BAB I
PENDAHULUAN

1.1 Latar belakang Masalah

Dalam permasalahan pengelolaan dan menejemen seringkali dijumpai kegiatan peramalan, pendugaan, perkiraan, dan lainnya. Salah satu metode yang dapat digunakan untuk menyelesaikan masalah tersebut adalah dengan menggunakan metode statistik. Metode statistika yang digunakan sangat bergantung pada struktur data atau banyaknya variabel yang akan diamati. Salah satu metode yang dipakai untuk banyaknya variabel lebih dari satu adalah analisis regresi.

Analisis regresi adalah suatu metodologi statistika untuk memprediksi nilai dari satu atau lebih variabel respon (variabel dependen) dari koleksi nilai variabel prediktor (variabel independen). Analisis ini juga dapat digunakan untuk memprediksi atau meramal pengaruh dari variabel prediktor (variabel independen) pada respon. Dalam analisis regresi pun dipelajari bagaimana variabel-variabel tersebut berhubungan dan dinyatakan dalam sebuah persamaan matematik.

Sayangnya, istilah regresi, diambil dari judul peper pertama dari F. Galton yang tidak menunjukkan atau menggambarkan pentingnya atau luasnya cakupan aplikasi dari metodologi ini. Dalam analisis regresi, ada dua jenis variabel yaitu variabel bebas atau variabel prediktor (dinotasikan dengan X) dan variabel tak bebas atau variabel respon (dinotasikan dengan Y). Untuk melihat hubungan antara variabel respon dan sejumlah variabel prediktor secara simultan dapat digunakan analisis regresi linier dengan variabel respon diukur sekurang-kurangnya dalam skala interval dam mempunyai distribusi normal.

Pada analisis regresi linier terbagi menjadi dua, yaitu analisis regresi linier sederhana dan analisis regresi linier berganda. Yang membedakan keduanya adalah hanya terletak pada variabel bebas atau variabel prediktornya, untuk analisis regresi linier sederhana variabel bebasnya hanya satu sedangkan untuk analisis regresi linier berganda banyaknya variabel bebas adalah lebih dari satu.

Tetapi bagaimana dengan banyaknya variabel tak bebas atau variabel respon yang lebih dari satu. Oleh karena itulah, kami mencoba untuk mempelajari lebih jauh tentang model regresi linier multivariat yang terdapat pada bab 7.

Pada makalah ini, kami akan mencoba mendiskusikan model regresi linier berganda untuk memprediksi respon tunggal. Model ini kemudian diperumum untuk membahas prediksi dari beberapa variabel dependen (variabel respon). Perlakuan penyingkatan kita menyoroti atau membahas asumsi-asumsi regresi dan konsekuensinya, formula alternatif dari model regresi, dan aplikasi umum dari teknik regresi pada kasus yang tampaknya berbeda.

1.2 Perumusan Masalah

Berdasarkan pemaparan diatas maka permasalahan yang akan dibahas dalam penulisan ini adalah bagaimana penjelasan secara terperinci mengenai model regresi linier multivariat pada bab7 tersebut.

1.3 Batasan Masalah

Dalam penulisan ini kami membatasi masalah sebagai berikut ;

Pemaparan mengenai model regresi linier multivariat hanya akan dibahas sesuai dengan yang telah kami sampaikan pada persentasi yang telah kami lakukan.

1.4 Tujuan Penulisan

Berdasarkan rumusan masalah diatas, maka tujuan penulisan ini adalah untuk mengetahui dan mempelajari lebih rinci mengenai model regresi linier multivariat.

BAB II
MODEL REGRESI LINIER MULTIVARIAT

Nama : Adzimattinur Luthfia

Nim : 055372

2.2 MODEL REGRESI LINEAR KLASIK

Model regresi linear dengan respon tunggal mempunyai bentuk

$$Y = \beta_0 + \beta_1 Z_1 + ... + \beta_r Z_r + \varepsilon$$

Dengan Y : variabel respon

$Z_1,...,Z_r$: variabel prediktor

$\beta_0, \beta_1,..., \beta_r Z_r$: parameter yang tidak diketahui

ε : nilai error (galat)

dengan n observasi independen pada Y dan nilai yang diasosiasi dari Z_i maka model lengkap regresi linier berbentuk

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 Z_{11} + \beta_2 Z_{12} + ... + \beta_r Z_{1r} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 Z_{21} + \beta_2 Z_{22} + ... + \beta_r Z_{2r} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 Z_{n1} + \beta_2 Z_{n2} + ... + \beta_r Z_{nr} + \varepsilon_n \end{aligned} \qquad (7\text{-}1)$$

Dimana errornya diasumsikan memiliki sifat :

$$\begin{aligned} &1.E(\varepsilon_j) = 0 \\ &2.Var(\varepsilon_j) = \sigma^2 \qquad \text{(konstan)} \\ &3.Cov(\varepsilon_j, \varepsilon_k) = 0, j \neq k \end{aligned} \qquad (7\text{-}2)$$

persamaan (7-1) dalam bentuk matriks adalah

$$\begin{bmatrix} Y \\ Y \\ \vdots \\ Y \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_r \end{bmatrix} \qquad (7\text{-}3)$$

atau

$$\underset{(n+1)}{Y} = \underset{(n\times(r+1))}{Z} \underset{((r+1)\times 1)}{\beta} + \underset{(n\times 1)}{\varepsilon}$$

dengan sifatnya : $1. E(\varepsilon) = 0$

$2. Cov(\varepsilon) = \sigma^2 \mathrm{I}$

contoh :

Tentukan bentuk matriks jika model regresi linear sesuai dengan situasi pada contoh 6.6

jawab :

Kita buat variabel boneka untuk mengatasi 3 rata-rata populasi,

$\mu_1 = \mu + \tau_1, \mu_2 = \mu + \tau_2, dan \mu_3 = \mu + \tau_3,$ Kita tentukan

$$z_1 = \begin{cases} 1; & \text{Jika observasi berasal dari populasi 1} \\ 0; & \text{Jika observasi berasal dari selain populasi 1} \end{cases}$$

$$z_2 = \begin{cases} 1; & \text{Jika observasi berasal dari populasi 2} \\ 0; & \text{Jika observasi berasal dari selain populasi 2} \end{cases}$$

$$z_3 = \begin{cases} 1; & \text{Jika observasi berasal dari populasi 3} \\ 0; & \text{Jika observasi berasal dari selain populasi 3} \end{cases}$$

Dan $\beta_0 = \mu, \beta_1 = \tau_1, \beta_2 = \tau_2, \beta_3 = \tau_3$ lalu $Y_j = \beta_0 + \beta_1 Z_{j1} + \beta_2 Z_{j2} + \beta_3 Z_{j3} + \varepsilon_j$

j = 1, 2, …, 8

Ketika kita menyusun nilai-nilai observasi dari 3 populasi dalam barisan, kita dapatkan vektor respon observasi dan matriks desain

$$\underset{(8x1)}{Y} = \begin{bmatrix} 9 \\ 6 \\ 9 \\ 0 \\ 2 \\ 3 \\ 1 \\ 2 \end{bmatrix}, \underset{(8x4)}{Z} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

2.3 PENAKSIR KUADRAT TERKECIL

Misal b adalah nilai taksiran untuk β. perhatikan perbedaan $y_j - b_0 - b_1 z_{j1} - ... - b_r z_{jr}$ antara y_j dan nilai $b_0 + b_1 z_{j1} + ... + b_r z_{jr}$ itu akan diharapkan jika b adalah vektor parameter sebenarnya. Selisih $y_j - b_0 - b_1 z_{j1} - ... - b_r z_{jr}$

tidak akan sama dengan nol karena nilai harapan respon berfluktuasi.

Metoda dari kuadrat terkecil memilih b untuk meminimumkan jumlah kuadrat S(b) = $\sum_{j=1}^{n}(y_j - b_0 - b_1 z_{j1} - ... - b_r z_{jr})^2 = (y - Zb)'(y - Zb)$

Koefisien b dipilih berdasarkan kriteria kuadrat terkecil, dan b disebut penaksir kuadrat terkecil dari $\boldsymbol{\beta}$ (b sering dinotasikan $\hat{\boldsymbol{\beta}}$).

Simpangan $\hat{\boldsymbol{\varepsilon}}_j = y_j - \hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\beta}}_1 z_{j1}$ disebut residu.

<u>Hasil 7.1</u>

Misal Z sebanyak $r+1 \leq n$. Penaksir kuadrat terkecil dari $\boldsymbol{\beta}$ adalah $\hat{\boldsymbol{\beta}} = (Z^{'}Z)^{-1}Z^{'}y$. Misal $\hat{y} = Z\hat{\boldsymbol{\beta}} = Hy$ diartikan nilai tertentu dari y, dengan $H = (Z^{'}Z)^{-1}Z^{'}$ disebut matriks Hat . Residunya :

$$\hat{\boldsymbol{\varepsilon}} = y - \hat{y} = \left[\mathrm{I} - Z(Z^{'}Z)^{-1}Z^{'}y\right] = (\mathrm{I} - H)y$$

Memenuhi $Z^{'}\hat{\varepsilon} = 0$ dan $y^{'}\hat{\varepsilon} = 0$. Juga

jumlah kuadrat residu kuadrat $= \sum_{j=1}^{n}(y_j - \hat{\beta}_0 - \hat{\beta}_1 Z_{j1} - ... - \hat{\beta}_r z_{jr})^2 = \hat{\varepsilon}^{'}\hat{\varepsilon}$

$$= y'\left[\mathrm{I} - Z(Z'Z)^{-1}Z'\right]y = y'y - y'Z\hat{\beta}$$

Hasil 7.1 menunjukkan bahwa penaksir kuadrat terkecil $\hat{\beta}$ dan residu dapat diperoleh dari desain matriks Z dan respon y dengan operasi matriks sederhana.

Contoh :

hitunglah $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varepsilon}}$ dan jumlah residu kuadrat untuk model $Y_j = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_{j1} + \boldsymbol{\varepsilon}_j$ yang cocok dengan data

z_1	0	1	2	3	4
y	1	4	3	8	9

Jawab :

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad Z' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} \quad Z'Z = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$$

$$(Z'Z)^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix} \qquad Z'y = \begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

Sehingga

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (Z'Z)^{-1} Z'y = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 25 \\ 70 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Dan persamaan yang tepat adalah

$$\hat{y} = 1 + 2z$$

Vektor nilai taksiran adalah

$$\hat{y} = Z\hat{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

maka

$$\hat{\varepsilon} = y - \hat{y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

jumlah kuadrat terkecilnya adalah

$$\hat{\varepsilon}'\hat{\varepsilon} = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} = 0^2 + 1^2 + (-2)^2 + 1^2 + 0^2 = 6$$

<u>JUMLAH DEKOMPOSISI KUADRAT</u>

$y'\hat{\varepsilon} = 0$, jadi jumlah respon total kuadrat $y'y = \sum_{j=1}^{n} y_j^2$ memenuhi

$$y'y = (\hat{y} + y - \hat{y})'(\hat{y} + y - \hat{y}) = (\hat{y} + \hat{\varepsilon})(\hat{y} + \hat{\varepsilon}) = \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon} \qquad (7\text{-}4)$$

karena kolom pertama dari Z adalah 1, kondisi $Z'\hat{\varepsilon} = 0$ memenuhi persamaan

$0 = 1'\hat{\varepsilon} = \sum_{j=1}^{n} \hat{\varepsilon}_j = \sum_{j=1}^{n} y_j - \sum_{j=1}^{n} \hat{y}_j$ atau $\bar{y} = \bar{\hat{y}}$

jika kedua sisi dari persegi (7-4) dikurangi $n\bar{y}^2 = n\bar{\hat{y}}^2$ diperoleh dekomposisi dasar dari jumlah rata-rata kuadrat $y'y - n\bar{y}^2 = \hat{y}'\hat{y} - n(\bar{\hat{y}})^2 + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$

atau $\sum_{j=1}^{n} (y_j - \bar{y})^2 = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2 + \sum_{j=1}^{n} \hat{\varepsilon}_j^2$

jumlah kuadrat diatas menyarankan kualitas dari model yang tepat dapat diukur dengan menghitung koefisien determinasi yaitu

$$R^2 = 1 - \frac{\sum_{j=1}^{n} \hat{\varepsilon}_j^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2} = \frac{\sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2}$$

GEOMETRI DARI KUADRAT TERKECIL

berdasarkan model regresi klasik

$$E(Y) = Z\beta = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} z_{11} \\ z_{21} \\ \vdots \\ z_{n1} \end{bmatrix} + \dots + \beta_r \begin{bmatrix} z_{1r} \\ z_{2r} \\ \vdots \\ z_{nr} \end{bmatrix}$$

E(Y) adalah sebuah kombinasi linear dari kolom Z. Seperti $\boldsymbol{\beta}, Z\boldsymbol{\beta}$ membentuk model bidang dari semua kombinasi linear. Biasanya vektor observasi y tidak akan berbaring di dalam model bidang karena nilai error ε , maka dari itu y bukanlah suatu kombinasi linear dari kolom Z.

Ketika observasi terjadi, solusi kuadrat terkecil diperoleh dari vektor simpangan

y - Zb = (vektor observasi)-(vektor pada model bidang)

panjang kudrat adalah S(b) kudrat. Seperti yang diilustrasikan pada gambar 7-1 (hal 293), nilai S(b) sekecil mungkin ketika b dipilih maka Zb adalah titik pada model bidang yang paling dekat ke y. titik terdekat ke y terjadi di ujung dari proyeksi tegak y pada bidang. Maka dari itu y, untuk pemilihan $b = \hat{\beta}, \hat{y} = Z\hat{\beta}$ yang merupakan proyeksi dari y pada bidang terdiri dari semua kombinasi linear dari kolom Z. vektor residu $\hat{\varepsilon} = y - \hat{y}$ adalah tegak terhadap bidang. Geometri ini terbentuk walaupun Z bukan rank penuh.

Ketika Z memiliki rank penuh, operasi proyeksi ditunjukkan secara analitik seperti perkalian oleh matrik $Z(Z'Z)^{-1}Z'$. untuk melihatnya, kita gunakan spektrum dekomposisi (2-16) untuk menulis

$Z'Z = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + ... + \lambda_{r+1} e_{r+1} e_{r+1}'$ dimana $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{r+1} > 0$ adalah nilai eigen dari Z'Z dan $e_1, e_2, ..., e_{r+1}$ adalah vektor eigen yang berkorespondensi.

Jika Z memiliki rank penuh maka $(Z'Z)^{-1} = \frac{1}{\lambda_1} e_1 e_1' + \frac{1}{\lambda_2} e_2 e_2' + ... + \frac{1}{\lambda_{r+1}} e_{r+1} e_{r+1}'$

Perhatikan $q_i = \lambda_i^{-\frac{1}{2}} Ze_i$ yang merupakan sebuah kombinasi linier dari kolom Z. Maka $q_i' q_k = \lambda_i^{-\frac{1}{2}} \lambda_{ki}^{-\frac{1}{2}} e_i' Z'Ze_k = 0$ jika $i \neq k$ atau 1 jika $i = k$. Maka dari itu, r+1 vektor secara berbalasan tegak dan memiliki unit panjang. Kombinasi linier dari kolom Z. Dan lagi $Z(Z'Z)^{-1}Z' = \sum_{i=1}^{r+1} \lambda_i^{-1} Ze_i e_1' Z' = \sum_{i=1}^{r+1} q_i q_i'$

Berdasarkan hasil 2A.2 dan definisi 2A.12 proyeksi dari y pada kombinasi linier dari $\{q_1, q_2, ..., q_{r+1}\}$ adalah $\sum_{i=1}^{r+1} (q_i' y) q_i = \left(\sum_{i=1}^{r+1} q_i q_i' \right) y = Z(Z'Z)^{-1} Z' y = Z\hat{\beta}$ Jadi perkalian dengan $Z(Z'Z)^{-1}Z'$ merencanakan sebuah vektor pada ruang yang dibentuk oleh kolom Z.

SIFAT SAMPLING DARI PENAKSIR KUADRAT TERKECIL KLASIK

Hasil 7.2

Berdasarkan model regresi linier umum pada (7-3), persamaan kudrat terkecil $\hat{\beta} = Z(Z'Z)^{-1} Z' y$ mempunyai $E(\hat{\beta}) = \beta$ dan $\text{cov}(\hat{\beta}) = \sigma^2 (Z'Z)^{-1}$.

Residu $\hat{\varepsilon}$ memiliki sifat $E(\hat{\varepsilon}) = 0$ dan $\text{cov}(\hat{\varepsilon}) = \sigma^2 (I - H)$ juga $E(\hat{\varepsilon}' \hat{\varepsilon}) = (n - r - 1)\sigma^2$,jadi membatasi

$$s^2 = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{n - (r+1)} = \frac{Y\left[I - Z(Z'Z)^{-1} Z'\right]Y}{n - r - 1} = \frac{Y'[I - H]Y}{n - r - 1}$$

Kita punyai $E(s^2) = \sigma^2$ dan lagi $\hat{\beta}$ dan $\hat{\varepsilon}$ tidak berkorelasi.

Persamaan kuadrat terkecil $\hat{\beta}$ memiliki varians minimum yang pertama kali ditetapkan oleh Gauss. Hasil ini mengenai penaksir "bagus" dari fungsi parametrik linear dari bentuk $c'\beta = c_0\beta_0 + c_1\beta_1 + ... + c_r\beta_r$ untuk setiap c.

Hasil 7.3 (Teorema Kuadrat Terkecil Gauss)

Misal $Y = Z\beta + \varepsilon$ dengan $E(\varepsilon) = 0$ dan $\text{cov}(\varepsilon) = \sigma^2 I$ dan Z memiliki rank penuh r+1. untuk setiap c, penaksir $c'\hat{\beta} = c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + ... + c_r\hat{\beta}_r$ dari $c'\beta$ memiliki varians sekecil mungkin diantara semua penaksir linear dari bentuk $a'Y = a_1Y_1 + a_2Y_2 + ... + a_nY_n$ yang tidak bias untuk $c'\beta$.

Hasil yang kuat ini menyatakan bahwa subtitusi dari $\hat{\beta}$ untuk β, menuju ke penaksir terbagus dari $c'\beta$ untuk setiap c.

2.4 KESIMPULAN TENTANG MODEL REGRESI

2.4.1 Kesimpulan mengenai parameter regresi.

Sebelum kita dapat menetapkan arti dari variabel utama dalam fungsi regresi $E(Y) = \beta_0 + \beta_1 z_1 + ... + \beta_r z_r$ kita harus menentukan distribusi samping dari $\hat{\beta}$ dan jumlah residu kuadrat $\hat{\varepsilon}'\hat{\varepsilon}$. Untuk itu kita asumsikan ε memiliki distribusi normal.

Hasil 7.4

Misal $Y = Z\beta + \varepsilon$ dimana Z memiliki rank penuh r+1 dan ε berdistribusi normal $N_n(0, \sigma^2 I)$. Penaksir maximum Likelihood dari β adalah sama dengan penaksir kuadrat terkecil $\hat{\beta}$. Dan lagi, $\hat{\beta} = (Z'Z)^{-1}Z'Y$ berdistribusi $N_{r+1}(\beta, \sigma^2(Z'Z)^{-1})$ dan didistribusikan secara independen dari residu $\hat{\varepsilon} = Y - Z\hat{\beta}$. Selanjutnya $n\sigma^2 = \hat{\varepsilon}'\hat{\varepsilon}$ berdistribusi $\sigma^2\chi_{n-r-1}$ dengan $\hat{\sigma}^2$ adalah penaksir maximum Likelihood dari σ^2

Ellipsoid kepercayaan untuk β sangat mudah disusun. Hal ini dapat dinyatakan dalam batas dari matriks penaksir covarian $s^2(Z'Z)^{-1}$ dengan $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-r-1)$

Hasil 7.5

Misal $Y = Z\beta + \varepsilon$ dimana Z memiliki rank penuh r+1 dan ε berdistribusi normal $N_n(0, \sigma^2 I)$. Daerah kepercayaan $100(1-\alpha)\%$ untuk β adalah $(\beta - \hat{\beta})' Z' Z(\beta - \hat{\beta}) \le (r+1)s^2 F_{r+1, n-r-1}(\alpha)$ juga, interval kepercayaan $100(1-\alpha)\%$ untuk β_i adalah $\hat{\beta}_i \pm \sqrt{Var(\hat{\beta}_i)}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}$, i= 0, 1,…, r

Dengan $\hat{V}ar(\hat{\beta}_i)$ adalah elemen diagonal dari $s^2(Z'Z)^{-1}$ yang berkorespondensi ke $\hat{\beta}_i$.

Ellipsoid kepercayaan adalah pusat pada penaksir maximum Likelihood $\hat{\beta}$ dan orientasinya dan ukuran ditentukan oleh nilai eigen dan vektor eigen dari $Z'Z$. Jika nilai eigen mendekati nol, ellips kepercayaan akan sangat panjang dalam arah dari vektor eigen yang berkorespondensi.

Para praktisi sering mengabaikan sifat kepercayaan dari taksiran interval pada hasil 7-5. mereka mengganti $(r+1)F_{r+1,n-r-1}$ dengan nilai t, $t_{n-r-1}(\alpha/2)$ dan menggunakan interval $\hat{\beta}_i \pm t_{n-r-1}(\alpha/2)\sqrt{\hat{V}ar(\hat{\beta}_i)}$ ketika mencari variabel prediktor utama.

Contoh:

Berdasarkan data pada tabel 7.1, model yang tepat adalah $Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \varepsilon_j$

Pada data ini digunakan metoda kudrat terkecil. Hasil perhitungan komputer adalah

$$(Z'Z)^{-1} = \begin{bmatrix} 1.9961 & & \\ -0.0896 & 0.0512 & \\ -0.0115 & -0.0172 & 0.0067 \end{bmatrix} \text{ dan } \hat{\beta} = (Z'Z)^{-1}Z'y = \begin{bmatrix} 11870.2 \\ 2634.4 \\ 45.2 \end{bmatrix}$$

Jadi persamaan yang tepat adalah $Y_j = 11870.2 + 2634.4 z_1 + 45.2 z_2$ dengan s = 3473.

Jika residu $\hat{\varepsilon}$ melewati pemeriksaan diagnosa yang dijelaskan pada seksi 7.6, persamaan yang tepat dapat digunakan untuk memprediksi harga jual dari rumah-rumah di sekitar berdasarkan ukuran dan nilai yang ditetapkan.

Kita misalkan 95% interval konfidensi untuk β_2 adalah

$\hat{\beta}_2 \pm t_{17}(0.025)\sqrt{\hat{V}ar(\hat{\beta}_2)} = 45.2 \pm 2.110(285)$ atau (-556647)

Karena interval konfidensi memuat $\beta_2 = 0$ variabel z_2 dapat dihilangkan dari model regresi dan analisis diulang dengan variabel prediktor tunggal z_1.

Dibanding ukuran tempat tinggal, kiranya nilai yang ditetapkan menambah sedikit pengaruh terhadap prediksi dari harga jual.

Nama : Realita Raymunda

Nim : 055800

2.4.2 Test rasio likelihood untuk parameter Regresi

Salah satu bagian dari analisis regresi terkait dengan menaksir pengaruh variabel prediktor pada variabel respon. Hipótesis nol menyatakan bahwa ada bagian dari Z_i yang tidak berpengaruh pada respon Y.variabel prediktor ini akan ditulis dengan $z_{q+1}, z_{q+2}, ..., z_r$. pernyataan yang menyebutkan $z_{q+1}, z_{q+2}, ..., z_r$ tidak mempengaruhi respon Y ditulis dalam hipótesis statistika:

$H_0 = \beta_{q+1} = \beta_{q+2} = ... = \beta_r = 0$

Aturlah $Z = \left[\underset{(n\times(q+1))}{Z_1} \vdots \underset{(n\times(r-q))}{Z_2} \right], \beta = \begin{bmatrix} \underset{((q+1)\times 1)}{\beta_{(1)}} \\ \hline \underset{((r-q)\times 1)}{\beta_{(2)}} \end{bmatrix}$

Maka model regresi umum dapat ditulis sebagai:

$$Y = Z\beta + \varepsilon = [Z_1 \vdots Z_2] \begin{bmatrix} \beta_{(1)} \\ \hline \beta_{(2)} \end{bmatrix} + \varepsilon = Z_1\beta_{(1)} + Z_2\beta_{(2)} + \varepsilon$$

Test rasio likelihood Ho berdasarkan pada:

Jumlah kuadrat ekstra

$$SS_{\mathrm{Res}}(Z_1) - SS_{\mathrm{Res}}(Z) = (y - Z_1\hat{\beta}_{(1)})'(y - Z_1\hat{\beta}_{(1)}) - (y - Z_1\hat{\beta})'(y - Z_1\hat{\beta}_{(1)})$$

Result 7.6

Misalkan Z full rank r+1 dan ε berdistibusi $N(0,\sigma^2 I)$. Test rasio likelihood $H_0 = \beta_{q+1} = \beta_{q+2} = ... = \beta_r = 0$ ekuivalent dengan dengan test Ho yang didasarkan pada jumlah kuadrat pada persamaan $SS_{\mathrm{Res}}(Z_1) - SS_{\mathrm{Res}}(Z) = (y - Z_1\hat{\beta}_{(1)})'(y - Z_1\hat{\beta}_{(1)}) - (y - Z_1\hat{\beta})'(y - Z_1\hat{\beta}_{(1)})$ dan $s^2 = (y - Z_1\hat{\beta})'(y - Z_1\hat{\beta}_{(1)})/(n - r - 1)$.

Test rasio likelihood menolak Ho jika:

$$\frac{(SS_{\mathrm{Res}}(Z_1) - SS_{\mathrm{Res}}(Z))/(r-q)}{s^2} > F_{r-q,n-r-1}(\alpha)$$

Dimana:

$$s^2 = (y - Z\hat{\boldsymbol{\beta}})'(y - Z\hat{\boldsymbol{\beta}})/(n - r - 1)$$

$F_{r-q,n-r-1}(\boldsymbol{\alpha})$ dimana r-q dan n-r-1 adalah derajat bebasnya.

Contoh 7.5

Laki-laki dan perempuan yang berlangganan menilai rata-rata pelayanan di tiga tempat pada sebuah daerah restoran yang luas. Rata-rata pelayanan dikonversikan pada sebuah nilai indeks. Data disediakan pada tabel 7.2 dibawah. Data mempunyai n = 18 pelanggan. Tiap data pada tabel dikategorikan sesuai dengan lokasi (1, 2, 3) dan jenis kelamin (laki-laki = 0, perempuan = 1). Tambahannya kombinasi antara lokasi satu dengan laki-laki ada lima respon, kombinasi lokasi dua dengan perempuan ada 2 respon. Kemudian diperkenalkan tiga variabel dummy untuk lokasi dan dua variabel dummy untuk jenis kelamin. Model regresi yang menghubungkan antara indeks pelayanan dengan lokasi, jenis kelamin dan kombinasinya dapat dibuat dalam suatu matriks:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Koefisien vektor β = $[\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{31}]$

Desain matriks diatas tidak full rank, oleh program komputer diperoleh:

$SS_{res}(Z) = 2977.4$

Rank (Z) = 6, n-Rank (Z) = 12

Model pertama dengan hanya menggunakan 6 kolom pertama dari Z, yaitu tanpa mempertimbangkan interkasi antara jenis kelamin dan lokasi kita peroleh Z1 dan

$SS_{res}(Z_1) = 3419.1$

Dengan n-rank (Z1) = 18-4 = 14

Hipotesisnya:

$H_0 : \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = ... = \boldsymbol{\gamma}_{31} = \boldsymbol{\gamma}_{32} = 0$

Kemudian kita menghitung nilai F

$$F = \frac{(SS_{res}(Z_1) - SS_{res}(Z))/(6-4)}{s^2}$$

$$F = \frac{(SS_{res}(Z_1) - SS_{res}(Z))/2}{SS_{res}(Z)/12}$$

$$F = \frac{(3419,1 - 2977,4)/2}{2977,4/12} = 0,89$$

Kesimpulannya, rata-rata pelayanan tidak dipengaruhi oleh interaksi dari lokasi dengan jenis kelamin.

2.5 Interferensi dari Fungsi Regresi yang diestimasi

Misalkan sebuah model regresi memenuhi model kecocokan regresi, maka dapat digunakan untuk memecahkan dua masalah prediksi. Misalkan z_0 = $\left[1, z_{01}, ..., z_{0r}\right]'$ merupakan nilai yang dipilih untuk variabel predictor. Maka z_0 dan $\hat{\beta}$ dapat digunakan untuk :

1. Mengestimasi fungsi regresi pada z_0

Misalkan Y_0 menyatakan nilai respon ketika variabel predictor memiliki nilai z_0 = $\left[1, z_{01}, ..., z_{0r}\right]'$. Menurut model 7.3, makan nilai ekspektasi dari Y_0 adalah :

$$E(Y_0 | Z_0) = \beta_0 + \beta_1 z_{01} + ... + \beta_r z_{0r} = z'_0\,\beta \text{ ………………(7-18)}$$

Estimasi nilai terkecilnya adalah $z_0'(Z'Z)^{-1} z_0 \sigma^2$.

Result 7.7

Untuk model regresi linier pada model 7.3, $z'_0\,\hat{\beta}$ merupakan estimator linier yang tidak bias dari $E(Y_0 | Z..)$ dengan nilai variansi minimum, Var ($z'_0\,\hat{\beta}$) = $z_0'(Z'Z)^{-1} z_0 \sigma^2$. Jika error ε berdisribusi normal, maka taraf kepercayaan $100(1-\alpha)\%$ untuk $E(Y_0 | Z_0)$ = $z'_0\,\beta$ adalah:

$$z'_0\,\beta \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{(z_0'(Z'Z)^{-1} z_0) s^2}$$

Dengan $t_{n-r-1}(\alpha/2)$ sebagai batas atas percentil ke $100(\alpha/2)$ dari distribusi t dan derajat bebas $n - r - 1$.

2. Meramalkan sebuah obsevasi baru pada z_o

Prediksi pada sebuah observasi, misalnya Y_o, pada $z_o = [1, z_{01}, ..., z_{0r}]$ lebih tidak pasti daripada mengestimasi nilai harapan dari Y_o. Sesuai model regresi pada (7.3)

$Y_o = z_0^{'}\beta + \varepsilon_0$

Atau

(Respon baru Y_o) = (nilai harapan baru Y_o pada z_o) + (error baru)

Dimana ε_0 berdistribusi $N(0, \sigma^2)$. Nilai ε mempengaruhi nilai penaksir $\hat{\beta}$ dan s^2 melalui nilai variabel respon Y, tetapi tidak mempengaruhi nilai ε_0

Result 7.8

Misalnya diberikan model regresi linier (7.3), sebuah nilai observasi baru Y_o mempunyai prediktor tidak bias

$z_0^{'}\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 z_{01} + ... + \hat{\beta}_r z_{0r}$

Variansi dari galat ramalan, $Y_0 - z_0^{'}\hat{\beta}$ adalah

Var($Y_0 - z_0^{'}\hat{\beta}$) = $\sigma^2(1 + z_0^{'}(Z'Z)^{-1}z_0)$

Ketika error ε berdistribusi normal, maka sebuah interval prediksi $100(1-\alpha)\%$ untuk Y_o diberikan sebagai berikut :

$$z_0^{'}\hat{\beta} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{s^2(1 + z_0^{'}(Z'Z)^{-1}z_0)}$$

Dengan $t_{n-r-1}(\alpha/2)$ sebagai batas atas percentil ke $100(\alpha/2)$ dari distribusi t dan derajat bebas $n-r-1$.

Interval prediksi untuk Y_o lebih luas dari interval kepercayaan untuk mengestimasi nilai dari fungsi regresi $E(Y_0 | Z_0)$ = $z'_0\beta$. Pertambahan ketidakpastian pada peramalan Y_o yang direpresentasikan oleh tambahan keberadaan s^2 pada pernyataan $s^2(1 + z_0^{'}(Z'Z)^{-1}z_0)$, datang dari keberadaan istilah error yang tidak dikenal atau diketahui ε_0

Contoh kasus

Sebuah perusahaan menyadari bahwa pembelian perangkat komputer haruslah terlebih dahulu menaksir kebutuhan masa depan mereka untuk menentukan perangkat tang tepat. Seorang ilmuwan komputer mengumpulakan data dari tujuh perusahaan di tempat yang sama sehingga persamaan peramalan dari permintaan perangkat keras komputer untuk inventaris manajemen dapat ditambah. Datanya disajikan dalam tabel 7.3

Dengan: z_1 = Pesanan pelanggan (dalam ribuan)

z_2 = Jumlah ítem add-delete (dalam ratusan)

Y = Waktu CPU (dalam jam)

Buatlah sebuah interval kepercayaan 95% untuk rata-rata waktu CPU, $E(Y_0 | Z_0)$ = $\beta_0 + \beta_1 z_{01} + \beta_2 z_{02}$ pada $z_0 = [1, 130, 7.5]'$. Buat juga interval prediksi 95% untuk permintaan baru fasilitas CPU yang berkorespondensi pada z_0 yang sama.

Tabel 7.3. Data Komputer

z_1 = Pesanan pelanggan	z_2 = Jumlah ítem add-delete	Y = Waktu CPU
123.5	2.108	141.5
146.1	9.213	168.9
133.9	1.905	154.8
128.5	0.815	146.5
151.5	1.061	172.8
136.2	8.603	160.1
92.0	1.125	108.5

Dengan software, diperoleh fungsi persamaan regresi diestimasi:

$$\hat{y} = 8.42 + 1.08 z_1 + 0.42 z_2$$

$$(Z'Z)^{-1} = \begin{bmatrix} 8.17969 & & \\ -0.06411 & 0.00052 & \\ 0.08831 & -0.00107 & 0.01440 \end{bmatrix}$$

Dengan s = 1.204.

$z_0^{'}\hat{\beta} = 8.42 + 1.08(130) + 0.42(7.5) = 151.97$

$s^2(z_0^{'}(Z'Z)^{-1}z_0)$ = 1.204(0.58928) = 0.71

t_4 (0.025) = 2.776

Jadi, interval kepercayaan untuk rata-rata waktu CPU pada z_o adalah

$z_0^{'}\hat{\beta} \pm t_4(0.025)s\sqrt{z_0^{'}(Z'Z)^{-1}z_0} = 151.97 \pm 2.776(0.71)$ = (150.00, 153.94)

Interval prediksi 95% waktu CPU pada fasilitas baru dengan syarat z_o :

$s\sqrt{1+z_0^{'}(Z'Z)^{-1}z_0}$ = (1.204)(1.16071) = 1.40

Maka: $z_0^{'}\hat{\beta} \pm t_4(0.025)s\sqrt{1+z_0^{'}(Z'Z)^{-1}z_0} = 151.97 \pm 2.776(0.40)$ = (1.48.08, 155.86)

2.6 Pengecekan Model dan Beberapa Hal Dalam Regresi

Apakah suatu model sudah cocok?

Asumsikan suatu model sudah benar, kita perlu mengestimasi terlebih dahulu fungsi regresi untuk membuat suatu keputusan. Tentulah sangat penting untuk memeriksa kecukupan model sebelum fungsi yang diestimasi menjadi keputusan yang tetap.

Semua informasi kekurangcocokan sampel terkandung pada Residual.

$\hat{\varepsilon}_1 = y_1 - \hat{\beta}_0 - \hat{\beta}_1 z_{11} - ... - \hat{\beta}_r z_{1r}$

$\hat{\varepsilon}_2 = y_2 - \hat{\beta}_0 - \hat{\beta}_1 z_{21} - ... - \hat{\beta}_r z_{2r}$

.

.

.

$\hat{\varepsilon}_n = y_n - \hat{\beta}_0 - \hat{\beta}_1 z_{n1} - ... - \hat{\beta}_r z_{nr}$

$\hat{\varepsilon} = [I - Z(Z'Z)^{-1}Z']y = [I - H]y$

Jika modelnya cocok, tiap residual $\hat{\varepsilon}_j$ adalah estimator dari ε_j yang diasumsikan merupakan variabel random normal dengan rata-rata nol dan variansi σ^2. Banyak statistikawan menggunakan diagnosa grafik untuk memeriksa residual yang didasarkan pada residual student. Persamaannya sebagai berikut:

$$\hat{\varepsilon}_j^* = \frac{\hat{\varepsilon}_j}{\sqrt{s^2(1-h_{jj})}}, \; j = 1, 2, ..., n$$

Kita mengharapkan residual student ini merupakan gambaran yang mendekati distribusi normal dengan rata-rata 0 dan variansi 1. Dengan menggunakan software statistika makan akan diperoleh beberapa grafik gambaran residual sebagai berikut (hal.309)

1. Plot residual, $\hat{\varepsilon}_j$,dengan nilai prediksi, $\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 z_{j1} + ... + \hat{\beta}_r z_{jr}$

   Kemungkinanannya akan tampak seperti pada gambar 7.2 a dan 7.2 b. ini menunjukkan model regresi kita ada yang kurang tepat. Bisa disebabkan oleh kesalahan penghitungan atau variabel intersepnya dikeluarkan dari model. Hal lain adalah kemungkinan variansi error yang tidak konstan yang menyebabkan residualnya membentuk seperti corong. Adanya fluktuasi yang besar pada nilai-nilai error. Untuk memperbaiki atau mengkoreksi makan dilakukan transormasi dan atau pendekatan bobot kuadrat terkecil. Tetapi kedua hal ini tidak dijelaskan lebih lanjut pada bahasan ini.

   Gambaran grafik yang ideal ditunjukkan pada gambar 7.2 d
2. Plot residual, $\hat{\varepsilon}_j$,dengan sebuah variabel prediksi, z_1 , produk dari variabel prediktor misalnya $z_1 z_2$ atau z_1^2. Jika hasil dari análisis ini menghasilkan grafik seperti gambar 7.2 c maka model regresi yamg kita peroleh masih belum baik. Situasi ini menyarankan kita untuk menambah variabel prediktor lain pada model kita.
3. Q-Q plot dan histogram. Untuk membaca hasil yamg diperoleh pada análisis ini kita bisa membaca análisis yang ada pada bab 4.6

4. Plot residual dengan waktu. Jika data yang kita peroleh sudah terurut secara kronologis, plot residual dengan waktu maka akan mungkin muncul formula yang sistematis. (dalam hal ini mungkin akan muncul asosiasi antara error). Tambahannya, residual yang bertambah seiring dengan waktu mengindikasikan keterikatan yang kuat

Beberapa permasalahan tambahan pada Regresi linier

1. Pemilihan variabel prediktor dari sebuah himpunan yang sangat besar

   Pada praktek sehari-hari, terkadang sangat sulit untuk membuat formula yang tepat untuk fungsi regresi liner secara langsung. Pertanyaannya adalah variabel predictor mana yang harus dimasukkan pada model? Bentuk regresi seperti apa yang harus dibentuk?

   Ketika kita memiliki sebuah himpunan variabel prediktor yang sangat besar (banyak), semua variabel ini tidak bisa dimasukkan dalam fungsi regresi. Program komputer menyediakan cara untuk memilih himpunan bagian variabel prediktor yang terbaik dari himpunan yang tersedia. Pada program komputer akan menyediakan gambar plot (C_p , p) dimana

   $$C_p = \left(\frac{\text{jumlah kuadrat residual dari subset model dengan p parameter termasuk intersept}}{\text{variansi residual untuk model penuh}}\right) -(n-2p)$$

   Model yang terbaik dapat dilihat dari koordinat (C_p , p) sekitar 45^0

2. Kolinier

   Jika Z tidak full rank, beberapa kombinasi linier misalnya Za, harus nol. Pada situasi ini, kolom-kolom dikatakan kolinier. Hal ini mengakibatkan Z'Z tidak memiliki invers. Pada kebanyakan model regresi keadaan Za tidak mungkin tepat sama dengan nol. Jadi akan muncul kombinasi linier kolom pada Z dengan nilai dipersekitaran nol. Hal ini akan menyebabkan kesulitan bagi kita untuk mendeteksi kesignifikanan koefisien parameter pada model regresi. Hal ini dapat diatasi dengan:

   1. Menghapus pasangan prediktor yang berkorelasi kuat

2. Menghubungkan variabel respon dengan komponen utama variabel prediktor.

3. Bias yang disebabkan oleh model yang kurang tepat.

   Misalkan beberapa variabel predictor yang penting dikeluarkan dari model regresi yang dianjurkan. Misalkan model yang tepat dengan $Z = \left[Z_1 \vdots Z_2 \right]$ dengan rank r + 1 dan

$$\underset{(n\times1)}{Y} = \left[\underset{(n\times(q+1))}{Z_1} \vdots \underset{(n\times(r-q))}{Z_2} \right] \left[\frac{\underset{((q+1)\times1)}{\beta_{(1)}}}{\underset{((r-q)\times1)}{\beta_{(2)}}} \right] + \underset{(n\times1)}{\varepsilon}$$

$$Y = Z_1\beta_{(1)} + Z_2\beta_{(2)} + \varepsilon$$

Dimana: $E(\varepsilon) = 0$
$Var(\varepsilon) = \sigma^2 I$

Bagaimanapun, penyelidik tanpa mengetahui telah memenuhi sebuah model hanya dengan menggunakan q variabel prediktor. Penaksir kuadrat terkecil dari β_1 adalah $\hat{\beta}_1$. $\hat{\beta}_1 = (Z_1^{'}Z_1)^{-1}Z_1^{'}Y$. Kemudian, tidak sama dengan situasi ketika modelnya benar,

$$E(\hat{\beta}_{(1)}) = (Z_1^{'}Z_1)^{-1}Z_1^{'}E(Y) = (Z_1^{'}Z_1)^{-1}Z_1^{'}(Z_1\beta_{(1)} + Z_2\beta_{(2)} + E(\varepsilon))$$

Jadi, $\hat{\beta}_1$ adalah penaksir bias dari β_1. Hal ini menyebabkan taksiran kuadrat terkecil dari $\hat{\beta}_1$ menjadi menyesatkan.

Nama : Adila Sandy Wulandari

Nim : 055518

2.7 Regresi Linier berganda multivariat

Regresi berganda multivariat merupakan hubungan antara *m* respon, $Y_1, Y_2, ..., Y_m$ dan variabel prediktornya $Z_1, Z_2, ..., Z_r$, masing-masing respon diasumsikan memenuhi model regresi :

$$Y_1 = \beta_{01} + \beta_{11} z_1 + \ldots + \beta_{r1} z_r + \varepsilon_1$$
$$Y_2 = \beta_{02} + \beta_{12} z_1 + \ldots + \beta_{r2} z_r + \varepsilon_2$$
$$\vdots$$
$$Y_m = \beta_{0m} + \beta_{1m} z_1 + \ldots + \beta_{rm} z_r + \varepsilon_m$$

Persamaan error $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_m]'$ dengan $E(\varepsilon) = 0$ dan $Var(\varepsilon) = \Sigma$.

Untuk percobaan ke j, variabel predictornya adalah $\left[z_{j0}, z_{j1}, ..., z_{jr}\right]$, himpunan persamaannya adalah $\left[Y_{j1}, Y_{j2}, ..., Y_{jm}\right]'$, dan himpunan errornya adalah $\varepsilon_j = \left[\varepsilon_{j1}, \varepsilon_{j2}, ..., \varepsilon_{jm}\right]'$. Dengan model matriknya :

$$\underset{(nx(r+1))}{Z} = \begin{bmatrix} z_{10} & z_{11} & \cdots & z_{1r} \\ z_{20} & z_{21} & \cdots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & \cdots & z_{nr} \end{bmatrix}$$

Dengan persamaan matriks

$$\underset{(nxm)}{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix} = \left[Y_{(1)} \;\vdots\; Y_{(2)} \;\vdots\; \cdots \;\vdots\; Y_{(m)} \right]$$

$$\underset{((r+1)xm)}{\beta} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} = \left[\beta_{(1)} \;\vdots\; \beta_{(2)} \;\vdots\; \cdots \;\vdots\; \beta_{(m)} \right]$$

dan $$\underset{(nxm)}{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{nm} \end{bmatrix} = \left[\varepsilon_{(1)} \ \vdots \ \varepsilon_{(2)} \ \vdots \ \dots \ \vdots \ \varepsilon_{(m)} \right]$$

$$= \begin{bmatrix} \varepsilon'_1 \\ \cdots \\ \varepsilon'_2 \\ \cdots \\ \vdots \\ \cdots \\ \varepsilon'_n \end{bmatrix}$$

Model regresi linier multivariatnya adalah

$$Y_{(nxm)} = Z_{(nx(r+1))} \beta_{((r+1)xm)} + \varepsilon_{(nxm)}$$

dengan

$$E(\varepsilon_{(i)}) = 0 \, ; \quad Cov(\varepsilon_{(i)}, \varepsilon_{(k)}) = \sigma_{ik} I \qquad i, k = 1, 2, \dots, m$$

Ket :

m = jumlah observasi ke j

β = parameter yang tidak diketahui

Untuk *i* respon, maka modelnya mengikuti :

$$Y_{(i)} = Z\beta_{(i)} + \varepsilon_{(i)} \qquad i = 1, 2, \dots, m$$

Seperti pada 1 respon $\hat{\beta}$ menjadi $\hat{\beta}_{(i)} = (Z'Z)^{-1} Z'Y_{(i)}$

Sehingga diperoleh :

Nilai prediksinya : $\hat{Y} = Z\hat{\beta} = Z(Z'Z)^{-1} Z'Y$

Residualnya : $\hat{\varepsilon} = Y - \hat{Y} = [I - Z(Z'Z)^{-1} Z']Y$

jumlah kuadrat residualnya dan cross-productnya : $\hat{\varepsilon}'\hat{\varepsilon} = Y'Y - \hat{\beta}'Z'Z\hat{\beta}$

contoh 7.8

Hitung nilai $\hat{\beta}, \hat{Y}$, dan $\hat{\varepsilon}$ dengan : $Y_{j1} = \beta_{01} + \beta_{11}Z_{j1} + \varepsilon_{j1}$

$Y_{j2} = \beta_{02} + \beta_{12}Z_{j1} + \varepsilon_{j2}$ **j= 1,2,…,5**

Digunakan data dua respon Y_1 dan Y_2 pada contoh 7.3 dengan datanya sebagai berikut :

z_1	0	1	2	3	4
y_1	1	4	3	8	9
y_2	-1	-1	2	3	2

Penyelesaian :

$$Y = \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad Z' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \quad (Z'Z)^{-1} = \begin{bmatrix} 6 & -2 \\ -2 & 1 \end{bmatrix}$$

dan $$Z'y_{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 20 \end{bmatrix}$$

Sehingga

$$\hat{\beta}_{(2)} = (Z'Z)^{-1}Z'y_{(2)} = \begin{bmatrix} 6 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 20 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Pada contoh 7.3

$$\hat{\beta}_{(1)} = (Z'Z)^{-1}Z'y_{(1)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Sehingga diperoleh $\hat{\beta}=\left[\hat{\beta}_{(1)} \vdots \hat{\beta}_{(2)}\right]=\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}=(Z'Z)^{-1}Z'\left[y_{(1)} \vdots y_{(2)}\right]$

Setelah melakukan perhitungan diatas diperoleh persamaan $\hat{y}_1=1+2z_1$ dan $\hat{y}_2=-1+z_1$

Matriks nilai taksiran adalah

$$\hat{Y}=Z\hat{\beta}=\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}=\begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix}$$

dan

$$\hat{\varepsilon}=Y-\hat{Y}=\begin{bmatrix} 0 & 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Sehingga

$$\hat{\varepsilon}'\hat{Y}=\begin{bmatrix} 0 & 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Karena

$$Y'Y=\begin{bmatrix} 1 & 4 & 3 & 8 & 9 \\ -1 & -1 & 2 & 3 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix}=\begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix}$$

$$\hat{Y}'\hat{Y}=\begin{bmatrix} 165 & 45 \\ 45 & 15 \end{bmatrix} \quad \text{dan} \quad \hat{\varepsilon}'\hat{\varepsilon}=\begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$$

Jadi sum of square dan cross-productsnya memenuhi : $Y'Y=\hat{Y}'\hat{Y}+\hat{\varepsilon}'\hat{\varepsilon}$

Latihan 7.9 halaman 351

Diberikan data dengan satu variabel predictor z_1 dan dua respon Y_1 dan Y_2

z_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

Dengan $$Y_{j1} = \beta_{01} + \beta_{11} Z_{j1} + \varepsilon_{j1}$$

$$Y_{j2} = \beta_{02} + \beta_{12} Z_{j1} + \varepsilon_{j2} \qquad \mathbf{j= 1, 2, 3, 4, 5}$$

Hitung matriks untuk $\hat{Y}$,dan residual $\hat{\varepsilon}$, dengan $Y = \begin{bmatrix} y_1 & \vdots & y_2 \end{bmatrix}$

Penyelesaian :

$$Y = \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad Z' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \quad (Z'Z)^{-1} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{8} \end{bmatrix}$$

$$Z' y_{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 \\ 5 \end{bmatrix}$$

dan

$$Z' y_{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ -1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 15 \end{bmatrix}$$

Sehingga

$$\hat{\beta}_{(1)} = (Z'Z)^{-1}Z'y_{(1)} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} 15 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{15}{6} \\ \frac{5}{8} \end{bmatrix}$$

$$\hat{\beta}_{(2)} = (Z'Z)^{-1}Z'y_{(2)} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} 0 \\ 15 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{15}{8} \end{bmatrix}$$

Sehingga diperoleh $\hat{\beta} = \left[\hat{\beta}_{(1)} \vdots \hat{\beta}_{(2)} \right] = \begin{bmatrix} \frac{15}{6} & 0 \\ \frac{5}{8} & \frac{15}{8} \end{bmatrix} = (Z'Z)^{-1}Z'\left[y_{(1)} \vdots y_{(2)} \right]$

Setelah melakukan perhitungan diatas diperoleh persamaan $\hat{y}_1 = \frac{15}{6} + \frac{5}{8} z_1$ dan $\hat{y}_2 = 0 + \frac{15}{8} z_1$

Matriks nilai taksiran adalah

$$\hat{Y} = Z\hat{\beta} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{15}{6} & 0 \\ \frac{5}{8} & \frac{15}{8} \end{bmatrix} = \begin{bmatrix} \frac{15}{12} & \frac{-15}{4} \\ \frac{45}{24} & \frac{-15}{8} \\ \frac{15}{6} & 0 \\ \frac{75}{24} & \frac{15}{8} \\ \frac{45}{12} & \frac{15}{4} \end{bmatrix}$$

dan

$$\hat{\varepsilon} = Y - \hat{Y} = \begin{bmatrix} \frac{45}{12} & \frac{27}{24} & \frac{3}{2} & \frac{-27}{24} & \frac{-33}{12} \\ \frac{3}{4} & \frac{7}{8} & -1 & \frac{1}{8} & \frac{-3}{4} \end{bmatrix}^{'}$$

Sehingga

$$\hat{\varepsilon}'\hat{Y} = \begin{bmatrix} \frac{45}{12} & \frac{27}{24} & \frac{3}{2} & \frac{-27}{24} & \frac{-33}{12} \\ \frac{3}{4} & \frac{7}{8} & -1 & \frac{1}{8} & \frac{-3}{4} \end{bmatrix} \begin{bmatrix} \frac{15}{12} & \frac{-15}{4} \\ \frac{45}{24} & \frac{-15}{8} \\ \frac{15}{6} & 0 \\ \frac{75}{24} & \frac{15}{8} \\ \frac{45}{12} & \frac{15}{4} \end{bmatrix} = \begin{bmatrix} \frac{-945}{288} & \frac{-2745}{96} \\ \frac{-765}{96} & \frac{-225}{32} \end{bmatrix}$$

maka

$$Y'Y = \begin{bmatrix} 5 & 3 & 4 & 2 & 1 \\ -3 & -1 & -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

Perkiraan Kuadrat Terkecil

Untuk perkiraan kuadrat terkecil determinan $\hat{\beta} = [\hat{\beta}_{(1)} \vdots \hat{\beta}_{(2)} \vdots \ldots \vdots \hat{\beta}_{(m)}]$ menurut model regresi berganda multivariate dengan full rank (Z) = r + 1 < n, adalah

$$E(\hat{\beta}_{(i)}) = \beta_{(i)} \text{ atau } \qquad E(\hat{\beta}) = \beta$$

Dan $$Cov(\hat{\beta}_{(i)}, \hat{\beta}_{(k)}) = \sigma_{ik} (Z'Z)^{-1} \qquad \boldsymbol{i, k = 1, 2, \ldots, r + 1}$$

Residual $\hat{\varepsilon} = [\hat{\varepsilon}_{(1)} \vdots \hat{\varepsilon}_{(2)} \vdots \ldots \vdots \hat{\varepsilon}_{(m)}] = Y - Z\hat{\beta}$ memenuhi

$E(\hat{\varepsilon}_{(i)}) = 0$ dan $E(\hat{\varepsilon}'_{(i)} \hat{\varepsilon}_{(k)}) = (n - r - 1)\sigma_{ik}$ jadi

$$E(\hat{\varepsilon}) = 0 \qquad \text{dan} \qquad E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n - r - 1)}\right) = \sum$$

Maka, $\hat{\varepsilon}$ dan $\hat{\beta}$ tidak berkorelasi.

Perkiraan Maximum Likelihood

Misal model regresi berganda multivariate

$$Y_{(nxm)} = Z_{(nx(r+1))}\beta_{((r+1)xm)} + \varepsilon_{(nxm)}$$

dengan full rank (Z) = r + 1, $n \geq (r+1) + m$ dan missal error ε berdistribusi normal. Maka $\hat{\beta} = (Z'Z)^{-1}Z'Y$ adalah perkiraan maksimum likelihood dari β dan $\hat{\beta}$ yang berdistribusi normal dengan

$E(\hat{\beta}) = \beta$ dan $Cov(\hat{\beta}_{(i)}, \hat{\beta}_{(k)}) = \sigma_{ik}(Z'Z)^{-1}$. $\hat{\beta}$ independent dari perkiraan maksimum likelihood dan definit positif $\sum$ diberikan oleh

$\hat{\sum} = \dfrac{\hat{\varepsilon}'\varepsilon}{n} = \dfrac{(Y - Z\hat{\beta})'(Y - Z\hat{\beta})}{n}$ dan $n\hat{\sum}$ adalah distribusi $W_{n-r-1}(.|\sum)$.

Tes rasio likelihood untuk parameter regresi

Tes ini merupakan rasio likelihood untuk banyak respon, dengan hipotesis bahwa respon tidak bergantung pada $Z_{q+1}, Z_{q+2}, \ldots, Z_r$, sehingga

$$H_0 : \beta_{(2)} = 0 \text{ dimana } \qquad \beta = \begin{bmatrix} \beta_{(1)} \\ ((q+1)xm) \\ \hline \beta_{(2)} \\ ((r-q)xm) \end{bmatrix}$$

dengan $Z = \begin{bmatrix} Z_1 & \vdots & Z_2 \\ (nx(q+1)) & & (nx(r-q)) \end{bmatrix}$, secara umum model dapat ditulis :

$$E(Y) = Z\beta = [Z_1 \vdots Z_2]\begin{bmatrix} \beta_{(1)} \\ \hline \beta_{(2)} \end{bmatrix} = Z_1\beta_{(1)} + Z_2\beta_{(2)}$$

dengan $H_0 : \beta_{(2)} = 0$, $Y = Z_1 \beta_{(1)} + \varepsilon$ dan tes rasio likelihood dari H_0 berdasarkan pada jumlah yang terkait dalam jumlah kuadrat ekstra dan coss-products $= (Y - Z_1\hat{\beta}_{(1)})'(Y - Z_1\hat{\beta}_{(1)}) - (Y - Z\hat{\beta})'(Y - Z\hat{\beta})$

$= n(\hat{\Sigma}_{(1)} - \hat{\Sigma})$

Dimana $\hat{\beta}_{(1)} = (Z'_1 Z_1)^{-1} Z'_1 Y$ dan $\hat{\Sigma}_1 = \dfrac{(Y - Z_1\hat{\beta}_{(1)})'(Y - Z_1\hat{\beta}_{(1)})}{n}$

Dari rasio likelihood (Λ) dapat memperlihatkan hubungan umum varian, jadi :

$$\Lambda = \frac{\max_{\beta_{(1)}, \Sigma} L(\beta_{(1)}, \Sigma)}{\max_{\beta, \Sigma} L(\beta, \Sigma)} = \frac{L(\hat{\beta}_{(1)}, \hat{\Sigma}_1)}{L(\hat{\beta}, \hat{\Sigma})} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|} \right)^{n/2}$$

Equivalent dengan statistic Wilks’Lambda :

$$\Lambda^{2/n} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}$$

dapat dipergunakan.

Hasil 7.11

Misal model regresi berganda multivariate

$$Y_{(nxm)} = Z_{(nx(r+1))} \beta_{((r+1)xm)} + \varepsilon_{(nxm)}$$

dengan full rank (Z) = r + 1, $n \geq (r+1) + m$ dan misal error ε berdistribusi normal. Dengan $H_0 : \beta_{(2)} = 0$, $n\hat{\Sigma}$ adalah distribusi $W_{n-r-1}(.| \Sigma)$ secara bebas adalah $n(\hat{\Sigma}_1 - \hat{\Sigma})$ dimana distribusinya $W_{r-q}(.| \Sigma)$. Tes rasio likelihood dari H_0 equivalent dengan tolak H_0 untuk besar nilai dari :

$$-2Ln\Lambda = -nLn\left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right) = -nLn\frac{|n\hat{\Sigma}|}{|n\hat{\Sigma}+n(\hat{\Sigma}_1-\hat{\Sigma})|}$$

untuk besar nilainya n buah maka statistiknya :

$$-\left[n-r-1-\frac{1}{2}(m-r+q+1)\right]Ln\left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right)$$

Menggunakan pendekatan chi kuadrat dengan derajat bebasnya m(r-q).

contoh 7.9

contoh ini merupakan lanjutan yang diberikan pada contoh sebelumnya yaitu pada contoh 7.5. Dengan menggunakan program computer, sehingga diperoleh :

$$\begin{pmatrix}\text{residual sum of squares}\\ \text{dan cross pruducts}\end{pmatrix} = n\hat{\Sigma} = \begin{bmatrix}2977,39 & 1021,72\\ 1021,72 & 2050,95\end{bmatrix}$$

$$\begin{pmatrix}\text{extrar sum of squares}\\ \text{dan cross pruducts}\end{pmatrix} = n(\hat{\Sigma}_1-\hat{\Sigma}) = \begin{bmatrix}441,76 & 246,16\\ 246,16 & 366,12\end{bmatrix}$$

Misal $\beta_{(2)}$ adalah matriks untuk interaksi parameter dua respon. Diketahui pada contoh sebelumnya bahwa nilai n= 18 yang dapat dikategorikan tidak terlalu besar, sehingga diperoleh hipotesis :

$$H_0 : \beta_{(2)} = 0$$

$$H_1 : \beta_{(2)} \neq 0$$

Dengan nilai alfa sebesar 0,05, dapat diuji :

$$-\left[n-r_1-1-\frac{1}{2}(m-r_1+q_1+1)\right]\ln\left(\frac{|n\hat{\Sigma}|}{|n\hat{\Sigma}+n(\hat{\Sigma}_1-\hat{\Sigma})|}\right)$$

$$= -\left[18 - 5 - 1 - \frac{1}{2}(2 - 5 + 3 + 1)\right]\ln(7605)$$

$$= 3,28$$

Dengan menggunakan pendekatan chi-kuadrat, diperoleh nilai pada tabel chi-kuadrat dengan derajat bebas sebesar m (r_1-q_1) = 2 (2) = 4 adalah 9,49.

Sehingga nilai hitung akan lebih kecil daripada nilai pada tabel yaitu $3,28 < \chi_4^2.(0,05) = 9,49$.

Untuk kriteria hitungnya maka H_0 ditolak pada nilai alfa sebesar 5%. Sehingga nilai $\beta_{(2)} \neq 0$, artinya nilai koefisien untuk $\beta_{(2)}$ berarti dan hubungan interaksi tidak dibutuhkan.

Nama : Siti Yunengsih

Nim : 055951

2.8 Konsep Dari Regresi Linier

Model regresi linier klasik menghubungkan antara suatu variabel terikat Y dan kumpulan variabel prediktor $z_1, z_2, \dots z_r$. Model regresi menganggap bahwa variabel acak Y bergantung pada variabel tetap z. Rata-rata nya diasumsikan sebagai fungsi linier dengan koefisien regresi $\beta_o, \beta_1, \dots, \beta_r$.

Anggaplah bahwa $Y, Z_1, Z_2, \dots, Z_r$ adalah variabel acak yang mempunyai distribusi sama tidak harus normal, dengan vektor rata-rata $\underset{(r+1)\times 1}{\mu}$ dan matrix covariant $\underset{(r+1)\times(r+1)}{\Sigma}$ partisi μ dan Σ kita tulis sebagai berikut

$$\mu = \begin{bmatrix} \underset{(1\times 1)}{\mu_Y} \\ \underset{(r\times 1)}{\mu_Z} \end{bmatrix} \text{ dan } \Sigma = \left[\begin{array}{c|c} \underset{(1\times 1)}{\sigma_{YY}} & \underset{(1\times r)}{\sigma_{ZY}'} \\ \underset{(r\times 1)}{\sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right] \text{ dengan } \sigma_{ZY} = \left[\sigma_{YZ_1}, \sigma_{YZ_2}, \dots, \sigma_{YZ_r}\right]'$$

Dalam memprediksi variabel terikat Y digunakan

prediktor linier = $b_0 + b_1 Z_1 + \dots + b_r Z_r = b_0 + b'Z$

dengan prediksi errornya yaitu

prediction error = Y - b_0 - b_1 Z_1 - … - $b_r Z_r$ =Y - b_0 - b'Z

karena error ini bersipat acak, biasanya untuk memilih b_0 dan b dengan meminimumkan

Mean square error = E(Y - b_0 - b'Z)^2

Mean square error ini bergantung pada distribusi bersama dari Y dan Z melalui parameter μ dan Σ

Akibat 7.12

Prediktor linier $\beta_0 + \beta' Z$ dengan koefisien

$$\beta = \Sigma_{ZZ}^{-1} \sigma_{ZY,} \qquad \beta_0 = \mu_Y - \beta' \mu_Z$$

Memilki rata-rata kuadrat minimum diantara semua prediktor linier respon Y dan memiliki mean square error yaitu

$$E(Y - \beta_0 - \beta' Z)^2 = E(Y - \mu_Y - \sigma_{ZY}' \Sigma_{ZZ}^{-1} (Z - \mu_Z))^2 = \sigma_{YY} - \sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}$$

Juga $\beta_0 + \beta' Z = \mu_Y + \sigma_{ZY}' \Sigma_{ZZ}^{-1} (Z - \mu_Z)$ adalah prediktor linier yang memiliki korelasi maksimum dengan Y

$$Corr(Y, \beta_0 + \beta' Z) = \max Corr(Y, b_0 + b' Z)$$

$$= \sqrt{\frac{\beta' \Sigma_{ZZ} \beta}{\sigma_{YY}}} = \sqrt{\frac{\sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}}$$

Korelasi antara variabel terikat Y dengan prediktor linier terbaiknya disebut koeffisien korelasi multiple populasi yang dinotasikan sebagai

$$\rho_{Y(Z)} = + \sqrt{\frac{\sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}}$$

kuadrat dari koeffisien ini $\rho_{Y(Z)}^2$ disebut koeffisien determinasi populasi, nilai dari koeffisien korelasi adalah akar kuadrat positif nya yaitu $0 \le \rho_{Y(Z)} \le 1$.

Koeffisien determinasi memiliki interpretasi penting. Dari akibat 7.12 mean square error menggunakan $\beta_0 + \beta' Z$ untuk meramalkan Y adalah

$$\sigma_{YY} - \sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY} = \sigma_{YY} - \sigma_{YY} \left(\frac{\sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}} \right) = \sigma_{YY} (1 - \rho_{Y(Z)}^2)$$

Jika $\rho^2_{Y(Z)} = 0$ tidak ada kekuatan prediksi dalam Z, perbedaan yang sangat besar jika $\rho^2_{Y(Z)} = 1$ mengakibatkan Y dapat diprediksi dengan tepat .

Contoh 7.11

Diberikan vektor rata-rata dan matrik kovarian dari Y, Z_1, Z_2

$$\mu = \begin{bmatrix} \mu_Y \\ \hline \mu_Z \end{bmatrix} = \begin{bmatrix} 5 \\ \hline 2 \\ 0 \end{bmatrix} \text{ dan } \Sigma = \left[\begin{array}{c|c} \sigma_{yy} & \sigma'_{ZY} \\ \hline \sigma_{ZY} & \Sigma_{ZY} \end{array}\right] = \left[\begin{array}{c|cc} 10 & 1 & -1 \\ \hline 1 & 7 & 3 \\ -1 & 3 & 2 \end{array}\right]$$

Tentukan a). prediktor Linier terbaik $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2$

b). mean square error

c). koeffisien korelasi multiple

penyelesaian

$$\beta = \Sigma_{ZZ}^{-1}\,\sigma_{ZY} = \begin{bmatrix} 7 & 3 \\ 3 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0,4 & -0,6 \\ -0,6 & 1,4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$\beta_0 = \mu_Y - \beta'\mu_Z = 5 - \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 3$$

a). Jadi prediktor linier terbaiknya adalah $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 = 3 + Z_1 - 2Z_2$

b). mean square errornya

$$\sigma_{YY} - \sigma'_{ZY}\,\Sigma_{ZZ}^{-1}\,\sigma_{ZY} = 10 - \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 0,4 & -0,6 \\ -0,6 & 1,4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 10 - 3 = 7$$

c). koeffisien korelasi multiplenya $\rho_{Y(Z)} = +\sqrt{\dfrac{\sigma'_{ZY}\,\Sigma_{ZZ}^{-1}\,\sigma_{ZY}}{\sigma_{YY}}} = \sqrt{\dfrac{3}{10}} = 0,548$

Pembatasan prediktor linier dekat dihubungkan dengan assumsi normalitas, khususnya

misalkan kita punya $\begin{bmatrix} Y \\ Z_1 \\ Z_2 \\ \vdots \\ Z_r \end{bmatrix}$ *berdistribusi* $N_{r-1}(\mu, \Sigma)$

maka distribusi bersyarat dari Y dengan memperhatikan nilai z_1, z_2, …,z_r adalah

$$N(\mu_Y + \sigma_{ZY}^{'} \Sigma_{ZZ}^{-1}(z-\mu_Z), \sigma_{YY} - \sigma_{ZY}^{'} \Sigma_{ZZ}^{-1} \sigma_{ZY})$$

rata-rata dari distribusi bersyarat ini adalah prediktor linier dalam akibat 7.12 adalah

$$\begin{aligned} E(Y|z_1, z_2, \dots, z_r) &= \mu_Y + \sigma_{ZY}^{'} \Sigma_{ZZ}^{-1}(z-\mu_z) \\ &= \beta_0 + \beta^{'} z \end{aligned}$$

dan kita menyimpulkan $E(Y|z_1, z_2, \dots, z_r)$ adalah prediktor linier tebaik dari Y ketika populasinya adalah $N_{r+1}(\mu, \Sigma)$. Ekspektasi bersyarat ini disebut fungsi regresi linier.

Ketika populasi tidak normal, fungsi regresi $E(Y|z_1, z_2, \dots, z_r)$ tidak harus berbentuk $\beta_0 + \beta^{'} z$. Namun, dapat ditunjukan bahwa $E(Y|z_1, z_2, \dots, z_r)$ apapun bentuknya, untuk memprediksi Y adalah dengan mean square error terkecil. Keuntungannya pengoptimalan diantara semua estimator yang dimiliki dengan prediktor linier adalah ketika populasinya normal.

Akibat 7.13

Anggaplah bahwa distribusi bersama dari Y dan Z adalah $N_{r+1}(\mu, \Sigma)$ misalkan

$$\hat{\mu} = \begin{bmatrix} \overline{Y} \\ \overline{Z} \end{bmatrix} \quad \text{dan} \quad S = \left[\begin{array}{c|c} S_{YY} & S_{ZY}^{'} \\ \hline S_{ZY} & S_{ZZ} \end{array}\right]$$

vektor rata-rata sampel dan matrik kovarian sampel berukuran n dari suatu populasi, penaksir maksimum likelihood dari koeffisien prediktor liniernya adalah

$$\hat{\beta} = S_{ZZ}^{-1} s_{ZY}, \quad \hat{\beta}_0 = \overline{Y} - s_{ZY}^{'} S_{ZZ}^{-1} \overline{Z} = \overline{Y} - \hat{\beta}^{'} \overline{Z}$$

akibatnya penaksir likelihood untuk fungsi liniernya adalah

$$\hat{\beta}_0 + \hat{\beta}^{'} z = \overline{Y} + s_{ZY}^{'} S_{ZZ}^{-1}(z - \overline{Z})$$

Penaksir maximum lilkelihood dari mean squre errornya $E\left[Y - \beta_0 - \beta^{'} Z\right]^2$ adalah

$$\hat{\sigma}_{YY \cdot Z} = \frac{n-1}{n}(s_{YY} - s_{ZY}^{'} S_{ZZ}^{-1} s_{ZY})$$

Biasanya dengan merubah pembagi dari n ke n-(r + 1) dalam estimator dari means square error diperoleh penaksir tak bias yaitu

$$\left(\frac{n-1}{n-r-1}\right)\left(s_{YY}-s_{ZY}^{'}S_{ZZ}^{-1}s_{ZY}\right)=\frac{\sum_{j=1}^{n}\left(Y_j-\hat{\beta}_0-\hat{\beta}^{'}Z_j\right)^2}{n-r-1}$$

Contoh 7.12

Hasil computer data contoh 7.6. dengan data 7 observasi pada Y (CPU Time), Z_1, Z_2 memberikan vektor rata-rata sampel dan matrik kovarian sampel yaitu

$$\hat{\mu}=\begin{bmatrix}\hat{y}\\ \overline{\overline{Z}}\end{bmatrix}=\begin{bmatrix}150,44\\ \hline 130,24\\ 3,547\end{bmatrix} \text{ dan } \Sigma=\left[\begin{array}{c|c}s_{yy} & s_{ZY}^{'}\\ \hline s_{ZY} & S_{ZY}\end{array}\right]=\left[\begin{array}{c|cc}467,913 & 418,763 & 35,983\\ \hline 418,763 & 377,200 & 28,034\\ 35,983 & 28,034 & 13,657\end{array}\right]$$

assumsikan berdistribusi normal bersama. Tentukan fungsi regresi dan mean square errornya.?

Penyelesaian

Dari akibat 7.13 penaksir maksimum likelihoodnya adalah

$$\hat{\beta}=S_{ZZ}^{-1}s_{ZY}=\begin{bmatrix}0,003128 & -0,006422\\ -0,006422 & 0,086404\end{bmatrix}\begin{bmatrix}418,763\\ 35,983\end{bmatrix}=\begin{bmatrix}1,079\\ 0,420\end{bmatrix}$$

$$\hat{\beta}_0=\overline{y}-\hat{\beta}^{'}\overline{z}=150,44-\begin{bmatrix}1,079 & 0,420\end{bmatrix}\begin{bmatrix}130,24\\ 3,547\end{bmatrix}=8,421$$

jadi fungsi regresinya adalah $\hat{\beta}_0+\hat{\beta}^{'}z=8,42-1,08z_1+0,42z_2$

mean square errornya adalah

$$\begin{aligned}\hat{\sigma}_{YY\cdot Z}&=\frac{n-1}{n}(s_{YY}-s_{ZY}^{'}S_{ZZ}^{-1}s_{ZY})\\ &=\left(\frac{6}{7}\right)\left(467,913-\begin{bmatrix}418,763 & 35,983\end{bmatrix}\begin{bmatrix}0,003128 & -0,006422\\ -0,006422 & 0,086404\end{bmatrix}\begin{bmatrix}418,763\\ 35,983\end{bmatrix}\right)\\ &=0,894\end{aligned}$$

Prediksi untuk beberapa variabel

Perluasan dari akibat sebelumnya untuk prediksi beberapa variabel terikat $Y_1, Y_2, \ldots, Y_m$ hampir dekat. Perluasan untuk populasi normal anggaplah bahwa $\begin{bmatrix} \underset{(m\times1)}{Y} \\ \hline \underset{(r\times1)}{Z} \end{bmatrix}$ berdistribusi $N_{m+r}(\mu, \Sigma)$ dengan $\mu = \begin{bmatrix} \underset{(m\times1)}{\mu_Y} \\ \hline \underset{(r\times1)}{\mu_Z} \end{bmatrix}$ dan

$$\Sigma = \left[\begin{array}{c|c} \underset{(m\times m)}{\Sigma_{YY}} & \underset{(m\times r)}{\Sigma_{YZ}} \\ \hline \underset{(r\times m)}{\Sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right]$$

Ekspektasi bersyarat dari $[Y_1, Y_2, \ldots, Y_m]$ atas sejumlah nilai variabel prediktor $z_1, z_2, \ldots, z_r$ adalah $E\left[Y \middle| z_1, z_2, \ldots, z_r\right] = \mu_Y + \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(z - \mu_Z)$ nilai harapan bersyarat ini, dianggap suatu fungsi atas $z_1, z_2, \ldots, z_r$ yang disebut dengan regresi multivariate dari vektor Y dalam Z. Fungsi ini terdiri dari m regresi univariat. Contohnya vektor rata-rata bersyarat dari komponen pertama adalah

$\mu_{Y_1} + \Sigma_{Y_1Z}\,\Sigma_{ZZ}^{-1}(z - \mu_Z) = E(Y_1 | z_1, z_2, \ldots, z_r)$ yang meminimumkan mean square error dari prediksi Y_1. Ukuran $m \times r$ matrik $\beta = \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}$ disebut matrik koeffisien regresi.

Kesalahan dari vektor prediksinya $Y - \mu_Y - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z)$ mempunyai kuadrat harapan dan matriks cross produk adalah

$$\begin{aligned}
\Sigma_{YY\cdot Z} &= E\left[Y - \mu_Y - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z)\right]\left[Y - \mu_Y - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z)\right]' \\
&= \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(\Sigma_{YZ})' - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY} + \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZZ}\,\Sigma_{ZZ}^{-1}(\Sigma_{YZ})' \\
&= \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}
\end{aligned}$$

karena μ dan Σ tidak diketahui secara khusus, maka harus diperkirakan dari sampel acak dalam urutan menyusun prediktor linier multivariate dan menentukan harapan kesalahan prediksi.

Akibat 7.14

Anggaplah Y dan Z berdistribusi $N_{m-r}(\mu, \Sigma)$. Regresi dari vektor Y dalam Z adalah

$$\beta_0 + \beta z = \mu_Y - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\mu_Z + \Sigma_{YZ}\Sigma_{ZZ}^{-1} z = \Sigma_{YZ}\Sigma_{ZZ}^{-1}(z - \mu_Z)$$

kuadrat harapan dan matriks cross produk untuk errornya adalah

$$E(Y - \beta_0 - \beta Z)(Y - \beta_0 - \beta Z)' = \Sigma_{YY \cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}$$

berdasarkan sampel acak ukuran n, estimator maximum likelihood untuk fungsi regresinya adalah

$$\hat{\beta}_0 + \hat{\beta} z = \bar{Y} + S_{YZ} S_{ZZ}^{-1}(z - \bar{Z})$$

Dan estimator likelihood dari $\Sigma_{YY \cdot Z}$ adalah

$$\hat{\Sigma}_{YY \cdot Z} = \left(\frac{n-1}{n}\right)(s_{YY} - S_{YZ} S_{ZZ}^{-1} S_{ZY})$$

Penaksir tak bias dari $\hat{\Sigma}_{YY \cdot Z}$ adalah

$$\left(\frac{n-1}{n-r-1}\right)(S_{YY} - S_{YZ} S_{ZZ}^{-1} S_{ZY}) = \frac{\sum_{j=1}^{n}(Y_j - \hat{\beta}_0 - \hat{\beta} Z_j)(Y_j - \hat{\beta}_0 - \hat{\beta} Z_j)}{n-r-1}$$

Contoh 7.13

Dari hasil komputer data contoh 7.6 dan contoh 7.10 untuk Y_1 (CPU time) dan Y_2 (Disc I/O Capacity)., diberikan Z_1 dan Z_2 diperoleh $\hat{\mu} = \left[\begin{array}{c} \hat{y} \\ \hline \bar{\bar{Z}} \end{array}\right] = \left[\begin{array}{c} 150,44 \\ 327,79 \\ \hline 130,24 \\ 3,547 \end{array}\right]$

$$\Sigma = \left[\begin{array}{c|c} S_{yy} & S_{ZY}' \\ \hline S_{ZY} & S_{ZY} \end{array}\right] = \left[\begin{array}{cc|cc} 467,913 & 1148,556 & 418,763 & 35,983 \\ 1148,556 & 3072,491 & 1008,976 & 140,558 \\ \hline 418,763 & 1008,976 & 377,200 & 28,034 \\ 35,983 & 140,558 & 28,034 & 13,657 \end{array}\right]$$

diasumsikan berdistribusi normal tentukan fungsi regresinya ?

$$\hat{\beta}_0+\hat{\beta}z=\overline{y}+S_{YZ}S_{ZZ}^{-1}(z-\overline{z})$$

$$=\begin{bmatrix}150,44\\327,79\end{bmatrix}+\begin{bmatrix}418,763 & 35,983\\1008,976 & 140,558\end{bmatrix}\times\begin{bmatrix}0,003128 & -0,006422\\-0,006422 & 0,086404\end{bmatrix}\begin{bmatrix}z_1-130,24\\z_2-3,547\end{bmatrix}$$

$$=\begin{bmatrix}150,44\\327,79\end{bmatrix}+\begin{bmatrix}1,079(z_1-130,24)+0,420(z_2-3,547)\\2,254(z_1-130,24)+5,665(z_2-3,547)\end{bmatrix}$$

sehingga predictor mean square error minimum dari Y_1 dan Y_2adalah

$$150,44+1,079(z_1-130,24)+0,420(z_2-3,547)=8,42+1,08z_1+0,42z_2$$

$$327,79+2,254(z_1-130,24)+5,665(z_2-3,547)=14,14+2,25z_1+5,67z_2$$

penaksir maksimum likelihood dari kuadrat harapan dan matrik cross produknya diberikan oleh

$$\Sigma_{YY\cdot Z}=\left(\frac{n-1}{n}\right)\left(S_{YY}-S_{YZ}S_{ZZ}^{-1}S_{ZY}\right)$$

$$=\left(\frac{6}{7}\right)\left(\begin{bmatrix}467,913 & 1148,536\\1148,536 & 3072,491\end{bmatrix}-\begin{bmatrix}418,763 & 35,983\\1008,976 & 140,558\end{bmatrix}\begin{bmatrix}0,003128 & -0,006422\\-0,006422 & 0,086404\end{bmatrix}\begin{bmatrix}418,763 & 1008,976\\35,983 & 140,558\end{bmatrix}\right)$$

$$=\left(\frac{6}{7}\right)\begin{bmatrix}1,043 & 1,042\\1,042 & 2,572\end{bmatrix}=\begin{bmatrix}0,894 & 0,893\\0,893 & 2,205\end{bmatrix}$$

hasil penaksiran pertama fungsi regresi $8,42+1,08z_1+0,42z_2$ memberikan mean square error 0,894 hasil yang sama dengan contoh 7.12 untuk kasus respon tunggal. Kita lihat bahwa data dapat diprediksi dari dari variable respon pertama memilki error yang lebih kecil dibandingkan dengan oleh respon kedua. Kovarian 0,893 menunjukan prediksi yang terlalu jauh dari CPU time yang cenderung ditemani oleh capasitas disk.

Akibat 7.14 menyatakan bahwa assumsi dari distribusi normal multivariate bersama untuk kumpulan $Y_1,Y_2,\ldots,Y_m\;\;,Z_1,Z_2,\ldots,Z_r$ mudah untuk memprediksi

persamaan

$$\begin{aligned}\hat{y}_1&=\hat{\beta}_{01}+\hat{\beta}_{11}z_1+\ldots+\hat{\beta}_{r1}z_r\\\hat{y}_2&=\hat{\beta}_{02}+\hat{\beta}_{12}z_1+\ldots+\hat{\beta}_{r2}z_r\\&\vdots\\\hat{y}_m&=\hat{\beta}_{0m}+\hat{\beta}_{1m}z_1+\ldots+\hat{\beta}_{rm}z_r\end{aligned}$$

Dengan catatan mengikuti

1. Nilai $z_1, z_2, \ldots, z_r$ yang sama digunakan untuk memprediksi tiap nilai Y_i.
2. $\hat{\beta}_{ik}$ diperkirakan untuk entri (i,k) pada matrik koeffisien regresi $\beta = \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}$ untuk $i,k \geq 1$.

Koefisien Korelasi Parsial

Anggaplah pasangan kesalahan
$$\begin{matrix} Y_1 - \mu_{Y_1} - \Sigma_{Y_1 Z}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z) \\ Y_2 - \mu_{Y_2} - \Sigma_{Y_2 Z}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z) \end{matrix}$$

diperoleh dari menggunakan prediktor linier terbaik Y_1 dan Y_2 hubungannya ditentukan dari matrik kovarian kesalahan $\Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}$

pengukuran hubungan antara Y_1 dan Y_2 setelah menghapus pengaruh dari $Z_1, Z_2, \ldots, Z_r$.

koeffisien korelasi parsial antara Y_1 dan Y_2 dengan menghapuskan

$Z_1, Z_2, \ldots, Z_r$ oleh $\rho_{Y_1Y_2\cdot Z} = \dfrac{\sigma_{Y_1Y_2\cdot Z}}{\sqrt{\sigma_{Y_1Y_1\cdot Z}}\sqrt{\sigma_{Y_2Y_2\cdot Z}}}$

yang diperkirakan oleh $r_{Y_1Y_2\cdot Z} = \dfrac{s_{Y_1Y_2\cdot Z}}{\sqrt{s_{Y_1Y_1\cdot Z}}\sqrt{s_{Y_2Y_2\cdot Z}}}$

Dimana $\sigma_{Y_iY_k\cdot Z}$ adalah entri (i,k) dalam matrik $\Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}$ hubungan koeffisien korelasi parsial sampel adalah

$$r_{Y_1Y_2\cdot Z} = \frac{s_{Y_1Y_2\cdot Z}}{\sqrt{s_{Y_1Y_1\cdot Z}}\sqrt{s_{Y_2Y_2\cdot Z}}}$$

Dengan $s_{Y_iY_k\cdot Z}$ dengan (i,k) elemen dari $S_{YY} - S_{YZ}S_{ZZ}^{-1}S_{ZY}$ dengan asumsi Y dan Z memiliki distribusi normal multivariate bersama. Koeffisien korelasi parsial sampel diatas adalah penaksir maximum likelihood untuk populasinya.

2.9 Membandingkan Dua Perumusan dari Model Regresi

Bentuk Rata-rata yang dikoreksi dari Model Regresi

Untuk beberapa variabel respon Y, model regresi multiple menegaskan bahwa

$$Y_j = \beta_0 + \beta_1 z_{1j} + \ldots + \beta_r z_{rj} + \varepsilon_j$$

Variabel prediktor dapat dipusatkan dengan mengurangi rata-ratanya.

Contohnya

$\beta_1 z_{1j} = \beta_1 (z_{1j} - \overline{z}_1) + \beta_1 \overline{z}_1$ dan kita dapat menulis

$$\begin{aligned} Y_j &= (\beta_0 + \beta_1 \overline{z}_1 + \ldots + \beta_r \overline{z}_r) + \beta_1 (z_{1j} - \overline{z}_1) + \ldots + \beta_r (z_{rj} - \overline{z}_r) + \varepsilon_j \\ &= \beta_* + \beta_1 (z_{1j} - \overline{z}_1) + \ldots + \beta_r (z_{rj} - \overline{z}_r) + \varepsilon_j \end{aligned}$$

Dengan $\beta_* = (\beta_0 + \beta_1 \overline{z}_1 + \ldots + \beta_r \overline{z}_r)$

Desain matrik rata-rata yang dikoreksi dihubungkan dengan pengulangan pembentukan parameter adalah

$$Z_c = \begin{bmatrix} 1 & z_{11} - \overline{z}_1 & \cdots & z_{1r} - \overline{z}_r \\ 1 & z_{21} - \overline{z}_1 & \cdots & z_{2r} - \overline{z}_r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} - \overline{z}_1 & \cdots & z_{nr} - \overline{z}_r \end{bmatrix}$$

Yang mana kolom r masing-masing tegak lurus terhadap kolom pertama karena

$$\sum_{j=1}^{n} 1(z_{ji} - \overline{z}_i) = 0, \quad i = 1, 2, \ldots., r$$

Selanjutnya tentukan $Z_c = \left[1 \middle| Z_{c2} \right]$ dengan $Z_{c2}^{'} 1 = 0$

Jadi

$$Z_c^{'} Z_c = \begin{bmatrix} I^{'} I & I^{'} Z_{c2} \\ Z_{c2}^{'} I & Z_{c2}^{'} Z_{c2} \end{bmatrix} = \begin{bmatrix} n & 0^{'} \\ 0 & Z_{c2}^{'} Z_{c2} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_* \\ \hline \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_r \end{bmatrix} = (Z_c^{'} Z_c)^{-1} Z_c^{'} y = \begin{bmatrix} \dfrac{1}{n} & 0^{'} \\ 0 & (Z_{c2}^{'} Z_{c2})^{-1} \end{bmatrix} \begin{bmatrix} I^{'} y \\ Z_{c2}^{'} y \end{bmatrix} = \begin{bmatrix} \overline{y} \\ \hline (Z_{c2}^{'} Z_{c2})^{-1} Z_{c2}^{'} y \end{bmatrix}$$

Dengan demikian koeffisien regresi $[\beta_1, \beta_2, \ \dots, \beta_r]'$ penaksir tak biasnya ditaksir oleh $\left(Z_{c2}'Z_{c2}\right)^{-1} Z_{c2}'y$ dan β_* ditaksir oleh $\bar{y}$. Karena koeffisien $\beta_1, \beta_2, \ \dots, \beta_r$ tetap tidak berubah oleh penggantian parameter penaksir terbaiknya dihitung dari desain matriks Z_c sama dengan yang dihitung desain matrik Z Sehingga, keadaan $\hat{\beta}_c = \left[\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_r\right]$ adalah predictor linier dari Y dapat ditulis sebagai $\hat{y} = \hat{\beta}_* + \hat{\beta}_c'\left(z - \bar{z}\right) = \bar{y} + y'Z_{c2}(Z_{c2}'Z_{c2})^{-1}(z - \bar{z})$ dengan

$\left(z - \bar{z}\right) = (z_1 - \bar{z}_1, z_2 - \bar{z}_2, \dots, z_r - \bar{z}_r)$ akhirnya

$$\begin{bmatrix} Var(\hat{\beta}_*) & Cov(\hat{\beta}_*, \hat{\beta}_c) \\ Cov(\hat{\beta}_c, \hat{\beta}_*) & Cov(\hat{\beta}_c) \end{bmatrix} = (Z_c'Z_c)^{-1}\sigma^2 = \begin{bmatrix} \dfrac{\sigma^2}{n} & 0' \\ 0 & (Z_{c2}'Z_{c2})^{-1}\sigma^2 \end{bmatrix}$$

Ulasan: Model Regresi Multiple Multivariate menghasilkan desain matrik rata-rata yang dikoreksi sama untuk setiap respon. Penaksiran kuadrat terkecil untuk koeffisien vector $\hat{\beta}_{(i)}$ untuk variable respon ke-i diberikan oleh

$$\hat{\beta}_{(i)} = \begin{bmatrix} \bar{y}_{(i)} \\ \hline (Z_{c2}'Z_{c2})^{-1}Z_{c2}'y_{(i)} \end{bmatrix}, \qquad i = 1, 2, \dots, m$$

Rumus-rumus yang berhubungan

Ketika variable $Y, Z_1, Z_2, \dots, Z_r$ berdistribusi normal bersama, kita menentukan bahwa prediktor penaksir dari Y adalah

$$\hat{\beta}_0 + \hat{\beta}'z = \bar{y} + s_{ZY}'S_{ZZ}^{-1}(z - \bar{z}) = \hat{\mu}_Y + \hat{\sigma}_{ZY}'\hat{\Sigma}_{ZZ}^{-1}(z - \mu_Z).$$

dari bentuk rata-rata yang dikoreksi pada model regresi penaksir linier terbaik dari prediktor Y adalah $\hat{y} = \hat{\beta}_* + \hat{\beta}_c'(z - \bar{z})$ dengan $\hat{\beta}_* = \bar{y} = \hat{\beta}_0$ dan dari persamaan sebelumnya $\hat{\beta}_c' = y'Z_{c2}(Z_{c2}'Z_{c2})^{-1}$ maka diperoleh hubungan

$$s_{ZY}'S_{ZZ}^{-1} = y'Z_{c2}(Z_{c2}'Z_{c2})^{-1}$$

oleh karena itu teori normal rata-rata bersyarat dan model regresi klasik memilki prediktor linier yang tepatnya sama.

Meskipun dua perumusan dari masalah prediksi linier menghasilkan persamaan predictor yang sama, pada dasarnya adalah berbeda, pada model regresi klasik variable input diassumsikan ditentukan oleh ekperiment, pada model regresi linier nilai dari variable predictor adalah variable acak yang diperoleh dihubungkan dengan nilai dari variable respon. Assumsi untuk pendekatan kedua lebih ketat tapi tapi menghasilkan predictor optimal diantara semua pilihan daripada melalui predictor linier yang jarang.

Rumus rumus yang berhubungan dengan regresi linier multivariat secara keseluruhan ádalah sebagai berikut :

Kasus Univariat

Terdapat satu variable respon Y untuk sejumlah data n

maka

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r} \\ 1 & z_{21} & \cdots & z_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \cdots & z_{1r} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

model persamaannya $\underset{(n\times1)}{Y} = \underset{(n\times(r+1)}{Z} \underset{((r+1)\times1)}{\beta} + \underset{(n\times1)}{\varepsilon}$

dengan metode kuadrat terkecil

penaksir : $\hat{\beta} = (Z'Z)^{-1}Z'y$

koefisien determinasi : $R^2 = \dfrac{\sum_{j=1}^{n}(\hat{y}_j - \overline{y})^2}{\sum_{j=1}^{n}(y_j - \overline{y})^2}$

interval kepercayaan : $\hat{\beta}_i \pm t_{n-r-1}\left(\dfrac{\alpha}{2}\right)\sqrt{\hat{V}ar(\hat{\beta}_i)}$

Test Hipotesis

$H_0 : \beta_i = 0 \qquad (\beta_1, \beta_2, \ldots, \beta_r)$

$H_1 : \beta_i \neq 0$

Statistik uji $F = \dfrac{SSR/r}{SSE/(n-r-1)}$

Dengan $SSR = \left(\hat{\beta}' Z' y - n\bar{y}^2\right)$ $SSE = \left(y' y - \hat{\beta}' Z' y\right)$

Kriteria tolak H_0 jika $F > F_{\alpha,r,n-r-1}$

(Rencerd;330)

Kasus Multivariat

Misalkan untuk variable respon sebanyak 2 atau terdapat Y_1 dan Y_2 dan 3 variabel predictor maka

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & z_{13} \\ 1 & z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & z_{n2} & z_{n3} \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \\ \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} \end{bmatrix}$$

jika teradapat m variable respon Y dan r variable predictor z, maka terdapat sejumlah persamaan model regresi :

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11} z_1 + ... + \beta_{r1} z_r + \varepsilon_1 \\ Y_2 &= \beta_{02} + \beta_{12} z_1 + ... + \beta_{r2} z_r + \varepsilon_2 \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ Y_m &= \beta_{0m} + \beta_{1m} z_1 + ... + \beta_{m1} z_r + \varepsilon_m \end{aligned}$$

dengan $\varepsilon = \left[\varepsilon_1, \varepsilon_2, ..., \varepsilon_m\right]'$ mempunyai $E(\varepsilon) = 0,$ $Var(\varepsilon) = \Sigma$

model Regresi Linear Multivariatnya adalah

$$\underset{(n\times m)}{Y} = \underset{(n\times(r+1))}{Z} \underset{((r+1)\times m)}{\beta} + \underset{(n\times m)}{\varepsilon}$$

dengan $E\left(\varepsilon_{(i)}\right) = 0,$ $Cov\left(\varepsilon_{(i)}, \varepsilon_{(k)}\right) = \sigma_{ik} I$ $i, k = 1, 2, ...\ m$

dengan menggunakan penaksiran kuadrat terkecil

penaksir : $\hat{\beta} = (Z'Z)^{-1} Z'Y$ dengan $\hat{\beta} = \left[\hat{\beta}_{(1)} \middle| \hat{\beta}_{(2)} \middle| \cdots \middle| \hat{\beta}_{(m)}\right]$

dan $\hat{Y} = Z\hat{\beta} = Z(Z'Z)^{-1} Z'Y$ dengan $\hat{Y} = \left[\hat{Y}_{(1)} \middle| \hat{Y}_{(2)} \middle| \cdots \middle| \hat{Y}_{(m)}\right]$

residualnya adalah $\varepsilon = Y - \hat{Y}$

dengan matrik kovariannya $\hat{\Sigma} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \dfrac{(Y - Z\hat{\beta})'(Y - Z\hat{\beta})}{n}$

interval kepercayaan :

$100(1-\alpha)\%$ confidence ellipsoid untuk $\beta' z_0$ adalah

$$\left(\beta' z_0 - \hat{\beta}' z_0\right)' \left(\frac{n\hat{\Sigma}}{n-r-1}\right)^{-1} \left(\beta' z_0 - \hat{\beta}' z_0\right) \le z_0' \left(Z'Z\right)^{-1} z_0 \left[\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m(\alpha)}\right]$$

$100(1-\alpha)\%$ interval kepercayaan simultan untuk $E(Y_{(i)}) = z_0' \beta_{(i)}$ adalah

$$z_0'\hat{\beta}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m(\alpha)}} \sqrt{z_0'(Z'Z)^{-1} z_0 \left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right)} \quad i = 1, 2, \ldots, m$$

Test Hipotesis

$H_0 :\ \beta_i = 0 \qquad (\beta_1, \beta_2, \ldots, \beta_r)$

$H_1 :\ \beta_i \neq 0$

statistik uji $\Lambda = \dfrac{|E|}{|E+H|}$ dengan $E = Y'Y - \hat{\beta}' Z'Y \qquad H = \hat{\beta}' Z'Y - n\bar{y}\,\bar{y}'$

kriteria Tolak H_0 jika $\Lambda \le \Lambda_{\alpha,m,r,n-r-1}$ dimana m menunjukan banyaknya variable Y, r menunjukan banyaknya variable Z.

Dalam tabel Wilks Lambda m menyatakan p, r menyatakan V_H dan n-r-1 menyatakan V_E

(Rencerd;344)

Konsep Regresi Linier

Untuk Kasus Univariat

Misalkan terdapat $Y,\ Z_1,\ Z_2,\ \ldots, Z_r$ dengan $\mu = \begin{bmatrix} \underset{(1\times 1)}{\mu_Y} \\ \hline \underset{(r\times 1)}{\mu_Z} \end{bmatrix}$ dan

$$\Sigma = \left[\begin{array}{c|c} \underset{(1\times 1)}{\sigma_{YY}} & \underset{(1\times r)}{\sigma_{ZY}'} \\ \hline \underset{(r\times 1)}{\sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right] \qquad \text{dimana } \sigma_{ZY} = \left[\sigma_{YZ_1}, \sigma_{YZ_2}, \ldots, \sigma_{YZ_r}\right]'$$

prediktor liniernya adalah $\beta_0 + \beta' Z$ dengan koefisien $\beta = \Sigma_{ZZ}^{-1} \sigma_{ZY,} \quad \beta_0 = \mu_Y - \beta' \mu_Z$

memiliki mean square error yaitu

$$E(Y - \beta_0 - \beta' Z)^2 = E(Y - \mu_Y - \sigma'_{ZY} \Sigma_{ZZ}^{-1} (Z - \mu_Z))^2 = \sigma_{YY} - \sigma'_{ZY} \Sigma_{ZZ}^{-1} \sigma_{ZY}$$

korelasi antara variabel terikat Y dengan prediktor linier terbaiknya disebut koeffisien korelasi multiple populasi yang dinotasikan sebagai

$$\rho_{Y(Z)} = + \sqrt{\frac{\sigma'_{ZY} \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}}$$

kuadrat dari koeffisien ini $\rho_{Y(Z)}^2$ disebut koeffisien determinasi populasi, nilai dari koeffisien korelasi adalah akar kuadrat positif nya yaitu $0 \le \rho_{Y(Z)} \le 1$.

Untuk Kasus Multivariat

Misalkan teradapat $Y_1, Y_2, \dots, Y_m \ , Z_1, Z_2, \dots, Z_r$ berdistribusi $N_{m+r}(\mu, \Sigma)$ dengan $\mu = \begin{bmatrix} \underset{(m\times 1)}{\mu_Y} \\ \hline \underset{(r\times 1)}{\mu_Z} \end{bmatrix}$ dan $\Sigma = \left[\begin{array}{c|c} \underset{(m\times m)}{\Sigma_{YY}} & \underset{(m\times r)}{\Sigma_{YZ}} \\ \hline \underset{(r\times m)}{\Sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right]$

regresi dari vektor Y dalam Z adalah

$$\beta_0 + \beta z = \mu_Y - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \mu_Z + \Sigma_{YZ} \Sigma_{ZZ}^{-1} z = \mu_Y + \Sigma_{YZ} \Sigma_{ZZ}^{-1} (z - \mu_Z)$$

kuadrat harapan dan matriks cross produk untuk errornya adalah

$$E(Y - \beta_0 - \beta Z)(Y - \beta_0 - \beta Z)' = \Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZY}$$

berdasarkan sampel acak ukuran n, estimator maximum likelihood untuk fungsi regresinya adalah

$$\hat{\beta}_0 + \hat{\beta} z = \overline{Y} + S_{YZ} S_{ZZ}^{-1} (z - \overline{Z})$$

dan estimator likelihood dari $\Sigma_{YY\cdot Z}$ adalah

$$\hat{\Sigma}_{YY\cdot Z} = \left(\frac{n-1}{n} \right) (s_{YY} - S_{YZ} S_{ZZ}^{-1} S_{ZY})$$

Koeffisien Korelasi Parsial

Anggaplah pasangan kesalahan $\begin{matrix} Y_1 - \mu_{Y_1} - \Sigma_{Y_1 Z} \Sigma_{ZZ}^{-1} (Z - \mu_Z) \\ Y_2 - \mu_{Y_2} - \Sigma_{Y_2 Z} \Sigma_{ZZ}^{-1} (Z - \mu_Z) \end{matrix}$

diperoleh dari menggunakan prediktor linier terbaik Y_1 dan Y_2 hubungannya ditentukan dari matrik kovarian kesalahan $\Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}$

koeffisien korelasi parsial sampel adalah

$$r_{Y_1Y_2\cdot Z} = \frac{s_{Y_1Y_2\cdot Z}}{\sqrt{s_{Y_1Y_1\cdot Z}}\sqrt{s_{Y_2Y_2\cdot Z}}}$$

Contoh :

Diberikan $\begin{aligned} z_1 &= 0,1,2,3,4 \\ y_1 &= 1,4,3,8,9 \\ y_2 &= -1,-1,2,3,2 \end{aligned}$ tentukan model persamaan regresi multivariatnya

Penyelesaian:

Akan ditentukan $\begin{aligned} Y_{j1} &= \beta_{01} + \beta_{11} z_{j1} + \varepsilon_{j1} \\ Y_{j2} &= \beta_{02} + \beta_{12} z_{j1} + \varepsilon_{j2} \end{aligned}$

Dari persoalan diatas maka dinyatakan dalam bentuk matriksnya adalah

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad Y_1 = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} \quad Y_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \\ 3 \\ 2 \end{bmatrix}$$

Selanjutnya cari $(Z'Z) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$

diperoleh
$$\begin{aligned} (Z'Z)^{-1} &= \frac{1}{150-100}\begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} \\ &= \frac{1}{50}\begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} \\ &= \begin{bmatrix} 0,6 & -0,2 \\ -0,2 & 0,1 \end{bmatrix} \end{aligned}$$

Selanjutnya akan ditentukan $\hat{\beta}_{(1)} = \left(Z'Z\right)^{-1} Z'Y_{(1)}$ dan $\hat{\beta}_{(2)} = \left(Z'Z\right)^{-1} Z'Y_{(2)}$

$$Z'Y_{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \end{bmatrix} \qquad Z'Y_{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 20 \end{bmatrix}$$

$$\hat{\beta}_{(1)} = \left(Z'Z\right)^{-1} Z'Y_{(1)} \qquad \hat{\beta}_{(2)} = \left(Z'Z\right)^{-1} Z'Y_{(2)}$$

$$= \begin{bmatrix} 0,6 & -0,2 \\ -0,2 & 0,1 \end{bmatrix} \begin{bmatrix} 25 \\ 70 \end{bmatrix} \qquad = \begin{bmatrix} 0,6 & -0,2 \\ -0,2 & 0,1 \end{bmatrix} \begin{bmatrix} 5 \\ 20 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Sehingga diperoleh $\hat{Y}_1 = 1 + 2z_1 \qquad \hat{Y}_2 = -1 + z_1$

Jadi matriks $\hat{\beta} = \left[\hat{\beta}_{(1)} \middle| \hat{\beta}_{(2)} \right] = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$

$$\hat{Y} = Z\hat{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix} \qquad \varepsilon = Y - \hat{Y} = \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix}$$

Penaksiran parameter

Hipotesis $\begin{array}{l} \beta_1 = 0 \\ \beta_1 \neq 0 \end{array}$

Statistik uji $\Lambda = \dfrac{|E|}{|E+H|}$ dengan $E = Y'Y - \hat{\beta}'Z'Y \qquad H = \hat{\beta}'Z'Y - n\bar{y}\,\bar{y}'$

$$Y'Y = \begin{bmatrix} 1 & 4 & 3 & 8 & 9 \\ -1 & -1 & 2 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix} = \begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix}$$

$$\hat{\beta}^{'} X^{'} Y = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & -1 \\ 3 & 2 \\ 8 & 3 \\ 9 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 25 & 5 \\ 70 & 20 \end{bmatrix}$$

$$= \begin{bmatrix} 165 & 45 \\ 45 & 15 \end{bmatrix}$$

$$n\bar{y}\,\bar{y}^{'} = 5\begin{bmatrix} 5 \\ 1 \end{bmatrix} \begin{bmatrix} 5 & 1 \end{bmatrix}$$

$$= 5\begin{bmatrix} 25 & 5 \\ 5 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 125 & 25 \\ 25 & 5 \end{bmatrix}$$

$$\Lambda = \frac{|E|}{|E+H|} = \frac{\left|Y^{'}Y - \hat{\beta}^{'} X^{'} Y\right|}{\left|Y^{'}Y - n\bar{y}\,\bar{y}^{'}\right|}$$

$$= \frac{\left|\begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix} - \begin{bmatrix} 165 & 45 \\ 45 & 15 \end{bmatrix}\right|}{\left|\begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix} - \begin{bmatrix} 125 & 25 \\ 25 & 5 \end{bmatrix}\right|}$$

$$= \frac{\begin{vmatrix} 6 & -2 \\ -2 & 4 \end{vmatrix}}{\begin{vmatrix} 46 & 18 \\ 18 & 14 \end{vmatrix}} = \frac{24-4}{644-324} = 0,0625$$

berdasarkan tabel Wilks lambda diperoleh

$\Lambda_{\alpha,m,r,n-r-1} = \Lambda_{0,05;2;1;3} = 0,050$ (Tabel A.9 Wilks Lambda;567)

kriteria Tolak H_0 jika $\Lambda \leq \Lambda_{\alpha,m,r,n-r-1}$ karena $\Lambda > \Lambda_{\alpha,m,r,n-r-1}$ yaitu 0,0625>0,050

kesimpulannya H_0 diterima, jadi koeffisien β_1 tidak berarti pada kedua persamaan diatas.

Nama : Siti Habsah

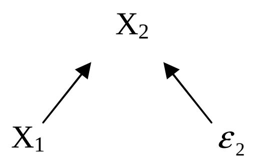
NIM : 055662

2.10 Analisis Jalur

Metode analisis jalur dikembangkan oleh ahli genetika Sewel Wright pada 1918-1921 untuk menjelaskan hubungan sebab akibat dalam genetika populasi. Aplikasi analisis jalurnya pada 1925 untuk mengawetkan dan memonopoli harga-harga turut memprakarsai penggunaan persamaan struktural dalam ekonomi. Tujuan análisis jalur (atau análisis persamaan struktural) untuk menyediakan penjelasan yang logis dari korelasi yang diobservasi dengan mengkonstruksi model hubungan sebab dan akibat antara variabel-variabel.

Koefisien korelasi signifikan yang tidak menunjukkan hubungan sebab akibat telah ditegaskan berkali-kali pada diskusi korelasi, seringkali dengan contoh menggelikan seperti asosiasi positif diantara penjualan permen karet dan dan angka kriminalitas. Tentu saja sebuah korelasi yang diobservasi tidak pernah bisa digunakan sebagai bukti hubungan sebab akibat. Argumen meyakinkan untuk sebab akibat dapat dikonstruksi dari inferensi statistik bersama dalil yang menyatakan hubungan yang dikembangkan dari ilmu pengetahuan dari subjek masalah dan pengertian yang berhubungan. Misalnya teori klasik tentang sifat-sifat harga, kenaikan harga jagung menaikkan permintaan dan menurunkan suplai. Dalam hal ini variabel permintaan dan suplai diperlakukan sebagai penyebab perubahan harga jagung.

Ketika satu variabel X_1 mendahului variabel lain pada suatu waktu, dapat disimpulakan X_1 menyebabkan X_2. Secara diagram kita dapat menulis $X_1 \rightarrow X_2$. Dengan mengikutsertakan error ϵ dalam hubungan, diagram jalurnya adalah



Dalam hubungan model linier, $X_2 = \beta_0 + \beta_1 X_1 + \varepsilon_2$ dimana sekarang X_1 adalah variabel penyebab yang tidak dipengaruhi oleh variabel lain. Gagasan hubungan sebab akibat antara X_1 dan X_2 mengharuskan semua faktor penyebab lain yang mungkin, dikesampingkan. Secara statistik, kita menetapkan bahwa X_1 dan ε_2 tidak berkorelasi, dimana ε_2 menunjukkan akibat bersama dari semua variabel tidak terukur yang dapat mempengaruhi X_1 dan X_2.

Lebih spesifik lagi, regresi $X_2 = \beta_0 + \beta_1 X_1 + \varepsilon_2$ ditulis dalam bentuk baku dengan notasi yang jelas

$$\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}} = \beta_1\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}\right) + \sqrt{\frac{\sigma_{\varepsilon\varepsilon}}{\sigma_{22}}}\frac{\varepsilon_2}{\sqrt{\sigma_{\varepsilon\varepsilon}}} \quad \text{atau}$$

$$Z_2 = p_{21}Z_1 + p_{2\varepsilon}\varepsilon \tag{7-71}$$

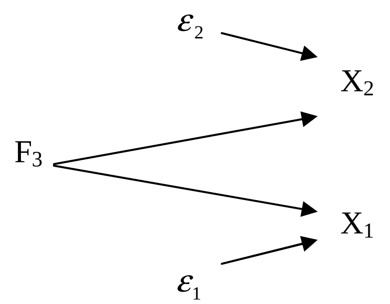
Walaupun error dalam bentuk baku, ε memiliki sebuah koefisien. Dalam model baku, parameter koefisien jalur biasa disebut p. Model dalam (7-71) mengakibatkan

$$\rho_{12} = Corr(X_2,X_1) = Corr(Z_2,Z_1) = p_{21}$$

$$1 = Var(Z_2) = p_{21}^2 Var(Z_1) + p_{2\varepsilon}^2 (Var\epsilon) = p_{21}^2 + p_{2\varepsilon}^2$$

Persamaan kedua menyatakan bahwa kesimpulan sementara diagram jalur itu sendiri lengkapnya ditentukan oleh variabel-variabel yang di tunjukkan karena konstribusi pada variansi Z_2 berjumlah satu.

Secara matematis, sama logisnya untuk merumuskan bahwa X_2 menyebabkan X_1 atau merumuskan model ketiga yang memuat sebuah faktor yang berhubungan, contohnys F_3 yang bertanggung jawab atas korelasi yang diobservasi antara X_1 dan X_2. Dalam kasus terakhir, korelasi antara X_1 dan X_2 adalah palsu dan bukan sebuah korelasi sebab akibat. Diagram jalurnya adalah



dimana kita memperhitungkan error lagi dalam hubungan. Dalam hubungan variabel-variabel baku, model linier yang diakibatkan oleh diagram jalur di atas menjadi

$$Z_1 = p_{13}F_3 + p_{1\epsilon_1}\epsilon_1$$

$$Z_2 = p_{23}F_3 + p_{2\epsilon_2}\epsilon_2 \qquad (7\text{-}72)$$

Dengan error baku ε_1 dan ε_2 tidak berkorelasi satu sama lain dengan F_3. Akibatnya, korelasi dihubungkan dengan koefisien jalur oleh

$$\rho_{13} = Corr(X_1, X_2) = Corr(Z_1, Z_2) = p_{13}p_{23}$$

$$1 = Var(Z_1) = p_{13}^2 + p_{1\epsilon_1}^2$$

$$1 = Var(Z_2) = p_{23}^2 + p_{2\epsilon_2}^2$$

dan

$$\rho_{13} = Corr(Z_1, F_3) = p_{13} \qquad \rho_{23} = Corr(Z_2, F_3) = p_{23}$$

Model sebab akibat yang dirumuskan dalam (7-72) berbeda dari model dalam (7-71) maka tidak mengejutkan bahwa hubungan antara korelasi dan koefisien jalur berbeda.

Analisis jalur berisi dua komponen utama: (1) diagram jalur, dan (2) dekomposisi korelasi yang diobservasi ke sejumlah hubungan koefisien jalur yang mewakili jalur-jalur sederhana dan gabungan.

2.10.1 Pengkonstruksian Diagram Jalur

Sebuah perbedaan dibuat diantara variabel-variabel yang tidak dipengaruhi oleh variabel-variabel lain dalam sistem (variabel eksogen) dan variabel-variabel yang dipengaruhi oleh variabel-variabel lain (variabel endogen). Dengan masing-masing variabel-variabel terikat terakhir dihubungkan sebuah residual. Aturan tertentu menentukan penggambaran sebuah diagram jalur. Tanda panah menunjukkan sebuah jalur. Diagram jalur dikonstruksi sebagai berikut:

1. Tanda panah lurus menunjukkan hubungan sebab antara variabel-variabel *exogenous* atau perantara dengan satu variabel terikat atau lebih

2. Tanda panah lurus juga menghubungkan kesalahan *(variabel residue)* dengan semua variabel *endogenous* masing-masing

3. Tanda panah kurva dengan ujung panah ganda digambar diantara masing-masing pasangan variabel bebas (endogen) yang memiliki korelasi tidak nol.

Tanda panah kurva untuk korelasi mengindikasikan koefisien korelasi alami simetris. Hubungan-hubungan lain yang langsung, seperti diindikasikan oleh tanda panah dengan ujung tunggal.

Ketika mengkonstruksi diagram jalur, biasanya menggunakan variabel-variabel yang telah baku yang memiliki rata-rata 0 dan variansi 1. Dalam konteks regresi berganda, modelnya adalah

$$\frac{Y-\mu_Y}{\sqrt{\sigma_{YY}}} = \beta_1 \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{YY}}}\left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}\right) + \beta_2 \frac{\sqrt{\sigma_{22}}}{\sqrt{\sigma_{YY}}}\left(\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}}\right) + \cdots + \beta_r \frac{\sqrt{\sigma_{rr}}}{\sqrt{\sigma_{YY}}}\left(\frac{X_r-\mu_r}{\sqrt{\sigma_{rr}}}\right) + \beta_1 \frac{\sqrt{\sigma_{\epsilon\epsilon}}}{\sqrt{\sigma_{YY}}}\left(\frac{\epsilon}{\sqrt{\sigma_{\epsilon\epsilon}}}\right)$$

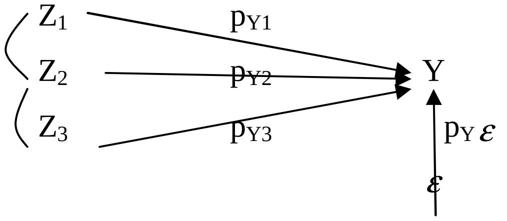
atau

$$Y_s = p_{Y1}Z_1 + p_{Y2}Z_2 + \cdots + p_{Yr}Z_r + p_{Y\epsilon}\epsilon_s \qquad (7\text{-}73)$$

dimana koefisien jalur, $p_{\gamma k} = \beta_k \sqrt{\sigma_{kk}} \big/ \sqrt{\sigma_{\gamma\gamma}}$ adalah koefisien regresi untuk prediktor baku dan $p_{\gamma\varepsilon} = \sqrt{\sigma_{\varepsilon\varepsilon}} \big/ \sqrt{\sigma_{\gamma\gamma}}$.

Untuk menilustrasikan pengkostruksian diagram jalur, pertama kita gambar diagram yang menjelaskan regresi berganda dengan variabel prediktor r = 3.

Ketika masing-masing Z_k diperlakukan sebagai variabel penyebab, korelasi antara pasangan variabel-variabel eksogen ditunjukkan oleh tanda panah berbentuk kurva dengan ujung ganda. Tanda panah lurus berangkat dari masing-masing variabel penyebab ke Y. Error ε dan masing-masing Z_k (diasumsikan) tidak berkorelasi sehingga tidak ada tanda panah yang menghubungkan variabel-variabel ini. Diagram jalur untuk variabel prediktor r = 3 diberikan dalam gambar 7.6

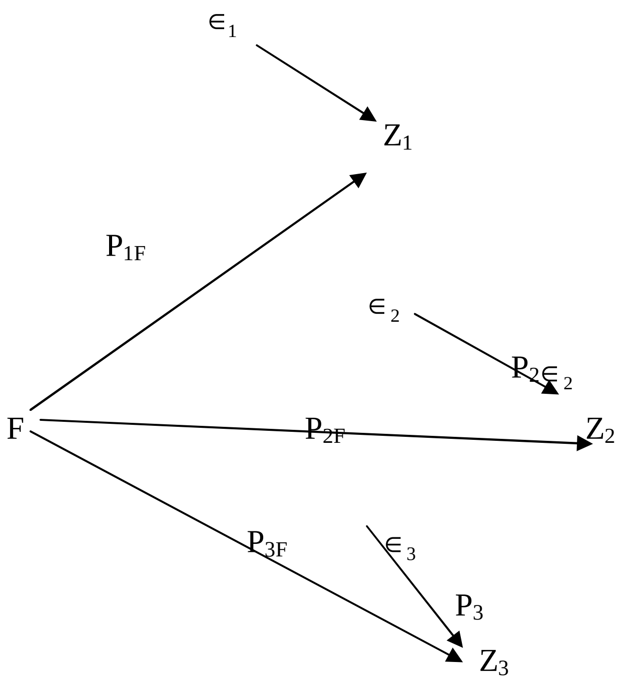


Gambar 7.6

Kesederhanaan lain, masih menarik, kondisi model analisis faktor dengan satu faktor biasa yang tidak diobservasi. Menurut model ini, faktor tunggal tidak diobservasi, F, bertanggung jawab atas korelasi antara variabel respon, model dapat ditulis dalam hubungan variabel-variabel baku F, ε_1, ε_2, ε_3, dan Z_1, Z_2, Z_3 sebagai

$$Z_1 = p_{1F} F + p_{1\epsilon_1} \epsilon_1$$
$$Z_2 = p_{2F} F + p_{2\epsilon_2} \epsilon_2$$
$$Z_3 = p_3 F + p_{3\epsilon_3} \epsilon_3 \qquad (7\text{-}74)$$

dimana F, ε_1, ε_2, dan ε_3 semuanya tidak berkorelasi. Diagram jalur ditunjukkan dalam gambar 7.7 .



Gambar 7.7

Pengkonstruksian diagram jalur dapat membantu peneliti berpikir benar tentang sebuah masalah dan menggambarkan komponen-komponen penting korelasi yang diobservasi.

2.10.2 Dekomposisi Korelasi yang Diobservasi

Estimasi koefisien jalur akan memungkinkan kita menaksir pengaruh langsung dan tidak langsung dimana satu variabel memiliki pengaruh pada variable lain. Dari model linier yang menyatakan hubungan sebab, kita dapat menemukan pernyataan yang menghubungkan koefisien jalur dan korelasi.

Contoh 7.16 (Analisis Jalur dari Model Regresi)

Dari bentuk baku model regresi berganda ([lihat (7-73)], korelasi antara Y dan masing-masing Z_k dapat di dekomposisi sebagai berikut

$$\rho_{Yk} = Corr(Y, Z_k) = Cov\left(\sum_{i=1}^{r} p_{Yi} Z_i , Z_k \right) = \sum_{i=1}^{r} p_{Yi} \rho_{ik} , \quad k = 1, 2, ..., r \qquad (7\text{-}75)$$

Juga, ketika diagram jalur memuat dirinya sendiri sehingga Y ditentukan oleh variabel-variabel dalam diagram, kita menemukan persamaan determinasi lengkap.

$$1 = Var(Y) = Var\left(\sum_{i=1}^{r} p_{Yi} Z_i + p_{Y\varepsilon}\varepsilon \right) = \sum_{i=1}^{r} \sum_{k=1}^{r} p_{Yi} \rho_{ik} p_{Yk} + p_{Y\varepsilon}^2$$

$$= \sum_{i=1}^{r} p_{Yi}^2 + 2 \sum_{i=1}^{r} \sum_{k=i+1}^{r} p_{Yi} \rho_{ik} p_{Yk} + p_{Y\varepsilon}^2 \qquad (7\text{-}76)$$

Variansi total Y	=	Proporsi variansi yg langsung diberikan oleh koefisien jalur	+	Proporsi variansi yg disebabkan interkorelasi antara variabel terikat	+	Proporsi variansi disebabkan error

Keadaan $\rho_{ZY} = [\rho_{Y1}, \rho_{Y2}, ..., \rho_{Yr}]^T$, matriks r x r $\rho_{ZZ} = \{\rho_{ik}\}$ dan $p_Y = [p_{Y1}, p_{Y2}, ..., p_{Yr}]^T$. Persamaan (7-75) dapat ditulis dalam notasi matriks sebagai $\rho_{ZY} = \rho_{ZZ} p_Y$, sehingga

$$p_Y = \rho_{ZZ}^{-1} \rho_{ZY}$$

Selain itu, error $p_{Y\varepsilon}\varepsilon$ dalam (7-73) memiliki variansi $p_{Y\varepsilon}^2 Var(\varepsilon) = p_{Y\varepsilon}^2$, yang berasal dari(7-76) menjadi

$$p_{Y\varepsilon}^2 = 1 - \rho'_{ZY} \rho_{ZZ}^{-1} \rho_{ZY} = 1 - \rho'_{ZY} p_Y$$

Kuadrat koefisien jalur $p_{Y\varepsilon}^2$ dihubungkan pada koefisien korelasi berganda karena

$$p_{Y\varepsilon}^2 = \frac{\left(1 - \rho'_{ZY}\ \rho_{ZZ}^{-1} \rho_{ZY}\right)}{1} = 1 - \rho_{Y(Z)}^2$$

Untuk data komputer contoh 7.6, kita mengajukan diagram jalur berikut berdasarkan dugaan hubungan sebab akibat antara Z_1, Z_2, dan Y:

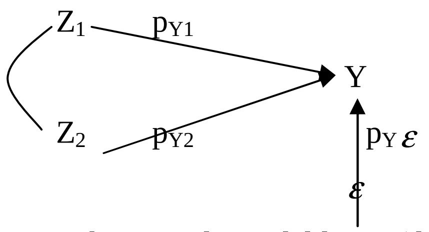


Diagram ini membawa pada model linier (dalam bentuk variabel-variabel baku)

$$Y = p_{Yi} Z_1 + p_{Y2} Z_2 + p_{Y\varepsilon} \varepsilon$$

Akibatnya, persamaan (7-75) dan (7-76) menjadi

$$\rho_{Y1} = p_{Y1}(1) + p_{Y2}\rho_{12}$$

$$\rho_{Y2} = p_{Y1}\rho_{12} + p_{Y2}(1)$$

dan

$$1 = Var(Y) = p_{Y1}^2 + p_{Y2}^2 + p_{Y\varepsilon}^2 + 2 p_{Y1}\rho_{12} p_{Y2}$$

sustitusi korelasi korelasi-korelasi contoh (lihat contoh 7.12 untuk S) $r_{Y1} = r_{YZ_1} = .997$, $r_{Y2} = r_{YZ_2} = .450$, dan $r_{12} = r_{Z_1 Z_2} = .391$ untuk banyaknya populasi yang berkorespondensi dia atas, kita dapat mengestimasi koefisien jalur p_{Y1} dan p_{Y2} dengan menyelesaikan

$$.997 = p_{Y1} + .391 p_{Y2}$$

$$.450 = .391 p_{Y1} + p_{Y2}$$

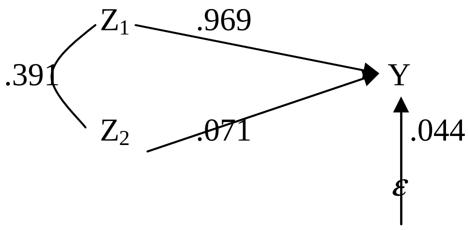
Secara ekivalen, kita dapat menggunakan

$$\hat{p}_Y = \begin{bmatrix} \hat{p}_{Y1} \\ \hat{p}_{Y2} \end{bmatrix} = \hat{\rho}_{ZZ}^{-1} \hat{\rho}_{ZY} = \begin{bmatrix} 1 & .391 \\ .391 & 1 \end{bmatrix}^{-1} \begin{bmatrix} .997 \\ .450 \end{bmatrix} = \begin{bmatrix} .969 \\ .071 \end{bmatrix}$$

Akhirnya

$$\hat{p}_{Y\varepsilon} = 1 - \hat{\rho}'_{ZY}\ \hat{p}_Y = 1 - \begin{bmatrix} .997 & .450 \end{bmatrix} \begin{bmatrix} .969 \\ .071 \end{bmatrix} = .002$$

Dengan demikian korelasi yang diobservasiantara respon Y = CPU time dan variabel prediktor Z_1 = permintaan dan Z_2 = penambahan-penghapusan item dapat didekmposisi ke dalam bagian-bagian yang mewakili pengaruh langsung dan tidak langsung. Contohnya, Z_1 secara langsung mempengaruhi Y (diwakili oleh koefisien jalur $\hat{p}_{Y1}$ dan juga mempengaruhi Y secara tidak langsung melalui Z_2 (ditunjukkan oleh hubungan produk $\hat{\rho}_{12}\hat{p}_{Y2}$. Dengan mensubstitusi bilangan-bilangan pada diagram jalur, kita punya



Tepat menggunakan sebuah tabel untuk menunjukkan pengaruh dekomposisi variabel-variabel prediktor pada respon.

	Indirect effect	Direct effect	Total effect
Z_1 (orders)	.028	.969	.997
Z_2 (add-del items)	.379	.071	.450

Perhatikan bahwa koefisien jalur mengukur pengaruh langsung Z_k pada Y adalah koefisien regresi untuk variabel-variabel baku.

Contoh 7.17 (Analisis Jalur dari Model Analisis Faktor dengan Satu Faktor Biasa)

Model faktor tunggal dalam (7-74) untuk 3 variabel respon menghasilkan hubungan untuk dekomposisi korelasi yang diobservasi.

$$\rho_{ik} = Corr(Z_i, Z_k) = Cov\left(p_{iF}F + p_{i\varepsilon_i}\varepsilon_i, p_{kF}F + p_{k\varepsilon_k}\varepsilon_k\right) = p_{iF}\, p_{kF}, \quad i \neq k = 1, 2, 3$$

dan persamaan determinasi lengkap

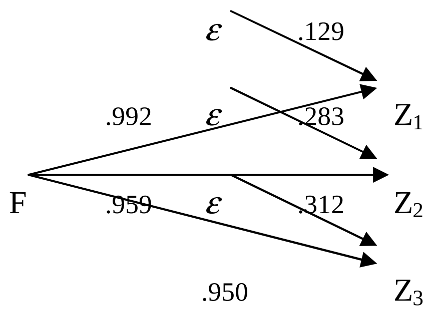
$$1 = Var(Z_k) = Var\left(p_{kF} + p_{k\varepsilon_k}\varepsilon_k\right) = p_{kF}^2 + p_{k\varepsilon_k}^2$$

Enam persamaan ini dengan mudah diselesaikan untukkoefisien jalur dalam bentuk korelasi yang diestimasi.

Contoh 8.4 memberikan matriks kovarian contoh S untuk dimensi tiga (of turtle shells), yang mana kita menentukan r_{12} = .951, r_{13} = .942, dan r_{14} = .911. dengan memngasumsikan faktor tunggal (prtumbuhan) menebabkan shell dimensions, kita bisa menulis

$$\left.\begin{aligned} .951 &= \hat{p}_{1F}\,\hat{p}_{2F} \\ .942 &= \hat{p}_{1F}\,\hat{p}_{3F} \\ .911 &= \hat{p}_{2F}\,\hat{p}_{3F} \end{aligned}\right\} \quad \text{jadi} \quad \frac{(.951)(.942)}{.911} = \frac{\hat{p}_{1F}\,\hat{p}_{2F}\,\hat{p}_{1F}\,\hat{p}_{3F}}{\hat{p}_{2F}\,\hat{p}_{3F}} = \hat{p}_{1F}^2$$

dan $\hat{p}_{1F} = .992$. Juga $\hat{p}_{1\varepsilon_1}^2 = 1 - \hat{p}_{1F}^2 .017$, dan $\hat{p}_{1\varepsilon_1} = .129$. Dengan cara yang sama, $\hat{p}_{2F} = \sqrt{(.951)(.911)/(.942)} = .959$, $\hat{p}_{2\varepsilon_2} = \sqrt{1-(.959)^2} = .283$, $\hat{p}_{3F} = .950$, dan $\hat{p}_{3\varepsilon_2} = .312$. Semua koefisien jalur untuk faktur biasa adalah besar dibandingkan koefisien jalur error. Ini menyatakan sebuah mekanisme sebab akibat kuat jika model sebab akibat ini tepat. Tambahan, koefisien jalur $\hat{p}_{kF}$ hampir sama, walaupun Z_1 = ln(length) dipengaruhi lebih sedikit oleh F. Diagram jalur dengan koefisien jalur yang diestimasi ditampilkan berikut.



Untuk menyimpulkan, analisis jalur mengambil teori-teori substansif untuk permintaan-permintaan sebab dan menggunakan diagram jalur untuk menemukan dekomposisi korelasi yang diobservasi terhadap pengaruh langsung dan tidak langsung. Koefisien-koefisien jalur membantu menentukan pentingnya pengaruh-pengaruh langsung dan tidak langsung. Kesimpulan analisis jalur akan bergantung hubungan sebab akibat yang diasumsikan

BAB III

KESIMPULAN

MODEL REGRESI LINIER MULTIVARIAT

Kasus Univariat

Terdapat satu variable respon Y untuk sejumlah data n maka

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r} \\ 1 & z_{21} & \cdots & z_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \cdots & z_{1r} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

model persamaannya $\underset{(n\times 1)}{Y} = \underset{(n\times(r+1))}{Z} \underset{((r+1)\times 1)}{\beta} + \underset{(n\times 1)}{\varepsilon}$

dengan metode kuadrat terkecil

penaksir : $\hat{\beta} = (Z'Z)^{-1}Z'y$

koefisien determinasi : $R^2 = \dfrac{\sum_{j=1}^{n} (\hat{y}_j - \overline{y})^2}{\sum_{j=1}^{n} (y_j - \overline{y})^2}$

interval kepercayaan : $\hat{\beta}_i \pm t_{n-r-1}\left(\dfrac{\alpha}{2}\right)\sqrt{\hat{V}ar(\hat{\beta}_i)}$

Test Hipotesis

$H_0 : \beta_i = 0 \qquad (\beta_1, \beta_2, \ldots, \beta_r)$

$H_1 : \beta_i \neq 0$

Statistik uji $\quad F = \dfrac{SSR/r}{SSE/(n-r-1)}$

Dengan $\quad SSR = \left(\hat{\beta}'Z'y - n\overline{y}^2\right) \quad SSE = \left(y'y - \hat{\beta}'Z'y\right)$

Kriteria tolak H_0 jika $F > F_{\alpha,r,n-r-1}$

(Rencerd;330)

Kasus Multivariat

Misalkan untuk variable respon sebanyak 2 atau terdapat Y_1 dan Y_2 dan 3 variabel predictor maka

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & z_{13} \\ 1 & z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & z_{n2} & z_{n3} \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \\ \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} \end{bmatrix}$$

jika teradapat m variable respon Y dan r variable predictor z, maka terdapat sejumlah persamaan model regresi :

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11} z_1 + \ldots + \beta_{r1} z_r + \varepsilon_1 \\ Y_2 &= \beta_{02} + \beta_{12} z_1 + \ldots + \beta_{r2} z_r + \varepsilon_2 \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ Y_m &= \beta_{0m} + \beta_{1m} z_1 + \ldots + \beta_{m1} z_r + \varepsilon_m \end{aligned}$$

dengan $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m]'$ mempunyai $E(\varepsilon) = 0, \qquad Var(\varepsilon) = \Sigma$

model Regresi Linear Multivariatnya adalah

$$\underset{(n\times m)}{Y} = \underset{(n\times(r+1))}{Z} \underset{((r+1)\times m)}{\beta} + \underset{(n\times m)}{\varepsilon}$$

dengan $E(\varepsilon_{(i)}) = 0, \qquad Cov(\varepsilon_{(i)}, \varepsilon_{(k)}) = \sigma_{ik} I \quad i,k = 1,2, \ldots m$

dengan menggunakan penaksiran kuadrat terkecil

penaksir : $\hat{\beta} = (Z'Z)^{-1} Z'Y$ dengan $\hat{\beta} = \left[\hat{\beta}_{(1)} \middle| \hat{\beta}_{(2)} \middle| \cdots \middle| \hat{\beta}_{(m)} \right]$

dan $\hat{Y} = Z\hat{\beta} = Z(Z'Z)^{-1} Z'Y$ dengan $\hat{Y} = \left[\hat{Y}_{(1)} \middle| \hat{Y}_{(2)} \middle| \cdots \middle| \hat{Y}_{(m)} \right]$

residualnya adalah $\varepsilon = Y - \hat{Y}$

dengan matrik kovariannya $\hat{\Sigma} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \dfrac{(Y - Z\hat{\beta})'(Y - Z\hat{\beta})}{n}$

interval kepercayaan :

$100(1-\alpha)\%$ confidence ellipsoid untuk $\beta' z_0$ adalah

$$\left(\beta' z_0 - \hat{\beta}' z_0\right)' \left(\frac{n\hat{\Sigma}}{n-r-1}\right)^{-1} \left(\beta' z_0 - \hat{\beta}' z_0\right) \le z_0' \left(Z'Z\right)^{-1} z_0 \left[\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m(\alpha)}\right]$$

$100(1-\alpha)\%$ interval kepercayaan simultan untuk $E(Y_{(i)}) = z_0^{'}\beta_{(i)}$ adalah

$$z_0^{'}\hat{\beta}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m(\alpha)}}\sqrt{z_0^{'}(Z^{'}Z)^{-1}z_0\left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right)} \quad i=1,2,\ldots,m$$

Test Hipotesis

$H_0:\ \beta_i = 0 \qquad (\beta_1, \beta_2, \ldots, \beta_r)$

$H_1:\ \beta_i \neq 0$

statistik uji $\Lambda = \dfrac{|E|}{|E+H|}$ dengan $E = Y^{'}Y - \hat{\beta}^{'}Z^{'}Y \qquad H = \hat{\beta}^{'}Z^{'}Y - n\bar{y}\,\bar{y}^{'}$

kriteria Tolak H_0 jika $\Lambda \le \Lambda_{\alpha,m,r,n-r-1}$ dimana m menunjukan banyaknya variable Y, r menunjukan banyaknya variable Z.

Dalam tabel Wilks Lambda m menyatakan p, r menyatakan V_H dan n-r-1 menyatakan V_E

(Rencerd;344)

Konsep Regresi Linier

Untuk Kasus Univariat

Misalkan terdapat $Y,\ Z_1,\ Z_2,\ \ldots, Z_r$ dengan $\mu = \begin{bmatrix} \underset{(1\times 1)}{\mu_Y} \\ \hline \underset{(r\times 1)}{\mu_Z} \end{bmatrix}$ dan

$$\Sigma = \left[\begin{array}{c|c} \underset{(1\times1)}{\sigma_{YY}} & \underset{(1\times r)}{\sigma_{ZY}^{'}} \\ \hline \underset{(r\times1)}{\sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right] \qquad \text{dimana } \sigma_{ZY} = \left[\ \sigma_{YZ_1},\ \sigma_{YZ_2},\ \ldots, \sigma_{YZ_r}\ \right]^{'}$$

prediktor liniernya adalah $\beta_0 + \beta^{'}Z$ dengan koefisien $\beta = \Sigma_{ZZ}^{-1}\,\sigma_{ZY},\quad \beta_0 = \mu_Y - \beta^{'}\mu_Z$

memiliki mean square error yaitu

$$E(Y-\beta_0-\beta^{'}Z)^2 = E(Y-\mu_Y-\sigma_{ZY}^{'}\,\Sigma_{ZZ}^{-1}(Z-\mu_Z))^2 = \sigma_{YY} - \sigma_{ZY}^{'}\,\Sigma_{ZZ}^{-1}\,\sigma_{ZY}$$

korelasi antara variabel terikat Y dengan prediktor linier terbaiknya disebut koeffisien korelasi multiple populasi yang dinotasikan sebagai

$$\rho_{Y(Z)} = +\sqrt{\frac{\sigma_{ZY}^{'}\,\Sigma_{ZZ}^{-1}\,\sigma_{ZY}}{\sigma_{YY}}}$$

kuadrat dari koeffisien ini $\rho^2_{Y(Z)}$ disebut koeffisien determinasi populasi, nilai dari koeffisien korelasi adalah akar kuadrat positif nya yaitu $0 \le \rho_{Y(Z)} \le 1$.

Untuk Kasus Multivariat

Misalkan teradapat $Y_1, Y_2, \ldots, Y_m \;, Z_1, Z_2, \ldots, Z_r$ berdistribusi $N_{m+r}(\mu, \Sigma)$

dengan $\mu = \begin{bmatrix} \underset{(m\times 1)}{\mu_Y} \\ \hline \underset{(r\times 1)}{\mu_Z} \end{bmatrix}$ dan $\Sigma = \left[\begin{array}{c|c} \underset{(m\times m)}{\Sigma_{YY}} & \underset{(m\times r)}{\Sigma_{YZ}} \\ \hline \underset{(r\times m)}{\Sigma_{ZY}} & \underset{(r\times r)}{\Sigma_{ZZ}} \end{array}\right]$

regresi dari vektor Y dalam Z adalah

$$\beta_0 + \beta z = \mu_Y - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\mu_Z + \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,z = \mu_Y + \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}(z - \mu_Z)$$

kuadrat harapan dan matriks cross produk untuk errornya adalah

$$E(Y - \beta_0 - \beta Z)(Y - \beta_0 - \beta Z)' = \Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}$$

berdasarkan sampel acak ukuran n, estimator maximum likelihood untuk fungsi regresinya adalah

$$\hat{\beta}_0 + \hat{\beta} z = \overline{Y} + S_{YZ} S_{ZZ}^{-1}(z - \overline{Z})$$

dan estimator likelihood dari $\Sigma_{YY\cdot Z}$ adalah

$$\hat{\Sigma}_{YY\cdot Z} = \left(\frac{n-1}{n}\right)(s_{YY} - S_{YZ} S_{ZZ}^{-1} S_{ZY})$$

Koeffisien Korelasi Parsial

Anggaplah pasangan kesalahan $\begin{array}{l} Y_1 - \mu_{Y_1} - \Sigma_{Y_1 Z}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z) \\ Y_2 - \mu_{Y_2} - \Sigma_{Y_2 Z}\,\Sigma_{ZZ}^{-1}(Z - \mu_Z) \end{array}$

diperoleh dari menggunakan prediktor linier terbaik Y_1 dan Y_2 hubungannya ditentukan dari matrik kovarian kesalahan $\Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{ZY}$

koeffisien korelasi parsial sampel adalah

$$r_{Y_1Y_2\cdot Z} = \frac{s_{Y_1Y_2\cdot Z}}{\sqrt{s_{Y_1Y_1\cdot Z}}\,\sqrt{s_{Y_2Y_2\cdot Z}}}$$

Analisis jalur

Tujuan análisis jalur (atau análisis persamaan struktural) untuk menyediakan penjelasan yang logis dari korelasi yang diobservasi dengan mengkonstruksi model hubungan sebab dan akibat antara variabel-variabel.

Analisis jalur berisi dua komponen utama: (1) diagram jalur, dan (2) dekomposisi korelasi yang diobservasi ke sejumlah hubungan koefisien jalur yang mewakili jalur-jalur sederhana dan gabungan.

Korelasi antara Y dan masing-masing Z_k dapat di dekomposisi sebagai berikut

$$\rho_{Yk} = Corr(Y, Z_k) = Cov\left(r\sum_{i=1} p_{Yi} Z_i , Z_k \right) = \sum_{i=1}^{r} p_{Yi} \rho_{ik} , \quad k = 1, 2, ..., r$$

dan persamaan determinasi lengkap

$$1 = Var(Y) = Var\left(r\sum_{i=1} p_{Yi} Z_i + p_{Y\varepsilon}\varepsilon \right) = \sum_{i=1}^{r} \sum_{k=1}^{r} p_{Yi} \rho_{ik} p_{Yk} + p_{Y\varepsilon}^{2}$$

$$= \sum_{i=1}^{r} p_{Yi}^{2} + 2\sum_{i=1}^{r} \sum_{k=i+1}^{r} p_{Yi} \rho_{ik} p_{Yk} + p_{Y\varepsilon}^{2}$$

Dari kedua persamaan tersebut kita dapat menentukan besar koefisian jalur.