BAB III

ROSNER REGRESSION METHOD DAN INTRACLASS CORRELATION COEFFICIENT

3.1 Pengertian ***Dilution Bias***

Pada kehidupan sehari-hari, seringkali ingin diketahui hubungan dari dua hal. Contohnya adalah hubungan antara rata-rata asupan garam dalam setahun dengan tekanan darah seseorang.

Untuk mengetahui hubungan tersebut dicari dengan menggunakan regresi linier dengan metode least square. Untuk mencari hubungan tersebut maka harus ditaksir parameter-parameter persamaan regresi.

Karena melibatkan variabel yang berupa rata-rata, maka dalam penaksiran parameter regresi harus menggunakan pengukuran yang berulang untuk individu yang sama, agar nilai taksiran nilai parameter regresi mendekati nilai parameter regresi yang sebenarnya. Namun karena keterbatasan sumber daya maka seringkali penaksiran parameter regresi menggunakan data yang diperoleh dengan hanya satu kali pengukuran saja untuk satu individu. Oleh karena itu hasil taksiran koefisien regresi yang dihasilkan akan menjadi tidak sesuai dengan koefisien regresi yang sebenarnya. Hal ini yang disebut *dilution bias*.

Jadi dilution bias adalah ketidaksesuaian koefisien regresi hasil taksiran, dengan koefisien regresi yang seharusnya dengan metode least square karena kurangnya pengulangan dalam pengambilan data.

3.2 Formulasi *dilution bias*

Misalkan ingin diketahui hubungan antara Y dengan X (X adalah variabel independen yang nilainya berupa rata-rata jangka panjang). Dimana hubungan yang seharusnya antara Y dan X adalah $y_i = \alpha * + \beta * x_i + \delta_i$, $\delta \sim N(0, \phi^2)$.

Karena X adalah variabel yang nilainya berupa rata-rata, maka untuk menaksir koefisien regresi antara Y dengan X harus digunakan pengukuran yang berulang untuk satu individu yang sama.

Namun pada saat pengukuran seringkali bukan nilai x yang digunakan untuk menaksir parameter regresi tersebut melainkan w, dimana w adalah nilai dari satu kali pengukuran suatu individu. Hubungan antara x dan w adalah $w_i = x_i + u_i \qquad u_i \sim N(0, \sigma_w^{\ 2})$, $x_i \sim N(\mu \ , \sigma_b^{\ 2})$

Karena X merupakan bagian dari W, maka besarnya pengaruh dari X terhadap W adalah sebesar $\sigma_b^{\ 2} / (\sigma_b^{\ 2} + \sigma_w^{\ 2})$, sehingga $X = W\sigma_b^{\ 2} / (\sigma_b^{\ 2} + \sigma_w^{\ 2})$.

Jadi besarnya faktor bias yang terjadi adalah $\sigma_b^{\ 2} / (\sigma_b^{\ 2} + \sigma_w^{\ 2})$.

Karena hubungan antara X dengan W adalah $X = W\sigma_b^{\ 2} /(\sigma_b^{\ 2} + \sigma_w^{\ 2})$ sehingga apabila nilai X tersebut disubstitusi pada model yang seharusnya antara Y dengan X, maka akan diperoleh hubungan antara Y dengan W adalah

$$y_i = \alpha + \beta w_i + \gamma_i \qquad \gamma_i \sim N(0, \psi^2) \tag{1}$$

dengan $\beta = \beta * \sigma_b^{\ 2} /(\sigma_b^{\ 2} + \sigma_w^{\ 2})$.

Komponen $\sigma_b^{\ 2} /(\sigma_b^{\ 2} + \sigma_w^{\ 2})$ pada persamaan ditas disebut faktor bias, oleh karena terdapat faktor bias tersebut sehingga pada saat penaksiran nilai $\widehat{\beta}$ yang didapat akan menjadi tidak sesuai dengan nilai $\widehat{\beta}$ yang sesungguhnya diinginkan .

Agar nilai β yang didapat sama dengan nilai $\beta *$, maka β harus dikalikan faktor koreksi λ, dimana $\lambda = \dfrac{\sigma_b^{\ 2} + \sigma_w^{\ 2}}{\sigma_b^{\ 2}}$ (2)

Karena nilai λ yang sebenarnya seringkali tidak ada, maka akan dicari nilai $\widehat{\lambda}$ yang akan digunakan untuk menggantikan nilai λ yang sesungguhnya. Dalam menaksir nilai $\widehat{\lambda}$ digunakan pengukuran yang berulang untuk individu-individu tertentu, dan individu yang akan diulang pengukurannya dipilih secara sembarang.

Terdapat beberapa metode untuk memperbaiki dilution bias, namun pada skripsi ini hanya akan dibahas 2 metode yaitu Rosner Regression Method dan Intraclass Correlation Coefiicient.

3.3 Rosner Regression Method (RRM)

Metode ini menggunakan hubungan antara pengukuran kedua dan pengukuran pertama untuk menaksir nilai $\hat{\lambda}$, dimana $\hat{\lambda}$ yang dihasilkan pada metode ini disebut $\hat{\lambda}_{REG}$ untuk membedakan dengan nilai $\hat{\lambda}$ yang akan didapat dari metode berikutnya.

Hubungan yang didapat tersebut yang akan digunakan untuk memperbaiki *dilution bias*. Pada metode ini hubungan antara pengukuran pertama (w_{i1}) dengan kedua (w_{i2}) didapat dengan meregresikan pengukuran kedua (w_{i2}) terhadap (w_{i1}). Sehingga diperoleh koefisien regresi $\hat{\beta}$ yang kemudian akan disebut $\hat{\beta}_{REG}$. Hal ini digunakan untuk membedakan $\hat{\beta}_{REG}$ dengan beberapa $\hat{\beta}$ yang sebelumnya dan $\hat{\lambda}_{REG}$ diestimasi dengan menggunakan $\hat{\lambda}_{REG} = \hat{\beta}_{REG}{}^{-1}$, sebagai kasus khusus dari persamaan (1) dengan $y_i = w_{i2}$ dan $\beta^* = 1$, maka hubungan regresi antara w_{12} dengan w_{11} adalah

$$w_{12} = \alpha + \frac{\sigma_b{}^2}{\sigma_b{}^2 + \sigma_w{}^2} w_{11} + \gamma_i \quad , \ \gamma_i \sim N(0, \psi^2) \qquad (4)$$

Sehingga $\widehat{\beta}_{REG}$ akan menjadi estimator dari $\frac{\sigma_b{}^2}{\sigma_b{}^2 + \sigma_w{}^2}$, sehingga estimator dari

$$\widehat{\lambda}_{REG} = \widehat{\beta}_{REG}{}^{-1} = \frac{\sum (wi1 - \overline{w}.1)^2}{\sum (wi1 - \overline{w}.1)(wi2 - \overline{w}.2)} \quad (5)$$

Apabila yang ingin dicari hanya taksiran faktor koreksinya, untuk memperbaiki model yang didapat dari satu kali pengamatan saja, maka $\widehat{\lambda}_{REG}$ sudah cukup. Namun karena pada skripsi ini ingin dibandingkan 2 metode,maka harus dicari variansi dari hasil metode ini.

Namun dari kedua metode yang dibahas pada skripsi ini, metode yang lebih baik untuk memperbaiki dilution bias adalah metode yang menaksir faktor koreksi yang memiliki variansi yang lebih kecil dibandingkan metode yang lain. Oleh karena itu akan dicari variansi dari $\widehat{\lambda}$ yang didapat untuk setiap metode.

Disamping membandingkan angka, yang kemungkinan berubah-ubah apabila digunakan pengukuran yang berbeda, maka dicari secara matematis, variansi $\widehat{\lambda}$ yang didapat dari setiap metode.

Untuk mengestimasi variansi $\widehat{\lambda}_{REG}$, maka harus mengambil variansi dari persamaan (4) untuk memperoleh var(γ) :

$$\text{var}(w_{12}) = \text{var}(\alpha) + \text{var}\left(\frac{\sigma_b{}^2}{\sigma_b{}^2 + \sigma_w{}^2} w_{11}\right) + \text{var}(\gamma_i)$$

Seperti sudah diketahui bahwa $\text{var}(w_{12}) = \sigma_b{}^2 + \sigma_w{}^2$, maka persamaan diatas menjadi:

$$\sigma_b{}^2 + \sigma_w{}^2 = 0 + \left(\frac{\sigma_b{}^2}{\sigma_b{}^2 + \sigma_w{}^2}\right)^2 \sigma_b{}^2 + \sigma_w{}^2 + \text{var}(\gamma_i)$$

$$\text{var}\ (\gamma) = \left(1 - \frac{\sigma_b{}^4}{(\sigma_b{}^2 + \sigma_w{}^2)^2}\right)\sigma_b{}^2 + \sigma_w{}^2$$

$$\text{var}\ (\gamma) = \left(\frac{(\sigma_b{}^2 + \sigma_w{}^2) - \sigma_b{}^4}{(\sigma_b{}^2 + \sigma_w{}^2)^2}\right)\sigma_b{}^2 + \sigma_w{}^2$$

$$\text{var}\ (\gamma) = \psi^2 = \frac{(\sigma_b{}^2 + \sigma_w{}^2) - \sigma_b{}^4}{\sigma_b{}^2 + \sigma_w{}^2} \tag{6}$$

Variansi dari $\hat{\beta}_{REG}$ dan $\hat{\lambda}_{REG}$ dapat diestimasi dari persamaan (6), dan karena besarnya variansi model adalah variansi error dan menggunakan least square, maka

$$\text{var}(\hat{\beta}_{REG}) = \frac{\left(\dfrac{(\sigma_b{}^2 + \sigma_w{}^2) - \sigma_b{}^4}{\sigma_b{}^2 + \sigma_w{}^2}\right)}{\sum (wi1 - \bar{w}.1)^2}$$

$$\approx \frac{(\sigma_b{}^2 + \sigma_w{}^2) - \sigma_b{}^4}{n(\sigma_b{}^2 + \sigma_w{}^2)}$$

Dan bila diekspresikan dalam bentuk λ dari persamaan (2) menjadi

$$= \frac{1 - 1/\lambda^2}{n}$$

Seperti sudah diketahui bahwa pada metode ini bahwa $\hat{\lambda}_{REG} = \hat{\beta}_{REG}{}^{-1}$ maka

$$\operatorname{var}(\hat{\lambda}_{REG}) = \operatorname{var}(\frac{1}{\hat{\beta}_{REG}}) \approx \frac{\operatorname{var}(\hat{\beta}_{REG})}{\beta^4} \approx \frac{\lambda^2(\lambda^2 - 1)}{n} \qquad (7)$$

3.3 Intraclass Correlation Coefficient (ICC)

Metode ini menaksir faktor koreksi $\hat{\lambda}$ yang kemudian akan disebut sebagai $\hat{\lambda}_\rho$ dengan menggunakan $\hat{\rho} = \hat{\sigma}_b{}^2 \big/ \left(\hat{\sigma}_b{}^2 + \hat{\sigma}_w{}^2\right)$. Penggunaan $\hat{\rho}$ karena diharapkan nilai $\hat{\rho}$ mendekati nilai ρ. Karena data yang digunakan untuk menaksir $\hat{\lambda}_\rho$ adalah data yang berulang maka akan terdapat variansi antar kelas (σ_w^2) sehingga akan terdapat intraclass correlation coefficient. seperti sudah diketahui bahwa dilution bias terjadi karena adanya faktor bias $\sigma_b{}^2 / (\sigma_b{}^2 + \sigma_w{}^2)$, maka nilai $\hat{\lambda}_\rho = \dfrac{1}{\hat{\rho}} = \left(\hat{\sigma}_b{}^2 + \hat{\sigma}_w{}^2\right) \big/ \hat{\sigma}_b{}^2$

Pada anova dengan pengulangan 2 kali, mean dari sum of squares antara individu dinyatakan dengan

$$\hat{s}_b{}^2 = 2 \sum \left(\overline{w}_{i.} - \overline{w}_{..}\right)^2 \big/ (n - 1)$$

dan Mean Square within individu dengan

$$\hat{s}_w{}^2 = \sum\sum\left(\bar{w}_{ij} - \bar{w}_{i.}\right)^2 \Big/ n$$

Karena S^2 adalah penaksir yang bias untuk σ^2, maka dicari taksiran tak bias dari untuk $\sigma_w{}^2$ dan $\sigma_b{}^2$.

Pada anova dengan 2 pengulangan terdapat n pengamatan pada masing-masing pengulangan, maka korelasi antara pasangan pasangan angka tersebut adalah $\rho = \sigma_b{}^2 \big/ \left(\sigma_b{}^2 + \sigma_w{}^2\right)$, sehingga untuk variansi Within dalam setiap kelas adalah

$$\sum_{i=1}^{2}\sum_{j=1}^{n}\left(\bar{w}_{ij} - \bar{w}_{i.}\right)^2 = 2\left(n-1\right)\sigma_w{}^2$$

$$SSW = 2\left(n-1\right)\hat{\sigma}_w{}^2$$

dengan $\hat{\sigma}_w{}^2$ adalah penaksir $\sigma_w{}^2$

Mean dari observasi pada setiap kelas terdiri dari dua bagian yaitu, bagian pertama dengan variansi antar individu dalam kelas ($\sigma_b{}^2$), dan bagian kedua adalah mean dari variansi antar kelas, dengan mean dari n nilai memiliki variansi $\sigma_w{}^2 / n$. Sehingga variansi dari mean dari setiap kelas adalah

$$\mathrm{n}\sum_{1}^{2}\left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 = \left(n-1\right)\left(2\sigma_b{}^2 + \sigma_w{}^2\right)$$

$$SSB = \left(n-1\right)\left(2\hat{\sigma}_b{}^2 + \hat{\sigma}_w{}^2\right)$$

sehingga $SSW/2(n-1)=\hat{\sigma}_w^{\,2}$ =Mean square Within, dan

$\hat{\sigma}_b^{\,2} = \dfrac{SSB/(n-1)-\hat{\sigma}_w^{\,2}}{2} = \dfrac{MSB-MSW}{2}$, dengan $\hat{\sigma}_b^{\,2}$ adalah penaksir $\sigma_b^{\,2}$ dan $\hat{\sigma}_w^{\,2}$ adalah penaksir $\sigma_w^{\,2}$

sehingga $\hat{\sigma}_b^{\,2} = \left(\hat{s}_b^{\,2}-\hat{s}_w^{\,2}\right)/2$ dan $\hat{\sigma}_w^{\,2} = \hat{s}_w^{\,2}$ mengestimasi $\sigma_b^{\,2}$ dan $\sigma_w^{\,2}$,

sehingga λ dapat diestimasi oleh $\hat{\lambda}_\rho$ dimana

$$\hat{\lambda}_\rho = \frac{1}{\hat{\rho}} = 1+\frac{\hat{\sigma}_w^{\,2}}{\hat{\sigma}_b^{\,2}}$$

$$\hat{\lambda}_\rho = 1+\frac{2\hat{s}_w^{\,2}}{\hat{s}_b^{\,2}-\hat{s}_w^{\,2}} \qquad (9)$$

Apabila yang ingin dicari hanya taksiran faktor koreksinya saja untuk memperbaiki model yang didapat dari satu kali pengamatan , maka hasil diatas sudah cukup, namun karena pada skripsi ini ingin dibandingkan 2 metode,maka harus dicari variansi dari hasil metode ini.

Seperti pada RRM agar dapat ditentukan metode yang lebih baik untuk memperbaiki dilution bias, maka dicari variansi dari $\hat{\lambda}_\rho$

Variansi dari $\hat{\lambda}_\rho$ dapat dihitung melalui hubungan antara $\hat{\lambda}_\rho$ dan F-statistik $\left(F=\left(s_b^{\,2}/s_w^{\,2}\right)\right)$, dan karena $\mathrm{SSB}/n\sigma_b^2+\sigma_w^2 \sim \chi^2_{n-1}$, dan $\mathrm{SSW}/\sigma_w^2 \sim \chi^2_{a(n-1)}$, maka $\dfrac{MSB}{n\sigma_b^2+\sigma_w^2}\Big/\dfrac{\mathrm{MSW}}{\sigma_w^2} \sim F^{a-1}_{a(n-1)}$ dan distribusi-F nya menjadi

$$F \sim \frac{2\sigma_b{}^2 + \sigma_w{}^2}{\sigma_w{}^2} F_{n-1,n}$$

$$= \left(1 + \frac{2}{\lambda - 1}\right) F_{n-1,n}$$ dari persamaan (2)

sehingga $$E(F) \approx 1 + \frac{2}{\lambda - 1} \tag{10}$$

dan

$$\operatorname{var}(F) = var\left(1 + \frac{2}{\lambda - 1}\right) \; var(F_{n-1,n})$$

$$\operatorname{var}(F) = \left(1 + \frac{2}{\lambda - 1}\right)^2 \frac{2n^{\,2}\left[1 + (n\; - 2)/n - 1\right]}{(n-2)^2 (n-4)}$$

$$\operatorname{var}(F) \approx \left(1 + \frac{2}{\lambda - 1}\right)^2 \frac{4}{n} \tag{11}$$

variansi dari $\hat{\lambda}$ dapat diturunkan dari persamaan (9)-(11) sebagai berikut

$$\hat{\lambda}_\rho = 1 + \frac{2s_w{}^2}{s_b{}^2 + s_w{}^2} = 1 + \frac{2}{F - 1}$$

dan

$$\operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \left\{\frac{2}{(F-1)^2}\right\}^2_{E(F)} \operatorname{var}(F)$$

$$\approx (\lambda - 1)^2 (\lambda + 1)^2 \frac{1}{n} \tag{12}$$

$$= \left(\lambda^2 - 1\right)^2 \frac{1}{n} \tag{13}$$

\

3.5 Perbandingan metode RRM dan ICC

Karena variansi taksiran $\hat{\lambda}$ dari RRM adalah $\operatorname{var}(\hat{\lambda}_{REG}) \approx \frac{\lambda^2(\lambda^2-1)}{n}$,dan variansi taksiran $\hat{\lambda}$ dari ICC adalah $\operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \left(\lambda^2-1\right)^2 \frac{1}{n}$,

Selisih variansinya sebesar

$$\operatorname{var}(\hat{\lambda}_{REG}) - \operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \frac{\lambda^2(\lambda^2-1)}{n} - \left(\lambda^2-1\right)^2 \frac{1}{n}$$

$$\operatorname{var}(\hat{\lambda}_{REG}) - \operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \frac{\lambda^2(\lambda^2-1)}{n} - \frac{\left(\lambda^2-1\right)^2}{n}$$

$$\operatorname{var}(\hat{\lambda}_{REG}) - \operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \frac{\lambda^4-\lambda^2}{n} - \frac{\left(\lambda^4-2\lambda^2-1\right)}{n}$$

$$\operatorname{var}\left(\hat{\lambda}_\rho\right) - \operatorname{var}(\hat{\lambda}_{REG}) \approx \frac{\lambda^4-2\lambda^2-1-\lambda^4-\lambda^2}{n}$$

$$\operatorname{var}(\hat{\lambda}_{REG}) - \operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \frac{\lambda^4-\lambda^2-\lambda^4+2\lambda^2+1}{n}$$

$$\operatorname{var}(\hat{\lambda}_{REG}) - \operatorname{var}\left(\hat{\lambda}_\rho\right) \approx \frac{\lambda^2+1}{n}$$

dapat dilihat $\operatorname{var}\left(\hat{\lambda}_\rho\right) < \operatorname{var}(\hat{\lambda}_{REG})$ sebesar $\frac{\lambda^2+1}{n}$, oleh karena itu metode ICC lebih baik dalam memperbaiki *dilution bias*.