Pertemuan 4-5

ANALISIS REGRESI SEDERHANA

Metode Kuadrat Terkecil (OLS)

- Persoalan penting dalam membuat garis regresi sampel adalah bagaimana kita bisa mendapatkan garis regresi yang baik yaitu sedekat mungkin dengan datanya sehingga akan menghasilkan prediksi yang baik.
- Metode yang sering digunakan dalam analisis regresi adalah metode ordinary least square (OLS),

Metode Kuadrat Terkecil (OLS)

- Misalkan kita akan menganalisis hubungan antara jumlah permintaan barang dengan tingkat harga sebelumnya. Spesifikasi modelnya sebagai berikut:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{e}_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Dimana :

- Y_i : jumlah permintaan barang
- X_i : harga barang
- e_i : variable gangguan

Metode Kuadrat Terkecil (OLS)

- Persamaan (1) merupakan prediksi, dan nilai prediksi ini akan berbeda dengan nilai aktualnya. Perbedaan nilai aktualdengan nilai prediksi disebut residual. Persamaan (1) dapat ditulis kembali sbb:

$$Y_i = \ddot{Y}_i + \hat{e}_i \dots\dots\dots\dots\dots\dots\dots\dots\dots .. (2)$$

Metode Kuadrat Terkecil (OLS)

- Dalam bentuk persamaan yang lain residual bisa ditulis sbb:

$$\hat{e}_i = Y_i - \ddot{Y}_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$\hat{e}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Metode Kuadrat Terkecil (OLS)

- Konsep residual dalam persamaan regresi sampel bisa dijelaskan melalui garis regresi sampel. Misalkan kita mempunyai 6 data sampel jumlah permintaan barang (Y) dan harganya (X). Secara grafis variable residual ini merupakan selisih jarak antara jumlah permintaan yang diprediksi dari garis regresi dengan data aktualnya. Untuk mendapatkan garis regresi yang baik kita harus meminimumlkan jumlah variable residual yaitu $\sum \hat{e}_i = \sum (Y_i - \hat{Y}_i)$ sekecil mungkin.

Metode Kuadrat Terkecil (OLS)

- Jika kita menjumlahkan semua residual, kita mungkin bisa mendapatkan jumlah total residual sekecil mungkin. Namun kemungkinan kita mendapatkan jumlah total residual adalah nol juga bisa terjadi. Padahal faktanya ada beberapa residual yang mungkin jaraknya jauh dari garis regresi baik di bawahnya maupun di atasnya. Hal ini terjadi karena kita memberi timbangan yang sama kepada setiap residual tanpa melihat jauh dekat jaraknya.

Metode Kuadrat Terkecil (OLS)

- Hal tersebut bisa dihindari dengan memberi timbangan kepada masing-masing residual. Salahsatu caranya adalah dengan mengkuadratkan masing-masing residual. Dengan mengkuadratkannya maka kita memberi timbangan yang lebih besar kepada residual yang mempunyai jarak yang lebar.
- Metode mencari nilai residual sekecil mungkin dengan menjumlahkan kuadrat residual ini disebut dengan metode kuadrat terkecil *(ordinary least squares).*

Metode Kuadrat Terkecil (OLS)

- Metode OLS yang akan menjamin jumlah residual kuadrat sekecil mungkin dapat dijelaskan sbb:

$$\sum \hat{e}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$\sum \hat{e}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_1)^2$$

Metode Kuadrat Terkecil (OLS)

- Untuk mendapatkan residual yang sekecil mungkin maka dilakukan diferensiasi. Proses diferensiasi menghasilkan estimasi $\hat{\beta}_0$ dan $\hat{\beta}_1$ sebagai berikut :

$$\hat{\beta}_1 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

$$= \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \qquad = \frac{\sum x_i y_i}{\sum x_i^2}$$

Metode Kuadrat Terkecil (OLS)

$$\hat{\beta}_0 = \frac{n \sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum Y_i}{n} - \hat{\beta}_1 \frac{\sum X_i}{n}$$

$$= \bar{Y} - \hat{\beta}_1 \bar{X}$$

- Dimana $\bar{Y}$ dan $\bar{X}$ adalah rata-rata serta n adalah jumlah observasi.

ASUMSI OLS

- Hubungan antara Y (variable dependen) dan X (variable independen) adalah linier dalam parameter
- Variabel X adalah variable stokastik yang nilainya tetap
- Nilai harapan (expected value) atau rata-rata dari variable gangguan e_i adalah nol. Secara simbolis dapat dinyatakan sbb:

$$E(e_i|X_i) = 0$$

ASUMSI OLS

- Varian dari variable gangguan e_i adalah sama (homoskedastisitas). Secara simbolis dinyatakan sbb:

$$
\begin{aligned}
Var(e_i|X_i) &= E[e_i - E(e_i|X_i)]^2 \\
&= E\left(e_i^2\middle|X_i\right) \\
&= \sigma^2
\end{aligned}
$$

ASUMSI OLS

- Tidak ada serial korelasi antara residual e_i atau residual e_i tidak saling berhubungan dengan e_j yang lain. Secara simbolis dinyatakan sebagai berikut :

$$Cov(e_i, e_j | X_i, X_j) = E[(e_i - E(e_i)|X_i)][(e_j - E(e_j)|X_j)]$$

$$= E(e_i|X_i)(e_j|X_j)$$

$$= 0$$

ASUMSI OLS

- Variabel gangguan e_i berdistribusi normal

$$e \sim N(0, \sigma^2)$$

- Asumsi 1-5 membentuk model yang disebut model regresi linier klasik (classical linear regression model). Apabila asumsi-asumsi tersebut terpenuhi maka metode OLS akan menghasilkan estimator yang tidak bias, linier dan mempunyai varian yang minimum *(best linear unbiased estimators =BLUE)*.

ASUMSI OLS

Suatu estimator $\hat{\beta}_1$ dikatakan mempunyai sifat yang BLUE jika memenuhi criteria sbb:

1. Estimator $\hat{\beta}_1$ adalah linier, yaitu linier terhadap variable stokastik Y sebagai variable dependen
2. Estimator $\hat{\beta}_1$ tidak bias, yaitu nilai rata-rata atau nilai harapan E($\hat{\beta}_1$) sama dengan nilai $\hat{\beta}_1$ yang sebenarnya.
3. Estimator $\hat{\beta}_1$ mempunyai varian yang minimum. Estimator yang tidak bias dengan varian minimum disebut estimator yang efisien.

STANDARD ERROR

- Standard error digunakan untuk mengukur ketepatan estimasi dari estimator $\hat{\beta}_0$ dan $\hat{\beta}_1$. Formulasinya adalah sebagai berikut :

$$Var(\hat{\beta}_0) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

$$Se(\hat{\beta}_0) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2} \sigma}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$$

STANDARD ERROR

$$Se(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

- Dimana var adalah varian, se adalah standard error dan σ^2 adalah varian konstan.

  $x_i = X_i - \bar{X}$

- Semua variable dalam perhitungan standard error di atas dapat diestimasi dari data yang ada kecuali σ^2.

STANDARD ERROR

- Nilai estimasi dari σ^2 dapat dihitung dengan formula sbb:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-k}$$

- Dimana n adalah jumlah observasi dan k adalah jumlah variable independen ditambah konstanta (intersep).

KOEFISIEN DETERMINASI

- Koefisien determinasi mengukur persentase total variasi Y yang dijelaskan oleh garis regresi. Konsep koefisien determinasi dapat dijelaskan sbb:

$$Y_i = \ddot{Y}_i + \hat{e}_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (1)$$

KOEFISIEN DETERMINASI

- Kedua sisi persamaan dikurangi dengan nilai rata-rata Y.

$$Y_i - \bar{Y} = \ddot{Y}_i + \hat{e}_i - \bar{Y} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .(2)$$

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (4)$$

KOEFISIEN DETERMINASI

- Persamaan (4) dapat pula dinyatakan sbb:

  TSS = ESS + RSS
- Berdasarkan rumusan di atas R^2 dapat didefinisikan sebagai rasio antara ESS dengan TSS.

$$R^2 = \frac{ESS}{TSS}$$

$$= \frac{\sum(Y_i - \bar{Y})^2}{\sum\left(Y_i - \hat{Y}_i\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$$

KOEFISIEN DETERMINASI

- Karena TSS = ESS + RSS, maka :

$$R^2 = \frac{ESS}{TSS}$$

$$= \frac{TSS - RSS}{TSS}$$

$$= 1 - \frac{RSS}{TSS}$$

$$= 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

Koefisien Korelasi (r)

$$r = \frac{cov(X_i, Y_i)}{\sqrt{var(X_i)var(Y_i)}} \ldots\ldots\ldots\ldots\ldots(1)$$

Dimana :

$$cov(X_i, Y_i) = \frac{\sum_{n=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$var(X_i) = \frac{\sum_{n=1}^{n}(X_i - \bar{X})^2}{n-1} \ldots\ldots\ldots\ldots\ldots(3)$$

$$var(Y_i) = \frac{\sum_{n=1}^{n}(Y_i - \bar{Y})^2}{n-1} \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Koefisien Korelasi (r)

- Dengan demikian formulasi korelasi dapat ditulis sebagai berikut :

$$r = \frac{\sum_{n=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{n=1}^{n}(X_i - \bar{X})^2 \sum_{n=1}^{n}(Y_i - \bar{Y})^2}} \dots\dots\dots\dots\dots (5)$$

Pengujian hipotesis

<u>Uji t</u>

Langkah-langkah uji t :

1. Membuat hipotesis melalui uji satu sisi atau dua sisi

   Misal uji hipotesis yang digunakan adalah uji hipotesis satu sisi

   H_0 : $\beta_1 \geq 0$

   H_a : $\beta_1 < 0$

2. Menghitung nilai statistic t (t hitung) dan mencari nilai t kritis dari table distribusi t pada α dan *degree of freedom* tertentu. Adapun nilai t hitung dapat dicari dengan formula sbb: $t = \dfrac{\hat{\beta}_1 - \beta_1^*}{se(\hat{\beta}_1)}$

Pengujian hipotesis

- Dimana β_1^* merupakan nilai pada hipotesis nul

3. Membandingkan nilai t hitung dengan t kritisnya. Keputusan menolak atau menerima H_o adalah sbb:

- Jika nilai t hitung > nilai t kritis maka H_o ditolak atau menerima H_a
- Jika nilai t hitung > nilai t kritis maka H_o diterima atau menolak H_a
- Jika kita menolak H_o atau menerima H_a berarti secara statistic variable independen signifikan mempengaruhi variable dependen dan sebaliknya jika menerima H_o dan menolak H_a berarti secara statistic variable independen tidak signifikan mempengaruhi variable dependen.

Uji normalitas

Metode yang dipakai untuk mendeteksi apakah residual mempunyai distribusi normal atau tidak adalah sbb:

1. Histogram residual
2. Uji Jarque-Bera

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

Beberapa Model Fungsi Regresi

Tipe Fungsi Regresi	Model Regresi
Linier	$Y = \beta_0 + \beta_1 X + e$
Log – linier	$\ln(Y) = \beta_0 + \beta_1 X + e$
Linier – log	$Y = \beta_0 + \beta_1 \ln(X) + e$
Log – log	$\ln(Y) = \beta_0 + \beta_1 \ln(X) + e$
Resiprokal	$Y = \beta_0 + \beta_1 \frac{1}{X} + e$
Log inverse	$\ln(Y) = \beta_0 + \beta_1 \frac{1}{X} + e$