasi pengaruh jenis kelamin, warna kulit dan masa kerja terhadap besarnya gaji per tahun (untuk masa kerja 15 tahun):

- Laki2 kulit putih E(Y | X, D_1=0, D_2=0) = $\alpha_o + \beta X$

  Y = 15,46 + 0,53 X = 15,46 + 0,53 (15) = Rp 23.410.000,-
- Wanita kulit putih E(Y | X, D_1=1, D_2=0) = $(\alpha_o + \alpha_1) + \beta X$

  Y = 13,48 + 0,53 X = 13,48 + 0,53 (15) = Rp 21.430.000,-
- Laki2 kulit hitam E(Y | X, D_1=0, D_2=1) = $(\alpha_o + \alpha_2) + \beta X$

  Y = 14,66 + 0,53 X = 14,66 + 0,53 (15) = Rp 22.610.000,-
- Wanita kulit hitam E(Y | X, D_1=1, D_2=1) = $(\alpha_o + \alpha_1 + \alpha_2) + \beta X$

  Y = 12,68 + 0,53 X = 12,68 + 0,53 (15) = Rp 20.630.000,-

Jadi kesimpulannya laki-laki tetap mempunyai ekspektasi lebih tinggi dari wanita dan kulit putih mempunyai ekspektasi lebih tinggi dari kulit hitam.

E. PERLUASAN MODEL REGRESI DALAM ANALISIS VARIABEL DUMMY

Dalam perluasannya tetap harus diperhatikan bahwa banyaknya dummy untuk setiap variabel kualitatif harus 1 (satu) lebih kecil dari jumlah kategorinya.

1. Regresi Lebih dari 1 Variabel Kuantitatif dan Ledih dari 2 Variabel Kualitatif Masing-Masing dengan 2 Kategori

Dari penelitian Shisko dan Rotsker tentang faktor-2 yang mempengaruhi upah pekerjaan sampingan dari 318 sampel buruh diperoleh regresi sbb:

$$Y = 37{,}07 + 0{,}403\ X_1 - 90{,}06\ D_1 + 75{,}51\ D_2 + 47{,}33\ D_3 + 113{,}64\ D_4 + 2{,}26\ X_2$$

t = (0,062) (24,47) (21,60) (23,42) (27,62) (0,94)

$R^2 = 0{,}95$

df = n - k - 1 = 318 - 7 = 311   t tabel 0,40 = 0,842

t tabel 0,05 = 1,960

t tabel 0,01 = 2,576

1 Variabel dependent: Y = upah pekerjaan sampingan (sen/jam) 6 variabel independent, terdiri dari:

a. 2 Variabel Kuantitatif X_1 = upah pekerjaan utama (sen/ jam)
   X_2 = umur (tahun)

b. 4 Variavel Kualitatif

D_1 = adalah ras (warna kulit)
= 0 jika putih
= 1 jika lainnya

D_2 = adalah urban (tinggal didaerah kota)
= 0 jika tinggal di non perkotaan
= 1 jika tinggal di perkotaan

D_3 = adalah Tingkat pendidikan
= 0 tidak lulus SLTA
= 1 lulus SLTA

D_4 = adalah daerah asal
= 0 jika bukan dari daerah Barat
= 1 dari daerah Barat

Semua variabel kualitatif signifikan pada tingkat kepercayaan 99 % artinya semua faktor tersebut mempengaruhi upah sampingan, sedangkan X_2 yakni umur buruh signifikan pada tingkat kepercayaan 80 %.

Interpretasinya:

Jika semua faktor lain konstant maka tingkat upah per jam diharapkan lebih tinggi sekitar 47 sen untuk buruh yang lulus SLTA dibandingkan yang berpendidikan lebih rendah.

Interpretasi nilai harapan (E) dari model regresi di atas dapat dijabarkan dalam beberapa regresi individual sbb:

a. Ekspektasi rata-2 tingkat upah pekerjaan sampingan/ jam dari buruh berkulit putih yang tidak tinggal di perkotaan, tidak berasal dari daerah Barat dan tidak lulus SLTA.
   $E(Y \mid X_1 X_2; D_1 = 0, D_2 = 0, D_3 = 0, D_4 = 0)$
   $Y = 37{,}07 + 0{,}403\ X_1 + 2{,}26\ X_2$
b. Ekspektasi rata2 tingkat upah pekerjaan sampingan/ jam dari buruh yang tidak berkulit putih, tinggal di daerah perkotaan, berasal dari Barat, dan lulus SLTA.
   $E(Y \mid X_1 X_2; D_1 = 1, D_2 = 1, D_3 = 1)$
   $Y = 183{,}49 + 0{,}403\ X_1 + 2{,}26\ X_2$
c. dan seterusnya.....

2. Variabel Dummy Dalam Analisis Musiman

Data musiman misalnya data semesteran, kuartalan, atau musim penghujan dan musim panas, awal tahun dan akhir tahun, musim panen, dan lain-lain. Sering kali komponen data musiman mengganggu analisis *time series*. Proses untuk membentuk komponen musiman dari data time series ini disebut *"seasional adjustment".*

Proses *seasional adjustment* dalam bidang ekonomi sangat penting antara lain dalam hubungan dengan: indeks harga konsumen, indeks harga wholesaler, indeks produksi industri, yang kebanyakan dinyatakan dalam musiman.

Ada beberapa metode untuk *seasional adjustment* salah satunya adalah dengan pendekatan variabel dummy.

Misal, regresi antara laba yang diterima (Y) dengan jumlah penjualan (X) untuk masing-2 kuartal (I, II, III, dan IV) pada perusahaan industri di USA th 1965 s/d 1970, lihat lampiran 5

D_1 = 1 jika kuartal 2

= 0 untuk lainnya (kuartal 1)

D_2 = 1 jika kuartal 3

= 0 untuk lainnya (kuartal 1)

D_3 = 1 jika kuartal 4

= 0 untuk lainnya (kuartal 1)

Secara arbitrary kuartal 1 ditetapkan sebagai kategori dasar

α_o = intersep kuartal 1 α_1 = intersep kuartal 2

α_2 = intersep kuartal 3 α_3 = intersep kuartal 4

Ekspektasi laba untuk jumlah penjualan kuartal 1

$E(Y \mid X, \; D_1=0, \; D_2=0, \; D_3=0) = \alpha_o + \beta X$

Ekspektasi laba untuk jumlah penjualan kuartal 2

$E(Y \mid X, \; D_1=1, \; D_2=0, \; D_3=0) = (\alpha_o + \alpha_1) + \beta X$

Ekspektasi laba untuk jumlah penjualan kuartal 3

$E(Y \mid X, \; D_1=0, \; D_2=1, \; D_3=0) = (\alpha_o + \alpha_2) + \beta X$

Ekspektasi laba untuk jumlah penjualan kuartal 4

$E(Y \mid X, \; D_1=0, \; D_2=0, \; D_3=1) = (\alpha_o + \alpha_3) + \beta X$

Contoh Soal:

Tahun Kuartal		Laba (Ribu $) Y	Penjualan (Ribu $) X	D_1	D_2	D_3
1965	1	10,503	114,862	0	0	0
	2	12,092	123,968	1	0	0
	3	10,834	121,454	0	1	0
	4	12,201	131,917	0	0	1
1966	1	12,245	129,911	0	0	0
	2	14,001	140,976	1	0	0
	3	12,213	137,828	0	1	0
	4	12,820	145,465	0	0	1
1967	1	11,349	136,989	0	0	0
	2	12,615	145,126	1	0	0
	3	11,014	141,536	0	1	0
	4	12,730	151,776	0	0	1
1968	1	12,539	148,862	0	0	0
	2	14,849	158,913	1	0	0
	3	13,203	155,727	0	1	0
	4	14,947	168,409	0	0	1
1969	1	14,151	162,781	0	0	0
	2	15,949	176,057	1	0	0
	3	14,024	172,419	0	1	0
	4	14,315	183,327	0	0	1
1970	1	12,381	170,415	0	0	0
	2	13,991	181,313	1	0	0
	3	12,174	176,712	0	1	0
	4	10,985	180,370	0	0	1

$Y = 6688{,}363 + 1322{,}892\ D_1 - 217{,}805\ D_2 + 183{,}856\ D_3 + 0{,}0382\ X$

Se = (638,474) (632,255) (654,292) (0,0115)

t = (2,072) (- 0,344) (0,281) (3,331)

p = (0,05212) (0,73426) (0,78175) (0,00351)

df = n - k - 1 = 24 - 5 = 19 $R^2 = 0{,}4256$

t tabel 0,10 = 1,729

t tabel 0,05 = 2,0930

t tabel 0,01 = 2,8609

Hanya koefisien arah penjualan (X) yang signifikan pada taraf kepercayaan 99 % sedangkan koefisien intersep dummynya tidak satupun yang signifikan pada taraf 95 %. Artinya pada taraf kepercayaan 95 % laba hanya dipengaruhi oleh penjualan dan tidak ada pola yang beraturan atau pola tertentu dalam faktor musiman yang mempengaruhi laba.

Jika digunakan $\alpha = 0,10$ maka koefisien intersep dummy kuartal 2 (D_1) akan signifikan (tepatnya pada taraf kepercayaan 94 % atau $\alpha = 0,052$) sedangkan koefisien dummy lainnya tetap non signifikan. Artinya:

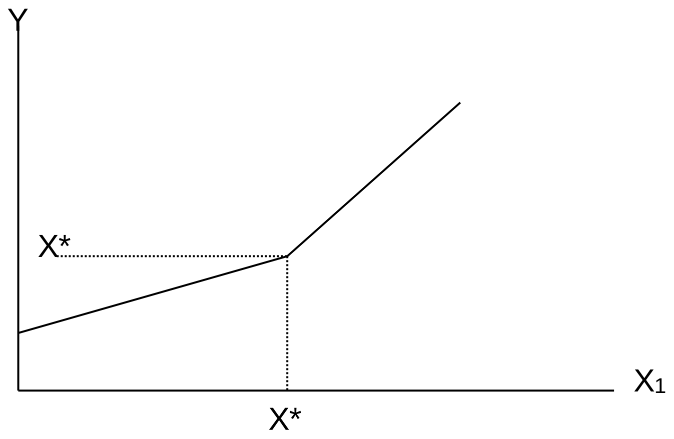
a. Pada taraf kepercayaan 94 % ada faktor yang bersifat musiman yang bekerja pada kuartal 2 yang ikut mempengaruhi laba.
b. Koefisien penjualan 0,0382 menyatakan bahwa dengan memperhitungkan pengaruh faktor musiman maka jika penjualan meningkat 1 satuan = 1000 dolar diharapkan rata2 laba (Y) akan meningkat sekitar 1000 * (0,0382) = 38,20 dolar atau sekitar 40 dollar.

Ekspektasi laba untuk jumlah penjualan kuartal 2 yang signifikan (jika penjualan $ 200.000,-) yakni:

$$
\begin{aligned}
E(Y \mid X,\ D_1=1,\ D_2=0,\ D_3=0) &= (\alpha_o + \alpha_1) + \beta X \\
Y &= 8011,255 + 0,0382\ X \\
&= 8011,255 + 0,0382\ (200) \\
&= \$\ 8.018.895,-
\end{aligned}
$$

3. Regresi Linear yang Patah

Kadang-kadang ditemui suatu fungsi regresi yang patah karena ada perobahan marginal yang besar dari Y akibat perobahan satu unit X pada tingkat tertentu (titik X*).



Katakan perobahan/ patahan tersebut terjadi pada titik (X*;Y*) Maka tentunya koefisien regresinya mulai titik tersebut akan berobah.

Regresi linear yang patah tersebut dapat terjadi karena ada faktor lain yang berpengaruh (pengaruh faktor eksternal).

Jika seandainya tidak terjadi patahan maka regresinya adalah:

$$Y = \alpha_o + \beta_1 X_1 + u$$

Karena ada patahan maka digunakan variabel dummy D dengan persamaan regresi adalah:

$$Y = \alpha_o + \beta_1 X_1 + \beta_2 (X_1 - X^*) D + u$$

dimana D = 1 jika X1 > X* setelah terjadi patahan

D = 0 jika X1 < X* sebelum terjadi patahan

X* = titik patah/ titik belok

Titik X* adalah konstanta yang ditentukan sebelumnya atau dengan bantuan diagram pencar.

Ekspektasi sebelum terjadi patahan

$E(Y \mid D = 0, X, X^*) = \alpha_o + \beta_1 X_1$

β_1 merupakan koefisien arah pada segmen I (sebelum terjadi pembelokan)

Ekspektasi setelah terjadi patahan

$$\begin{aligned} E(Y \mid D = 1, X, X^*) &= \alpha_o + \beta_1 X_1 + \beta_2 X_1 - \beta_2 X^* \\ &= \alpha_o - \beta_2 X^* + \beta_1 X_1 + \beta_2 X_1 \\ &= (\alpha_o - \beta_2 X^*) + (\beta_1 + \beta_2) X_1 \end{aligned}$$

$(\beta_1 + \beta_2)$ merupakan koefisien arah regresi setelah terjadi patahan.

Pengujian signifikansi apakah garis regresi tersebut patah adalah dengan menguji hipotesis

Ho: $\beta_2 = 0$

Ha: $\beta_2 \neq 0$

Pengujian dengan uji $t = \dfrac{b_2}{se\ (b_2)}$

Jika Ho ditolak berarti terjadi patahan

Jika Ho diterima berarti tidak terjadi patahan

Ekspektasi regresi patahan dilakukan apabila Ho ditolak.

Misal dari hasil penelitian pengaruh penjualan (X dalam unit) terhadap komisi penjualan (Y dalam juta rupiah) diperoleh data berikut (hasil analisis lihat lampiran 6).

N	Y	X_1	$(X_1 – X^*)$ D
1	20	2	0
2	32	3	0
3	40	4	0
4	47	5	0
5	60	6	0
6	110	7	1
7	220	8	2
8	290	9	3
9	370	10	4
10	410	11	5

$X^* = 6$

Persamaan regressinya adalah:

Y = 3,1412 + 9,0529 X1 + 65,2765 D1

R^2 = 0,9913

t = (2,341) (10,595)

p = (0,05176) (0,00001)

Pengujian signifikansi apakah garis regresi tersebut patah adalah dengan menguji hipotesis

Ho: β2 = 0

Ha: β2 ≠ 0

Uji t = $\frac{b_2}{se\ (b_2)}$ = 10,595 signifikan (Ho ditolak)

Karena Ho ditolak berarti terjadi patahan sehingga dapat dilakukan ekspektasi regresi patahan (b_2).

Dari persamaan tersebut tampak bahwa probabilitas X_1 = 0,05176 atau X_1 signifikan pada taraf kepercayaan 94 % dan probabilitas D_1 = 0,00001 atau D_1 signifikan pada taraf kepercayaan 99 %.

Ekspektasi komisi penjualan setelah terjadi patahan (D=1), jika rata-rata unit penjualan = 6,5

$$Y = \alpha_o + \beta_1 X_1 + \beta_2 (X_1 - X^*)D + u$$

$$
\begin{aligned}
E(Y \mid D = 1, X, X^*) &= \alpha_o + \beta_1 X_1 + \beta_2 (X_1 - X^*) \\
&= \alpha_o + \beta_1 X_1 + \beta_2 X_1 - \beta_2 X^* \\
&= \alpha_o - \beta_2 X^* + \beta_1 X_1 + \beta_2 X_1 \\
&= (\alpha_o - \beta_2 X^*) + (\beta_1 + \beta_2)\, X_1 \\
&= -\ 388{,}5178 + 74{,}3294\ X_1 \\
&= -\ 388{,}5178 + 74{,}3294\ (6{,}5) = \text{Rp } 94.623{,}30
\end{aligned}
$$

Jika unit penjualan = 15

$$
\begin{aligned}
E(Y \mid D = 1, X, X^*) &= -\ 388{,}5178 + 74{,}3294\ X_1 \\
&= -\ 388{,}5178 + 74{,}3294\ (15) = \text{Rp } 726.423{,}20
\end{aligned}
$$

BAB VI

MULTIKOLINIERITAS

A. PENDAHULUAN

Jika asumsi regresi linear dapat dipenuhi maka melalui metode OLS *(ordinary least square)* dapat dihasilkan koefisien regresi yang BLUE *(best linear unbiased estimator)* yakni: koefisien regresi yang linear, tidak bias dan memiliki varians yang minimum. Multikolinearitas merupakan salah satu pelanggaran asumsi model regresi linear klasik bahwa: "Seyogianya tidak terdapat multikolinearitas antar variabel independent". Multikolinearitas artinya terdapat hubungan linear yang sempurna (r = 1) diantara beberapa atau semua variabel independent dalam model yang dianalisis. Kondisi hubungan linear yang sempurna adalah korelasi antar variabel tersebut = 1. Hal ini terjadi jika:

$$\alpha 1\ X1 + \alpha_2\ X_2 + \dots + \alpha k\ Xk = 0$$

Multikolinearitas telah diartikan lebih luas termasuk kolinearitas yang tinggi walaupun tidak sempurna, karena itu:

$$\alpha 1\ X1 + \alpha 2\ X2 + \dots + \alpha k\ Xk + vi = 0$$

Jadi antara X satu dengan X lainnya tidak merupakan kombinasi linear yang pasti karena ditentukan juga oleh unsur kesalahan yang variabel atau stokastik vi.

Contoh $X_2 = \alpha X_1$ dimana ($\alpha \neq 0$)

X_1	X_2	X_2^*
10	50	52
15	75	75
18	90	97
24	120	129
30	150	152

$X_2 = 5\ X_1$ terjadi kolinearitas sempurna karena $r_{12} = 1$ X_2^* diperoleh dengan menambah bilang- an 2, 0, 7, 9, 2 ke X_2 sehingga X_1 dan X_2^* berkolinearitas tidak sempurna $r_{12}^* = 0{,}9959$.

$$r_{12} = \frac{\sum X_1 X_2 - (\sum X_1)(\sum X_2)/ n}{\sqrt{[\sum X_1^2 - (\sum X_1)^2/n]\ [\sum X_2^2 - (\sum X_2)^2/n]}}$$

$$= \frac{10625 - (97)(485) / 5}{\sqrt{[2125 - (97)^2/5]\ [53125 - (485)^2/5)}}$$

$$= \frac{1216}{\sqrt{[243{,}2][6080]}} = \frac{1216}{1216} = 1$$

Catatan:

1. Semua variabel independent X yang termasuk dalam model mempunyai pengaruh terpisah atau independent atas variabel tak bebas Y; atau X diasumsikan tetap atau nonstokastik.

   Jadi multikolinearitas merupakan fenomena sampel, jangan sampai sampel kita menyesatkan analisis. Misal:

Y Konsumsi = $b_o + b_1\ X_1$ Pendapatan + $b_2\ X_2$ Kekayaan

   Pendapatan dan kekayaan mungkin berkorelasi sempurna atau sangat berkorelasi dimana sampel (orang) yang lebih kaya cenderung mempunyai pendapatan lebih tinggi (atau sebaliknya) sehingga sulit melihat pengaruh terpisah dari pendapatan dan kekayaan atas belanja konsumsi.

2. Multikolinearitas hanya membahas hubungan linear, tidak termasuk non linear, karena itu regresi

$$Y = b_o + b_1X + b_2 X_2$$

atau

$$Y = b_o + b_1X_1 + b_2 X_2$$

dimana $X_2 = X_2$ yang secara fungsional berhubungan dengan $X_1 = X$ tetapi hubungannya nonlinear tidak menyalahi asumsi multikolinearitas.

Contoh fungsi $TC = aQ_2 + bQ + c$ tidak terjadi kolinearitas

Alasan tidak terdapat multikolinearitas:

1. Multikolinearitas sempurna menyebabkan koefisien regresi tak dapat ditentukan dan kesalahan standarnya tak terhingga

   Misal:

$$Y = b_o + b_1 X_1 + b_2 X_2 \quad \text{dimana } X_2 = \alpha X_1$$

$$b_1 = \frac{(\sum Y_i X_{1i})(\sum X_{2i}^2) - (\sum Y_i X_{2i})(\sum X_{1i} X_{2i})}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i} X_{2i})^2}$$

$$b_1 = \frac{(\sum Y_i X_{1i})(\alpha^2 \sum X_{1i}^2) - (\alpha \sum Y_i X_{1i})(\alpha \sum X_{1i} X_{1i})}{(\sum X_{1i}^2)(\alpha^2 \sum X_{1i}^2) - \alpha^2 (\sum X_{1i} X_{1i})^2}$$

$$b_1 = \frac{(\sum Y_i X_{1i})(\alpha^2 \sum X_{1i}^2) - (\sum Y_i X_{1i})(\alpha^2 \sum X_{1i}^2)}{\alpha^2 (\sum X_{1i}^2)^2 - \alpha^2 (\sum X_{1i}^2)^2} = \frac{0}{0}$$

$$V(b_1) = \frac{\sum X_{2i}^2}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i} X_{2i})^2} \sigma^2$$

$$V(b_1) = \frac{\alpha^2 \sum X_{1i}^2}{(\sum X_{1i}^2)(\alpha^2 \sum X_{1i}^2) - (\alpha \sum X_{1i} X_{1i})^2} \sigma^2$$

$$V(b_1) = \frac{\alpha^2 \sum X_{1i}^2}{\alpha^2 (\sum X_{1i}^2)^2 - \alpha^2 (\sum X_{1i}^2)^2} \sigma^2$$

$$V(b_1) = \frac{1}{(\sum X_{1i}^2) - (\sum X_{1i}^2)} \sigma^2$$

$$V(b_1) = \frac{\sigma^2}{0} = \infty$$ dan se (b_1) tak terdifinisi/ tak terhingga

Identik dengan b_1 maka V(b_2) dan se (b_2) juga tak terdifinisikan

2. Pada multikolinearitas tak sempurna walaupun koefisien regresinya dapat ditentukan namun memiliki kesalahan standar yakni se(bi) yang besar. Ini berarti koefisien regresi tidak dapat ditaksir dengan ketepatan yang tinggi.

Akibat2 multikolinearitas

1. Dalam kasus multikolinearitas sempurna koefisien regresi tak dapat ditentukan dan varians akan tak terhingga.
2. Multikolinearitas tidak sempurna tetapi cukup tinggi dapat menyebabkan:
   a. Estimator OLS dapat ditentukan namun se (b_i) cukup besar. Tingkat kolinearitas semakin tinggi standar errornya makin besar.
   b. se yang besar menyebabkan convident interval makin melebar sehingga peluang untuk menerima hipotesis yang salah (tipe II error) semakin besar.

c. Estimator OLS (koefisien regresi) dan se (b_i) sensitif sehingga mudah berobah dengan sedikit perobahan data.

Y	X_1	X_2	Y	X_1	X_2
1	10	52	1	10	52
2	15	75	2	15	75
3	18	97	3	18	97
4	24	129	4	24	152
5	30	152	5	30	129
	(a)			(b)	

Hasil perhitungan tabel (a):

$Y = -0{,}9607 - 0{,}0060\ X_1 + 0{,}0404\ X_2$

p = (0,9578) (0,1749) adj.R^2 = 0,9919

t = (- 0,06) (2,066) r_{12} = 0,9959

df = n-k-1 = 2 t tabel 0,05 = 4,3032

t tabel 0,01 = 9,9248

Hasil perhitungan tabel (b):

$Y = -0{,}9701 + 0{,}1787\ X_1 + 0{,}0050\ X_2$

p = (0,0267) (0,4782) adj.R^2 = 0,9814

t = (6,000) (0,865) r_{12} = 0,8859

Tampak bahwa koefisien regresi maupun se (b_i) berobah. Koefisien b_1 yang tadinya non signifikan menjadi signifikan pada taraf yang sama (0,05), r_{12} juga berobah dari 0,9959 menjadi 0,8859

d. Multikolinearitas menyebabkan R^2 yang tinggi namun tidak ada satupun koefisien regresi yang signifikan.

Dari hasil analisis tabel (a) di atas tampak bahwa:

1) walaupun koefisien korelasi sangat besar (99 %) namun tak satupun koefisien regresi yang signifikan.
2) Selain nonsignifikan, variabel X_1 juga bernilai negatif
3) Walaupun X_1 dan X_2 nonsignifikan namun uji F-nya sangat signifikan (karena R^2 yang tinggi) atau kita tidak menolak hipotesis secara simultan. Ini berarti tidak mungkin dapat mengisolasi pengaruh individu dari variabel X_1 dan X_2 karena ada gejala kolinearitas yang ekstrim.

Dengan adanya multikolinearitas kita tidak bisa memisahkan pengaruh X1 dan X_2 secara individual (atau X_1 dan X_2 tidak independent).

Cara Mendeteksi Multikolinearitas

1. Nilai R^2 cukup tinggi.
2. Hasil uji F (anava) sangat signifikan tetapi tidak satupun koefisien regresi yang signifikan dari hasil uji t parsial.
3. Gunakan uji Fj terhadap nilai R^2 dari setiap pasangan variabel X dengan rumus:

$$Fj = \frac{R^2 x_1, x_2 ...,xk^{(k-2)}}{(1 - R^2 x_1, x_2 ...,xk^{(n-k-1)}}$$

F $_{tabel}$ {α = 0,05; db = (k-2)(n-k-1)}

n = jumlah sampel

k = jumlah variabel X

R^2 x_1, x_2 ...,xk = koefisien determinasi untuk jumlah variabel X sebanyak k

Jika Fj > F tabel atau signifikan berarti Xj tertentu berkorelasi dengan X lainnya sehingga perlu dipertimbangkan untuk dikeluarkan dari model.

Cara Menanggulangi Multikolinearitas

1. Apriori terhadap informasi.

   Artinya hubungan antar variabel independent dipertimbangkan berdasarkan teori dan kenyataan hubungan yang ada.

   Contoh pengaruh pendapatan (X_1) dan kekayaan (X_2) terhadap pola konsumsi (Y) dengan model regresi:

   $$\boxed{Y = \beta o + \beta_1 X_1 + \beta_2 X_2}$$

   Secara apriori pendapatan dan kekayaan mempunyai hubungan yang erat misal $\beta_2 = 0{,}01\ \beta_1$

   Untuk menghindari multikolinearitas maka hubungan tersebut dapat disubtitusikan ke dalam model regresi:

   $Y = \beta o + \beta_1 X_1 + \beta_2 X_2$

   $Y = \beta o + \beta_1 X_1 + 0{,}01\beta_1 X_2$

   $Y = \beta o + \beta_1 (X_1 + 0{,}01\ X_2)$

   $Y = \beta o + \beta_1 Xi$ dimana $Xi = (X_1 + 0{,}01\ X_2)$

   Setelah diperoleh β_1 dapat dihitung $\beta_2 = 0{,}01\beta_1$

2. Pooling data (penggabungan data)

   Artinya menggabungkan data cross sectional dan time series mengestimasi elastisitas harga β_1 dan elastisitas pendapatan konsumen β_2 terhadap penjulan mobil per tahun (Y) dengan model regresi:

   $$\boxed{\ln Y_i = \beta_o + \beta_1 \ln P_i + \beta_2 \ln I_i + u_i}$$

   dimana Yi = jumlah mobil yang terjual

   Pi = rata-rata harga mobil

   Ii = pendapatan konsumen

Secara apriori harga dan pendapatan mempunyai kolineariti yang tinggi karena itu seyogianya tidak dianalisis secara langsung. Jalan keluarnya dengan menggunakan data cross sectional untuk estimasi yang realistik bagi elastisitas pendapatan β_2, sebab dari data tersebut untuk suatu titik waktu harga tidak banyak bervariasi. Jadi elastisitas pendapatan β_2 diestimasi secara cross sectional dengan rumus:

$\boxed{\ln Y = bo + \beta_2 \ln I}$ (Y dan I data cross sectional)

β_2 digunakan untuk estimasi regresi time series dengan rumus:

$\boxed{Y^* t = \beta o + \beta_1 \ln Pt + ut}$ (Pt adalah data time series)

dimana $Y^* = \ln Y - \beta_2 \ln I$ (Y dan I data time series) yang digunakan untuk mengestimasi elastisitas harga β_1.

3. Mengeleminir variabel yang menyebabkan bias spesifik

   Cara yang paling mudah untuk mengatasi multikolineariti adalah membuang salah satu variabel yang berkorelasi dengan variabel eksplanatori lainnya.

4. Mentransformasikan data variabel

   Biasanya untuk data time series, misal untuk pola konsumsi, pendapatan dan kekayaan dimana terjadi kolinearitan antara pendapatan dan kekayaan karena cenderung memiliki ketergantungan dalam arah yang sama.

   Untuk mengurangi ketergantungan tersebut adalah dengan cara mengurangi data antar dua waktu berurutan atau meregresikan selisih dua titik waktu yang berurutan pada data asli sbb:

Jika model regresi:

$$Y_t = \beta_o + \beta_1 X_1t + \beta_2 X_{2t} + u_t$$

yang berlaku untuk waktu ke t harus juga berlaku untuk waktu ke (t-1) dengan model regresi:

$$Y_{t-1} = \beta o + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + u_{t-1}$$

Apabila kedua model tersebut dikurangkan akan diperoleh model regresi estimator sbb:

$$Y_t - Y_{t-1} = \beta_1 (X_{1t} - X_{1t-1}) + \beta_2 (X_{2t} - X_{2t-1}) + v_t$$

dimana $v_t = u_t - u_{t-1}$