

HAFTA 13

REGRESYON ANALİZİNE DEVAM

$Y_i = \alpha_0 + \alpha_1 x_i + e_i$, $i = 1, 2, 3, \dots, n$ regresyon modelini $\underline{y} = X \underline{\beta} + \underline{e}$ gibi yazalım. Bu modele regresyon modeli denebilmesi için temel varsayımları tekrar hatırlayalım. Önce X tasarım matrisinin biliniyor olması gerekir. Ayrıca, hata terimlerinin ilişkisiz olması gerekir. Yani, bütün i ler için $E(e_i) = 0$ ve $Var(e_i) = \sigma^2$ ve $i \neq j$ için $Cov(e_i, e_j) = 0$ olmalıdır. Regresyon analizi yapılacak verilerin bu varsayımları sağlayıp sağlamadığı kontrol edilmelidir. Bu varsayımlarla birlikte aşağıdaki soruların da yanıtlanması gerekir.

- i) Bütün hata terimleri aynı dağılımlı mıdır?
- ii) Bütün gözlemler için, $E(\underline{y}) = X \underline{\beta}$ midir?
- iii) Model doğru mudur? Yani, ihmal edilen başka bir açıklayıcı değişken var mıdır?
- iv) Hata terimlerinin varyansı bütün i ler için aynı mıdır?
- v) Hata terimleri korelasyonlu mudur?

İstatistiki sonuç çıkarımlar için bu varsayımların geçerliliği kontrol edilmelidir. Bu varsayımları kontrol etmek için bir çok istatistiki yöntem başvurulabilir. Bir çok istatistiki paket program (SPSS, SAS gibi) bu varsayımların kontrolü için bazı test istatistiklerinin değerlerini ve grafiklerini verir. Değişik kaynaklarda değişik yöntemlerle de karşılaşılabılır. Burada, temel varsayımlar hata terimleri üzerinde yapıldığı için, regresyondan elde edilen artıklar kullanılarak birkaç özel durumu kısaca açıklayacağız.

$\underline{y} = X \underline{\beta} + \underline{e}$ şeklinde verilen regresyon modeli için $\underline{\beta}$ nın en küçük kareler tahmin edicisi $\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{y}$ olup kestirim ve artık vektörleri,

$$\hat{\underline{y}} = X \hat{\underline{\beta}} = X (X'X)^{-1} X' \underline{y} = P_X \underline{y}, \quad \hat{\underline{e}} = \underline{y} - \hat{\underline{y}} = (I_n - P_X) \underline{y}$$

şeklinde dir. Regresyon modelinin doğruluğu varsayımı altında \hat{e}_i , e_i ler için birer kestirimdir. e_i ler beklenen değeri 0, varyansı σ^2 olan normal dağılıma sahip bağımsız rasgele değişkenlerdir. Ancak, kestirimlerin (\hat{e}), beklenen değeri 0, varyansı $\sigma^2 (I_n - P_X)$ dir. Yani, kestirimler farklı varyanslara sahip olup bazılarının arasındaki korelasyonlar sıfırdan farklıdır. Bu nedenle, t - türü artıklara dayalı artık grafikleri ile model varsayımları kontrol edilmektedir. $Var(\hat{\underline{e}}) = \sigma^2 (I_n - P_X)$ olup standart hataları $s(\hat{e}_i) = \sqrt{MSE(1 - h_{ii})}$ dir. Burada, h_{ii} , $I_n - P_X$ matrisinin i .sıra ve i .sütun elemanıdır. Buradan, t - türü artıklar $e_i^* = \hat{e}_i / s(\hat{e}_i)$ şeklinde olup bunların bağımsız $N(0, \sigma^2)$ dağılıma uygunluğuna bakılır. Bu artıklar ile oluşturulan grafiklerde herhangi bir ilişki (pattern) gözlenmemesi gerekir (Şekil (10.2.1a)). Bu artık grafiklerinden bazılarını göz atalım.

Şekil (10.2.1b) de verilen grafiğe benzer bir artık grafiği gözleendiğinde nümerik hesaplamalar, özellikle, kesim noktasının (intercept) modele dahil edilmediği kontrol edilmelidir.

Artıkların açıklayıcı değişkenlere karşı grafiğinde Şekil (10.2.1a) gölenen bir örüntüye benzer bir grafik modelin doğru olmadığı, bazı açıklayıcı değişkenlerin unutulmuş olabileceğini belirtir. Yani, bazı açıklayıcı değişkenler ihmal edilmiş olabilir. Örneğin, model doğrusal olmayıp x_i^2 gibi yeni bir açıklayıcı değişkenin eklenmesi gerekebilir.

Kestirimlerin, t – türü artıklara karşı grafiği çizildiğinde Şekil (10.2.1e) de verilen grafiğe benzer bir örüntü (patern) göze çarparsa, Y_i lerin varyansı beklenen değerin bir fonksiyonudur. Yani, beklenen değer değiştikçe varyans da değişir. Böyle bir durumda, varyans sabit olacak şekilde bir dönüşüm veya ağırlıklı en küçük kareler yöntemi (belki her ikisi de) dikkate alınmalıdır.

Artıkların grafikleri, y_i gözlem değerlerine karşı çizilmez. Bu t – türü artıkların grafikleri \hat{y}_i kestirim değerlerine, x_i açıklayıcı değişkenin değerlerine göre hazırlandığı gibi, hata terimlerinin kendi içinde otokorelasyonlu olup olmadığına bakmak için e_i^* lerin i ye ve/veya e_{i-1}^* lere karşı grafiklerine bakılır. Hata terimlerinin grafiklerinde Şekil (10.2.1c veya 10.2.1d) de verilen grafiklere benzer bir örüntü (yapı, patern) hata terimlerinin korelasyonlu olabileceğine işaret eder. Hata terimlerinin ilişkisiz olduğunu kontrol etmek için yukarıdaki grafiklerin yanında, Durbin-Watson test istatistiğinin değeri de kullanılır. Örneğin, e_i^* lerin e_{i-1}^* lere karşı grafiği (birinci dereceden otokorelasyon için) veya e_i^* lerin i lere karşı grafiğinde bir ilişki (patern) gözleniyorsa hata terimleri otokorelasyonludur.

Hata terimlerinin birinci dereceden otokorelasyonlu olup olmadığını sınamak için Durbin-Watson test istatistiğinin değeri

$$d = \left[\sum_{i=1}^n \hat{e}_i^2 \right]^{-1} \left[\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2 \right]$$

formülü ile hesaplanır. Hata terimlerinin otokorelasyonlu olmadığı hipotezi altında bu istatistiğin payının beklenen değeri yaklaşık $2n\sigma^2$, paydasının beklenen değeri de $(n-1)\sigma^2$ olup, oran yaklaşık olarak 2 dir. Hesaplanan değer 2 den uzaklaştıkça hata terimlerinin birinci dereceden otokorelasyonlu olduğu söylenir. Bu istatistiğin kritik değerleri için tablolar düzenlenmiştir. Hesaplanan Durbin-Watson istatistiğinin değeri tablo değeri ile karşılaştırılarak, hata terimlerinin otokorelasyonlu olup olmadığına karar verilir. Bir çok istatistikî paket program bu istatistiğin değerini hesaplamaktadır. Örnek (10.2.1) deki hata terimleri arasındaki birinci dereceden otokorelasyon için Durbin-Watson istatistiğinin değeri,

$$d = \left[\sum_{i=1}^n \hat{e}_i^2 \right]^{-1} \left[\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2 \right] = \frac{(-81-44)^2 + (24+81)^2 + (19-24)^2 + (-6-19)^2}{SSE}$$

$$= \frac{125^2 + 105^2 + 5^2 + 25^2}{9470} = \frac{27300}{9470} \cong 2.883$$

olarak hesaplanmıştır.

İstatistiki sonuç çıkarım için, hata terimlerinin normal dağılıma sahip olması gerekir. Normallliği sınamak için bir çok yöntem bulunmaktadır. Bir çok istatistiki paket program (SPSS, SAS gibi) normallüğün sınanması için Kolmogorov-Simironov test istatistiğinin değeri ile, $P-P$ çiziti (plot), $Q-Q$ çiziti (plot), Box-Cox çiziti ve normal olasılık grafiklerini verir.

Bunlardan normal olasılık grafiğinin hazırlanmasını inceleyelim. Bunun için regresyondan elde edilen artıklar küçükten büyüğe sıralanır. Bunları $\hat{e}_{(i)}$ ile gösterelim. Normallik varsayımı altında, $E(\hat{e}_{(i)}) \approx \sqrt{MSE} z((i-0.375)/(n+0.25))$ dir.

Buradan, normal dağılım tablosundan bulunacak z -değerlerinin $\hat{e}_{(i)}$ lara karşı grafiği çizilir. Bu grafik Şekil (10.2.1f) de verilen grafiğe benzer doğrusal bir ilişkiyi gösteriyorsa, hata terimlerinin normal olduğu sezgisel olarak söylenebilir.

Örnek (10.2.1) deki veriler için normal olasılık grafiği için z -değerleri,

$$i = 1: z((1-0.375)/(5+.25)) = z(0.119), P(Z < -1.18) \cong 0.119 \Rightarrow z_1 = -1.18$$

$$i = 2: z((2-0.375)/(5+.25)) = z(0.3095), P(Z < -0.5) \cong 0.3095 \Rightarrow z_2 = -0.5$$

$$i = 3: z((3-0.375)/(5+.25)) = z(0.5), P(Z < 0) \cong 0.5 \Rightarrow z_3 = 0$$

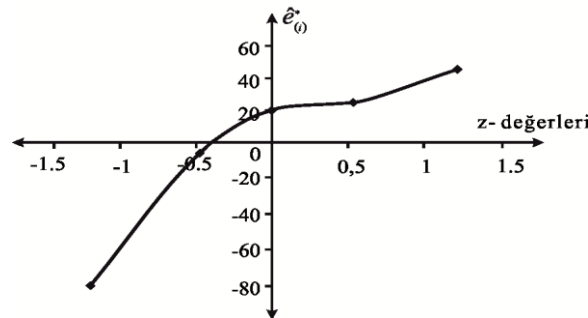
$$i = 4: z((4-0.375)/(5+.25)) = z(0.69), P(Z < 0.5) \cong 0.69 \Rightarrow z_4 = 0.5$$

$$i = 5: z((5-0.375)/(5+.25)) = z(0.881), P(Z < 1.18) \cong 0.881 \Rightarrow z_5 = 1.18$$

şeklinde hesaplanmıştır. Burada sıralanmış değerler ile z -değerleri

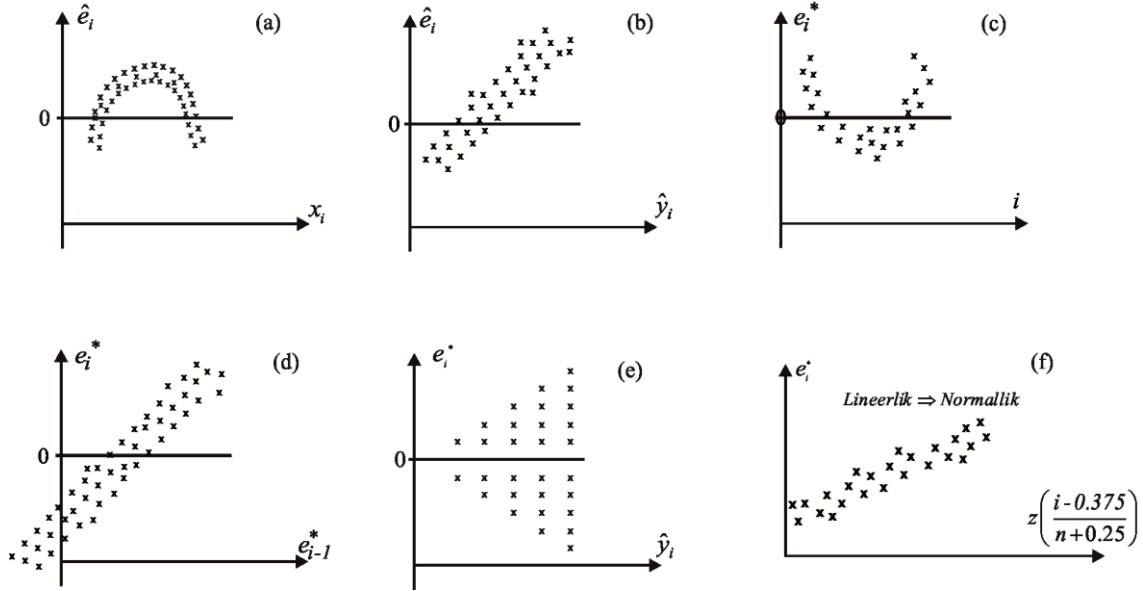
$\hat{e}_{(i)}$	-81	-6	19	24	44
z -değeri	-1.18	-0.5	0.0	0.5	1.18

şeklinde özetlenmiş, bu sıralanmış artıklar ile normal dağılım tablosundan bulunan z -değerlerinin grafiği aşağıdadır.



Şekil 10.2.1a Artıklara ilişkin normal olasılık grafiği

Normal olasılık grafiği tam bir lineerliği göstermese de, verilerin azlığı da dikkate alındığında verilerin normal dağılımdan geldiği sezgisel olarak söylenebilir.



Şekil 10.2.1 Artık Grafikleri

$Y_i = \alpha_0 + \alpha_1 x_i + e_i$, $i = 1, 2, 3, \dots, n$ şeklinde verilen regresyon modelini tekrar göz önüne alalım. Buradaki varsayımlardan biri, hata terimlerinin aynı varyanslı olmasıdır. Burada, $E(Y_i) = \alpha_0 + \alpha_1 x_i$ ve $Var(Y_i) = Var(e_i) = \sigma^2$ dir. Bazen, Y_i lerin varyansı beklenen değerin bir fonksiyonu olabilir. Y lerin beklenen değeri sabit değildir, x ler değiştikçe beklenen değer de değişir (Şekil (10.2.1e)). Örneğin, Y_i ler Poisson dağılımına sahip ise, Poisson dağılımının beklenen değeri ile varyansı aynıdır. Yani, beklenen değer değiştikçe varyans da değişir. Benzer şekilde Y_i ler Binom dağılımlı rasgele değişkenler ise, varyans yine beklenen değerin bir fonksiyonudur. Her iki durumda da, Y_i lerin varyansı, yani hata terimlerinin varyansı sabit değildir. Kısaca, Y_i lerin varyansı ile beklenen değerleri arasında, $Var(Y) = f(E(Y))$ şeklinde bir ilişki olabilir (burada f biliniyor). Böyle bir durumda, regresyon varsayımları sağlanmaz. Poisson örneği için, $Var(Y) = E(Y)$ olup $f(z) = z$, $Y \sim Binom(n, p)$ ise, $E(Y) = np$ ve $Var(Y) = np(1-p)$ için $f(z) = z(1-z)$ şeklinde bir ilişki vardır.

Bu problemin giderilmesi için dönüşüm yapmak gerekir. Y_i lerin varyansı ile beklenen değeri arasında $Var(Y) = f(E(Y))$ şeklinde bir ilişki var olsun. $Var(g(Y))$ sabit olacak şekilde bir g dönüşümünü belirlemek istiyoruz. Bunun için $Var(g(Y)) = c$ olacak şekilde bir g fonksiyonu belirlensin. Fonksiyonun $E(Y)$ noktası komşuluğundaki birinci dereceden Taylor serisi açılımı, $g(Y) \approx g(E(Y)) + g'(E(Y))(Y - E(Y))$ olup,

$$c = \text{Var}(g(Y)) \approx [g'(E(Y))]^2 \text{Var}(Y - E(Y)) \\ = [g'(E(Y))]^2 \text{Var}(Y) = [g'(E(Y))]^2 f(E(Y))$$

yazılabilir. $y = E(Y)$ denirse, $c = [g'(y)]^2 f(y)$ eşitliğinden aranan dönüşüm,

$$g(y) = \int (1/\sqrt{f(y)}) dy$$

şeklinde elde edilir.

Buna göre, $Y \sim \text{Poisson}(\lambda)$ ise $\text{Var}(Y) = E(Y) = \lambda$ olup, $f(y) = y$ alınır. Buradan, varyansı sabitlemek için yapılması gereken dönüşüm,

$$g(y) = \int (1/\sqrt{f(y)}) dy = \int (1/\sqrt{y}) dy = 2\sqrt{y}$$

olduğundan $g(y) = \sqrt{y}$ dönüşümü yapılmalıdır. Benzer şekilde, $Y \sim \text{Binom}(n, p)$ ise, $f(y) = y(1-y)$ seçildiğinde aranan dönüşüm, yukarıdaki integralin sonucuna göre elde edilir.

Bu integral için $y = \sin^2(\theta)$ ise $dy = 2\sin(\theta)\cos(\theta)$ olup

$$g(y) = \int (1/\sqrt{f(y)}) dy = \int (1/\sqrt{y(1-y)}) dy = \int \frac{2\sin(\theta)\cos(\theta)}{\sqrt{\sin^2(\theta)(1-\sin^2(\theta))}} d\theta = 2 \int d\theta = \theta$$

olup varyansın sabitlenmesi için yapılması gereken dönüşüm, ters dönüşüm yardımı ile $g(y) = \arcsin(\sqrt{y})$ şeklinde olmalıdır. Varyansın sabitlenmesi için, Box-Cox dönüşümleri gibi literatürde başka yöntemlere de rastlanabilir.

Verilerin normal dağılımdan geldiğini sınamak için en çok kullanılan yöntemlerden biri de *Kolmogorov-Smirnov* testidir. Bu yöntem sadece normallik için değil, herhangi bir dağılıma uygunluğunu sınamak için de kullanılır. Ancak, normalliğin öneminden dolayı genellikle normal dağılıma uygunluğun sınanmasında kullanılmaktadır. Verilerin dağılım fonksiyonu $F_0(x)$ olarak belirlenen bir dağılıma uygunluğunu test etmek için, $H_0 : F(x) = F_0(x)$ yokluk hipotezi $H_a : F(x) \neq F_0(x)$ alternatif hipotezine karşı test edilir (alternatif hipotez $H_a : F(x) < F_0(x)$ veya $H_a : F(x) > F_0(x)$ de olabilir). Bu hipotez testi problemi için $S(x)$ birikimli sıklıkları göstermek üzere $D = \sup |S(x) - F_0(x)|$ istatistiğinin değeri hesaplanır. Bu test istatistiğinin kritik değerleri istatistik kitaplarında bulunmaktadır (Daniel (1990)). Hesaplanan değer kritik değerden küçük ise, H_0 hipotezi red edilemez. Bu testin bir uygulaması aşağıdaki örnekte gösterilmiştir. Burada, $F_0(x) = P(X \leq x)$ olup, $S(x)$ ler birikimli sıklıklardır.

Örnek 10.2.2 Bir yaşından küçük çocukların ağırlıkları ile ilgili veriler aşağıda verilmiştir. Burada X çocukların yaşı (ay), Y de ağırlıklarıdır (kilogram).

X	0	0	0	3	3	3	6	6	6	9	9	9	12	12	12
Y	2.6	3.2	3.4	3.8	3.9	4.2	5.2	5.5	6.1	7.1	7.4	7.5	7.9	8.4	8.4

Bu verilere ilişkin bazı özet bilgiler de,

$$\sum_{i=1}^n x_i = 90, \sum_{i=1}^n x_i^2 = 810, \sum_{i=1}^n x_i y_i = 630.9, \sum_{i=1}^n y_i = 84.6, \sum_{i=1}^n y_i^2 = 535.3, n = 15$$

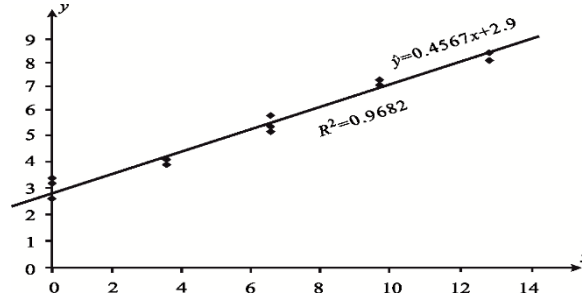
şeklinde hesaplanmıştır. Bu verilere,

$$Y_i = \alpha_0 + \alpha_1 x_i + e_i, i = 1, 2, 3, \dots, n$$

şeklinde basit doğrusal regresyon modelinin uygun olduğunu varsayalım. Önce kestirim denklemini bulup ANOVA tablosunu hazırlayalım.

Parametrelerin en küçük kareler tahmileri,

$\hat{\alpha}_1 \cong 0.45\bar{6}$, $\hat{\alpha}_0 = \bar{y}_n - \hat{\alpha}_1 \bar{x}_n = 5.64 - (0.45\bar{6})(6) = 2.9$ olarak hesaplanmıştır. Kestirim doğrusu ile veri serpilme grafiği Şekil (10.2.2) de verilmiştir.



Şekil 10.2.2 Örnek (10.2.2) deki verilerin serpilme grafiği ile kestirim doğrusunun grafiği

Kestirim denklemi $\hat{Y}_i = 2.90 + (0.456)x_i$ şeklinde olup, kestirimler ile artık değerler de hesaplanarak aşağıda tablo halinde verilmiştir.

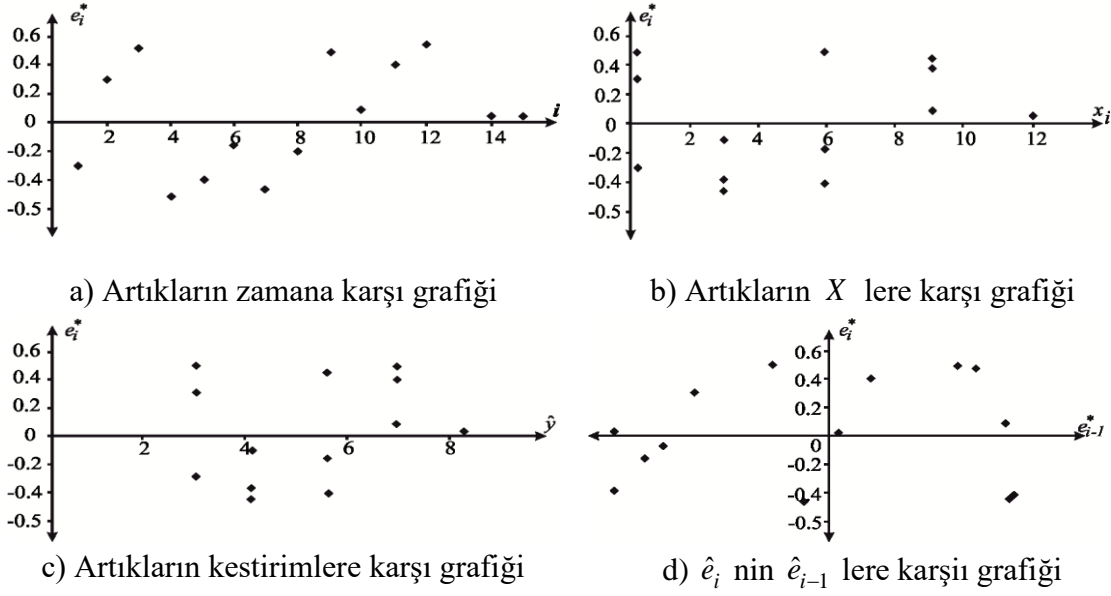
X	0	0	0	3	3	3	6	6	6	9	9	9	12	12	12
Y	2.6	3.2	3.4	3.8	3.9	4.2	5.2	5.5	6.1	7.1	7.4	7.5	7.9	8.4	8.4
\hat{Y}	2.9	2.9	2.9	4.27	4.27	4.27	5.64	5.64	5.64	7.01	7.01	7.01	8.38	8.38	8.38
\hat{e}	-0.3	0.3	0.5	-0.47	-	-	-	-	0.46	0.09	0.39	0.49	-	0.02	0.02
					0.37	0.07	0.44	0.14					0.48		

Bu kestirim ve artık değerler kullanılarak, ANOVA tablosu aşağıdaki gibi hazırlanmış ve $R^2 = SSR / SST = 56.307 / 58.156 = 0.9682$ gözlenmiştir.

Kaynak	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kareler Toplamı	F Değeri
Regresyon	1	56.307	56.307	395.885
Artıklar	13	1.849	0.14223	
Toplam	14	58.156		

Artıkların grafiklerinde herhangi bir ilişki (patern) görülmemektedir. Artıklara ait grafikler Şekil (10.2.3) de verilmiştir. Buna göre, hata terimlerinin regresyon varsayımlarını sağladığı sezgisel olarak söylenebilir. Diğer taraftan, istatistikî sonuç çıkarımlar için hata terimlerinin

normal dağılıma sahip olması gerekir. Bir önceki örnekten, normal olasılık grafiğinin nasıl hazırlandığını biliyoruz. Şimdi, Kolmogorov-Smirnov yöntemini kullanarak, verilerin normallik varsayımını inceleyelim.



Şekil 10.2.3 Örnek (10.2.2) deki verilere ait artıkların grafikleri

$\hat{e}_{(i)}$	Frekans	Birikiml i Frekans	$S(x)$	$z_i = \hat{e}_{(i)} / \sigma$	$F_0(x) = P(Z \leq z_i)$	$ S(x) - F_0(x) $
-0.48	1	1	1/15=0.0667	-1.32	0.0934	0.0267
-0.47	1	2	2/15=0.1333	-1.29	0.0985	0.0348
-0.44	1	3	3/15=0.2000	-1.21	0.1131	0.0869
-0.37	1	4	4/15=0.2667	-1.02	0.1538	0.1129
-0.30	1	5	5/15=0.3333	-0.83	0.2032	D=0.1301
-0.14	1	6	6/15=0.400	-0.39	0.3482	0.0518
-0.07	1	7	7/15=0.4667	-0.19	0.4246	0.0421
0.02	2	9	9/15=0.6000	0.06	0.5240	0.0760
0.09	1	10	10/15=0.666 7	0.25	0.5988	0.0679
0.30	1	11	11/15=0.733 3	0.83	0.7967	0.0634
0.39	1	12	12/15=0.800 0	1.07	0.8577	0.0577
0.46	1	13	13/15=0.866 7	1.27	0.8980	0.0313
0.49	1	14	14/15=0.933 3	1.35	0.9115	0.0218
0.50	1	15	15/15=1.000 0	1.38	0.9162	0.0838

Regresyon modeli $Y_i = \alpha_0 + \alpha_1 x_i + e_i$, $i = 1, 2, 3, \dots, n$ şeklinde verilmiş olsun. $\sigma^2 = 0.132$ için hata terimlerinin $N(0, \sigma^2)$ dağılımına uygunluğunu sınavalım. Bunun için $D = \sup |S(x) - F_0(x)|$ test istatistiğinin değerinin hesaplanması gerekir. $S(x)$ değerlerini hesaplamak için veriler küçükten büyüğe doğru sıralanır. Tekrar eden veriler bir defa yazılır.

$F_0(x)$ değerlerinin hesaplanması için ($e_i \sim N(0, \sigma^2 = 0.132)$ dağılımına uygunluk test edilecektir) $z = (x - \mu) / \sigma$ şeklinde z -değerleri ile bu z -değerlerine karşılık gelen olasılıklar hesaplanarak yukarıdaki tabloda verilmiştir.

Örnek 10.2.3 Aşağıdaki verilere $Y_i = \alpha_0 + \alpha_1 x_i + e_i$, $i = 1, 2, \dots, n$ şeklinde bir modelin uygun olduğunu varsayalım.

x	-1	0	1	2	3
Y	0	-1	1	2	3

Tablodaki özet bilgilerden parametrelerin en küçük kareler tahmin değerleri

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^5 x_i y_i - n \bar{x}_n \bar{y}_n}{\sum_{i=1}^5 x_i^2 - n \bar{x}_n^2} = \frac{14 - 5}{15 - 5} = \frac{9}{10} = 0.9 \quad \text{ve} \quad \hat{\alpha}_0 = \bar{y}_n - \hat{\alpha}_1 \bar{x}_n = 1 - 0.9(1) = 0.1$$

olup kestirim denklemi $\hat{Y}_i = 0.1 + 0.9x_i$ dir. Tablodaki değerlerden görüldüğü gibi,

$$\sum_{i=1}^5 \hat{e}_i = \sum_{i=1}^5 x_i \hat{e}_i = \sum_{i=1}^5 \hat{y}_i \hat{e}_i = 0, \quad \sum_{i=1}^5 y_i \hat{e}_i = 1.9 \neq 0 \quad \text{ve} \quad \sum_{i=1}^5 Y_i = \sum_{i=1}^5 \hat{Y}_i = 5$$

dir (Problem (10.5.1)). Hesaplanan artık ve kestirim değerleri aşağıdaki tabloda verilmiştir.

	x	Y	xy	x^2	\hat{Y}	\hat{e}	$x\hat{e}$	$\hat{e}\hat{y}$	\hat{Y}^2	\hat{e}^2	Y^2
	-1	0	0	1	-0.8	0.8	-0.8	-0.64	0.64	0.64	0
	0	-1	0	0	0.1	-1.1	0	-0.11	0.01	1.21	1
	1	1	1	1	1	0	0	0	1	0	1
	2	2	4	4	1.9	0.1	0.2	0.19	3.61	0.01	4
	3	3	9	9	2.8	0.2	0.6	0.56	7.84	0.04	9
Σ	5	5	14	15	5	0	0	0	13.1	1.9	15

Buna göre, $R^2 = 0.81$ olup ANOVA tablosu,

Kaynak	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kareler Toplamı	F Değeri
Regresyon	1	8.1	8.1	12.79
Artıklar	3	1.9	0.63	
Toplam	4	10		

şeklinde oluşturulmuştur. Verilerin matris formunda yazılması ile,

$$X'X = \begin{bmatrix} 5 & 5 \\ 5 & 15 \end{bmatrix}, (X'X)^{-1} = \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.1 \end{bmatrix}, X'\tilde{y} = \begin{bmatrix} 5 \\ 14 \end{bmatrix},$$

$$\hat{\beta} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = (X'X)^{-1} X'\tilde{y} = \begin{bmatrix} 0.3 & -0.1 \\ -0.1 & 0.1 \end{bmatrix} \begin{bmatrix} 5 \\ 14 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix},$$

değerleri hesaplanmıştır. Buradan, parametrelerin en küçük kareler standart hataları ile t -istatistiklerinin değerleri de

$$s(\hat{\alpha}_0) = \sqrt{MSE(X'X)^{-1}_{1,1}} = \sqrt{0.19} \cong 0.436, t_h = \hat{\alpha}_0 / s(\hat{\alpha}_0) = 0.1 / 0.436 = 0.229$$

$$s(\hat{\alpha}_1) = \sqrt{MSE(X'X)^{-1}_{2,2}} = \sqrt{0.0633} \cong 0.25, t_h = \hat{\alpha}_1 / s(\hat{\alpha}_1) = 0.9 / 0.25 = 3.6$$

şeklinde hesaplanmıştır. $\alpha = 0.05$ için $t_3(\alpha/2) = t_3(0.025) = 3.182$ olup, $H_0 : \alpha_0 = 0$ yokluk hipotezi %5 anlam düzeyinde red edilemez. Ancak, $H_0 : \alpha_1 = 0$ hipotezi %5 anlam düzeyinde red edilir. Buna rağmen, $H_0 : \alpha_0 = \alpha_1 = 0$ hipotezini test etmek için,

$$F_h = \frac{SSR(full)/2}{MSE(full)} = \frac{13.1/2}{1.9/3} \cong 10.34$$

olup $F_h > F^{1-\alpha}(2,3) = 9.55$ dir. Yani, $H_0 : \alpha_0 = \alpha_1 = 0$ hipotezi %5 anlam düzeyinde red edilir. $H_0 : \alpha_0 = \alpha_1 = 0$ hipotezi %5 anlam düzeyinde red edilmesine rağmen $H_0 : \alpha_0 = 0$ hipotezi red edilememektedir

Bu parametrelere ait %95 lik güven aralıkları:

$$\alpha_0 \text{ için \%95 lik güven aralığı: } \hat{\alpha}_0 \pm t_3(\alpha/2) \Leftrightarrow 0.1 \pm (3.182)(0.436) \Leftrightarrow (-1.29, 1.49)$$

$$\alpha_1 \text{ için \%95 lik güven aralığı: } \hat{\alpha}_1 \pm t_3(\alpha/2) \Leftrightarrow 0.9 \pm (3.182)(0.25) \Leftrightarrow (0.1, 1.7)$$

şeklinde \oplus

Yukarıda, regresyon parametreleri için ayrı ayrı güven aralıkları oluşturuldu. Şimdi, aynı parametreler için *ortak güven aralıklarının* (veya güven bölgelerinin) oluşturulmasını inceleyelim. Bu güven bölgeleri için birkaç değişik yöntem üzerinde duracağız. Bunun için basit doğrusal regresyon modelini $\tilde{y} = X\beta + \varepsilon$ şeklinde yazalım. $\varepsilon \sim MN(0, \sigma^2 I_n)$ olup

$$\hat{\beta} = (X'X)^{-1} X'\tilde{y}, \hat{\beta} \sim MN(\beta, \sigma^2 (X'X)^{-1})$$

ve $\hat{\beta} = (\hat{\alpha}_0, \hat{\alpha}_1)'$ olmak üzere,

$$Var(\hat{\alpha}_0) = \sigma^2 (X'X)^{-1}_{1,1}, Var(\hat{\alpha}_1) = \sigma^2 (X'X)^{-1}_{2,2} \text{ ve } Cov(\hat{\alpha}_0, \hat{\alpha}_1) = \sigma^2 (X'X)^{-1}_{1,2}$$

olduğunu hatırlayalım. $p = rank(X)$ olmak üzere,

$$\frac{(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta) / p}{MSE(full)} \sim F(p, n - p)$$

dir. Buradan, $\underline{\delta} = (\delta_0, \delta_1)' = \hat{\beta} - \beta = (\hat{\alpha}_0 - \alpha_0, \hat{\alpha}_1 - \alpha_1)'$ olmak üzere

$$\frac{\underline{\delta}'(X'X)\underline{\delta} / p}{MSE(full)} = F^{1-\alpha}(p, n - p)$$

diyelim. Yani, $\underline{\delta}'(X'X)\underline{\delta} = p MSE F^{1-\alpha}(p, n - p)$ eşitliğinden α_0 ve α_1 parametreleri için %95 ortak güven aralığını, yani güven bölgesini oluşturmaya çalışalım. Buradan,

$$(\hat{\alpha}_0 - \alpha_0, \hat{\alpha}_1 - \alpha_1)'(X'X)(\hat{\alpha}_0 - \alpha_0, \hat{\alpha}_1 - \alpha_1)' = p MSE F^{1-\alpha}(p, n - p)$$

eşitliğinin çözülmesi ile

$$n\delta_0^2 + 2\delta_0\delta_1 \sum_{i=1}^n x_i + \delta_1^2 \sum_{i=1}^n x_i^2 = p MSE F^{1-\alpha}(p, n - p)$$

elde edilir. $\delta_1 = 0$ yazarak α_0 için, $\delta_0 = 0$ yazarak da α_1 için güven aralıkları oluşturulur.

Örnek 10.2.3 (Devam) Şimdi, α_0 ve α_1 parametreleri için %95 lik güven bölgesini oluşturalım. $p MSE F^{1-\alpha}(p, n - p) = 2(1.9/3)(9.55) = 12.097$ olmak üzere $\delta_1 = 0$ ise

$$5\delta_0^2 = 12.097 \Rightarrow \delta_0 = \pm 1.56 \Rightarrow (0.1 - \alpha_0) = \pm 1.56 \Rightarrow \alpha_0 = 0.1 \pm 1.56$$

dan α_0 için %95 lik güven bölgesinin sınırları $(-1.46, 1.66)$ olarak bulunur. Benzer şekilde, $\delta_0 = 0$ denirse,

$$15\delta_1^2 = 12.097 \Rightarrow \delta_1 = \pm 0.898 \Rightarrow (0.9 - \alpha_1) = \pm 1.56 \Rightarrow \alpha_1 = 0.9 \pm 0.898$$

ifadesinden, α_1 için %95 lik güven bölgesinin sınırları da $(0.02, 1.798)$ olarak bulunur. Buna göre, ikinci güven bölgesi birinciye göre daha geniştir \oplus

Yukarıda ortak güven aralıkları, $\delta_0 = 0$ (veya $\delta_1 = 0$) yazılarak oluşturuldu. Şimdi $\delta_0 = 0$ yerine, $(\hat{\alpha}_0 - \alpha_0) / s(\hat{\alpha}_0) \sim t_{n-p}$ olmasından dolayı $\delta_0 = s(\hat{\alpha}_0)t_{n-p}(\alpha/2)$ (veya $\delta_1 = 0$ yerine $\delta_1 = s(\hat{\alpha}_1)t_{n-p}(\alpha/2)$) yazalım. O zaman,

$$n\delta_0^2 + 2\delta_0\delta_1 \sum_{i=1}^n x_i + \delta_1^2 \sum_{i=1}^n x_i^2 = p MSE F^{1-\alpha}(p, n - p)$$

ikinci dereceden polinomun çözümünden başka bir güven aralığı (elipsoid veya kuadratik) elde edilir. Şimdi bunu yukarıdaki örnek üzerinde uygulayalım.

Örnek 10.2.3 (Devam) Şimdi, $\delta_1 = 0$ yerine $\delta_1 = s(\hat{\alpha}_1)t_{n-p}(\alpha/2)$ terimlerini alarak α_0 parametresi için %95 lik güven aralığı yazmaya çalışalım. Her bir durumda da,

$$n\delta_0^2 + 2\delta_0\delta_1 \sum_{i=1}^n x_i + \delta_1^2 \sum_{i=1}^n x_i^2 = p \text{MSE} F^{1-\alpha}(p, n-p)$$

ikinci dereceden denklemin çözümlerini bulalım. Burada, $p \text{MSE} F^{1-\alpha}(p, n-p) = 12.097$ dir.

a) $\delta_1 = 0$ yerine $\delta_1 = s(\hat{\alpha}_1)t_{n-p}(\alpha/2)$ değerleri yazıldığında $t_3(0.025) = -3.182$ olup $\delta_1 = (0.25)(-3.182) \cong -0.8$ ve $5\delta_0^2 + 10\delta_0\delta_1 + 15\delta_1^2 = 12.097$ dir. Buradan $\delta_1 = -0.8$ yazılırsa $5\delta_0^2 - 8\delta_0 - 2.497 = 0$ denkleminin çözümleri $\delta_0 = 1.867$ ve $\delta_0 = -0.267$ olur.

$$\left. \begin{array}{l} \delta_0 = 1.867 \Rightarrow 1.867 = \hat{\alpha}_0 - \alpha_0 \Rightarrow 1.867 = 0.1 - \alpha_0 \Rightarrow \alpha_0 = -1.767 \\ \delta_0 = -0.267 \Rightarrow -0.267 = \hat{\alpha}_0 - \alpha_0 \Rightarrow -0.267 = 0.1 - \alpha_0 \Rightarrow \alpha_0 = 0.367 \end{array} \right\} \Rightarrow \boxed{(-1.767, 0.367)}$$

ifadesinden α_0 için elipsoid türü güven aralığının sınır noktalarından biri $(-1.767, 0.367)$ olur.

b) Şimdi, $\delta_1 = 0$ yerine $\delta_1 = s(\hat{\alpha}_1)t_{n-p}(1-\alpha/2)$ değerini alalım. $t_3(0.975) = 3.182$ olup $\delta_1 = (0.25)(3.182) \cong 0.8$ ve $5\delta_0^2 + 10\delta_0\delta_1 + 15\delta_1^2 = 12.097$ eşitliğinde $\delta_1 = 0.8$ yazılırsa $5\delta_0^2 + 8\delta_0 - 2.497 = 0$ ikinci derece polinomunun çözümleri $\delta_0 = -1.867$ ve $\delta_0 = 0.267$ olur. Buradan da,

$$\left. \begin{array}{l} \delta_0 = -1.867 \Rightarrow -1.867 = \hat{\alpha}_0 - \alpha_0 \Rightarrow -1.867 = 0.1 - \alpha_0 \Rightarrow \alpha_0 = 1.967 \\ \delta_0 = 0.267 \Rightarrow 0.267 = \hat{\alpha}_0 - \alpha_0 \Rightarrow 0.267 = 0.1 - \alpha_0 \Rightarrow \alpha_0 = -0.167 \end{array} \right\} \Rightarrow \boxed{(-0.167, 1.967)}$$

ifadesinden α_0 için elipsoid türü güven aralığının sınır noktalarından biri $(-0.167, 1.967)$ olur.

c) $\delta_0 = 0$ yerine $\delta_0 = s(\hat{\alpha}_0)t_{n-p}(\alpha/2)$ yazıldığında ise $t_3(0.025) = -3.182$ olup $\delta_0 = (0.436)(-3.182) \cong -1.387$ ve $5\delta_0^2 + 10\delta_0\delta_1 + 15\delta_1^2 = 12.097$ eşitliğinde $\delta_0 = -1.387$ yazılırsa $15\delta_0^2 - 13.87\delta_1 - 2.48 = 0$ ikinci derece polinomunun çözümleri $\delta_1 = 1.078$ ve $\delta_1 = -0.153$ dür. Buna göre de,

$$\left. \begin{array}{l} \delta_1 = 1.078 \Rightarrow 1.078 = \hat{\alpha}_1 - \alpha_1 \Rightarrow 1.078 = 0.9 - \alpha_1 \Rightarrow \alpha_1 = -0.178 \\ \delta_1 = -0.153 \Rightarrow -0.153 = \hat{\alpha}_1 - \alpha_1 \Rightarrow -0.153 = 0.9 - \alpha_1 \Rightarrow \alpha_1 = 1.053 \end{array} \right\} \Rightarrow \boxed{(-0.178, 1.053)}$$

ifadesinden α_0 için elipsoid türü güven aralığının sınır noktalarından biri $(-0.178, 1.053)$ dür.

d) Son olarak $\delta_0 = 0$ yerine $\delta_0 = s(\hat{\alpha}_0)t_{n-p}(1-\alpha/2)$ alalım. $t_3(0.975) = 3.182$ olup $\delta_0 = (0.436)(3.182) \cong 1.387$ ve $5\delta_0^2 + 10\delta_0\delta_1 + 15\delta_1^2 = 12.097$ eşitliğinde $\delta_0 = 1.387$ yazılırsa $15\delta_0^2 + 13.87\delta_1 - 2.48 = 0$ ikinci derece polinomunun çözümleri $\delta_1 = -1.078$ ve $\delta_1 = 0.153$ olur. Dolayısı ile,

$$\left. \begin{array}{l} \delta_1 = -1.078 \Rightarrow -1.078 = \hat{\alpha}_1 - \alpha_1 \Rightarrow -1.078 = 0.9 - \alpha_1 \Rightarrow \alpha_1 = 1.978 \\ \delta_1 = 0.153 \Rightarrow 0.153 = \hat{\alpha}_1 - \alpha_1 \Rightarrow 0.153 = 0.9 - \alpha_1 \Rightarrow \alpha_1 = 0.747 \end{array} \right\} \Rightarrow \boxed{(0.747, 1.978)}$$

ifadesinden α_0 için elipsoid türü güven aralığının sınır noktalarından biri de $(0.747, 1.978)$ olur.

Bu noktalar, koordinat düzleminde α_0 ve α_1 için elipsoid türü güven bölgesine karşılık gelir

⊕