İki Değişkenli Bağlanım – Çıkarsama Sorunu

Çıkarsamaya İlişkin Konular

Ekonometri 1 – Konu 17

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011

Ders Planı

Varyans Çözümlemesi

- "Varyans çözümlemesi" (analysis of variance) ya da kısaca "VARÇÖZ" (ANOVA), istatistiksel çıkarsama sorununa tamamlayıcı ve aydınlatıcı bir yaklaşım sunar.
- Aşağıdaki özdeşliği anımsayalım:

$$\begin{aligned}
\sum y_i^2 &= \sum \hat{y}_i{}^2 + \sum \hat{u}_i{}^2 \\
\sum y_i^2 &= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i{}^2 \\
\text{TKT} &= \text{BKT} + \text{KKT}
\end{aligned}$$

- VARÇÖZ yaklaşımının temelinde TKT'nin bu iki parçasının incelenmesi yatar.

Varyans Çözümlemesi

- BKT 1 sd ile ve KKT de iki değişkenli model için $(n-2)$ sd ile ki-kare dağılımlıdır.
- O halde, toplamların kendi sd'lerine bölünmesi ile bulunan "ortalama kareleri toplamı" (mean sum of squares) ya da kısaca "OKT" (MSS) değerlerini kullanarak şunu yazabiliriz:

$$
\begin{aligned}
F &= \frac{\text{BKT'nin OKT'si}}{\text{KKT'nin OKT'si}} \\
&= \frac{(\hat{\beta}_2^2 \sum x_i^2)/1}{\sum \hat{u}_i{}^2/(n-2)}
\end{aligned}
$$

- Yukarıdaki değişken, hata teriminin normalliği varsayımı altında pay 1 ve payda $(n-2)$ sd ile F dağılımına uyar.

Varyans Çözümlemesi

- Tanımladığımız F oranından nasıl yararlanabileceğimizi görmek için aşağıdaki eşitliklere bakalım:

$$
\begin{aligned}
E\left((\hat{\beta}_2^2 \sum x_i^2)/1\right) &= \quad \ldots \quad = \sigma^2 + \beta_2^2 \sum x_i^2 \\
E\left(\sum \hat{u}_i^2/(n-2)\right) &= E(\hat{\sigma}^2) = \sigma^2
\end{aligned}
$$

- β_2 ve σ^2 gerçek anakütle katsayılarıdır.
- Eğer β_2 sıfır ise eşitliklerin her ikisi de aynı çıkar.
- Demek ki, F oranı bize $H_0 : \beta_2 = 0$ sıfır önsavını sınamada kullanılabilecek bir sınama istatistiği vermektedir.
- Dikkat: Bu durum iki değişkenli bağlanım için geçerlidir. F oranının çoklu bağlanımdaki yorumu farklıdır.

Varyans Çözümlemesi Örnek

Varyans çözümlemesine örnek olarak, Türkiye gelir-tüketim örneğimize dönelim.

- F değeri 1213,49 olarak hesaplanmaktadır.
- Anlamlılık düzeyi %5 iken, 1 ve 18 sd için kritik F değeri 4,41 olarak verilidir.
- Elimizdeki F istatistiği kritik değerden büyük olduğu için, $\beta_2 = 0$ önsavını reddederek Türkiye'de gelirin, özel tüketim harcamaları üzerinde etkili olduğunu söyleyebiliriz.
- Bu noktada, k sd ile t dağılımına uyan değişkenin karesinin de 1 ve k sd ile F dağılımına uyduğunu da anımsayalım.
- $H_0 : \beta_2 = 0$ altında tahmin edilen t değeri 34,84'tür.

Varyans Çözümlemesi Örnek

- Yuvarlama hatalarını bir yana bırakırsak $t^2 = (34{,}84)^2 = F$ eşitliğinin geçerli olduğunu görüyoruz.
- Bu nedenle, iki değişkenli bağlanım için F sınamasına aslında gerek yoktur.
- Şimdilik F ve t sınamalarının $\beta_2 = 0$ sınamasının iki farklı ve birbirini tamamlayıcı yolu olduğunu söyleyebiliriz.
- F sınamasının önemini ve farklı uygulamalarını çoklu bağlanım konusu içerisinde ele alacağız.

Ortalama Kestirimi

- Örneklem katsayıları yanında tekil $\hat{Y}_i$ değerleri için de aralık tahmini ve önsav sınaması yapılabilir.
- Örnek olarak, aşağıdaki örneklem bağlanımına bakalım:

$$\hat{Y}_i = 25 + 2X_i$$

- Katsayı tahminlerine dayanarak $E(Y|X_0 = 100)$ kestirimini yapmak istediğimizi varsayalım.
- Bu "ortalama kestirimi" (mean prediction) şöyle bulunur:

$$\begin{aligned}\hat{Y}_0 &= \hat{\beta}_1 + \hat{\beta}_2 X_0 \\ &= 25 + 2(100) = 225\end{aligned}$$

- $\hat{Y}_0$ burada $E(Y|X_0)$ tahmincisidir.

Ortalama Kestirimi

- $\hat{Y}_0$'nın, bir tahminci olmasından dolayı, kendi gerçek değerinden farklı çıkması söz konusudur.
- $\hat{Y}_0$ tahmincisinin aşağıda gösterilen ortalama ve varyans ile normal dağılımlı olduğu kanıtlanabilir:

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0 \qquad \mathrm{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

- Bilinmeyen σ^2 yerine yansız tahminci $\hat{\sigma}^2$ koyulduğunda ise bulunan değişken $(n-2)$ sd ile t dağılımına uyacaktır:

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\text{öh}(\hat{Y}_0)}$$

- Öyleyse, t dağılımını kullanarak $E(Y_0|X_0)$ güven aralığını bulabilir ve bunu önsav sınaması yapmada kullanabiliriz.

Ortalama Kestirimi

- $E(Y_0|X_0)$ güven aralığının tüm X'ler için hesaplanması ile anakütle bağlanım işlevine ilişkin bir "güven kuşağı" (confidence band) elde edilebilir.
- Bu güven kuşağı $X_0 = \bar{X}$ olduğunda en dar noktadadır. X_0 değeri $\bar{X}$'den uzaklaştıkça kemer de genişler.
- Dolayısıyla, örneklem ortalaması $\bar{X}$'den uzaklaşıldıkça örneklem bağlanımının kestirim yeteneği de azalacaktır.

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Bağlanım Bulgularının Değerlendirilmesi

Verilere yakıştırılan model sonuçları yorumlanırken aşağıdaki üç ölçüt göz önüne alınmalıdır:

1. Tahmin edilen katsayıların işaretlerinin kuramsal ya da önsel bilgilere dayalı beklentilerle uyumluluğu,
2. Kuramsal ilişkinin istatistiksel olarak anlamlı olup olmadığı,
3. Bağlanım modelinin güvenilirliği ve kuramsal ilişkiyi açıklayabilme derecesi.

Bağlanım Bulgularının Değerlendirilmesi

Türkiye gelir-tüketim örneği için gretl bağlanım çıktısı şöyledir:

gretl: model 1

Dosya Düzenle Sınamalar Kaydet Çizitler Çözümleme LaTeX

```
Model 1: SEK (OLS), kullanılan gözlemler: 1987-2006 (T = 20)
Bağımlı değişken: tuketim

             katsayı     ölç. hata    t-oranı    p-değeri
  ---------------------------------------------------------
  const      8,03438     1,85509       4,331     0,0004    ***
  gsyih      0,593351    0,0170331    34,84      5,68e-18  ***

Bağımlı değişken ort    71,20574   Bağımlı değişken ö.s.   14,07370
Kalıntı kareleri top    55,00622   Bağlanım ö.h.            1,748114
R-kare                  0,985384   Ayarlamalı R-kare        0,984572
F(1, 18)                1213,489   P-değeri(F)              5,68e-18
Log-olabilirlik        -38,49591   Akaike ölçütü           80,99182
Schwarz ölçütü          82,98329   Hannan-Quinn            81,38058
ro                      0,613231   Durbin-Watson            0,748726
Not: ö.s. ve ö.h. ölçünlü sapma ve ölçünlü hatayı göstermektedir.
```

Bağlanım Bulgularının Değerlendirilmesi

Bağlanım bulgularını incelediğimizde şunları görürüz:

- Keynesci tüketim kuramı çerçevesinde β_1 otonom tüketimi, β_2 ise marjinal tüketim eğilimi MTüE'yi göstermektedir.
- Sıfır gelir gerçek hayatta gözlenen bir durum olmadığı için β_1'e otonom tüketim anlamı yüklemekten kaçınılmalıdır.
- β_2 ise önsel beklentilere uygun şekilde 1'den küçük ve 0,59 olarak tahmin edilmiştir. Buna göre, Türkiye'de milli gelir 1 TL arttığında tüketim de 59 kuruş artmaktadır.
- t istatistikleri ilgili anakütle değerinin sıfır olduğu varsayımı altında bulunmuştur. β_2 için $34{,}84 = 0{,}59335/0{,}017033$'tür.
- $\hat{\beta}_2$'ya ait p-değeri de 18 sd ile 34,84 ya da daha yüksek bir t değeri bulma olasılığını $5{,}68 \times e^{-18}$ olarak vermektedir. Demek ki MTüE'nin sıfırdan farklı olduğunu söyleyebiliriz.
- Yaklaşık 0,98 büyüklüğündeki r^2 değeri, özel tüketim harcamalarındaki değişimin %98 oranında milli gelirdeki değişim ile açıklanabildiğini söylemektedir.

Bağlanım Bulgularının Değerlendirilmesi

- Bağlanım sonuçlarının güvenilir olduğuna karar verebilmek için modelimizin KNDBM varsayımlarını sağladığını da onaylamak zorundayız.
- Şu an tüm KNDBM'nin varsayımlarını denetleyemesek de u_i hata teriminin normalliği varsayımına bakabiliriz.
- Yazında çeşitli normallik sınamaları bulunmaktadır. Biz bunlardan ki-kare "yakışmanın iyiliği" (goodness of fit) ve Jarque-Bera normallik sınamalarını ele alacağız.
- Bu sınamaların ikisi de $\hat{u}_i$ kalıntılarını ve ki-kare olasılık dağılımını temel almaktadır.

χ^2 Yakışmanın İyiliği Sınaması

χ^2 yakışmanın iyiliği sınamasının adımları şöyledir:

1. Bağlanım işlevi bulunur ve $\hat{u}_i$ kalıntıları elde edilir.
2. $\hat{u}_i$'nın örneklem ölçünlü sapması hesaplanır.
3. Örneklem büyüklüğüne göre bir "kap" (bin) sayısı belirlenir. Kalıntılar büyüklük sırasına sokulur ve sıfırdan kaç ölçünlü sapma uzaklıkta olduklarına göre bu kaplara bölüştürülür.
4. Gözlenen sıklıklar (G_i) ile normal dağılım için beklenen sıklıklar (B_i) arasındaki farkların kareleri alınır, beklenen sıklıklara bölünür ve bunların toplamı hesaplanır:

$$\chi^2 = \sum_{i=1}^{k} \frac{(G_i - B_i)^2}{B_i}$$

$k =$ kap sayısı iken, yukarıdaki değişken $(k-3)$ (normal dağılıma karşı sınadığımız için) sd ile χ^2 dağılımına uyar.

5. Eğer p değeri yüksekse H_0 : normallik önsavı reddedilmez.

Jarque-Bera Normallik Sınaması

JB sınaması bir “kavuşmazsal” (asymptotic) ya da büyük örneklem sınamasıdır. Şu şekilde yapılır:

1. Öncelikle SEK kalıntılarının “çarpıklık” (skewness) ve “basıklık” (kurtosis) ölçüleri bulunur.
2. Daha sonra aşağıdaki istatistik hesaplanır:

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

   Burada S çarpıklığı, K ise basıklığı göstermektedir.

   Jarque ve Bera, 1987 tarihli bir çalışmalarında, kalıntıların normal dağıldığı varsayımı altında JB istatistiğinin büyük örneklemde 2 sd ile χ^2 dağılımlı olduğunu göstermişlerdir.
3. Eğer hesaplanan sınama istatistiğine ait p değeri yüksekse H_0 : normallik önsavı reddedilmez.

Jarque-Bera Normallik Sınaması

- Ki-kare sınamasının çekici yanı, "yığınsal dağılım işlevi" (cumulative distribution function) hesaplanabilen her türlü dağılım için yakışmayı sınamak için kullanılabilmesidir.
- Sakıncası ise kap sayısının nesnel bir ölçütü olmadığı için hesaplanan χ^2 değerinin farklılık gösterebilmesidir.
- SEK yönteminin istatistiksel özelliklerinden dolayı, büyük örneklemlerde normallik sınaması çoğu zaman gerekmez.
- Bağlanım ile ilgili olarak normallik sınaması daha çok bir küçük örneklem konusudur.
- Diğer yandan, JB kavuşmazsal bir sınama olduğu için küçük örneklemlerde ki-kare dağılımından sapmaktadır.
- Örnek olarak, $n = 70$ gibi çok da küçük sayılamayacak örneklemlerde bile bulunan JB p değeri yanıltıcı olabilir.
- Bu yüzden Jarque-Bera yerine Doornik-Hansen sınaması yazın "literature" içinde yeğlenmektedir.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Sıfır noktasından geçen bağlanım