Çoklu Bağlanım – Çıkarsama Sorunu

F Sınamaları

Ekonometri 1 – Konu 25

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011

Ders Planı

Bağlanımın Bütününün Anlamlılık Sınaması

- Türkiye örneğimize dönelim ve β_2 ve β_3'ün aynı anda sıfır olduğunu öneren $H_0 : \beta_2 = \beta_3 = 0$ önsavını ele alalım.
- Bu sıfır önsavının sınanmasına, bağlanıma ilişkin "bütününün anlamlılığı" (overall significance) sınaması adı verilir.
- Bu sınama tekil anlamlılık sınamalarından farklıdır.
- Bunun nedeni şudur: $\hat{\beta}_2$ ve $\hat{\beta}_3$ gibi farklı katsayılar için tekil anlamlılık sınaması yaparken, her bir sınamanın farklı ve bağımsız bir örnekleme dayandığı varsayılır.
- Diğer yandan, verili bir örneklemde $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = 0$ geçerli olmayabilir. Diğer bir deyişle, $\hat{\beta}_2$ ile $\hat{\beta}_3$ ilişkili olabilirler.
- Bu durumda, $\hat{\beta}_2$ ile $\hat{\beta}_3$'nın aynı anda $[\hat{\beta}_2 \pm\ t_{\alpha/2}\text{öh}(\hat{\beta}_2)]$ ve $[\hat{\beta}_3 \pm\ t_{\alpha/2}\text{öh}(\hat{\beta}_3)]$ aralıklarında bulunma olasılığı $(1-\alpha)^2$ değildir.

Bağlanımın Bütününün Anlamlılık Sınaması

- Anakütle kısmi bağlanım katsayılarının aynı anda sıfır olduğu yönündeki ortak önsavı sınamak için varyans çözümlemesi yöntemi kullanılabilir:

$$\begin{aligned} \sum y_i^2 &= \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} && + \sum \hat{u}_i{}^2 \\ \text{TKT} &= \text{BKT} && + \text{KKT} \end{aligned}$$

- Buna göre aşağıdaki VARÇÖZ çizelgesini düzenleyebiliriz:

Değişimin Kaynağı	KT	sd	OKT (KT/sd)
Bağlanımdan (BKT)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	$k-1$	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{k-1}$
Kalıntılardan (KKT)	$\sum \hat{u}_i{}^2$	$n-k$	$\frac{\sum \hat{u}_i{}^2}{n-k} = \hat{\sigma}^2$
Toplamlarından (TKT)	$\sum y_i^2$	$n-1$	

- Burada k, sabit terim ile birlikte tahmin edilen toplam anakütle katsayılarının sayısıdır.

Bağlanımın Bütününün Anlamlılık Sınaması

- Üçlü model için, hata teriminin normal dağıldığı varsayımı ve $\beta_2 = \beta_3 = 0$ sıfır önsavı altında şu istatistik hesaplanır:

$$F = \frac{(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})/(k-1)}{\sum \hat{u}_i^2/(n-k)} = \frac{\text{BKT}/sd}{\text{KKT}/sd}$$

- Yukarıda verilen değişkenin $(k-1)$ ve $(n-k)$ sd ile F dağılımına uyduğu gösterilebilir.
- Buna göre, hesaplanan F istatistiğinin p değeri yeterince küçükse H_0 reddedilir.

Bütünün Anlamlılık Sınaması Açıklayıcı Örnek

Bağlanımın bütününün anlamlılığının sınanmasına örnek olarak Türkiye için gelir, işsizlik, ve göç alma modelimize dönelim ve aşağıdaki VARÇÖZ çizelgesini oluşturalım:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	205,022	2	102,511
Kalıntılar	448,080	78	5,74461
Toplam	653,102	80	

- F değeri çizelgeden aşağıdaki gibi hesaplanır:

$$F = \frac{102{,}511}{5{,}74461} = 17{,}8448$$

- Yüzde 5 anlamlılık düzeyinde ve 2 ile 78 sd için kritik değer $F_{0{,}05}(2,78) = 3{,}1138$'dir.
- Hesaplanan F değeri anlamlı olduğu için H_0 reddedilir.

Çoklueşdoğrusallığın Etkisi

- Göstermiş olduğumuz yöntemle hesaplanan F istatistiği çoğu zaman yüksek çıkar.
- Tüm bağlanım katsayıları tek tek istatistiksel olarak anlamlı değilken, F değerinin anlamlı çıkması olasıdır.
- Bu durum açıklayıcı değişkenler kendi aralarında yüksek derecede ilinti gösteriyorsa karşımıza çıkabilir.
- Bu sorunu ileride çoklueşdoğrusallık başlığı altında ayrıntılı biçimde ele alacağız.
- Şimdilik F ve t sınama sonuçlarını yorumlarken dikkatli olmak gerektiğini vurgulamakla yetiniyoruz.

R^2 ve F Arasındaki İlişki

- Belirleme katsayısı R^2 ile varyans çözümlemesindeki F değeri arasında yakın bir ilişki vardır.
- k değişkenli durumda ve $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ sıfır önsavı altında şu gösterilebilir:

$$
\begin{aligned}
F &= \frac{n-k}{k-1}\frac{\text{BKT}}{\text{KKT}} \\
&= \frac{n-k}{k-1}\frac{\text{BKT}}{\text{TKT}-\text{BKT}} \\
&= \frac{n-k}{k-1}\frac{\text{BKT}/\text{TKT}}{1-(\text{BKT}/\text{TKT})} \\
&= \frac{n-k}{k-1}\frac{R^2}{1-R^2}
\end{aligned}
$$

- Burada $R^2 = \text{BKT}/\text{TKT}$ tanımı kullanılmıştır.
- Eşitliğe göre R^2 ile F aynı yönde değişirler.
- Tahmin edilen bağlanımın bütün olarak anlamlılığının ölçüsü olan F demek ki aynı zamanda $H_0 : R^2 = 0$ sınamasına eşdeğerdir.

R^2 ve F Arasındaki İlişki

- F sınamasının R^2 cinsinden gösterilmesinin üstün yanı hesaplama kolaylığıdır. Tek gereken R^2 değeridir.
- VARÇÖZ çizelgesini R^2 ile aşağıdaki gibi düzenleyebiliriz:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	$R^2(\sum y_i^2)$	$k-1$	$R^2(\sum y_i^2)/(k-1)$
Kalıntılar	$(1-R^2)(\sum y_i^2)$	$n-k$	$(1-R^2)(\sum y_i^2)/(n-k)$
Toplam	$\sum y_i^2$	$n-1$	

Bir Açıklayıcı Değişkenin Marjinal Katkısı

- Bir açıklayıcı değişkenin marjinal katkısına bakmak için, X_2 ve X_3 gibi iki değişkeni modele sırayla ekleyelim.
- Burada görmek istediğimiz, eklenen değişkenin KKT'yi eskiye oranla ne ölçüde azalttığıdır.
- Çoğu görgül çalışmada, çeşitli olası X değişkenleri içinden KKT'yi çok azaltmayanları modele eklememek yeğlenebilir.
- Aynı şekilde KKT'yi önemli ölçüde azaltan, diğer bir deyişle R^2'yi "anlamlı" biçimde yükselten değişkenler de modelden çıkartılmak istenmez.
- Bu yüzden bir X değişkeninin marjinal katkısı uygulamada önemli bir konudur.

Bir Açıklayıcı Değişkenin Marjinal Katkısı

- Ek bir açıklayıcı değişkenin KKT'yi anlamlı biçimde azaltıp azaltmadığını bulmak için yine varyans çözümlemesinden yararlanılabilir.
- Türkiye'de illerin aldığı göç örneğimize dönelim ve şu ikili bağlanımı tahmin edelim:

$$\begin{array}{llll} \hat{Y}_i = & 4{,}8832 & + 1{,}8800\, X_{2i} & \\ \text{öh} & (0{,}6197) & (0{,}3717) & \\ t & (7{,}8797) & (5{,}0574) & r^2 = 0{,}2446 \end{array}$$

- Bulgular, kişi başına düşen GSYH'yi gösteren X_2'nin Y'yi anlamlı biçimde etkilediğini göstermektedir.

Bir Açıklayıcı Değişkenin Marjinal Katkısı

- İkili bağlanıma ait VARÇÖZ çizelgesi aşağıdaki gibidir:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	159,732	1	159,732
Kalıntılar	493,370	79	6,24519
Toplam	653,102	80	

- Şimdi, ildeki erkek işsizlik oranını gösteren X_3 değişkenini modele eklemek istediğimizi varsayalım.
- Üçlü bağlanıma ait VARÇÖZ çizelgesi ise şöyle idi:

Değişimin Kaynağı	KT	sd	OKT
Bağlanım	205,022	2	102,511
Kalıntılar	448,080	78	5,74461
Toplam	653,102	80	

- KKT'deki $(493{,}370 - 448{,}080 = 45{,}290)$ birimlik azalmanın istatistiksel olarak anlamlı olup olmadığını bulmak istiyoruz.

Bir Açıklayıcı Değişkenin Marjinal Katkısı

- X_2'nin katkısı biliniyorken, X_3'ün marjinal katkısı şu sınama istatistiği ile ölçülebilir:

$$F = \frac{Q_3/sd}{Q_4/sd} = \frac{(\text{KKT}_{eski} - \text{KKT}_{yeni})/m}{\text{KKT}_{yeni}/(n-k)}$$

- Burada *m*, yeni modele eklenen değişken sayısını gösterir.
- Elimizdeki örnek için *F* istatistiği şu şekilde hesaplanır:

$$F = \frac{(493{,}370 - 448{,}080)/1}{448{,}080/78} = 7{,}884$$

- Bulunan istatistik anlamlıdır. İldeki erkek işsizlik oranını eklemek KKT'yi anlamlı biçimde azaltmaktadır.

Bir Açıklayıcı Değişkenin Marjinal Katkısı

- Eldeki F oranı R^2 değerlerini kullanarak da bulunabilir:

$$F = \frac{(R^2_{yeni} - R^2_{eski})/m}{(1 - R^2_{yeni})/(n-k)}$$

- Örneğimiz için:

$$F = \frac{(0{,}3139 - 0{,}2446)/1}{(1 - 0{,}3139)/78} = 7{,}878$$

- Bu da yuvarlama hataları dışında önceki değer ile aynıdır.

Yeni Bir Değişken Ne Zaman Eklenmeli?

- Araştırmacılar çoğu zaman aynı bağımlı değişkeni içeren ama açıklayıcı değişkenleri farklı olan modeller arasında seçim yapmak durumunda kalırlar.
- Böyle durumlardaki genel eğilim en yüksek $\bar{R}^2$'yi seçmek yönündedir.
- Diğer yandan, yeni eklenen bir değişkenin katsayısının t değeri mutlak olarak 1'den büyük olduğu sürece $\bar{R}^2$ artar.
- Diğer bir deyişle yeni eklenen bir değişkene ilişkin $F(= t^2)$ değeri 1'den büyükse, bağlanım $\bar{R}^2$ değeri de yükselir.
- Demek ki $\bar{R}^2$ değerini yükselttiği halde KKT'yi istatistiksel olarak anlamlı ölçüde azaltmayan bir ek değişkenin modele eklenmesi konusunda dikkatli olunmalıdır.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Sınırlamalı enküçük kareler yöntemi