HAFTA 2

Hataların Olasılık Dağılımı:

Klasik doğrusal regresyon modelinde hata terimleri ortalaması sıfır ve varyansı σ^2 olan normal dağılıma sahip olduğu varsayılır. Çıkarsama yapılmak istendiğinde bilinen en küçük kareler yöntemi ile tahmin edilen regresyon katsayılarının nokta tahminleri yeterli değildir. Bu nedenle hata terimlerinin olasılık dağılımının bilinmesi gerekir. Çünkü $\hat{\beta}_0$ ve $\hat{\beta}_1$ nokta tahmin edicileri hata teriminin fonksiyonu olduğundan $\hat{\beta}_0$ ve $\hat{\beta}_1$ ‘nın istatistiki sonuç çıkarımları için dağılımlarının bilinmesi gerekir.

- Neden normal dağılım varsayımı yapılır?
  1. Merkezi limit teoremi gereği çok sayıda bağımsız ve aynı dağılıma sahip rasgele değişken var ise bu değişkenlerin sayısı “sonsuza doğru” arttıkça bu değişkenlerin toplamlarının dağılımı (birkaç aykırılık dışında) normal dağılıma yakınsar.
  2. Merkezi limit teoreminin bir başka özelliği değişken sayısı “çok büyük” olmasa ya da bu değişkenler “tam bağımsız dağılmasalar da” toplamlarının (ya da ortalamalarının) dağılımı yine normal dağılıma yakınsar.
  3. Normal dağılıma sahip rasgele değişkenlerin lineer fonksiyonlarının dağılımı da normaldir. Buna göre hata terimleri normal dağılıma sahipse $\hat{\beta}_0$ ve $\hat{\beta}_1$ rasgele değişkenleri hata terimlerinin lineer fonksiyonları olduğuna göre normal dağılıma sahiptirler.

$$Y = \beta_0 + \beta_1 x + \varepsilon = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \Rightarrow \hat{Y} = X\hat{\boldsymbol{\beta}} = P_X Y$$

$\boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \sigma^2 I)$ ve $Y|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta},\ \sigma^2 I)$ olacaktır. Bu bilgilerle;

$$E(\hat{Y}) = E(P_X Y) = P_X E(Y) = \underbrace{P_X \mathbf{X}}_{\mathbf{X}}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}$$

$$Cov(\hat{Y}) = Cov(P_X Y) = P_X Cov(Y) P_X' = P_X \sigma^2 I P_X' = \sigma^2 \underbrace{P_X P_X'}_{\text{idempotent}} = \sigma^2 P_X$$

$$\Rightarrow \hat{Y} \sim N(\mathbf{X}\boldsymbol{\beta},\ \sigma^2 P_X),$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y \sim N(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$E(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(Y) = \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\boldsymbol{\beta} = \boldsymbol{\beta}$$

$$Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underbrace{Cov(Y)}_{\sigma^2\mathbf{I}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2 I\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

$$= \sigma^2 \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Buradan;

$$E(\hat{\boldsymbol{\varepsilon}}) = E[(I - P_X)Y] = (I - P_X)E(Y) = \underbrace{(I - P_X)\mathbf{X}}_{\mathbf{0}}\boldsymbol{\beta} = \mathbf{0}$$

$$Cov(\hat{\boldsymbol{\varepsilon}}) = Cov[(I - P_X)Y] = (I - P_X)\underbrace{Cov(Y)}_{\sigma^2 I}(I - P_X)'$$

$$= (I - P_X)\sigma^2 I(I - P_X)' = \sigma^2 \underbrace{(I - P_X)(I - P_X)'}_{\text{idempotent}} = \sigma^2 (I - P_X)$$

Gauss-Markov teoremi gereği hata terimlerinin sıfır ortalamalı ve sabit varyansa sahip oldukları varsayımı altında en küçük kareler regresyon katsayılarının tahmin edicileri en iyi lineer yansız tahmin edicilerdir.

(BLUE = DESTE)

doğrusal en küçük sapmasız tahmin edici

Normallik varsayımının test edilmesi:

Regresyon analizinde hata terimlerinin normallik varsayımının test edilmesi için bir çok yöntemler önerilmiştir. Burada;

1. Ki-kare (χ^2) uyum iyiliği testi:

Açıklanan değişken ile açıklayıcı değişken arasındaki kestirim denklemi bulunarak artıklar $(\hat{\varepsilon})$ hesaplanır. Hata terimi ε'nın tahmin değerleri yani artıkların örnek ortalaması ve standart sapması bulunur. $\hat{\varepsilon}$ değerleri küçükten büyüğe doğru sıralaması yapılıp, sıfırdan kaç standart sapma uzaklıkta olduklarına göre kümelere ayrılır. Bu bir örnekle açıklanırsa: Keynes'in tüketim fonksiyonu örneğinden,

$$\hat{Y}_i = 23.906 + 0.5182x_i; \quad i = 1, 2, \ldots, 10$$

$\hat{Y}_i$	$\hat{\varepsilon}$
65.362	4.638
75.726	-10.726
86.090	3.91
96.454	-1.454
106.818	3.182
117.182	-2.182
127.546	2.454

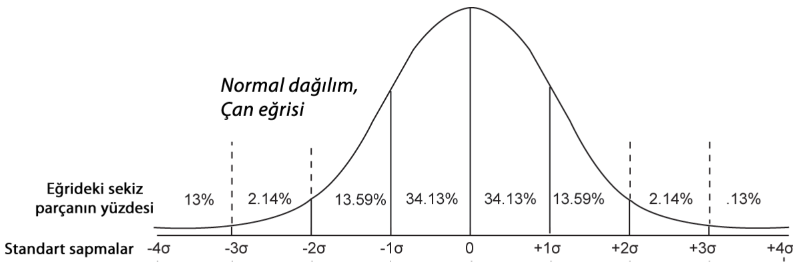
$$\sum_{i=1}^{10} \hat{\varepsilon}_i = 0 \qquad \sum_{i=1}^{10} \hat{\varepsilon}_i^2 = 299.091$$

$$S_{\hat{\varepsilon}}^2 = \frac{1}{10-1}\sum_{i=1}^{10} \hat{\varepsilon}_i^2 = \frac{299.091}{9} = 33.2323$$

$$S_{\hat{\varepsilon}} = 5.765; \qquad 2S_{\hat{\varepsilon}} = 11.529$$

137.910	2.09
148.274	6.726
158.638	-8.638

Sınıf aralığı	Sıklık (G_i)	E_i	$(G_i - E_i)^2 / E_i$
$\leq -2S_{\hat{\varepsilon}}$	0	0.025*10=0.25	0.25
$-2S_{\hat{\varepsilon}} < \hat{\varepsilon} \leq -S_{\hat{\varepsilon}}$	2	0.135*10=1.35	0.313
$-S_{\hat{\varepsilon}} < \hat{\varepsilon} \leq 0$	2	0.34*10=3.4	0.576
$0 < \hat{\varepsilon} \leq S_{\hat{\varepsilon}}$	5	0.34*10=3.4	0.753
$S_{\hat{\varepsilon}} < \hat{\varepsilon} \leq 2S_{\hat{\varepsilon}}$	1	0.135*10=1.35	0.091
$2S_{\hat{\varepsilon}} < \hat{\varepsilon}$	0	0.025*10=0.25	0.25



(Referans: https://upload.wikimedia.org/wikipedia/commons/b/bb/Normal_distribution_and_scales_tr.gif)

H_0 : ε_i 'ler normal dağılıma sahiptir.

H_1 : ε_i 'ler normal dağılıma sahip değildir.

H_0 yokluk hipotezinin H_1 alternatif hipotezine karşı test edilmesi için Ki-kare uyum iyiliği test istatistiği:

$$\chi_t^2 = \sum_{i=1}^{n} \frac{(G_i - E_i)^2}{E_i} \sim \chi_{n-1}^2 \text{ olup } \chi_t^2 = 2.233$$

$\alpha = 0.05$ anlamlılık düzeyinde Ki-kare kritik değeri $\chi^2_{9\,(0.05)} = 16.919$ ve

$\chi_t^2 = 2.233 < \chi^2_{9\,(0.05)} = 16.919$ olduğundan H_0 hipotezi red edilemez. Sonuç olarak ε_i 'ler normal dağılıma sahiptir.

2. Jarque-Bera (JB) normallik testi:

Bu test büyük örneklem testidir. Bu test için önce $\hat{\varepsilon}_i$ 'ların çarpıklık ve basıklık ölçüleri hesaplanır. Örneğimizden;

$$\sum_{i=1}^{10} \hat{\varepsilon}_i = 0, \quad \sum_{i=1}^{10} \hat{\varepsilon}_i^2 = 299.091, \quad \sum_{i=1}^{10} \hat{\varepsilon}_i^3 = -1372.034, \quad \sum_{i=1}^{10} \hat{\varepsilon}_i^4 = 21731.262, \quad S_{\hat{\varepsilon}} = 5.765;$$

Çarpıklık (Skewness):

$$S=\frac{\sum_{i=1}^{n}\left(\hat{\varepsilon}_i-\bar{\hat{\varepsilon}}\right)^3}{(n-1)S_{\hat{\varepsilon}}^3}=\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^3}{(9)S_{\hat{\varepsilon}}^3}=\frac{-1372.034}{(9)(5.765)^3}=-0.7957$$

Basıklık (Kurtosis):

$$K=\frac{\sum_{i=1}^{n}\left(\hat{\varepsilon}_i-\bar{\hat{\varepsilon}}\right)^4}{(n-1)S_{\hat{\varepsilon}}^4}=\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^4}{(9)S_{\hat{\varepsilon}}^4}=\frac{21731.262}{(9)(5.765)^4}=2.186$$

bulunur. JB test istatistiği:

$JB=n\left[\frac{S^2}{6}+\frac{(K-3)^2}{24}\right]\sim\chi_2^2$ olup, $JB=n\left[\frac{(-0.7957)^2}{6}+\frac{(2.186-3)^2}{24}\right]=1.3313$ dir.

$\alpha=0.05$ anlamlılık düzeyinde Ki-kare kritik değeri $\chi^2_{2\,(0.05)}=5.99147$ ve

$JB=1.3313<\chi^2_{2\,(0.05)}=5.99147$ olduğundan H_0 hipotezi red edilemez. Sonuç olarak ε_i ’ler normal dağılıma sahiptir.