

Tobit Model ve Bir Uygulama

Şeyma KOÇ¹ , Mustafa ŞAHİN¹ 

¹Kahramanmaraş Sütçü İmam Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Kahramanmaraş

✉ : koçkutlu1@gmail.com

ÖZET

Bu çalışmada sınırlı bağımlı değişkenli modelleri açıklamada yaygın biçimde kullanılan Tobit Model incelenmiştir. Çalışmanın amacı bağımlı değişkeni sınırlı olan veri setlerinin parametre tahmin basamaklarını hem teorik olarak hem de bir uygulama ile anlatmaya çalışmaktır. Literatür tarama yöntemi kullanılarak tobit model ana çizgileri ile anlatılmış, toplanan bilgiler bir örnek ile sayısallaştırılarak daha anlaşılır hale getirilmiştir. Çalışmanın sonucunda analiz süreçlerinde sürekli ve kesikli karma veri yapısı ile karşılaşan araştırmacılara tobit regresyon modelinin teorisi sunulmuş ve istatistiksel çıkarıma sürecinde Tobit model kullanımını önerilmiştir.

DOI:10.18016/ksudobil.285929

Makale Tarihçesi

Received : 15.01.2017

Accepted : 15.03.2017

Anahtar Kelimeler

Sansürlü Veri,
Kırpılmış Veri,
Tobit Analiz

Araştırma Makalesi

Tobit Model and an Application

ABSTRACT

In this study, Tobit Model which is widely used to describe the limited dependent variable models was examined. The aim of the study was to try to explain parameter estimation steps of data sets, which dependent variable is limited, both theoretically and by an application. By using the literature survey method, main lines of Tobit model were explained and the gather information was digitized with a sample to make it more understandable. As a result of the study, the theory of the tobit regression model was presented and the use of the Tobit model was proposed in the statistical inference process.

Article History

Geliş : 15.01.2017

Kabul : 15.03.2017

Keywords

Censored Data,
Truncated Data,
Tobit Analysis

Research Article

To Cited : Koç Ş, Şahin M 2018. Tobit Model ve Bir Uygulama. KSÜ Tarım ve Doğa Derg 21(1):73-80, DOI:10.18016/ksudobil.285929.

GİRİŞ

İstatistik biliminde kullanılan temel analizlerden biri olan regresyon analizi, bir bağımlı (açıklanan) değişkenin, bir veya daha fazla bağımsız (açıklayıcı) değişkenle ilişkisinin matematiksel bir fonksiyonla ifade edilmesi amacı ile kullanılmaktadır. Ancak bazı deney süreçlerinde örnekteki bağımlı değişkenlerin tüm değerlerine ulaşamaz. Bağımlı değişkenin aldığı değerler sınırlı olduğu için bazı yazarlar tarafından bu modellere “*Bağımlı Değişkeni Sınırlı Modeller*” adı verilmiştir. Sınırlı değişkenler olarak da adlandırılan bu tür durumlarla sıklıkla karşılaşan araştırmacılar, bu değişkenlerin sınırlanmış yapısını, sansürlenmiş (Censored) ve kırpılmış (Truncated) veriler şeklinde ikiye ayırmak zorunda kalmışlardır (Kennedy, 2006).

Sınırlı bağımlı değişkenli modelleri açıklamada yaygın

biçimde kullanılan Tobit model, gerek ilk kullanıcısı olan James Tobin'e ithafen, gerekse probit modele çok benzemesinden bu ismi almıştır. Tobin, ilk kez 1958 yılında dayanıklı tüketim malları üzerine hane halkı harcamalarını analiz ederken kimi ailelerin dayanıklı tüketim malı harcaması gibi bir harcama kaleminin olmaması sebebiyle bağımlı değişkeni negatif çıkan bir regresyon türü ile karşılaşmıştır. Tobin çalışmasında, harcamanın hiçbir zaman negatif olmayacağı gerçeğinden hareketle hane halkı geliri, belli bir düzeyi geçene kadar bu değişkene sıfır değeri atamıştır. O yıllarda tanımladığı bu model sansürlü regresyon modeline klasik bir örnektir (Tobin, 1958). Bu çalışmada kullanılan model aslında negatif olmayan bağımlı değişkenli birer doğrusal regresyon modelidir. Fiyat, ücret, harcama yaşam süresi(ömür),

miktar, bir sistemin ya da bir makinenin dayanıklılığı... vb negatif değerler almayan değişkenler üzerine kurulu modeller literatürde 'sınırlı bağımlı değişkenli' modeller olarak bir çok disiplinde yaygın olarak kullanılmaktadır.

MATERYAL ve METOT

Materyal

Bu çalışmada, Kahramanmaraş'ta Hane Halklarının Gıda Tüketim Talebi Ekonometrik Analizi konulu anket verilerinden tesadüfi olarak seçilen 200 deneye ait veri seti kullanılmıştır (Akbaş, 2005). Verilerden hareketle ailedeki fert sayısı, aile reisinin aylık geliri, kişinin eğitim düzeyi gibi bilgileri bağımsız değişken olarak kullanılarak, bu değişkenlerin kişilerin aylık balık tüketim harcaması üzerine etkisi belirlenmeye çalışılmıştır. (Koç,2013). Değişkenlerin analizinde SAS (Statistical Analysis System) paket programı kullanılmıştır (SAS, 2008). Esas olarak üç ana pencereden oluşan SAS uluslararası bir istatistiksel paket programdır (Orhan ve ark., 2004). Uygulamada kullanılan bağımlı değişken ve bağımsız değişkenler ve analiz sürecinde kullanılan kısaltılmış kodlamalar aşağıda verilmiştir.

Bağımlı Değişken:

Aylık balık tüketimi için harcanan miktar(TL)/BLK

Bağımsız Değişkenler:

1-Ailedeki fert sayısı /FS

2-Aile reisinin aylık geliri(TL)/GEL

3- Kişinin eğitim düzeyi/EGT

Gölge değişken: İlköğretim mezunu olmak, lise mezunu olmak, üniversite mezunu olmak/ D_i

Çalışmada regresyon denklemi tahmin edildikten sonra bu denklemin gerçek değerleri ne kadar yansıttığına dair model uygunluğu analizleri yapılmıştır.

Tahmin değerlerinin gerçek gözlem değerlerini ne kadar temsil ettiğinin bir ölçüsü olan belirleme katsayısı (R²) klasik regresyon modellerine göre sınırlı bağımlı değişkenli modellerde çok daha düşüktür (Tatoğlu, 2005). Kullanışlı bir çıkarsama yapmaya imkân tanımamasından dolayı bu modellerin analizinde R² nin yerine pseudo R² kullanılmaktadır (Eren, 2012). Bu modellerin doğruluk düzeyini test etmek için kullanılan pseudo R² için çeşitli öneriler vardır (Tatoğlu, 2005; Joost ve Kalbermatten, 2010). Ayrıca tobit modeli de kapsayan birçok modelde modelin uygunluk ölçüsü anlamlı değildir. Bir yorum yapılmak istenirse yaklaşık bir değer olarak gözlenen değerlerle tahmin değerleri arasındaki korelasyona bakılabilir. Bununla birlikte parametrelere konan sınırlama testi de pseudo R² lerin teorik aralıkları dışında kaldığı durumlarda kullanılmaktadır. Bunlar; tahmin edilen modelin gerçeği ne ölçüde tanımladığına dair elverişli ölçüler sunan Olabilirlik oranı(LR), Wald, Lagrange çarpanı(LM) testleridir. (Gujarati,1999; Chay ve Powell, 2001). 'Kısıtlamaların geçerliliği testleri' ya da 'parametrelere konan

sınırlama testi' olarak da bilinen bu testler modele hangi parametrenin alınacağı konusunda araştırmacıyı aydınlatır.

Metot

Sınırlı Bağımlı Değişkenli Modeller

Bağımlı değişkenin aldığı değerlerin sınırlı olduğu modellerle sıklıkla karşılaşan araştırmacılar, bu değişkenlerin sınırlanmış yapısını, sansürlenmiş (Censored) ve kırılmış (Truncated) veriler şeklinde ikiye ayırmak zorunda kalmışlardır. Çünkü veri setinin kırılması ve sansürlenmesi birbiri ile kolaylıkla karıştırılmaktadır. Dolayısıyla gerek verilerin elde edilmesi ve ardından modelin kurulması esnasında, gerekse tahmin ve çıkarsama süreçlerinde hassasiyet gösterilmelidir. Bu yüzden kimi zaman veri setinin başından, kimi zaman sonundan kimi zaman da her iki taraftan uç değerlerin kırılması ile ortaya çıkan bu örneklem süreçlerinin analizleri için de çeşitli metotlar geliştirilmiştir. Sansürlü ve kırılmış verileri birbirinden ayırt etmekte titizlik gösterilmelidir. Zira bu değişkenler kolayca birbirine karıştırılmaktadır. Eğer belirli bir aralığın dışındaki gözlemler tamamen kaybedilmekte ise kırılmış (truncated) model, en azından bağımsız değişkenler gözlenebiliyorsa sansürlü (censored) model söz konusu olmaktadır (Emeç ve ark., 2001).

Yapılan çalışmada bazı veriler sistematik olarak deneyden çıkarılıyorsa oluşan model kesilmiş ya da kırılmış model olacaktır. Kesikli modeller, belli bir oranda sınırlanmış gözlemleri olan bağımlı ve bağımsız değişkenler içeren modellerdir. Bundan dolayı bu modellere '*kırılmış modeller*' de denilmektedir. Hem bağımlı hem de bağımsız değişkenlere ait sınırlı ya da kayıp gözlemler vardır. Örneğin, çalışan hanımların ücret hadlerinin belirlenmesi için yapılan bir çalışmada veri seti, maaşları beklentilerine eşit ya da yüksek hanımlardan oluşur. Maaşları, beklentilerinin altında olan hanımlar çalışmamayı tercih edeceklerdir (Kennedy, 2006). Araba kullanma tercihi ile ilgili bir çalışmada ise örneklem 18 yaş altı için kırılmıştır.

Sansürlenmiş örneklemde bağımsız değişken verilerinin varlığına karşın bağımlı değişkenin bazı değerlerine ulaşamaz. Veri setinden sistematik olarak bir dışlanma söz konusu değildir ancak bağımlı değişkende gözlenemeyen kayıp veriler vardır. Örneğin et tüketiminin, gelir düzeyi ve ailedeki kişi sayısı ile ilişkisinin araştırıldığı bir çalışmada, hiç et tüketmeyen vejeteryenler için bağımsız değişken (ailedeki kişi sayısı, gelir düzeyi) değerleri mevcut iken bağımlı değişken (et tüketimi) değerleri yoktur. Bu durumda veri seti sansürlüdür. Tesadüfi olarak seçilen bir lisede, öğrencilerin sigaraya başlama yaşı ile yaşadığı yer (kent merkezi, ilçe, köy) ile ilişkisinin araştırıldığı bir çalışmada hiç sigara içmeyen öğrencilerin varlığı kaçınılmazdır. Tüm öğrencilerin bağımsız değişken olan yaşanan yer bilgisi mevcut

iken, bazı öğrenciler hiç sigaraya başlamadıkları için bu öğrencilere ait bağımlı değişken olan sigaraya başlama yaşı bilgisi yoktur. Bağımsız değişkenler gözleniyorken bağımlı değişkenler için veriye ulaşılamamaktadır.

Sansürlenmiş Değişken Modeli (Tobit Model)

1958 de Nobel ödüllü iktisatçı James Tobin tarafından geliştirilen, bağımlı değişkene ait bilginin yalnızca bazı gözlemler için bulunduğu sansürlü örneklem modeli olarak bilinir. En küçük kareler regresyonunun parametrik olmayan alternatifidir (Liao, 1994). Probit modelin bir uzantısı olduğundan tobin ve probit isimlerinden Tobit türetilmiştir. Model gizli bir y^* değişkenin varlığını kabul etmektedir. Bu değişken x_i açıklayıcı değişkenine β parametreler vektörü ile bağlıdır. Bu değişkenin dağılımı normal, ortalaması μ , varyansı σ^2 dir. y gözlenen değişken c keyfi sansür noktası olmak üzere model, $u \sim N(0, \sigma^2)$ ve $y^* \sim N(\mu, \sigma^2)$ varsayımlarına bağlıdır. Latent değişkenin c katsayısından büyük, küçük ya da eşit olma durumuna göre model,

$$\begin{aligned} y &= c \text{ eğer } y^* \leq c \text{ ise,} \\ y &= y^* \text{ eğer } y^* > c \text{ ise.} \end{aligned}$$

şeklinde kurulur. Bu ifade sıfır noktasında sansürlemenin genelleştirilmiş şeklidir. Eğer c değeri sıfıra eşitlenirse 'Standart Tobit Model' elde edilir (Carson ve Sun, 2007). Genel formül ile standart tobit model latent değişkene bağlı olarak aşağıdaki gibi ifade edilir.

$$\begin{aligned} y &= 0 \text{ eğer } y^* \leq 0 \text{ ise,} \\ y &= y^* \text{ eğer } y^* > 0 \text{ ise.} \\ y^* &= \beta'x_i + u_i \end{aligned} \quad (1)$$

Sansürlenmiş regresyon modeli kesiklik ve süreklilik vasfını birlikte taşımaktadır. Bu değişkenler aslında birer sürekli değişkenlerdir ancak veri setinin bir kısmı hakkında gözlem sahibi olunmadığı ya da sınırlandırıldığı için, özetle sansüre tabi tutulduğundan eksik kalan gözlemlerin yerine sıfır ataması yapılmaktadır. Bundan dolayı gözlem setinin kesikli özelliğe sahip olduğu ileri sürülmektedir (Maddala, 1987). Her ne kadar her iki yapıdaki (truncated, censored) bağımlı değişkenler sınırlı bağımlı değişkenler kategorisinde yer alsın da oluşturdukları örneklem farklıdır. Bu da söz konusu regresyon modelinde parametre tahminini ve çıkarımları ilgilendirmektedir.

Sansürlü Normal Değişkenin Olabilirlik Fonksiyonu

Bu değişkenin analizi yapılırken latent değişken y^* in yalnızca $y^* > 0$ olduğunda gözlemlendiği unutmamalıdır. y^* 'in sıfıra eşit ve sıfırdan küçük olduğu durumlara sıfır ataması yapılır. Olabilirlik fonksiyonu türetilirken $y^* \leq c$ ve $y^* > c$ olmak üzere iki aşamalı olarak düşünülmektedir. Gösterim kolaylığı sağlama açısından $P(\text{sansürlü})$ ve $P(\text{sansürsüz})$ gösterimleri kullanılacaktır. Burada, y ,

sürekli değişkenin olasılık yoğunluk fonksiyonu, $f(y)$: y değişkeninin olasılık yoğunluk fonksiyonunu, c : keyfi sansür katsayısını, d : gösterge değişkenini ifade etmek üzere fonksiyon,

$$f(y) = [f(y^*)]^{d_1} [F(c)]^{1-d_1} \quad (2)$$

biçiminde ifade edilir. Burada d gösterge değişkeni, $y > c$ olduğunda 1'e eşittir. $d=1$ iken y 'nin yoğunluğu $y > c$ (sansürsüz gözlemler) için y^* ile aynıdır. d değişkeni $y=c$ olduğunda 0'dır. $d=0$ iken y 'nin yoğunluğu $y^* \leq c$ nin gözlenme olasılığına eşittir. $y^* \leq c$ ve $y^* > c$ iken olasılıklar aşağıdaki gibi olacaktır (Chay ve Powell, 2001).

$P(\text{sansürlü}); y^* \leq c$ iken olasılık,

$$\begin{aligned} P(y_i = c) &= P(y^* \leq c) = P(\beta'x_i + u_i \leq c) \\ &= P(u_i \leq c - \beta'x_i) = P\left(\frac{u_i}{\sigma} \leq \frac{c - \beta'x_i}{\sigma}\right) \\ &= \Phi\left(\frac{c - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu - c}{\sigma}\right) \end{aligned} \quad (3)$$

$P(\text{sansürsüz}); y^* > c$ iken olasılık

$$P(y^* > c) = 1 - \Phi\left(\frac{c - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - c}{\sigma}\right) \quad (4)$$

Bu ifadelerle göre olabilirlik fonksiyonu $\mu = \beta'x_i$ dönüşümü yapılarak,

$$\begin{aligned} L &= \prod_{i=1}^N \left[-\Phi\left(\frac{\beta'x_i - c}{\sigma}\right) \right]^{1-d_i} \left[\frac{1}{\sigma} \phi\left(\frac{y_i - \beta'x_i}{\sigma}\right) \right]^{d_i} \\ L &= \prod_{y_i=0} \left[1 - \Phi\left(\frac{\beta'x_i}{\sigma}\right) \right] \prod_{y_i>0} \frac{1}{\sigma} \phi\left(\frac{y_i - \beta'x_i}{\sigma}\right) \\ \ln L &= \sum_{y_i=0} \ln \left[1 - \Phi\left(\frac{\beta'x_i}{\sigma}\right) \right] + \sum_{y_i>0} \ln \frac{1}{\sigma} \phi\left(\frac{y_i - \beta'x_i}{\sigma}\right) \end{aligned} \quad (5)$$

şeklinde formüle edilir (Park, 2003; Pettinicchi, 2012). $\ln L$ fonksiyonuna benzer biçimde Eşitlik 5'nin logaritması alınarak log-olabilirlik fonksiyonu da Eşitlik 7' deki gibi elde edilebilir.

$$\log L = \sum_{y_i=0} \log \left[1 - \Phi\left(\frac{\beta'x_i}{\sigma}\right) \right] + \sum_{y_i>0} -\frac{1}{2} \left[\log(2\pi) + \log \sigma^2 + \left(\frac{y_i - \beta'x_i}{\sigma}\right)^2 \right] \quad (7)$$

Bu oldukça karışık fonksiyon Olsen (1978)'in $\gamma = \frac{\beta}{\sigma}$ ve $\theta = \frac{1}{\sigma}$ dönüşümleri ile önemli ölçüde basitleştirilerek Eşitlik 8' deki şeklini almaktadır.

$$\log L = \sum_{y_i>0} -\frac{1}{2} [\log(2\pi) - \log \theta^2 + (\theta y_i - \gamma'x_i)^2] + \sum_{y_i=0} \log [1 - \Phi(\gamma'x_i)] \quad (8)$$

Sansürlenmiş Normal Değişkenin Momentleri

$P(\text{sansürlü})$ ve $P(\text{sansürsüz})$ olasılıklar kullanılarak beklenen değer ve varyans formülleri aşağıda verilmiştir.

$$E[y] = \left\{ \begin{array}{l} P(\text{sansürsüz}). E[y|y > c] \\ + P(\text{sansürlü}). E[y|y \leq c] \end{array} \right\} \quad (9)$$

$$E[y] = (1 - \Phi)(\mu + \sigma\lambda) + \Phi c \quad (10)$$

$$\text{Var}[y] = \sigma^2(1 - \Phi)[(1 - \delta) + (\alpha - \lambda)^2\Phi] \quad (11)$$

$c=0$ olduğu durumda $E[y] = (1 - \Phi)(\mu + \sigma\lambda)$ şeklindedir.

Burada, Φ , λ , δ ifadeleri, $\Phi = \Phi(\alpha) = \Phi\left(\frac{c-\mu}{\sigma}\right)$, $\lambda = \frac{\phi}{1-\Phi}$, $\delta = \lambda^2 - \lambda\alpha$ şeklinde tanımlanır (Lu,2012).

Tobit Modelde Beklenen Değer Ve Marjinal Etki

Sigelman ve Zeng (1999)'e göre standart tobit model için üç farklı beklenen değer hesabı vardır. Bu beklenen değerlerden hangisinin kullanılacağı araştırmacının amacına ve veri setinin hangi kısmı ile ilgilendiğine bağlı olarak değişiklik göstermektedir. Beklenen değer hesaplamaları genellikle istatistik paket programlarının içeriğinde mevcuttur. Farklı kaynaklarda bu duruma ilişkin farklı yorumlara rastlanılmaktadır. Greene (2003, s.764) beklenen değer kavramını şu şekilde açıklamıştır:

“Bu konuda herhangi bir mutabakat sağlanılmamıştır. Yalnızca latent değişkeni temel alan araştırmacının $E[y^*]$ ' den ziyade $E[y]$ beklenen değerini kullanması daha kullanışlıdır.”

Çizelge1. Beklenen değer ve marjinal etki için özet formüller

Beklenen değer		Marjinal etki
latent değişken y^* için	$E[y^*] = X_i\beta$	$\frac{\partial E[y_i^* x_i]}{\partial x_i} = \beta$
y tesadüfi değişkeni için	$E[y] = \Phi\left(\frac{X_i\beta}{\sigma}\right)[X_i\beta + \sigma\lambda(\alpha)]$	$\frac{\partial E[y_i x_i]}{\partial x_i} = \beta\Phi(\beta'x_i/\sigma)$
$y y > 0$ değişkeni için	$E[y y > 0] = X_i\beta + \sigma\lambda(\alpha)$	$\frac{\partial E[y_i x_i, y_i > 0]}{\partial x_i} = \beta \left\{ 1 - \lambda\left(\frac{\beta'x_i}{\sigma}\right) \left[\frac{\beta'x_i}{\sigma} + \lambda\left(\frac{\beta'x_i}{\sigma}\right) \right] \right\}$

Bu düzey diğerlerinin karşılaştırılması için bir başvuru düzeyidir. Eğer temel kategorik değişken, örneğin, ilköğretim düzeyi olarak belirlenirse farklı gölge değişken tanımlama yoluna gidilmiş olur. Yorumlamaların değişeceğine dikkat edilmelidir.

Çalışmada temel düzeyin D_1 değişkeni olarak seçildiği tahmin Çizelgesi de alternatif olarak sunulmuştur. Aylık balık harcama miktarının FS, GEL, EGT değişkenlerine göre regresyonunda üç eğitim düzeyinin $\alpha_1, \alpha_2, \alpha_3$ şeklinde üç sabit terimi vardır. Bunlardan α_1, α_2 sırayla ilköğretim ve lise düzeyinin sabit terimleri α_3 ise temel düzeyin sabit terimini ifade eder. Model şu şekilde kurulur,

$$BLK = \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 + \beta_1 FS + \beta_2 GEL + u_i$$

Çalışmada kullanılan değişken sayıları, ortalama, standart sapma, minimum ve maksimum değerleri gibi özet istatistikler Çizelge 2' de verilmiştir.

Çalışmada gözlem sayısı 200 dür. Bunlardan 127 tanesi sansürlü, 73 tanesi soldan sansürlüdür. Verilerin derlendiği süreçte 73 gözlemden aylık balık tüketimi harcaması sıfır olarak kaydedilmiştir. Harcamada alt sınır sıfır alınarak soldan sansürleme yapılmıştır. SAS istatistik paket programında

Wooldridge (2002, s.520) ise beklenen değerlerden hangisinin kullanılacağı konusu şu şekilde ifade edilmiştir.

“Eğer araştırmacı, bağımsız değişken veya değişkenlerin, sansürlü olsun ya da olmasın açıklayıcı değişken üzerine etkisini ölçüyorsa $E[y]$ beklenen değerini kullanmalı, sansürlenmemiş gözlemlerle ilgileniyorsa $E[y|y > 0]$ formülünü kullanmalıdır.”

Çizelge 1'de beklenen değer ve marjinal etki için özet formüller verilmiştir.

BULGULAR ve TARTIŞMA

Regresyon modelinde biri nitel (üç düzeyli), ikisi nicel olmak üzere üç bağımsız değişken kullanılmıştır. Gölge değişkenler (dummy variable) olarak bilinen nitel değişkenli regresyon modellerinin genel kuralı olan “m düzeyli değişken varsa buna karşılık m-1 tane gölge değişken kullanılır” kriteri gereğince üç düzeyli nitel EGT değişkeni için iki gölge değişken kullanılmalıdır. D_3 değişkeni temel düzey olarak keyfi seçilmiştir.

LİFEREG Prosedürü kullanılarak bulunan analiz sonuçları Çizelge 3' de verilmiştir.

Çalışmada kullanılan bağımsız değişkenlerden kişinin eğitim durumuna ait EGT bağımsız değişkeni nitel değişkendir ve üç düzeyi vardır. Bu sebeple analiz aşamasında temel kategori, keyfi olarak sırasıyla önce üniversite mezunu olmak (Çizelge 4), sonra ilköğretim mezunu olmak (Çizelge 5) düzeyleri seçilerek parametre tahmin çizelgesi iki farklı şekilde sunulmuştur.

Çizelge 4'de parametre tahmin Çizelgesinin son satırında bulunan SCALE değeri hata varyansını temsil eder. SAS istatistik programında sansürlü değişken analizinde kullanılan QLIM prosedürü için bu sütun _SİGMA şeklinde görülmektedir. Çıkan sonuçlara göre fert sayısı ve eğitim bağımsız değişkenlerin katsayılarının anlamlılığını ölçmek için kullanılan ki-kare test istatistiklerinin p değeri 0.05 anlamlılık düzeyinden büyük çıkmıştır. Bu durumda bu değişkenlerin katsayılarının istatistiksel olarak anlamsız olduğu söylenebilir. Gelir değişkeninin p değeri olan 0.0193 değeri 0.05 ten düşük çıktığı için istatistiksel olarak önemli bulunmuştur.

Çizelge 2. Tanımlayıcı istatistikler

EGT	Değişkenler	N	\bar{X}	S	Min	Max
İlköğretim	BLK	149	28.71	37.94	0.00	280.00
	FS	149	4.38	2.09	1.00	17.00
	GEL	149	601.14	497.67	50.30	3063.00
Lise	BLK	38	27.68	33.17	0.00	135.00
	FS	38	4.26	1.81	2.00	9.00
	GEL	38	761.62	550.38	261.00	3000
Üniversite	BLK	13	49.42	111.39	0.00	415.00
	FS	13	3.46	0.52	3.00	4.00
	GEL	13	1127.05	508.60	514.20	2175.00

Çizelge 3. Tobit analiz sonuçları

LIFEREG Prosedürü	
Bağımlı değişken	Balık harcaması
Gözlem sayısı	200
Sansürlü gözlem sayısı	127
Sağdan sansürlü gözlem sayısı	0
Soldan sansürlü gözlem sayısı	73
Log Likelihood değeri	-759.6424412
İterasyon sayısı	28
AIC	1531
Schwarz kriteri	1551

Çizelge 4. Parametre tahmini (1)(Temel düzey: Üniversite mezunu)

PARAMETRE	sd	β	S	95% Güven aralıkları		Ki-Kare	Pr>Ki-Kare
SABİT	1	10.5622	21.6416	-31.8546	52.9789	0.24	0.6255
FS	1	0.8554	2.4062	-3.8607	5.5715	0.13	0.7222
GEL	1	0.0213	0.0091	-0.0035	0.0392	5.48	0.0193*
EGT	D ₁	-14.6929	19.3625	-52.6427	23.2569	0.58	0.4480
EGT	D ₂	-20.7474	21.0797	-62.0629	20.5680	0.97	0.3250
EGT	D ₃	0
SCALE	1	61.2757	4.1047	53.7364	69.8728	.	.

Çizelge 5. Parametre tahmini (2)(Temel düzey: İlköğretim mezunu)

PARAMETRE	sd	β	S	t	Pr>t
SABİT	1	-4.1306	12.5220	-0.33	0.7415
FS	1	0.8554	2.4062	0.36	0.7222
GEL	1	0.0213	0.0091	2.34	0.0193*
EGT	D ₁	0	.	.	.
EGT	D ₂	-6.0545	21.0338	-0.50	0.6149
EGT	D ₃	14.6928	19.3624	0.76	0.4480
_SIGMA	1	61.2757	4.1047	14.93	<.0001

Çizelge 4'deki katsayılar kullanılarak oluşturulan model,

$$E(Y|X) = 10.5622 + 0.8554FS + 0.0213GEL - 14.6929(D_1) - 20.7474(D_2)$$

şeklinde ifade edilebilir. Bu katsayıların önem testi klasik yollarla yapılabilir. Sonuçların yorumlanması ise EKK regresyonun yorumuna benzerdir.

Örneğin, ailedeki fert sayısındaki bir birimlik artış diğer değişkenlerin sabit tutulması halinde balık tüketimi harcama miktarında 0,8554 lık artışa sebep olur. Gelir değişkenindeki bir birimlik artış ise balık tüketimi harcama miktarında 0,0213 lık artışa sebep olur. Bütün değişkenler sabit tutulduğunda ise kişilerin aylık balık harcama miktarları ortalama 10-

11TL civarında olması beklenmektedir.

Eğitim değişkeninin yorumu diğer nicel değişkenlerden farklıdır. Bu değişken üç düzeyi bulunan nitel bir değişkendir. Çalışmada üniversite değişkeni temel düzey olarak alınmıştır. Bu da nitel değişken yorumu sırasında araştırmacıya bir kıyas kolaylığı sağlar. Örneğin eğitim değişkeninin düzeylerinden biri olan ilköğretim mezunu olmak düzeyinin katsayısı -14,6929 dur. Tüm değişkenler sabit tutularak bir ilköğretim mezunu olan kişinin aylık balık harcama miktarı üniversite mezunu olan bir başkasından 4,1307 TL (10,5622-14,6929=-4,1307) daha azdır. Eğitim değişkeninin diğer düzeyi lise mezunu olmak düzeyinin katsayısı ise -20,7474 TL dir. Tüm değişkenler sabit

tutularak bir lise mezunu olan kişinin aylık balık harcama miktarları üniversite mezunu olan bir başkasından 10,1852 TL (10,5622-20,7474=-10,1852) daha azdır. İlköğretim mezunu olmak temel düzey değişkeni olarak alınsaydı parametrelerin katsayıları, bu katsayılarla ilişkin p değerlerinin değişeceği kaçınılmazdı. Bu çalışmada araştırmacıya bir alternatif sunmak amacıyla temel düzey ilköğretim mezunu alınarak yapılan parametre tahmin çizelgesi de Çizelge 5'de verilmiştir. Çizelge 5'de tahmin edilen parametreler ile model kurulup katsayılar yerine koyulduğunda aynı sonuçların çıkması beklenir.

Çizelge 5' de çıkan sonuçlara göre katsayıların anlamlılığını ölçmek için kullanılan t testi istatistiğinin p değeri fert sayısı ve eğitim değişkeni için $\alpha=0.05$ anlamlılık düzeyinde istatistiksel olarak önemsiz, gelir değişkeni ise $\alpha=0.05$ anlamlılık düzeyinde istatistiksel olarak önemli çıkmıştır.

Bu katsayıları kullanılarak oluşturulan model,

$$E(Y|X) = -4.1306 + 0.8554x_{FS} + 0.0213x_{GEL} - 6.0545x_{(D_2)} - 146928x_{(D_3)}$$

şeklinindedir. Eğitim değişkeninin diğer düzeyi lise mezunu olmak düzeyinin katsayısı -6.0545 TL dir. Tüm değişkenler sabit tutularak bir lise mezunu olan

kişinin aylık balık harcama miktarları üniversite mezunu olan bir başkasından 10,1851 TL (-4.1306-6.0545=-10,1851) daha azdır. Bu yorum Çizelge 4 kullanılarak bulunan sonuçla hemen hemen aynıdır.

Modelin tahmininden sonra uyum iyiliği ölçülerine bakılmıştır. Bağımlı değişkenin nitel olduğu analizlerde belirlilik katsayısının değeri oldukça düşük çıkmaktadır. Bu nedenle, bu modellerde R^2 yerine pseudo R^2 kullanılmaktadır. Bu pseudo R^2 değerleri alışılmış R^2 değerlerinden farklıdır ve 0 ile 1 arasında yer almayabilir. Bu durumun nedeni modelin olabilirlik süreci ile tahmin edilmesinden kaynaklanmaktadır. Dolayısıyla belirlilik katsayısının negatif ya da birden büyük değer alması yadırganılacak bir durum değildir (Maddala, 1987; Kennedy, 2006). Çalışmada modelin uygunluk ölçüsünün yorumlanması adına Akaike bilgi kriteri, Schwarz kriterinin yanı sıra Tobit model için önerilen çeşitli pseudo R^2 değerleri hesaplanmış alternatif olarak gözlenen değerlerle tahmin edilen değerler arasındaki korelasyona da bakılmıştır (Çizelge 6). Bunun yanı sıra parametrelerin tek tek ve birlikte kısıtlılık testleri Wald, LR ve LM testleri ile analiz edilmiştir. Sonuçlar sırasıyla Çizelge 7, Çizelge 8, Çizelge 9'da verilmiştir.

Çizelge 6. Gözlenen ve tahmin edilen harcama arasındaki korelasyon

Değişken adı	Tahmini balık harcaması	P değeri
Balık harcaması	0.248***	0.0004

Çizelge 7. Kısıtlamaların geçerliliği testi (1)

Sınırlama	Model	Test adı	Test istatistiği	P
Fert sayısı=0	$\hat{y} = \hat{b}_0 + \hat{gel}x_2 + \hat{egt}x_3$	WALD	72.88***	<.0001
		LR	0.00	0.9821
		LM	4.99***	<.0001

Çizelge 8. Kısıtlamaların geçerliliği testi (2)

Sınırlama	Model	Test adı	Test istatistiği	P
Gelir=0	$\hat{y} = \hat{b}_0 + \hat{fs}x_1 + \hat{egt}x_3$	WALD	.	.
		LR	0.00	0.9609
		LM	834E-14***	<.0001

Çizelge 9. Kısıtlamaların geçerliliği testi (3)

Sınırlama	Model	Test adı	Test istatistiği	P
Egt=0	$\hat{y} = \hat{b}_0 + \hat{fs}x_1 + \hat{gel}x_2$	WALD	72.88***	<.0001
		LR	0.00	0.9791
		LM	0.00	1.0000

Çizelge 10. Kısıtlamaların geçerliliği testi (4)

Sınırlama	Model	Test adı	Test istatistiği	P değeri
fs=0,gel=0	$\hat{y} = \hat{b}_0 + \hat{egt}x_3$	WALD	3470.3***	<.0001
		LR	0.01	0.9970
		LM	838E-14***	<.0001
fs=0,egt=0	$\hat{y} = \hat{b}_0 + \hat{gel}x_2$	WALD	.	.
		LR	0.00	0.9988
		LM	5.06E-8***	<.0001
gel=0,egt=0	$\hat{y} = \hat{b}_0 + \hat{egt}x_3$	WALD	.	.
		LR	0.01	0.9969

		LM	84E-15***	<.0001
--	--	----	-----------	--------

Gözlenen değerlerle tahmin edilen değerler arasındaki korelasyon 0.248 bulunmuştur ($p < 0.001$). Bu değer karesi $0.248^2 = 0.0615$ 'tir. Bu değere bakıldığında modelde tüm bağımsız değişkenler beraberce, bağımlı değişkendeki değişkenliğin %6'sını açıklamaktadır. Bağımsız değişkenlerin bağımlı değişkeni açıklama gücü olarak ifade edilen belirleme katsayısı için %6'lık bir değer düşük bir düzeydir.

Wald ve LM test istatistiklerinin olasılık değerleri çok küçük çıkmıştır. Bu durumda H_0 hipotezi reddedilir. Bu testlere göre fert sayısı değişkeni modelde bulunmalıdır. Sınırlama geçersizdir. LR test istatistiğinin olasılık değeri istatistiksel olarak anlamlı bulunmamıştır. Modele getirilebilecek farklı sınırlama çeşitleri ve hipotezleri aşağıdaki gibi açıklanmaya çalışılmıştır.

Gelir değişkenine getirilen sınırlama, LR, Wald ve LM testleri ile sınanmıştır (Çizelge 8).

Çizelge 8' den LM test istatistiğinin olasılık değeri istatistiksel olarak anlamlı bulunmuştur. Yani H_0 hipotezi LM testine göre reddedilir. Gelir değişkenine uygulanan sınırlama LM testine göre geçersizdir.

Eğitim değişkenine getirilen sınırlama LR, Wald ve LM testleri ile sınanmıştır (Çizelge 9)

Wald test istatistiğinin olasılık değeri istatistiksel olarak önemli bulunmuştur ($P < 0.001$). Yani H_0 hipotezi Wald testine göre reddedilir. Eğitim değişkenine uygulanan sınırlama Wald testine göre geçersizdir.

Fert sayısı – gelir, fert sayısı-egitim ve gelir-egitim değişken birleşimlerine getirilen sınırlamalar Çizelge 10'da toplu olarak verilmiştir.

Bu sonuçlara göre; fert sayısı ve gelir değişkenine getirilen sınırlama Wald ve LM test sonuçlarına göre geçersizdir. Yani bu testlere göre değişkenler modele katılmalıdır. Fert sayısı ve eğitim değişkenine getirilen sınırlamalar LM test sonuçlarına göre geçersizdir. Bu sonuçtan hareketle fert sayısı ve eğitim değişkeni beraberce bağımlı değişkeni açıklamada istatistiksel olarak anlamlıdır, modele alınmalıdır denilebilir. Gelir ve eğitim değişkenine getirilen sınırlamalar da LM test sonuçlarına göre geçersizdir. Burada da aynı şekilde bu değişkenlerin modele katılması gerektiği sonucuna varılabilir.

SONUÇ

Sonuç olarak, yaşam analizi, ömür modellemeleri, makinelerin bozulma süreleri, tüketim harcamaları gibi bağımlı değişkenin negatif olmadığı deneylerin parametre tahminlerinde Tobit Regresyon Analizini kullanmak araştırmacıyı etkin sonuçlara götürür. Hazırlanan bu çalışma ile analizlerinde sürekli ve kesikli verilerin karma yapısı ile karşılaşan araştırmacılara istatistiksel çıkarsama sürecine yardımcı bilgiler sunulmuştur.

Ayrıca, incelenen Türkçe kaynaklarda modelin analiz basamakları hakkında çok fazla bilgiye rastlanılmaması, teorik ve pratik bilgiye beraberce ulaşılamaması ve yapılan araştırmalarda konunun ayrıntısından ziyade sadece sonuçlarına yer verilmesi söz konusudur. Geleceğe yönelik olarak araştırmacıların, eserlerinde bu konulara yer vermeleri Türkiye'deki istatistik biliminin gelişmesine katkıda bulunacaktır.

KAYNAKLAR

- Akbay C 2005. Kahramanmaraş'ta Hane halklarının Gıda Tüketim Talebi Ekonometrik Analizi. KSÜ Fen ve Mühendislik Dergisi, 8(1): 114-121.
- Carson RT, Sun Y 2007. The Tobit model with a non-zero threshold. *Econometrics Journal*, 10: 488-502.
- Chay YK, Powell LJ 2001. Semiparametric Censored Regression Model. *Journal of Economic Perspectives*, American Economic Association, 15(4): 29-42.
- Orhan H, Efe E, Şahin M 2004. SAS Yazılımı İle İstatistiksel Analizler. Tuğra Ofset, ISBN:975-270-435-2, Isparta, 122s.
- Emeç H, Üçdoğruk Ş, Akın F, 2001. Türkiye Hanehalkı Eğlence Kültür Harcamalarında Tobit Modelin Kullanımı. G.Ü. İ.İ.B.F Dergisi, (3):13-26.
- Eren M 2012. Sınırlı Bağımlı Değişkenli Modeller Ve Ülkelerin Gelişmişlik Düzeyleri Üzerine Uygulama. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri, Erzurum, Yüksek Lisans Tezi, 68s
- Gujarati DN 1999. Temel Ekonometri, Çevirenler: Ümit Şenesen-G.Günlük Şenesen, Litaratür Yayınları No:33, İstanbul, 848s
- Greene HW 2003. *Econometric Analysis*. Prentise Hall, ISBN: 01 306 61899, New Jersey, 1026s.
- Joost S, Kalbermatten M 2010. MatSAM versiyon 2 Beta http://www.econogene.eu/software/sam/download/matsamv2/MatSAMv2_documentation.pdf (erişim tarihi: 02.03.2013)
- Kennedy P 2006. *Ekonometri Kılavuzu, Çevirenler: Muzaffer Sarımeşeli-Şenay Açıkgöz*, Gazi Kitapevi, Ankara, 754s.
- Koç Ş 2013. Tobit Regresyon Analizi Ve Bir Uygulama. KSÜ Fen Bilimleri Enstitüsü, Zootekni, K.Maraş, Yüksek Lisans Tezi, 57s.
- Liao TF 1994. *Interpreting Probability Models: Logit, Probit, And Other Generalized Linear Models*. Sage Publications, California, 87s.
- Maddala GS 1987. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, 1004s.
- Olsen RJ 1978. Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model. *Econometrica*, 46(5): 1211-1215
- Park SY 2003. Unbiasedness or Statistical Efficiency: Comparison between One-stage Tobit of MLE and

- Two-step Tobit of OLS. *International Journal of Human Ecology*, 4(2): 77-86.
- Pettinicchi Y 2012. *Emprical Economics*. http://www.virgo.unive.it/ypetti/Lecture3_Pettinicchi.pdf (erişim tarihi: 7.11.2012)
- SAS, 2008. SISITEINFO NAME='VAN 100.YIL UNIVERSITESI' TE=93053001
- Sigelman L, Zeng L 1999. Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models. *Political Analysis*, 8 (2): 167-182.
- Tatoğlu FY 2005. *Sermaye Piyasası'nda Riskin Sınırlı Bağımlı Değişkenli Panel Veri Modelleri İle Analizi*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri, İstanbul, Doktora Tezi,199s.
- Tobin J 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 46(1): 24-36.
- Wooldridge J 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, 735s