

MODUL 5

ASUMSI KLASIK

Formula atau rumus regresi diturunkan dari suatu asumsi data tertentu. Dengan demikian tidak semua data dapat diterapkan regresi. Jika data tidak memenuhi asumsi regresi, maka penerapan regresi akan menghasilkan estimasi yang bias. Jika data memenuhi asumsi regresi maka estimasi (β) diperoleh akan bersifat BLUE yang merupakan singkatan dari: *Best, Linear, Unbiased, Estimator*.

Best artinya yang terbaik, dalam arti garis regresi merupakan estimasi atau ramalan yang baik dari suatu sebaran data. Garis regresi merupakan cara memahami pola hubungan antara dua seri data atau lebih. Garis regresi adalah *best* jika garis itu menghasilkan *error* yang terkecil. *Error* itu sendiri adalah perbedaan antara nilai observasi dan nilai yang diramalkan oleh garis regresi. Jika *best* disertai sifat *unbiased* maka estimator regresi disebut efisien.

Linear. Estimator β disebut *linear* jika estimator itu merupakan fungsi linear dari sampel.

$$\text{Rata-rata } \bar{X} = \frac{1}{n} \sum X = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

Adalah estimator yang *linear*, karena merupakan fungsi *linear* dari nilai-nilai X. Nilai2 OLS juga merupakan klas estimator yang linear.

Unbiased. Suatu estimator dikatakan *unbiased* jika nilai harapan dari estimator β sama dengan nilai yang benar dari β .

$$\text{Rata-rata } \beta = \beta$$

$$\text{Bias} = \text{Rata-rata } \beta - \beta$$

Metode OLS (*Ordinary Least Square*) yang dirumuskan di atas merupakan klas penaksir yang memiliki sifat BLUE. OLS akan memiliki sifat BLUE jika memenuhi asumsi-asumsinya, dari mana penurunan formula OLS diturunkan. Gujarati (1995) mendaftar 10 asumsi yang mejadi syarat penerapan OLS.

Asumsi 1: *Linear Regression Model*. Model regresi merupakan hubungan linear dalam parameter.

$$Y = a + bX + e$$

Untuk model regresi $Y = a + bX + cX^2 + e$

Walaupun variabel X dikuadratkan tetap merupakan regresi yang linear dalam parameter, sehingga OLS masih dapat diterapkan.

Asumsi 2: Nilai X adalah tetap dalam sampling yang diulang-ulang (*X fixed in repeated sampling*). Tepatnya bahwa nilai X adalah *nonstochastic* (tidak random).

Asumsi 3: variabel pengganggu e memiliki rata-rata nol (*zero mean of disturbance*). Ini berarti garis regresi pada nilai X tertentu tepat di tengah-tengah sehingga rata-rata *error* yang di atas regresi dan di bawah garis regresi kalau dijumlahkan hasilnya nol.

Asumsi 4: *Homoscedasticity* atau variabel pengganggu e memiliki *variance* yang sama sepanjang observasi dari berbagai nilai X. Ini berarti data Y pada setiap nilai X tertentu memiliki rentangan yang sama.

Homoscedasticity

Heteroscedasticity

Kedua gambar di atas menunjukkan situasi yang diasumsikan yaitu keadaan *homoscedasticity*, dan keadaan yang sebaliknya, yaitu kondisi *heteroscedasticity*.

Asumsi 5: No autocorrelation between the disturbance (tidak ada *otokoreasi* antara variabel e pada setiap nilai X_i dan X_j).

$$E(e|X_i)(e|X_j) = 0$$

Korelasi e_t dan e_{t-1} cukup rendah. Jelasnya perhatikan seri data pada contoh di atas.

Y	X ₁	X ₂	e_t	e_{t-1}
10.00	2.00	1.20	-1.4518	
12.00	2.20	1.40	-0.2035	-1.4518
14.00	2.30	2.00	0.4741	-0.2035
15.00	2.20	2.30	1.0927	0.4741
16.00	2.40	2.60	1.1517	1.0927
16.00	2.80	2.80	0.0269	1.1517
17.00	2.70	3.50	-0.1117	0.0269
18.00	3.00	4.00	0.6179	-0.1117
18.00	3.00	4.20	0.9965	0.6179
20.00	3.40	4.00	0.6360	0.9965
				0.6360

Jika korelasi e_t dan e_{t-1} rendah maka berarti tidak terdapat *otokorelasi* dari e .

Asumsi 6: variabel X dan disturbance e tidak berkorelasi. Ini berarti kita memisahkan pengaruh X atas Y dan pengaruh variabel e atas Y. Jika X dan e berkorelasi maka pengaruh keduanya akan tumpang tindih (sulit dipisahkan pengaruh masing-masing atas Y).

Asumsi ini pasti terpenuhi jika X adalah variabel *non random* atau *nonstochastic*.

Asumsi 7: Jumlah observasi atau besar sampel n harus lebih dari jumlah parameter yang diestimasi. Bahkan untuk menjamin terpenuhinya asumsi yang lain, sebaliknya n besar sampel harus cukup besar.

Asumsi 8: Variabel X harus memiliki variabilitas. Jadi tidak bias dilakukan regresi jika nilai X selalu sama sepanjang observasi.

Asumsi 9: Model regresi secara benar terspesifikasi. Tidak ada spesifikasi yang bias. Artinya, kita sudah memasukkan variabel yang direkomendasikan oleh teori dengan tepat. Atau juga kita tidak memasukkan variabel yang sembarangan yang tidak jelas kaitannya. Spesifikasi ini juga menyangkut bentuk fungsi apakah parameter linear, dan juga bentuk X linear (pangkat 1) atau kuadratik (berbentuk kurve U), atau kubik (bentuk S).

Asumsi 10: Tidak ada *multikolinearitas* antara variabel penjelas X_1 , X_2 dan X_n . Jelasnya korelasi antar variabel penjelas tidak boleh sempurna atau sangat tinggi.

Dari asumsi 10 di atas tidak semuanya perlu diuji. Sebagian cukup hanya diasumsikan sedangkan sebagian yang lain memerlukan test.

Penyimpangan masing-masing asumsi juga tidak sama impaknya terhadap regresi. Penyimpangan atau tidak terpenuhinya asumsi *multikolinearitas* (asumsi 10) tidak mengganggu sepanjang uji t sudah signifikan. Hal ini disebabkan oleh membesarnya *standar error* pada kasus *multikolinearitas*, sehingga jika $t = b/sb$ menjadi cenderung kecil sehingga jika t masih signifikan, maka *multikolinearitas* tidak perlu di atasi.

Sebaliknya, penyimpangan asumsi *homocedasticity* dan *autokorelasi* menyebabkan b pada sb sehingga $t = b/sb$ menjadi tidak menentu. Walaupun t sudah signifikan atau tidak signifikan tidak dapat memberi informasi yang sesungguhnya.

Untuk memenuhi asumsi-asumsi di atas estimasi regresi hendaknya dilengkapi dengan uji-uji yang diperlukan. Berikut beberapa teknik uji regresi yang disebut dengan uji asumsi klasik.