Normallik Varsayımı ve Ençok Olabilirlik Yöntemi

Normallik Varsayımı ve İlişkin Dağılımlar

Ekonometri 1 – Konu 12

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

İşbu belge, "Creative Commons Attribution-Non-Commercial ShareAlike 3.0 Unported" (CC BY-NC-SA 3.0) lisansı altında bir açık ders malzemesi olarak genel kullanıma sunulmuştur. Eserin ilk sahibinin belirtilmesi ve geçerli lisansın korunması koşulu ile özgürce kullanılabilir, çoğaltılabilir ve değiştirilebilir. Creative Commons örgütü ve "CC-BY-NC-SA" lisansı ile ilgili ayrıntılı bilgi "http://creativecommons.org" adresinde bulunmaktadır. Bu ekonometri ders notları setinin tamamına "http://www.acikders.org.tr" adresinden ulaşılabilir.

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 BY-NC-SA

Ders Planı

Hata Teriminin Olasılık Dağılımı

Ekonometrik çözümlemede amaç yalnızca ÖBİ'yi hesaplamak değil, aynı zamanda ABİ'ye ilişkin çıkarsama ve çeşitli önsav sınamaları da yapabilmektir.

Bu doğrultuda u_i hatalarının olasılık dağılımının bilinmesi ya da belirlenmesi iki nedenden dolayı önemlidir:

1. SEK bir katsayı hesaplama yöntemidir. ÖBİ'den ABİ'ye yönelik çıkarsama yapmada tek başına işe yaramaz.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri Y'nin doğrusal işlevi ve Y'nin kendisi de u_i'lerin bir doğrusal işlevidir. Öyleyse, $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın örneklem dağılımları u_i'lerin olasılık dağılımına ilişkin varsayımlara dayanmaktadır.

Hata Teriminin Olasılık Dağılımı

- Daha önce ele alınan klasik doğrusal bağlanım modeli (KDBM), u_i hata teriminin olasılık dağılımı ile ilgili herhangi bir varsayımda bulunmaz.
- Almaşık olarak, "Klasik Normal Doğrusal Bağlanım Modeli" (Classical Normal Linear Regression Model) ya da kısaca "KNDBM" (CNLRM) ise u_i'lerin normal dağıldığını varsayar.

Normallik Varsayımı

KNDBM, her bir u_i'nin aşağıdaki değerlerle normal dağıldığı varsayımını getirir:

$$\begin{array}{ll} \text{Ortalama:} & E(u_i) = 0 \\ \text{Varyans:} & E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2 \\ \text{Kovaryans:} & \operatorname{cov}(u_i, u_j) = 0, \qquad i \neq j \end{array}$$

- Bu varsayımlar kısaca $u_i \sim N(0, \sigma^2)$ şeklinde de gösterilir.
- Buradaki ($\sim$), "dağılımlı" (distributed) anlamına gelir. N ise normal dağılımı göstermektedir.

Normallik Varsayımı

- İki rastsal değişkenin kovaryansının sıfır olması bu iki değişkenin bağımsız olduğunu gösterir.
- Bu nedenle $u_i \sim N(0, \sigma^2)$ yerine $u_i \sim \text{NBD}(0, \sigma^2)$ de yazılabilir.
- Burada NBD "normal ve bağımsız dağılımlı" (normally and independently distributed, kısaca NID) demektir.

Normallik Varsayımının Nedenleri

Normallik varsayımının nedenleri şunlardır:

1. Merkezi limit kanıtsavına göre bağımsız ve özdeş dağılımlı (BÖD) rastsal değişkenlerin toplam dağılımı, değişken sayısı sonsuza yaklaştıkça normale yakınsar.
2. Merkezi limit kanıtsavına göre değişken sayısı çok fazla olmasa ya da değişkenler tam bağımsız dağılmasalar bile toplamları normal dağılabilir.
3. Normal dağılan değişkenlerin doğrusal işlevleri de normal dağılır. (Örnek: $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\sigma}^2$)
4. Normal dağılım yalnızca iki katsayı (ortalama ve varyans) içeren basit, istatistiksel özellikleri iyi bilinen bir dağılımdır.

Normallik Varsayımı Altında SEK Tahmincileri

$\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincileri ile $\hat{\sigma}^2$ varyans tahmini, normallik varsayımı altında şu istatistiksel özellikleri taşırlar:

1. Yansızdırlar. Diğer bir deyişle beklenen değerleri gerçek değerlerine eşittir. Örnek: $E(\hat{\beta}_1) = \beta_1$.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri, doğrusal ve doğrusal-dışı tüm yansız tahminciler içinde enaz varyanslıdır.
3. Tutarlıdırlar. Örneklem sonsuza doğru büyürken gerçek değerlerine yakınsarlar.
4. $\hat{\beta}_1$ şöyle dağılır: $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$.
5. $\hat{\beta}_2$ şöyle dağılır: $\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$.
6. $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincileri $\hat{\sigma}^2$'dan bağımsız olarak dağılırlar.
7. $(n-2)\hat{\sigma}^2/\sigma^2$ değeri ise $n-2$ sd ile χ^2 dağılımlıdır. Bu bilgi, σ^2'ye ilişkin çıkarsamalarda $\hat{\sigma}^2$'den yararlanabilmek içindir.

Normallik Varsayımı Altında SEK Tahmincileri

- Normallik varsayımı yardımı ile $\hat{\beta}_1$ (normal), $\hat{\beta}_2$ (normal) ve $\hat{\sigma}^2$ (ki-kare ile ilgili) örneklem dağılımı bilgilerine ulaşıyoruz.
- Bu durum güven aralıklarını belirlemek ve önsav sınaması yapabilmek için önemlidir.
- Dikkat: Y_i'ler u_i'lerin bir işlevi olduğuna göre, $u_i \sim N(0, \sigma^2)$ varsayımı altında $Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$ olur.

Normal Dağılımla İlişkili Dağılımlar

- F, t ve ki-kare (χ^2) olasılık dağılımları temelde normal dağılımla ilişkilidirler.
- Bu dağılımların normal dağılımla ilişkilerini özetleyen yedi kanıtsav bulunmaktadır.
- Bu kanıtsavların uygulamadaki önemi büyüktür. Yararları ileride daha iyi anlaşılacaktır.

Kanıtsav 1

Kanıtsav 1

$Z_1, Z_2, \ldots, Z_n$ değişkenleri $Z_i \sim N(\mu_i, \sigma_i^2)$ olan normal ve bağımsız dağılımlı rastsal değişkenler olsun. Bu durumda $Z = \sum k_i Z_i$ toplamı da aşağıda verilen ortalama ve varyans ile normal dağılır.

$$E(Z) = \sum k_i \mu_i$$
$$\text{var}(Z) = \sum k_i^2 \sigma_i^2$$

- Kısaca normal rastsal değişkenlerin doğrusal bileşimleri de normal dağılır.
- Örnek: $Z_1 \sim N(10, 2)$ ve $Z_2 \sim N(8; 1{,}5)$ varsayalım ve Z rd'si de $Z = 0{,}8Z_1 + 0{,}2Z_2$ olsun. Buna göre Z
$$0{,}8(10) + 0{,}2(8) = 9{,}6 \text{ ortalama ve}$$
$$0{,}64(2) + 0{,}04(1{,}5) = 1{,}34 \text{ varyans ile}$$
$Z \sim N(9{,}6; 1{,}34)$ olur.

Kanıtsav 2

Kanıtsav 2

$Z_1, Z_2, \ldots, Z_n$ rastsal değişkenleri normal dağılımlı olsun ancak bağımsız olmasın. Bu durumda $Z = \sum k_i Z_i$ toplamı da aşağıda verilen ortalama ve varyans ile normal dağılır.

$$E(Z) = \sum k_i \mu_i$$
$$\mathrm{var}(Z) = \left[\sum k_i^2 \sigma_i^2 + 2 \sum k_i k_j \mathrm{cov}(Z_i, Z_j), i \neq j\right]$$

- Örnek: $Z_1 \sim N(6, 2)$, $Z_2 \sim N(7, 3)$, $\mathrm{cov}(Z_1, Z_2) = 0{,}8$ olsun. Z rd'si de $Z = 0{,}6Z_1 + 0{,}4Z_2$ olsun. Buna göre Z
$$0{,}6(6) + 0{,}4(7) = 6{,}4 \text{ ortalama ve}$$
$$0{,}36(2) + 0{,}16(3) + 2(0{,}6)(0{,}4)(0{,}8) = 1{,}58 \text{ varyans ile}$$
$Z \sim N(6{,}4; 1{,}58)$ olur.

Kanıtsav 3

Kanıtsav 3

$Z_1, Z_2, \ldots, Z_n$ değişkenleri $Z_i \sim N(0, 1)$ ölçünlü normal ve aynı zamanda bağımsız rastsal değişkenler olsun. Bu durumda $\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ toplamı da n serbestlik derecesi ile ki-kare dağılımına uyar.

- Kısaca ölçünlü normal dağılımlı bağımsız rd'lerin kareleri toplamı, toplam terim sayısına eşit serbestlik derecesi ile ki-kare dağılımına uyar.
- Bir yazım kolaylığı olarak sd'si k olan χ^2 dağılımı χ_k^2 diye gösterilebilir.
- Örnek: $Z_1, Z_2, Z_3 \sim NBD(0, 1)$ olsun. Öyleyse şu geçerlidir:

$$Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

Kanıtsav 4

Kanıtsav 4

$Z_1, Z_2, \ldots, Z_n$ değişkenleri her birinin sd'si k_i olmak üzere χ^2 dağılımlı ve bağımsız rastsal değişkenler olsun. Bu durumda bunların toplamı olan $\sum Z_i = Z_1 + Z_2 + \cdots + Z_n$ rastsal değişkeni de sd'si $k = \sum k_i$ olan ki-kare dağılımına uyar.

- Örnek: Z_1 ve Z_2, sd'leri sırasıyla 7 ve 9 olan iki bağımsız ki-kare değişkeni olsun. Buna göre $Z_1 + Z_2$ de serbestlik derecesi $7 + 9 = 16$ olan bir χ^2_{16} değişkenidir.

Kanıtsav 5

Kanıtsav 5

$Z_1 \sim N(0,1)$ ölçünlü normal ve Z_2 de Z_1'den bağımsız ve k sd ile χ^2 dağılımına uyan bir rastsal değişken olsun. Bu durumda $t = Z_1/(\sqrt{Z_2/k})$ şeklinde tanımlanan değişken de k sd ile Student t dağılımına uyar.

- Serbestlik derecesi k sonsuza doğru yaklaştıkça Student t dağılımı ölçünlü normal dağılıma yakınsar.
- Bir yazım kolaylığı olarak k sd'li t dağılımı t_k diye gösterilir.

Kanıtsav 6

Kanıtsav 6

Z_1 ve Z_2, serbestlik dereceleri sırasıyla k_1 ve k_2 olan ve bağımsız dağılımlı birer ki-kare değişkeni olsun. Bu durumda $F = (Z_1/k_1)/(Z_2/k_2)$ olarak tanımlanan F değişkeni de k_1 pay ve k_2 payda serbestlik derecesi ile F dağılımına uyar.

- Diğer bir deyişle, F değişkeni kendi sd'lerine bölünmüş iki bağımsız χ^2 değişkeni arasındaki oranı gösterir.
- Bir yazım kolaylığı olarak, sd'leri k_1 ve k_2 olan F dağılımlı değişken F_{k_1,k_2} diye gösterilir.

Kanıtsav 7

Kanıtsav 7

Serbestlik derecesi k olan t rastsal değişkeninin karesi, pay sd'si $k_1 = 1$ ve payda sd'si $k_2 = k$ ile F dağılımlıdır.

- Örnek: $F_{1,4} = (t_4)^2$ 'dir.
- Örnek: $(t_{25})^2 = F_{1,25}$ olur.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Ençok olabilirlik yöntemi