BAB III

MENYELESAIKAN MASALAH REGRESI INVERS DENGAN METODE GRAYBILL

3.1 Pengertian

Masalah regresi invers dengan bentuk linear dapat dijumpai dalam berbagai bidang kehidupan, diantaranya dalam bidang ekonomi, kesehatan, fisika, kimia dan masih banyak lagi. Sebagai ilustrasi, berikut ini merupakan beberapa contoh masalah regresi invers dengan bentuk linear :

Bidang ekonomi

Misalkan diketahui bahwa terdapat hubungan linear antara jumlah tenaga kerja pada suatu perusahaan sebagai variabel bebas X dengan jumlah produksi yang dihasilkan suatu perusahaan tersebut sebagai variabel terikat Y. Jika ingin diketahui jumlah tenaga kerja yang dibutuhkan agar tercapai jumlah produksi yang diinginkan oleh suatu perusahaan, maka ini merupakan masalah regresi invers.

Bidang Kesehatan

Misalkan diketahui bahwa terdapat hubungan linear antara jumlah kadar suatu obat penurun tekanan darah yang diberikan kepada seorang pasien dalam suatu selang waktu tertentu sebagai variabel bebas X dengan penurunan tekanan darah yang terjadi pada pasien tersebut sebagai variabel terikat Y. Jika ingin diketahui jumlah kadar obat yang harus diberikan kepada seorang pasien dalam selang waktu tertentu agar diperoleh penurunan tekanan darah yang diharapkan, maka ini merupakan masalah regresi invers.

Bidang Kehutanan

Misalkan diketahui bahwa terdapat hubungan linear antara nilai dugaan umur yang diperoleh dari perhitungan lingkaran umur pada sebuah pohon sebagai variabel bebas X dengan dugaan umur yang diperoleh melalui proses *carbon-dating* sebagai variabel terikat Y. Diketahui pula bahwa ternyata perhitungan dengan melihat lingkaran umur lebih akurat. Karena sesuatu hal, penyelidikan terhadap lingkaran umur tidak dapat dilakukan sehingga dipilih proses *carbon-dating*. Untuk mengetahui umur pohon yang lebih akurat maka kita harus mencari umur pohon berdasarkan lingkaran umur, maka ini merupakan masalah regresi invers.

Berdasarkan beberapa ilustrasi di atas, maka masalah regresi invers didefinisikan sebagai berikut :

Definisi 3.1 :

Misalkan diketahui bahwa X dan Y mempunyai hubungan linear sederhana (diasumsikan β_0 dan β diketahui)

$$Y = \mu(x) + \varepsilon = \beta_0 + \beta X + \varepsilon \tag{3.1}$$

Jika terdapat nilai dari X, katakanlah x_0 yang tidak diketahui dan nilai dari Y (atau sampel acak berukuran k untuk nilai dari Y) yang berkorespondensi dengan x_0 dapat diamati, maka masalah untuk menentukan nilai x_0 disebut sebagai masalah regresi invers dengan bentuk linear sederhana (Graybill, 1976).

Brown (dalam Thonnard, 2006) membagi masalah regresi invers ini kedalam dua kelompok, yaitu :

1. Masalah regresi invers alami atau acak, yaitu apabila nilai x yang ingin diketahui berasal dari sampel acak.
2. Masalah regresi invers terkontrol, yaitu apabila nilai x yang ingin diketahui tertentu dan galat hanya terjadi pada nilai y.

3.2 Penyelesaian Masalah Regresi Invers

Terdapat beberapa cara untuk menyelesaikan masalah regesi invers. Namun dalam tugas akhir ini hanya akan dibahas penyelesaian masalah regresi invers beserta interval kepercayaannya dengan menggunakan metode yang dikemukakan oleh Franklin A. Graybill (1976), dengan terlebih dulu membahas mengenai dua metode penyelesaian masalah regresi invers yang paling banyak digunakan, yaitu metode klasik dan metode invers.

3.2.1 Metode Klasik dan Metode Invers

Masalah regresi invers sering diselesaikan dengan dua cara yang secara umum berbeda, yaitu dengan metode klasik dan metode invers.

Metode Klasik

Misalkan diketahui model linear berbentuk :

$$Y = \beta_0 + \beta X + \varepsilon,\ \varepsilon \sim N\left(0, \sigma^2\right)$$

dengan β_0 dan β ditaksir oleh $\hat{\beta}_0$ dan $\hat{\beta}$ berdasarkan metode kuadrat terkecil atau metode kemungkinan maksimum.

Jika diberikan nilai dari variabel X, misalkan x_0 (tidak diketahui) dan dapat diamati nilai Y (atau sampel acak berukuran k untuk nilai dari Y) yang berkorespondensi dengan x_0, katakanlah y_0. Maka metode klasik menggunakan nilai y_0 (atau rata-rata dari k-buah nilai y_0) untuk menaksir x_0 dengan rumus

$$\hat{X}_0 = \frac{\overline{Y}_0 - \hat{\beta}_0}{\hat{\beta}} \tag{3.2}$$

Parameter-parameter yang terdapat pada penaksir klasik dari regresi invers ini merupakan parameter yang terdapat pada model regresi linear sederhananya, sehingga pengujiannya cukup dilakukan pada model regresi linear sederhananya saja.

Metode Invers

Misalkan diketahui model linear berbentuk :

$$Y = \beta_0 + \beta X + \varepsilon,\ \ \varepsilon \sim N\left(0, \sigma^2\right)$$

dengan β_0 dan β_1 ditaksir oleh $\hat{\beta}_0$ dan $\hat{\beta}_1$ berdasarkan metode kudrat terkecil atau metode kemungkinan maksimum.

Jika diberikan nilai dari variabel X, misalkan x_0 (tidak diketahui) dan dapat diamati nilai Y (atau sampel acak berukuran k untuk nilai dari Y) yang berkorespondensi dengan x_0, katakanlah y_0. Maka penyelesaian regresi invers dengan metode invers adalah dengan membuat regresi X atas Y, yaitu dengan membuat model

$$X = \alpha_0 + \alpha Y + \varepsilon,\ \ \varepsilon \sim N\left(0, \sigma^2\right). \tag{3.3}$$

Penaksir untuk X_0 dengan metode invers adalah

$$\hat{X}_0 = \hat{\alpha}_0 + \hat{\alpha}\bar{Y}_0 \tag{3.4}$$

dengan α_0 dan α ditaksir oleh $\hat{\alpha}_0$ dan $\hat{\alpha}$ berdasarkan metode kuadrat terkecil atau metode kemungkinan maksimum.

Pengujian yang harus dilakukan pada penaksir invers dari regresi invers ini adalah dengan melakukan uji-uji pada regresi linear sederhana X atas Y.

Beberapa statistikawan tertarik untuk membandingkan kedua metode ini dengan tujuan menemukan metode yang lebih baik untuk digunakan.

3.2.2 Metode Graybill

Metode untuk menyelesaikan masalah regresi invers ini diungkapkan oleh Franklin A. Graybill pada tahun 1976 dalam sebuah bukunya yang berjudul "*Theory and Application of the Linear Models*". Metode ini merupakan metode untuk mencari penaksir interval dari X_0 dengan derajat kepercayaan $1-\alpha$, karena seperti telah diungkapkan pada bab sebelumnya bahwa penaksiran interval dibutuhkan karena seringkali penaksiran titik untuk suatu parameter memberikan hasil yang kurang memuaskan.

Penaksiran X_0 beserta interval kepercayaannya dengan menggunakan Metode Graybill ini ditentukan berdasarkan sampel acak berpasangan berukuran $n+k$ dengan $k \geq 1$, yaitu :

$$(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n),(X_0,Y_{n+1}),(X_0,Y_{n+2}),\ldots,(X_0,Y_{n+k})$$

dengan X_0 merupakan variabel bebas yang tidak diketahui nilainya dan Y_{n+i} untuk $i = 1, 2, .., k$ diambil dari distribusi yang sama pada X_0.

Secara umum, dari sampel acak berukuran $n + k$ ini harus dapat dimodelkan :

$$\left.\begin{array}{ll} Y_i = \beta_0 + \beta X_i + \varepsilon_i & i = 1, 2, ..., n \\ Y_i = \beta_0 + \beta X_0 + \varepsilon_i & i = n+1, n+2, ..., n+k \end{array}\right\} \varepsilon_i \sim NID\left(0, \sigma^2\right) \qquad (3.5)$$

Uraian mengenai Metode Graybill pada tugas akhir ini akan menggunakan model regresi linear dengan bentuk terpusat, yaitu dengan menggunakan variabel bebas $\left(X - \overline{X}\right)$ sebagai pengganti variabel bebas X. Jadi, jika diketahui model regresi linear sederhana $Y = \beta_0 + \beta X + \varepsilon$ dengan $\varepsilon \sim N(0, \sigma^2)$ maka model regresi linear sederhana dengan bentuk terpusatnya adalah :

$$Y = \tau_0 + \tau(X - \overline{X}) + \varepsilon \text{ dengan } \varepsilon \sim N(0, \sigma^2). \qquad (3.6)$$

Hal ini dilakukan untuk menyederhanakan bentuk dalam penurunan rumus dan pembuktian teorema (Thonnard, 2006).

3.2.2.1 Prosedur Penyelesaian

Penyelesaian masalah regresi invers beserta interval kepercayaannya akan diuraikan berdasarkan langkah-langkah penaksiran interval dengan pendekatan/teknik besaran pivot sebagai berikut :

1. Mengambil sampel acak

   Ambil sampel acak berpasangan berukuran $n + k$, yaitu :

$$\left(X_1, Y_1\right), \left(X_2, Y_2\right), ..., \left(X_n, Y_n\right), \left(X_0, Y_{n+1}\right), \left(X_0, Y_{n+2}\right), ..., \left(X_0, Y_{n+k}\right)$$

dimana X_0 merupakan variabel bebas yang tidak diketahui nilainya dan Y_{n+i} untuk $i = 1, 2, .., k$ diambil dari distribusi yang sama pada X_0.

2. Menentukan penaksir titik untuk X_0

Graybill memilih menggunakan metode klasik untuk menentukan penaksir titik X_0 dengan alasan tidak dapat dipenuhinya asumsi pada penaksir invers, yaitu :

a. Jika menggunakan penaksir invers atau dengan kata lain memodelkan regresi linear X atas Y, maka Y haruslah memenuhi asumsi bahwa Y merupakan sebuah nilai tertentu dan bukan merupakan variabel acak.

b. Jika menggunakan penaksir invers atau dengan kata lain memodelkan regresi linear X atas Y, maka X haruslah memenuhi asumsi bahwa X merupakan variabel acak.

Dalam menentukan penaksir titik untuk X_0 ini, Graybill melakukan langkah-langkah sebagai berikut :

a. Membuat model $Y_i = \tau_0 + \tau\left(X_i - \overline{X}\right) + \varepsilon_i$, $i = 1, 2, ..., n$ berdasarkan sampel acak berukuran n dan menaksir nilai τ_0 dan τ oleh $\hat{\tau}_0$ dan $\hat{\tau}$ berdasarkan metode kemungkinan maksimum, yaitu :

$$\hat{\tau}_0 = \overline{Y} \qquad (3.7) \qquad \text{dan} \qquad \hat{\tau} = \frac{\sum_{i=1}^{n} Y_i\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \qquad (3.8)$$

(Bukti lihat lampiran 1.E butir ke-i dan ke-ii)

b. Menentukan penaksir titik untuk X_0 dengan metode klasik, yaitu :

$$\hat{X}_0 = \overline{X} + \frac{\overline{Y}_0 - \overline{Y}}{\hat{\tau}}, \qquad \text{jika } \hat{\tau} \neq 0 \tag{3.9}$$

(Bukti lihat lampiran 1.E butir ke-iii)

3. Membentuk besaran pivot

Besaran pivot haruslah memuat penaksir titik $\left(\hat{X}_0\right)$ dan yang ditaksirnya $\left(X_0\right)$, tidak memuat parameter lain yang tidak diketahui serta distribusinya diketahui. Dalam menentukan besaran pivot yang diinginkan, diperlukan pengkajian yang mendalam terhadap sampel dan beberapa teorema pendukung sehingga dapat diperoleh informasi-informasi yang dibutuhkan.

Pertama, akan dicari penaksir dari varians model regresi $\left(\hat{\sigma}^2\right)$, penaksir varians model regresi ini diperoleh berdasarkan sampel acak berukuran $n+k$ dengan menggunakan metode kemungkinan maksimum, yaitu :

$$\hat{\sigma}^2_{KM} = \frac{1}{n+k}\left[\sum_{i=1}^{n}\left(Y_i - \hat{\tau}_0 - \hat{\tau}\left(X_i - \overline{X}\right)\right)^2 + \sum_{i=n+1}^{n+k}\left(Y_i - \overline{Y}_0\right)^2\right]$$

dengan $\overline{Y}_0 = \frac{1}{k}\sum_{i=n+1}^{n+k} Y_i$.

(Bukti lengkap lihat lampiran 1.E butir ke-iv)

Karena setelah diuji nilai ekspektasi dari $\hat{\sigma}^2_{KM} = \frac{n+k-3}{n+k}\sigma^2$, maka ini merupakan penaksir bias. Penaksir tak biasnya adalah :

$$\hat{\sigma}^2 = \frac{1}{n+k-3}\left[\sum_{i=1}^{n}\left(Y_i - \hat{\tau}_0 - \hat{\tau}\left(X_i - \overline{X}\right)\right)^2 + \sum_{i=n+1}^{n+k}\left(Y_i - \overline{Y}_0\right)^2\right] \tag{3.10}$$

Sehingga jika $k=1$ maka penaksir untuk variansinya adalah :

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \hat{\tau}_0 - \hat{\tau}\left(X_i - \overline{X}\right)\right)^2 \qquad (3.11)$$

Beberapa teorema dan definisi yang diperlukan untuk menentukan besaran pivot ini adalah :

Teorema 3.1 :

Misalkan $\left(X_i, Y_i\right)$, $i = 1, 2, ..., n$ mempunyai hubungan linear sederhana $Y_i = \tau_0 + \tau\left(X_i - \overline{X}\right) + \varepsilon_i$, $\varepsilon_i \sim NID\left(0, \sigma^2\right)$. Misalkan Y_{n+i}, $i = 1, 2, ..., k$, $k \geq 1$ merupakan pengamatan acak dari distribusi normal dengan rata-rata $\tau_0 + \tau\left(X_0 - \overline{X}\right)$ (X_0 tidak diketahui) dan varians σ^2. Jika semua variabel acak Y_i saling bebas, maka penaksir kemungkinan maksimum dari τ_0, τ, X_0 dan σ^2 (dengan koreksi ketakbiasan) masing-masing diberikan oleh persamaan (3.7), (3.8), (3.9) dan (3.10).

(Bukti lihat lampiran 1.E).

Teorema 3.2 :

Perhatikan model yang diberikan oleh teorema 3.1 dan penaksir-penaksirnya yang diberikan oleh persamaan (3.7), (3.8), (3.9) dan (3.10), maka berlaku :

(i). $\hat{\tau}_0$ dan $\hat{\tau}$ saling bebas

(ii). $\hat{\tau}_0 \sim N\left(\tau_0, \frac{\sigma^2}{n}\right)$ dan $\hat{\tau} \sim N\left(\tau, \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$

(iii). $U = (n+k-3)\dfrac{\hat{\sigma}^2}{\sigma^2}$ berdistribusi χ^2_{n+k-3}

(iv). $\hat{\sigma}^2$ saling bebas dengan $\left(\hat{\tau}_0, \hat{\tau}, \bar{X}_0\right)$

(Bukti lihat lampiran 1.F)

Teorema 3.3 :

Jika X adalah variabel acak berdistribusi normal dengan rata-rata μ dan varians σ^2 $\left(X \sim N\left(\mu, \sigma^2\right)\right)$, maka $Z = \dfrac{X-\mu}{\sigma}$ berdistribusi normal baku $\left(Z \sim N(0,1)\right)$.

(Bukti lihat lampiran 1.G)

Definisi 3.2 :

Jika W variabel acak berdistribusi normal baku $\left(W \sim N(0,1)\right)$ dan V variabel acak berdistribusi chi kuadrat dengan derajat kebebasan r $\left(V \sim \chi^2_r\right)$. Jika W dan V bebas secara stokastik maka $T = \dfrac{W}{\sqrt{V/r}}$ berdistribusi t dengan derajat kebebasan r $(T \sim t_r)$.

Berdasarkan teorema dan definisi di atas, dapat ditunjukkan bahwa :

$$\begin{aligned}
\mathrm{E}\left(\bar{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \bar{X}\right)\right) &= \mathrm{E}\left(\bar{Y}_0\right) - \mathrm{E}\left(\hat{\tau}_0\right) - \left(X_0 - \bar{X}\right)\mathrm{E}\left(\hat{\tau}\right) \\
&= \left(\tau_0 + \tau\left(X_0 - \bar{X}\right)\right) - \tau_0 - \tau\left(X_0 - \bar{X}\right) \\
&= 0
\end{aligned}$$

$$Var\left(\overline{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \overline{X}\right)\right) = Var\left(\overline{Y}_0\right) + Var\left(\hat{\tau}_0\right) + \left(X_0 - \overline{X}\right)^2 Var\left(\hat{\tau}\right)$$
$$- 2Cov\left(\overline{Y}_0, \hat{\tau}_0\right) - 2Cov\left(\overline{Y}_0, \hat{\tau}\left(X_0 - \overline{X}\right)\right)$$
$$+ 2Cov\left(\hat{\tau}_0, \hat{\tau}\left(X_0 - \overline{X}\right)\right)$$

$$= \frac{\sigma^2}{k} + \frac{\sigma^2}{n} + \left(X_0 - \overline{X}\right)^2 \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} - 2(0) - 2(0) + 2(0)$$

$$= \sigma^2 \left(\frac{1}{k} + \frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \right)$$

$$= \sigma^2 A^2$$

Dengan $A^2 = \frac{1}{k} + \frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$

Karena $\overline{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \overline{X}\right)$ merupakan kombinasi dari komponen-komponen yang berdistribusi normal, maka $\overline{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \overline{X}\right)$ pun akan berdistribusi normal dengan rata-rata 0 dan variansi $\sigma^2 A^2$ $\left(\overline{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \overline{X}\right) \sim N\left(0, \sigma^2 A^2\right)\right)$. Sehingga berdasarkan teorema 3.3 diperoleh :

$$Z = \frac{\overline{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \overline{X}\right) - 0}{\sqrt{\sigma^2 A^2}} \sim N(0,1) \qquad (3.12)$$

Berdasarkan teorema 3.2 butir ke-iii diperoleh :

$$
\begin{aligned}
U &= (n+k-3)\frac{\hat{\sigma}^2}{\sigma^2} \\
&= (n+k-3)\frac{\frac{1}{n+k-3}\left[\sum_{i=1}^{n}\left(Y_i-\hat{\tau}_0-\hat{\tau}\left(X_i-\overline{X}\right)\right)^2+\sum_{i=n+1}^{n+k}\left(Y_i-\overline{Y}_0\right)^2\right]}{\sigma^2} \\
&= \frac{\sum_{i=1}^{n}\left(Y_i-\hat{\tau}_0-\hat{\tau}\left(X_i-\overline{X}\right)\right)^2+\sum_{i=n+1}^{n+k}\left(Y_i-\overline{Y}_0\right)^2}{\sigma^2} \sim \chi^2_{n+k-3}
\end{aligned}
\tag{3.13}
$$

Berdasarkan teorema 3.2 butir ke-iv dapat disimpulkan bahwa Z dan U saling bebas, sebab σ^2 saling bebas dengan $\left(\hat{\tau}_0, \hat{\tau}, \hat{X}_0\right)$. Karena $Z \sim N(0,1)$ dan $U \sim \chi^2_{n+k-3}$ saling bebas, maka berdasarkan definisi 3.2 diperoleh :

$$
T = \frac{Z}{\sqrt{\frac{U}{n+k-3}}} \sim t_{n+k-3} \tag{3.14}
$$

Akan diperlihatkan bahwa T memenuhi syarat sebagai besaran pivot, yaitu memuat penaksir titik $\left(\hat{X}_0\right)$ dan yang ditaksirnya $\left(X_0\right)$, tidak memuat parameter lain yang tidak diketahui serta distribusinya diketahui.

$$
\begin{aligned}
T = \frac{Z}{\sqrt{\frac{U}{n+k-3}}} = \frac{\frac{\overline{Y}_0-\hat{\tau}_0-\hat{\tau}\left(X_0-\overline{X}\right)}{\sqrt{\sigma^2 A^2}}}{\sqrt{\frac{(n+k-3)\frac{\hat{\sigma}^2}{\sigma^2}}{n+k-3}}} &= \frac{\overline{Y}_0-\hat{\tau}_0-\hat{\tau}\left(X_0-\overline{X}\right)}{\hat{\sigma}A} \\
&= \frac{\overline{Y}_0-\overline{Y}-\hat{\tau}X_0+\hat{\tau}\overline{X}}{\hat{\sigma}A}
\end{aligned}
$$

$$= \frac{\hat{\tau}\left(\left(\frac{\overline{Y}_0 - \overline{Y}}{\hat{\tau}} + \overline{X}\right) - X_0\right)}{\hat{\sigma} A}$$

$$= \frac{\hat{\tau}\left(\hat{X}_0 - X_0\right)}{\hat{\sigma} A}$$

dengan $A^2 = \frac{1}{k} + \frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$.

Sehingga dipilih T sebagai besaran pivot.

4. Dengan mengambil derajat kepercayaan sebesar $\gamma = 1 - \alpha$ dan dengan menyatakan $P\left[-t_{\frac{1}{2}\alpha;n+k-3} \le T \le t_{\frac{1}{2}\alpha;n+k-3}\right] = 1 - \alpha$, maka akan dicari penaksir interval dari X_0.

$$\Leftrightarrow P\left[-t_{\frac{1}{2}\alpha;n+k-3} \le T \le t_{\frac{1}{2}\alpha;n+k-3}\right] = 1 - \alpha$$

$$\Leftrightarrow P\left[-t_{\frac{1}{2}\alpha;n+k-3} \le \frac{\hat{\tau}\left(\hat{X}_0 - X_0\right)}{\hat{\sigma} A} \le t_{\frac{1}{2}\alpha;n+k-3}\right] = 1 - \alpha$$

$$\Leftrightarrow P\left[-\hat{\sigma} A t_{\frac{1}{2}\alpha;n+k-3} - \hat{\tau}\hat{X}_0 \le -\hat{\tau} X_0 \le \hat{\sigma} A t_{\frac{1}{2}\alpha;n+k-3} - \hat{\tau}\hat{X}_0\right] = 1 - \alpha$$

Karena nilai $\hat{\tau}$ dapat bernilai positif atau negatif maka tidak dapat diperoleh penaksir interval untuk X_0 yang diinginkan. Oleh sebab itu akan dilakukan pengkuadratan terhadap besaran pivot.

$$P\left[-t_{\frac{1}{2}\alpha;n+k-3} \le T \le t_{\frac{1}{2}\alpha;n+k-3}\right] = 1 - \alpha$$

$$\Leftrightarrow P\left[T^2 \le t^2_{\alpha/2;n+k-3}\right] = 1-\alpha$$

$$\Leftrightarrow P\left[\frac{\left(\bar{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \bar{X}\right)\right)^2}{\hat{\sigma}^2 A^2} \le t^2_{\alpha/2;n+k-3}\right] = 1-\alpha$$

$$\Leftrightarrow P\left[\left(\bar{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \bar{X}\right)\right)^2 - \hat{\sigma}^2 A^2 t^2_{\alpha/2;n+k-3} \le 0\right] = 1-\alpha \qquad (3.15)$$

Perhatikan bentuk $\left(\bar{Y}_0 - \hat{\tau}_0 - \hat{\tau}\left(X_0 - \bar{X}\right)\right)^2 - \hat{\sigma}^2 A^2 t^2_{\alpha/2;n+k-3} \le 0$ dengan X_0 merupakan satu-satunya parameter yang tidak diketahui. Bila dijabarkan dengan mensubstitusikan $A^2 = \frac{1}{k} + \frac{1}{n} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ dapat diperoleh :

$$\left[\hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]\left(X_0 - \bar{X}\right)^2 + \left[-2\hat{\tau}\left(\bar{Y}_0 - \hat{\tau}_0\right)\right]\left(X_0 - \bar{X}\right)$$
$$+\left[\left(\bar{Y}_0 - \hat{\tau}_0\right)^2 - \hat{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\frac{1}{k} + \frac{1}{n}\right)\right] \le 0 \qquad (3.16)$$

yang merupakan pertidaksamaan kuadrat $a\left(X_0 - \bar{X}\right)^2 + b\left(X_0 - \bar{X}\right) + c \le 0$

dengan : $a = \hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

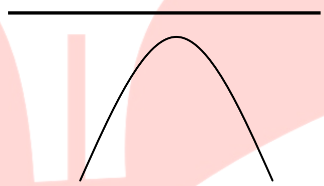
$b = -2\hat{\tau}\left(\bar{Y}_0 - \hat{\tau}_0\right)$

$c = \left(\bar{Y}_0 - \hat{\tau}_0\right)^2 - \hat{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\frac{1}{k} + \frac{1}{n}\right)$

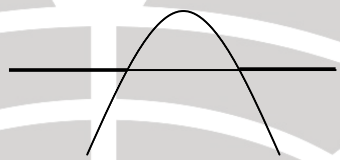
Jika nilai X_0 yang memenuhi pertidaksamaan (3.16) berupa interval, maka nilai tersebut merupakan kondisi untuk X_0 dengan interval kepercayaan sebesar $100(1-\alpha)\%$.

Bentuk pertidaksamaan kuadrat di atas akan memiliki salah satu dari empat kemungkinan solusi berikut :

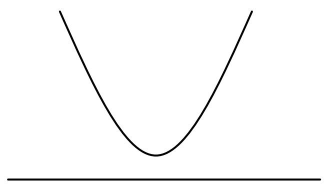
a. Saat $a<0$ dan diskriminan (b^2-4ac) < 0, maka pertidaksamaan akan kurang dari 0 untuk semua nilai X_0 sehingga selang kepercayaan untuk X_0 adalah $-\infty < x_0 < \infty$.



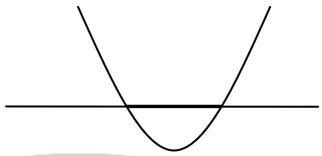
b. Saat $a<0$ dan diskriminan (b^2-4ac) > 0, maka solusi untuk X_0 tak terbatas.



c. Saat $a>0$ dan diskriminan (b^2-4ac) < 0 maka pertidaksamaan akan lebih dari 0 untuk semua nilai X_0 sehingga tidak terdapat solusi untuk X_0.



d. Saat $a > 0$ dan diskriminan $\left(b^2 - 4ac\right) > 0$ maka akan diperoleh selang kepercayaan untuk X_0 yang diinginkan.



Berdasarkan empat kemungkinan tersebut, dapat disimpulkan bahwa interval kepercayaan untuk X_0 akan ada saat $a > 0$ dan diskriminan $\left(b^2 - 4ac\right) > 0$. Sehingga batas atas dan batas bawah dari selang kepecayaan untuk X_0 dapat ditentukan dengan menyelesaikan (mencari akar kuadrat) persamaan kuadrat yang berkaitan dengan pertidaksamaan (3.16), yaitu :

$$\left[\hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]\left(X_0 - \bar{X}\right)^2 + \left[-2\hat{\tau}\left(\bar{Y}_0 - \hat{\tau}_0\right)\right]\left(X_0 - \bar{X}\right) + \left[\left(\bar{Y}_0 - \hat{\tau}_0\right)^2 - \hat{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\frac{1}{k} + \frac{1}{n}\right)\right] = 0 \qquad (3.17)$$

Persamaan kuadrat (3.17) akan diselesaikan dengan menggunakan rumus berikut :

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (3.18)$$

Sehingga akar-akar dari persamaan (3.17) adalah sebagai berikut :

$$\left(X_0 - \bar{X}\right)_1 = \frac{\hat{\tau}(\bar{Y}_0 - \bar{Y})}{a} - \frac{\hat{\sigma} t_{\frac{\alpha}{2};n+k-3}}{a}\sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{\left(\bar{Y}_0 - \bar{Y}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \qquad (3.19)$$

dan

$$\left(X_0 - \bar{X}\right)_2 = \frac{\hat{\tau}(\bar{Y}_0 - \bar{Y})}{a} + \frac{\hat{\sigma} t_{\frac{\alpha}{2};n+k-3}}{a} \sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{\left(\bar{Y}_0 - \bar{Y}\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \tag{3.20}$$

Dengan $a = \hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$.

(Bukti lengkap lihat lampiran 1.H)

Berdasarkan penyelesaian persamaan kuadrat (3.17) maka interval kepercayaan sebesar $100(1-\alpha)\%$ untuk X_0 akan memiliki batas bawah :

$$\bar{X} + \frac{\hat{\tau}(\bar{Y}_0 - \bar{Y})}{a} - \frac{\hat{\sigma} t_{\frac{\alpha}{2};n+k-3}}{a} \sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{\left(\bar{Y}_0 - \bar{Y}\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}, \tag{3.21}$$

dan batas atas :

$$\bar{X} + \frac{\hat{\tau}(\bar{Y}_0 - \bar{Y})}{a} + \frac{\hat{\sigma} t_{\frac{\alpha}{2};n+k-3}}{a} \sqrt{a\left(\frac{1}{n} + \frac{1}{k}\right) + \frac{\left(\bar{Y}_0 - \bar{Y}\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}. \tag{3.22}$$

Dengan $a = \hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$

3.2.2.2 Pengujian Asumsi Pada Metode Graybill

Pengujian asumsi harus dilakukan untuk memperoleh kesimpulan yang dapat dipertanggungjawabkan secara teori. Pada metode Graybill untuk menyelesaikan masalah regresi invers beserta interval kepercayaannya ini, pengujian yang harus dilakukan adalah melakukan uji-uji asumsi pada regresi linear sederhana Y atas X (karena Graybill memilih metode klasik untuk mencari penaksir titik X_0).

Penyelidikan mengenai keberadaan asumsi lain yang harus dipenuhi pada Metode Graybill haruslah dilakukan. Telah disinggung sebelumnya bahwa interval kepercayaan untuk X_0 akan ada pada saat $a > 0$ dan diskriminan $\left(b^2 - 4ac\right) > 0$. Dengan menguraikan syarat $a > 0$ dan $\left(b^2 - 4ac\right) > 0$, maka akan diperoleh syarat adanya interval kepercayaan untuk X_0 sebagai berikut :

- $a > 0$

$$\Leftrightarrow \hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} > 0$$

$$\Leftrightarrow \frac{\hat{\tau}^2 \sum_{i=1}^{n}(X_i - \bar{X})^2}{\hat{\sigma}^2} > t^2_{\alpha/2;n+k-3} = F_{\alpha;1,n+k-3}$$

(Bukti bahwa distribusi t_r^2 sama dengan distribusi $F_{1,r}$ lihat pada lampiran 1.A butir ke-(ii)).

- $\left(b^2 - 4ac\right) > 0$

$$\Leftrightarrow \left(-2\hat{\tau}\left(\bar{Y}_0 - \hat{\tau}_0\right)\right)^2 - 4\left(\hat{\tau}^2 - \frac{\hat{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)\left(\left(\bar{Y}_0 - \hat{\tau}_0\right)^2 - \hat{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\frac{1}{k} + \frac{1}{n}\right)\right) > 0$$

$$\Leftrightarrow 4\hat{\tau}^2\left(\bar{Y}_0-\hat{\tau}_0\right)^2-4\hat{\tau}^2\left(\bar{Y}_0-\hat{\tau}_0\right)^2+4\hat{\tau}^2\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\frac{1}{k}+\frac{1}{n}\right)+$$

$$4\left(\bar{Y}_0-\hat{\tau}_0\right)^2\frac{\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}-4\frac{\tilde{\sigma}^4 t^4_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\left(\frac{1}{k}+\frac{1}{n}\right)>0$$

$$\Leftrightarrow 4\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\hat{\tau}^2\left(\frac{1}{k}+\frac{1}{n}\right)+\frac{\left(\bar{Y}_0-\hat{\tau}_0\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}-\frac{\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\left(\frac{1}{k}+\frac{1}{n}\right)\right)>0$$

$$\Leftrightarrow \tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\hat{\tau}^2\left(\frac{1}{k}+\frac{1}{n}\right)+\frac{\left(\bar{Y}_0-\hat{\tau}_0\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}-\frac{\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\left(\frac{1}{k}+\frac{1}{n}\right)\right)>0$$

$$\Leftrightarrow \tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(\left(\hat{\tau}^2-\frac{\tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)\left(\frac{1}{k}+\frac{1}{n}\right)+\frac{\left(\bar{Y}_0-\hat{\tau}_0\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)>0$$

$$\Leftrightarrow \tilde{\sigma}^2 t^2_{\alpha/2;n+k-3}\left(a\left(\frac{1}{k}+\frac{1}{n}\right)+\frac{\left(\bar{Y}_0-\hat{\tau}_0\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)>0$$

Syarat di atas akan terpenuhi saat $a>0$, sehingga interval untuk X_0 akan ada jika $a>0$, yaitu $\frac{\hat{\tau}^2\sum_{i=1}^{n}(X_i-\bar{X})^2}{\tilde{\sigma}^2}>t^2_{\alpha/2;n+k-3}=F_{\alpha;1,n+k-3}$ yang tidak lain merupakan kriteria pengujian untuk menolak H_0 pada uji keberartian koefisien regresi (lihat uji keberartian koefisien regresi secara manual pada subbab 2.1).

Sehingga dapat disimpulkan bahwa interval untuk X_0 akan ada jika koefisien regresi berarti.

Dari uraian di atas, dapat disimpulkan bahwa masalah regresi invers dapat memiliki penyelesaian X_0 beserta interval kepercayaannya dengan menggunakan Metode Graybill apabila asumsi-asumsi pada bentuk regresi linear sederhananya terpenuhi.

Langkah-langkah metode Graybill untuk menyelesaikan masalah regresi invers beserta interval kepercayaannya, yaitu mencari penaksir interval dari X_0 dengan koefisien kepercayaan $1-\alpha$ (dengan mengambil sampel acak berpasangan berukuran $n+k$) dapat diringkaskan sebagai berikut :

1. Gunakan statistik pada persamaan (3.7) dan (3.8) untuk menentukan persamaan regresi linear sederhana taksiran dengan bentuk terpusat.
2. Lakukan pengujian terhadap asumsi-asumsi pada bentuk regresi linear sederhana dengan bentuk terpusat. Jika semua asumsi terpenuhi, maka terdapat penyelesaian untuk X_0 beserta selang intervalnya.
3. Tentukan penaksir titik untuk X_0 menggunakan metode klasik berdasarkan persamaan (3.9).
4. Susun interval kepercayaan sebesar $1-\alpha$ untuk X_0. Interval kepercayaan ini akan memiliki batas-batas (3.21) dan (3.22).