BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka yang digunakan dalam skripsi ini adalah referensi buku-buku dan beberapa penelitian yang sudah ada dengan pokok bahasan yang relevan. Adapun penelitian-penelitian tersebut antara lain sebagai berikut :

1. Penelitian yang dilakukan oleh Herindani (2016) dengan judul "Perbandingan Metode Resampling Bootstrap dan Jacknife Untuk Estimasi Parameter Model Regresi Linear Berganda". Penelitian ini mengatakan bahwa Metode *resampling* Bootstrap dan Jackknife dapat dijadikan alternative untuk estimasi parameter model regresi linear, apabila hasil estimasi dari metode kuadrat terkecil dengan sampel asli, tidak memenuhi asumsi. Metode *resampling* Bootstrap dan Jackknife memberikan hasil estimasi dengan *standard error* yang lebih kecil daripada hasil estimasi menggunakan metode kuadrat terkecil sampel asli.
2. Penelitian oleh Istiari (2017) dengan judul "Perbandingan Metode Resampling Bootstrap Residual Dan Bootstrap Pairs Untuk Menduga Parameter Model Regresi Linear Berganda". Penelitian ini menjelaskan tentang perbandingan Metode Resampling Bootstrap Residual Dan Bootstrap Pairs. untuk alternatif menduga parameter model regresi linear berganda, apabila hasil pendugaan dengan menggunakan metode kuadrat terkecil dengan sampel asli tidak memenuhi asumsi klasik atau ukuran sampel kecil. Metode *resampling bootstrap* dibagi menjadi 2 yaitu metode *resampling*

bootstrap residual dan metode *resampling bootstrap pairs*. Metode *resampling bootstrap* residual dan *pairs* memberikan hasil penduga dengan standard error yang lebih kecil dibandingkan dengan hasil penduga menggunakan metode kuadrat terkecil sampel asli.

Dari uraian diatas, dapat diketahui bahwa metode resampling bootstrap residual untuk regresi masih dilakukan dalam penelitian yang berbeda. Perbedaan penelitian dalam skripsi ini dengan penelitian sebelumnya yaitu pokok bahasan dalam skripsi ini adalah membandingkan metode *resampling* bootstrap residual dalam mengestimasi parameter untuk model regresi linear berganda.

Beberapa landasan teori yang digunakan dalam penelitian ini adalah sebagai berikut:

2.1 Analisis Regresi Berganda dengan Metode Ordinary Least Square

Metode *Ordinary Least Square* (OLS) atau sering disebut metode kuadrat terkecil adalah metode yang bertujuan untuk meminimumkan jumlah kuadrat galat (*Sum Square Error).*

Persamaan umum regresi linier berganda dengan metode kuadrat terkecil adalah:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon_i \qquad (2.1)$$

Dengan

$\hat{Y}$: nilai ramalan atau prediksi dari peubah atau variabel tak bebas (respon)

X_i : nilai pengamatan dari variabel bebas (prediktor), dengan i = 1,2,...,n

β_0 : intersep atau konstanta

β_i : slope atau koefisien kemiringan model regresi, dengan i = 1,2,...,n

ε_i : variabel kesalahan (galat)

Pada analisis regresi berganda dengan metode OLS terdapat prosedur yang harus dilakukan yaitu :

1. **Mencari persamaan regresi**

Pada analisis regresi berganda yang pertama harus dilakukan adalah mencari persamaannya terlebih dahulu dengan memasukkan variabel Y sebagai variabel dependen dan variabel X sebagai variabel independen.

2. **Uji Sign. F**

Selanjutnya melakukan uji signifikansi regresi (Uji F), dilakukan untuk mengetahui apakah ada variabel independen yang berpengaruh signifikan terhadap variabel dependen secara serentak (simultan).

Menurut Ghozali (2009), uji statistik *F* menunjukkan apakah semua variabel independen atau bebas yang dimasukkan dalam model mempunyai pengaruh secara bersama-sama terhadap variabel dependen atau terikat. Uji statistik *F* digunakan untuk mengetahui pengaruh semua variabel independen yang dimasukkan dalam model regresi secara bersama-sama terhadap variabel dependen yang diuji pada tingkat signifikan 0,05.

a.) Hipotesis:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

(semua parameter secara simultan sama dengan nol)

H_1: $\beta_0 \neq 0$, unuruk suatu $k \epsilon$ $(1,2, \ldots p)$

(minimal ada satu parameter secara simultan tidak sama dengan nol)

b.) Statistik uji :

$$F_{hitung} = \frac{\frac{ESS}{p-1}}{\frac{TSS}{n-p}} = \frac{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{p-1}}{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-p}} \text{ dan nilai } F_{tabel} = F_{(\alpha;(p-1,n-p))} \quad (2.2)$$

Dengan *ESS* adalah *Explained Sum of Squares*, *TSS* adalah *Total Sum of Squares*, *n*= jumlah observasi dan *p*= jumlah variabel independen.

c.) Aturan pengambilan keputusan :

H_0 ditolak jika $F_{hitung} > F_{tabel}$ atau dengan $p - value < \alpha$.

d.) Dikatakan memenuhi uji statistik *F* apabila H_0 ditolak yang artinya minimal ada satu parameter secara simultan tidak sama dengan nol dengan kata lain semua variabel independen secara simultan berpengaruh signifikan terhadap variabel dependen.

3. Uji Sign. t

Setelah dilakukan uji signifikansi secara serentak (simultan) maka selanjutnya melakukan uji signifikansi secara individu (parsial) dengan melakukan uji signifikasi regresi (uji t).

Uji statistik *t* menunjukkan seberapa jauh pengaruh satu variabel penjelas atau independen secara individual dalam menerangkan variasi variabel dependen dan digunakan untuk mengetahui ada atau tidaknya pengaruh masing-masing variabel independen secara individual terhadap variabel dependen yang diuji pada tingkat signifikansi 0,05 Ghozali (2009)

a.) Hipotesis:

H_0: $\beta_k = 0$ (parameter suatu variabel sama dengan nol)

H_1: $\beta_k \neq 0$ (parameter suatu variabel tidak sama dengan nol)

b.) Statistik uji :

$$t_{hitung} = \frac{\beta_k}{se(\beta_k)} \text{ dan nilai } t_{tabel} = t_{(\alpha;n-p)} \qquad (2.3)$$

Dengan n = ukuran sampel dan p = banyaknya variabel independen.

c.) Aturan pengambilan keputusan:

H_0 ditolak jika $t_{hitung} > t_{tabel}$ atau dengan $p - value < \alpha$.

d.) Dikatakan memenuhi uji statistik t apabila H_0 ditolak yang artinya parameter suatu variabel tidak sama dengan nol dengan kata lain x_k, untuk suatu indeks k berpengaruh signifikan terhadap y.

1.2 Metode *Bootstrap*

Metode *bootstrap* mendapatkan sampelnya dengan cara *sampling* dengan pengembalian dari sampel asli. Prosedur *Bootstrap* untuk regresi linear dapat dilakukan dengan 2 cara, yaitu *Bootstrap Residual* dan *Bootstrap Pairs* data terhadap estimasi model dari sampel asli (Efron dan Tibshirani, 1998). Secara umum, misalkan memiliki sampel berukuran n, yaitu x_1, x_2, ..., x_n yang diambil secara random dari populasi dengan fungsi distribusi F yg tidak diketahui.

Parameter bernilai real dari fungsi distribusi F pada populasi dinotasikan dengan $\theta = \theta(F)$. Estimator untuk θ yang merupakan fungsi bernilai real dari x_1, x_2, ..., x_n dinotasikan $\hat{\theta}$. Fungsi distribusi F biasanya tidak diketahui, tetapi nilai sampel random x_1, x_2, ..., x_n dari F diketahui, maka F dapat diestimasi menggunakan fungsi distribusi empiris $\widehat{F_n}(x)$. Menurut Efron dan Tbshirani(1993), sampel *Bootstrap* dapat diperoleh melalui pengambilan scara random berukuran n dari distribusi empiris $\widehat{F_n}(x)$ dengan pengembalian diperoleh $x_1^*, x_2^*, \dots, x_n^*$. Estimator dari parameter θ yang diperoleh berdasarkan sampel *Bootstrap* ini dinyatakan sebagai estimator *Bootstrap*.

1.2.1 Metode Resampling Bootstrap residual

Apabila regresor adalah variabel yang fix, metode bootstrap yang dipakai adalah dengan melakukan resampling pada *residual* (hasil bentukan model *OLS*, pada sampel). *Residual* dalam regresi artinya selisih nilai antara *Y* sebenarnya dengan *Y* estimasi($\hat{Y}$). Dari nilai residual ini selanjutnya diestimasikan parameter model regresi. Proses ini dilakukan berulang sampai sebanyak B kali.

1.2.2 Metode *Resampling Bootstrap pairs*

Apabila regresor adalah random, metode bootstrap yang dipakai adalah dengan melakukan resampling pada observasi dengan probabilitas setiap observasi akan terambil sebanyak 1/n untuk jumlah sampel (i) = 1,2,…, n dan untuk sejumlah variabel (j) = 1,2,...,k. Resampling dilakukan sebanyak B kali. Dimana jumlah B diisyaratkan cukup besar, hingga diperoleh estimasi parameter yang konvergen atau bahkan sampai sejumlah n pangkat n sampel. Dengan jumlah

B yang cukup besar ini, diharapkan estimasi parameter regresi yang dihasilkan akan lebih kuat (robust).

1.2.3 Konsep Dasar Metode *Bootstrap*

Menurut Efron dan Tibshirani (1993), metode *bootstrap* adalah metode berbasis komputer yang didasarkan pada simulasi data untuk keperluan statistik. Metode *bootstrap* digunakan untuk mencari distribusi sampling dari suatu estimator dengan prosedur *resampling* dengan pengembalian dari data asli, dilakukan dengan mengambil sampel dari sampel asli dengan ukuran yang sama dengan sampel asli atau kurang dari sampel asli. Metode penyampelan ini biasa disebut dengan *resampling bootstrap*.

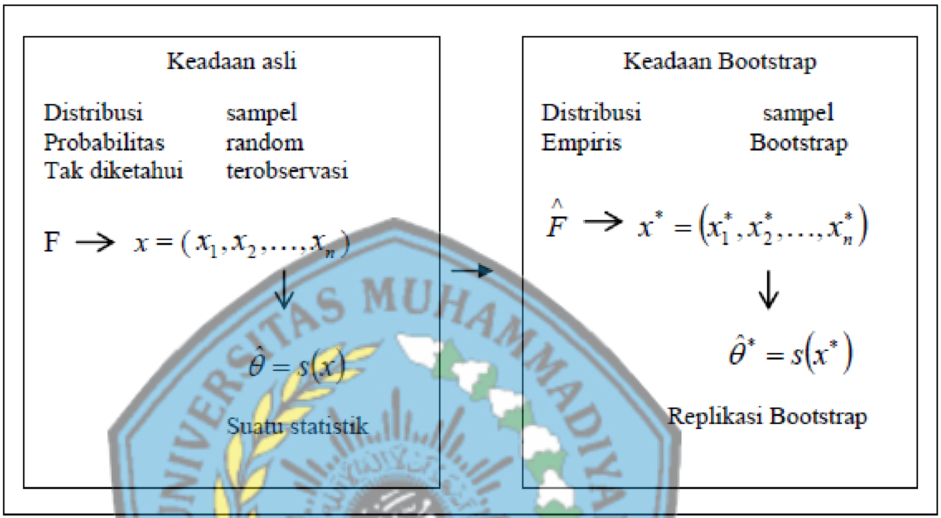
1. Pembentukan sampel *Bootstrap*

Metode *bootstrap* bergantung pada estimasi-estimasi dari sampel *bootstrap*. $\hat{F}$ adalah suatu distribusi empiris yang memberi probabilitas 1/*n* untuk setiap nilai observasi $x_i, i = 1,2, \dots, n$. Sampel *bootstrap* didefinisikan sebagai suatu sampel random berukuran *n* yang didapat dari $\hat{F}$, katakan $x^* = (x_1{}^*, x_2{}^*, \dots, x_n{}^*)$, atau dapat dinyatakan seperti berikut :

$$x_1{}^*, x_2{}^*, \dots, x_n{}^* \sim F \qquad (2.4)$$

Notasi bintang mengindikasikan bahwa x^* bukanlah data sebenarnya pada data **x**, akan tetapi merupakan versi dari *x* yang telah mengalami *resampling*. Data hasil *resampling* tersebut yang disebut sebagai data *bootstrap* $x_1{}^*, x_2{}^*, \dots, x_n{}^*$ yaitu sampel *random* berukuran *n* yang diambil dengan pengembalian dari $x = (x_1, x_2, \ldots, x_n)$, maka kemudian mungkin saja di dapat hasil *resampling* seperti berikut : $x_1{}^* = x_2, x_2{}^* = x_1, x_2{}^* = x_2, \dots x_n{}^* = x_4$. Data *resampling bootstrap*

$(x_1^*, x_2^*, \ldots, x_n^*)$ merupakan anggota-anggota dari data asli $(x_1, x_2, \ldots, x_n)$. Suatu nilai x_i dapat muncul satu kali, juga dapat beberapa atau mungkin tidak muncul sama sekali. Hal itu disebabkan karena teknik pengambilan dengan cara pengembalian.



Gambar 2.1 Prinsip Korespondensi *Bootstrap* (*Sumber :* Efron dan Tibshirani (1993))

Gambar 2.1 menjelaskan penggunaan metode *bootstrap* dalam sampel tunggal. Pada keadaan asli, F adalah suatu distribusi probabilitas yang tidak diketahui, memberikan data $x = (x_1, x_2, \ldots, x_n)$ dengan penyampelan secara random. Dari x tersebut dapat dihitung suatu statistik $\hat{\theta} = s(x)$. Pada keadaan *bootstrap* $\hat{F}$ menghasilkan x^* dengan penyampelan random yang memberikan hasil $\hat{\theta}^* = s(x^*)$. Hanya terdapat satu nilai observasi dari $\hat{\theta}$ tetapi mampu menghasilkan banyak replikasi *bootstrap* $\hat{\theta}^*$.

2. Estimasi *Standar Error* Metode *Bootstrap*

Menurut Efron dan Tibshirani (1993) *standard error* merupakan ukuran yang paling sederhana untuk mengukur keakuratan dalam perhitungan statistika. Dalam model regresi, *standard error* adalah kesalahan baku atau simpangan baku dari *e*, dimana *e* adalah *residual* atau selisih antara *y* dengan *y* hasil estimasi $(\hat{y})$.

Menurut Efron dan Tibshirani (1993) *bootstrap* merupakan sebuah metode yang berbasiskan komputer untuk mengestimasi *standard error* dari $\hat{\theta}$ (nilai dugaan bagi parameter populasi θ). *Bootstrap* menyediakan keakuratan estimasi dengan menggunakan prinsip *plug-in* yaitu menggantikan populasi dengan sampel yang dianggap mewakili. Estimasi *bootstrap* untuk *standar error* tidak memerlukan perhitungan teori dan selalu tersedia walaupun sekompleks apapun perhitungan statistika untuk estimator $\hat{\theta}$, artinya prosedur *bootstrap* untuk *standar error* selalu sama untuk semua bentuk distribusi data. Algoritma *bootstrap* untuk mengestimasi *standard error* adalah sebagai berikut :

1. Pilih sampel *bootstrap*, yaitu sampel yang telah di *resampling* dari sampel aslib, dinotasikan dengan $x_1^*, x_2^*, \ldots, x_B^*$ masing-masing berisi nilai data yang telah disampling secara random dengan pengembalian dari sampel *x*.
2. Evaluasi hasil *bootstrap* yang telah diperoleh untuk masing-masing sampel *bootstrap*. $\hat{\theta}_b^* = x_b^*$, dengan $b = 1, 2, \ldots, B$
3. Mengestimasi *standard error* dari $se_{F(\hat{\theta})}$ dengan sampel *standard deviasi* dari replikasi *B*.

$$\widehat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b{}^{*} - \hat{\theta}^{*})^2} \qquad (2.5)$$

Dengan $\hat{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*$

3. Interval Konfidensi *Bootstrap* Pendekatan Normal

Pada interval konfidensi *bootstrap* pendekatan normal sebenarnya sama dengan interval konfidensi klasik untuk keperluan tersebut dibutuhkan estimasi *standard error* berdasarkan sampel *bootstrap*. Interval konfidensi *bootstrap* pendekatan normal $1 - \alpha$ dituliskan sebagai :

$$\hat{\theta}^* - Z_{\alpha/2}\widehat{se}_B \leq \theta \leq \hat{\theta}^* - Z_{1-\alpha/2}\widehat{se}_B \qquad (2.6)$$

Dengan $Z_{\alpha/2}$ merupakan nilai Z yang luas daerah sebelah kanan di bawah kurva normal baku dengan $\alpha/2$.

1.2.4 Prosedur *Resampling Bootstrap*

Sesuai dengan konsep dasar *bootstrap* yaitu melakukan *resampling* dengan cara pengembalian. Prosedur *resampling* dapat dituliskan sebagai berikut :

1. Mengkonstruksi distribusi empiris $\hat{F}^n$ dari suatu sampel dengan memberikan probabilitas $\frac{1}{n}$ pada setiap x_i, dimana i= 1, 2, . . ., n.
2. Mengambil sampel *bootstrap* berukuran n secara random dengan pengembalian dari distribusi empiris $\hat{F}^n$ disebut sebagai sampel *bootstrap* pertama x_1^*.
3. Menghitung statistik θ yang diinginkan dari sampel *bootstrap* x_1^*. disebut sebagai $\hat{\theta}_1^*$.
4. Mengulangi langkah 2 dan 3 hingga B kali, diperoleh $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$.

5. Mengkonstruksi suatu distribusi probabilitas dari $\hat{\theta}_b^*$ untuk b = 1, 2, … ,b, dengan memberikan probabilitas $\frac{1}{B}$ pada setiap $\theta_1^*, \theta_2^*, \ldots, \theta_B^*$. Distribusi tersebut merupakan estimator *bootstrap* untuk distribusi sampling $\hat{\theta}^*$ dan dinotasikan $\hat{F}^*$.
6. Pendekatan estimasi *bootstrap* untuk $\hat{\theta}$ adalah mean dari distribusi $\hat{F}^*$, yaitu:

$$\hat{\theta}^* = \frac{1}{B}\sum_{b}^{B}\hat{\theta}_b^* \qquad (2.7)$$

Pendekatan *bootstrap* jika diulang lebih dari satu kali akan memberikan hasil yang berbeda, hal ini karena yang dilakukan adalah suatu simulasi. Jika dapat dilakukan menggunakan semua kemungkinan sampel yaitu n^n maka hasilnya akan sama. Untuk keperluan estimasi bias, menurut Efron dan Tibshirani (1993), pendekatan metode *bootstrap* memberikan hasil estimasi untuk estimator sebagai:

$$\widehat{bias}_B = \hat{\theta}^* - \hat{\theta} \qquad (2.8)$$

Dimana $\hat{\theta}$ adalah estimator berdasarkan sampel asli.

1.2.5 *Bootstrap* Untuk Mengestimasi Parameter Model Regresi Linear

Boostrap untuk regresi dibagi menjadi 2 pendekatan berdasarkan variabel independennya yaitu variabel tetap dan variabel random. Apabila variabel independennya adalah variabel tetap maka digunakan metode Boostrap *resampling* residual. Apabila variabel independennya adalah variabel random maka digunakan metode *bootstrap resampling pairs*.

1. *Bootstrap Resampling* Residual untuk Mengestimasi Parameter Model Regresi Linear

Dimisalkan sebuah model regresi linear dengan variabel dependen y dan variabel independen x_k untuk $k = 1,2, \ldots, p$. Apabila variabel independen bersifat tetap maka digunakan *resampling* residual. Kemudian diambil sampel *random* berukuran n untuk data berpasangan yang dapat ditulis sebagai : $w_i = (y_i, x_{1i}, x_{2i}, \ldots, x_{pi})$ untuk $i = 1,2, \ldots, n$. Karena ukuran sampel yang diambil n, maka akan diperoleh observasi w_1, w_2, w_n. Estimasi koefisien regresi dilakukan berdasar sampel asli dengan metode kuadrat terkecil diperoleh :

$$\hat{\beta} = (x^T x)^{-1} x^T y \qquad (2.9)$$

Dengan

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, x = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \qquad (2.10)$$

Berdasarkan estimasi tersebut dapat ditentukan model fit untuk persamaan

regresi, yaitu

$$\hat{y} = x\hat{\beta}$$

Setelah itu masing-masing observasi dapat dihitung nilai residualnya, yaitu dengan rumus $e_i = y_i - \hat{y}_i$ untuk i=1, 2, . . ., n diperoleh $e = (e_1, e_2, \ldots, e_n)$. Kemudian diberikan distribusi empiris pada data residual, yaitu dengan memberikan probabilitas $\frac{1}{n}$ untuk setiap data. *Resampling* dilakukan pada data residual dengan mengambil sampel berukuran *n* dari data residual dengan pengembalian. Hal ini berarti bahwa setiap data dapat muncul lebih dari satu kali. Proses ini diulang sebanyak *B* kali, akan diperoleh e^{*1}, $e^{*2}, \ldots, e^{*B}$. Dipandang sampel pertama, $e^{*1} = e_1{}^{*1}, e_2{}^{*1}, \ldots, e_n{}^{*1})'$, dapat dihitung nilai *bootstrap* untuk *y*, yaitu:

$$y^{*1} = x\hat{\beta} + e^{*1}$$

Untuk *B* sampel yang akan diperoleh $\boldsymbol{y}^{*b}$ dengan b= 1, 2, . . ., *B*. Estimasi koefisien regresi dengan metode kuadrat terkecil berdasarkan sampel *bootstrap* adalah :

$$\hat{\beta} = (x^T x)^{-1} x^T y^{*b} \qquad (2.11)$$

Untuk *b* = 1,2,. . ., *B* dengan,

$$\hat{\beta}^{*b} = \begin{bmatrix} \hat{\beta}_0^{*b} \\ \hat{\beta}_1^{*b} \\ \hat{\beta}_2^{*b} \\ \vdots \\ \hat{\beta}_p^{*b} \end{bmatrix}, y^{*b} = \begin{bmatrix} y_1^{*b} \\ y_2^{*b} \\ \vdots \\ y_n^{*b} \end{bmatrix}, x = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \qquad (2.12)$$

Dengan memberikan distribusi empiris untuk $\boldsymbol{\beta}^{*b}$, maka pendekatan *bootstrap* untuk estimasi koefisien regresi merupakan mean dari distribusi empiris tersebut, yaitu :

$$\hat{\beta}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{*b} \qquad (2.13)$$

Hasil estimasi pada persamaan digunakan untuk mengestimasi koefisien regresi, sehingga diperoleh pendekatan *bootstrap* untuk estimasi model regresi linear dapat dinyatakan sebagai :

$$\hat{y} = \hat{\beta}^{*} x$$

Prosedur pendekatan *bootstrap* untuk regresi berdasarkan pada *resampling* residual menurut Sahinler dan Topuz (2007) dituliskan sebagai berikut :

1. Menentukan fit model regresi berdasarkan pada sampel asli dengan menggunakan metode kuadrat terkecil, $\hat{y}$.
2. Menghitung nilai dari residual $e_i = y_i - \hat{y}_i$, diperoleh $(e_1, e_2, \dots, e_n)$
3. Mengambil sampel *bootstrap* berukuran *n* dari $e_1, e_2, \dots, e_n$ dengan pengembalian, misalkan diperoleh sampel *bootstrap* pertama $e^{*1} = \left(e_1^{*1}, e_2^{*1}, \dots, e_n^{*1}\right)^T$
4. Menghitung nilai *bootstrap* **y** dengan menambahkan e^{*i} pada fit model regresi
$$y^{*1} = \hat{\beta}_x + e^i$$
5. Diperoleh estimasi metode kuadrat terkecil untuk sampel *bootstrap* yang pertama
$$\hat{\beta}^{*1} = (x^T x)^{-1} x^T y^{*1}$$
6. Proses di atas dilakukan sebanyak *B* kali, diperoleh $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$
7. Dikosntruksikan distribusi empiris untuk $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$ yaitu $\hat{F}^*$
8. Pendekatan *bootstrap* untuk menduga koefisien regresi linear merupakan

    mean dari distribusi empiris yaitu :
$$\hat{\beta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{*b}$$

2. *Bootstrap Resampling Pairs* untuk Mengestimasi Parameter Regresi Linear

Dimisalkan sebuah model regresi linear dengan variabel dependen *y* dan variabel independen sebanyak *p*, x_k untuk *k=1, 2, . . ., p*. Model regresi untuk populasi yang diberikan adalah

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \quad (2.14)$$

Kemudian diambil sampel *random* berukuran n untuk data berpasangan yang dapat ditulis sebagai $w_i = (y_i, x_{1i}, x_{2i}, \dots, x_{pi})$ untuk *i*=1, 2, . . ., *n*. Karena ukuran sampel yang diambil *n*, maka akan diperoleh observasi $w_1, w_2, \dots, w_n$. Dalam melakukan inferensi, distribusi yang tidak diketahui dan kecilnya ukuran sampel dapat diatasi dengan metode *bootstrap*. Untuk *n* yang kecil perlu diadakan *resampling* sebanyak B kali. *Resampling* data menggunakan metode *bootstrap* dilakukan melalui suatu pengambilan sampel berukuran n secara *random* dari sampel asli dengan pengembalian. *Resampling* dengan cara seperti ini akan diperoleh sampel *bootstrap* yang berukuran sama dengan sampel asli. Karena dilakukan dengan pengembalian maka sangat dimungkinkan setiap observasi dapat muncul lebih dari satu kali atau tidak muncul sama sekali. Misalkan data sampel asli yang dimiliki adalah $w = (w_1, w_2, \dots, w_n)'$, maka dengan metode *bootstrap* diperoleh sampel *bootstrap* $w_1^*, w_2^*, \dots, w_B^*$. Misalkan sampel *bootstrap* yang muncul adalah :

Sampel *bootstrap* ke-1, $w_1^* = (w_2, w_1, w_2, \dots, w_2)'$

Sampel *bootstrap* ke-2, $w_2^* = (w_3, w_3, w_1, \dots, w_2)'$

Sampel *bootstrap* ke-3, $w_3^* = \cdots$

Sampel *bootstrap* ke-*B*, $w_B^* = (w_2, w_1, w_3, \dots, w_1)'$

Pada setiap sampel *bootstrap* yang telah diperoleh digunakan untuk mengestimasi koefisien regresi dengan metode kuadrat terkecil. Misalkan pada sampel *bootstrap* pertama $w^{*1} = ({w_1}^{*1}, {w_2}^{*1}, \dots, {w_n}^{*1})'$ yang dapat dijabarkan seperti berikut :

$${w_1}^{*1} = ({y_1}^{*1}, {x_{11}}^{*1}, {x_{21}}^{*1}, \dots, {x_{p1}}^{*1})$$

$${w_2}^{*1} = ({y_2}^{*1}, {x_{12}}^{*1}, {x_{22}}^{*1}, \dots, {x_{p2}}^{*1})$$

$${w_3}^{*1} = ({y_3}^{*1}, {x_{13}}^{*1}, {x_{23}}^{*1}, \dots, {x_{p3}}^{*1})$$

$$\vdots$$

$${w_n}^{*1} = ({y_n}^{*1}, {x_{1n}}^{*1}, {x_{2n}}^{*1}, \dots, {x_{pn}}^{*1}) \qquad (2.15)$$

Dimana x_{ki}^{*1} untuk k=1, 2, . . ., p adalah variabel independen yang terpilih bersesuaian dengan y_i^{*1}. Metode kuadrat terkecil memberikan hasil sebagai :

$$\hat{\beta}^{*1} = ({x^{*1}}^T x^{*1})^{-1} {x^{*1}}^T y^{*1}$$

Untuk b= 1, 2, . . ., B dengan,

$$\hat{\beta}^{*1} = \begin{bmatrix} \hat{\beta}_0^{*1} \\ \hat{\beta}_1^{*1} \\ \hat{\beta}_2^{*1} \\ \vdots \\ \hat{\beta}_p^{*1} \end{bmatrix}, y^{*1} = \begin{bmatrix} y_1^{*1} \\ y_2^{*2} \\ \vdots \\ y_n^{*1} \end{bmatrix}, x^{*1} = \begin{bmatrix} 1 & x_{11}^{*1} & x_{21}^{*1} & \cdots & x_{pi}^{*1} \\ 1 & x_{12}^{*1} & x_{22}^{*1} & \cdots & x_{p2}^{*1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n}^{*1} & x_{2n}^{*1} & \cdots & x_{pn}^{*1} \end{bmatrix} \qquad (2.16)$$

Estimasi kuadrat terkecil pada persamaan dilakukan untuk semua data sampel *bootstrap* $w^{*1}, w^{*2}, \dots, w^{*B}$. Selanjutnya dapat diperoleh estimasi koefisien regresi $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$. Dikonstruksikan distribusi empiris untuk $\hat{\beta}^{*B}$, *b=1, 2,*

. . ., B yaitu $\hat{F}^*$. Pendekatan *bootstrap* untuk menduga koefisien regresi adalah mean dari distribusi $\hat{F}^*$, dapat dituliskan sebagai berikut :

$$\hat{\beta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{*b}$$

Hasil estimasi pada persamaan digunakan sebagai penduga koefisien regresi sehingga diperoleh pendekatan *bootstrap* untuk estimasi model regresi linear sebagai :

$$\hat{y} = \hat{\beta}^* x$$

Misalkan $\hat{F}_n$ fungsi distribusi empiris untuk sampel asli, menurut Sahinler dan Topuz (2007), prosedur pendekatan *bootstrap* estimasi parameter model regresi linear berdasarkan pada *resampling* observasi dapat dituliskan sebagai berikut :

1. Mengambil sampel *random* sederhana berukuran n dengan pengembalian dari sampel asli berdasarkan fungsi distribusi empiris $\hat{F}_n$sebut w^{*1} sebagai sampel *bootstrap* pertama.
2. Melakukan estimasi koefisien regresi menggunakan metode kuadrat terkecil pada sampel w^{*1} diperoleh $\hat{\beta}^{*1}$ sebagai :

$$\hat{\beta}^{*1} = (x^{*1T}x^{*1})^{-1}x^{*1T}y^{*1}$$

3. Proses di atas dilakukan sebanyak B kali, diperoleh $\hat{\beta}^{*1}$, $\hat{\beta}^{*2}$, …, $\hat{\beta}^{*B}$
4. Dikosntruksikan distribusi empiris untuk $\hat{\beta}^{*1}$, $\hat{\beta}^{*2}$, …, $\hat{\beta}^{*B}$ yaitu $\hat{F}^*$
5. Pendekatan *bootstrap* untuk estimasi koefisien regresi linear merupakan mean dari distribusi empiris $\hat{F}^*$ yaitu

$$\hat{\beta}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{*b}$$

6. *Bootstrap* untuk estimasi model regresi linear diberikan sebagai :

$$\hat{y} = \hat{\beta}^{*}x$$