

HİYERARŞİK LOGARİTMİK DOĞRUSAL MODELLER ARASINDAN EN UYGUN MODELİN SEÇİMİ

Ayşe OĞUZLAR

Uludağ Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, Yardımcı Doçent Dr.

THE SELECTION OF THE BEST MODEL OUT OF HIERARCHICAL LOG LINEAR MODELS

Abstract: Contingency tables or cross tables are the oldest and most widely used tools by social scientist. The reason of widely usage of contingency tables are to be simplicity and necessity of non parametric assumptions. Contingency tables are named according to number of variables such as 2 way, 3 way or multi way. Chi-square test which is used for analysis of 2 way contingency tables are not sufficient for analysis of 3 way or multi way tables. Most widely used method for 3 way or multi way tables is the hierarchical log linear models. In this study DİE 2002 III. term household manpower survey was used. The 3, 3 way contingency tables which were created from DİE survey. For this 3, 3 way contingency tables, all hierarchical log linear models were created. Then for these 3 tables, the best model selection phase was explained and interpreted.

Keywords: Contingency Tables, Hierarchical Log Linear Models.

HİYERARŞİK LOGARİTMİK DOĞRUSAL MODELLER ARASINDAN EN UYGUN MODELİN SEÇİMİ

Özet: Kontenjans tabloları veya çapraz tablolar sosyal bilimcilerin kullandıkları en eski ve en çok kullanılan araçlardır. Kontenjans tablolarının yaygın kullanılma nedenleri, basit oluşları ve zayıf parametrik varsayımlar gerektirmeleridir. Kontenjans tabloları içerdikleri değişken sayısına bağlı olarak 2 yönlü, 3 yönlü veya çok yönlü olarak isimlendirilirler. İki yönlü kontenjans tablolarının analizinde kullanılan ki-kare testleri, 3 yönlü veya çok yönlü tabloların analizinde yetersiz kalmaktadırlar. Üç yönlü veya çok yönlü tablolar için en çok kullanılan yöntem hiyerarşik logaritmik doğrusal modellerdir. Bu çalışmada, DİE 2002 3. döneminde yapılan hanehalkı işgücü anketi kullanılmıştır. DİE' nin yaptığı bu alan araştırmasından 3 tane 3 yönlü kontenjans tablosu oluşturulmuştur. Bu 3 tane 3 yönlü kontenjans tablosu için, tüm hiyerarşik logaritmik doğrusal modeller oluşturulmuştur. Ardından bu 3 tablo için en iyi modelin seçimi aşaması açıklanmış ve yorumlanmıştır.

Anahtar Kelimeler: Kontenjans Tabloları, Hiyerarşik Logaritmik Doğrusal Modeller.

I. GİRİŞ

Kontenjans tabloları veya çapraz tablolar, sosyal bilimcilerin sıkça kullandıkları eski araçlardan biridir. Kontenjans tablolarının sık kullanılma gerekçesi basitliği ve parametrik olmayan veya zayıf parametrik varsayımları gerektirmeleridir [1]. İki değişken için oluşturulan iki yönlü tabloların analizinde kullanılan ki-kare testi, üç veya daha fazla değişkeni içeren sırasıyla üç yönlü veya çok yönlü çapraz tabloların analizinde yetersiz kalmaktadır. Üç değişkeni veya üçten fazla değişkeni aynı anda içeren üç yönlü veya çok yönlü tabloların analizinde kullanılabilecek yaklaşım, logaritmik doğrusal modellerin kullanımınıdır. Çalışmamızda öncelikle logaritmik doğrusal modeller hakkında genel bilgi verilmeye çalışılacaktır. Ardından üç yönlü tabloların analizi için logaritmik doğrusal modeller açıklanarak çok yönlü tablolar için kısaca bilgi verilecek ve en iyi modelin seçimi konusuna açıklık getirilmeye çalışılacaktır. Uygulama bölümünde ise DİE 2002 yılı 3. dönemi için yapılan hanehalkı işgücü anketi verilerinden hareketle oluşturulmuş üç yönlü üç tablo logaritmik doğrusal modellerle analiz edilmeye çalışılacaktır.

II. HİYERARŞİK LOGARİTMİK DOĞRUSAL MODELLER

Yapılan çalışmalarda kullanılan verilerin ölçüm düzeyleri, kullanılacak analizi belirlemektedir. Düşük ölçüm düzeyleri olarak tanımlanan sınıflayıcı (nominal) ve sıralayıcı (ordinal) ölçek düzeyleri ile yapılan çalışmalarda, diğer bir deyişle kategorik değişkenlerle çalışıldığı durumlarda logaritmik doğrusal modeller kullanılabilmektedir. Logaritmik doğrusal modellerle iki veya daha çok sayıda değişkenin birbirine bağımlı olup olmadığını ve aralarındaki etkileşimleri test etmek, değişkenler arasında neden-sonuç ilişkisi kurmaksızın mümkün olabilmektedir. Diğer bir deyişle logaritmik doğrusal modeller değişkenler arasındaki birlikteliği göstermektedirler. Eğer değişkenlerden biri bağımlı değişken olarak atfedilebiliyorsa ve diğer değişkenlerde bu değişkeni açıklamak için bağımsız değişken olarak ele alınabiliyorsa logaritmik doğrusal modeller logit modeline dönüşürler [2]. Logaritmik doğrusal modeller iki veya daha fazla kategorik değişken arasındaki koşullu ilişkilerin analizinde kullanılabilmektedir [3].

Basitlik sağlması açısından öncelikle iki yönlü kontenjans tablolarının analizinde logaritmik doğrusal modellerin kullanımı açıklanmaya çalışılacaktır. Bunun ardından üç yönlü tablolar için genelleştirme yapılacaktır.

Her biri iki farklı düzeye sahip iki değişkenli bir kontenjans tablosu için (2x2) geleneksel logaritmik doğrusal model aşağıdaki şekildedir:

$$\ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad (1)$$

Bu eşitlikteki;

$\ln(F_{ij})$: Kontenjans tablosundaki ij hücresi için beklenen frekansın doğal logaritması

μ : Beklenen frekansların doğal logaritmasının ortalaması

λ_i^A : A değişkeninin ana etkisi

λ_j^B : B değişkeninin ana etkisi

λ_{ij}^{AB} : A ve B değişkenlerinin etkileşim etkisini göstermektedir.

(1) nolu eşitlikteki model doymuş (saturated) model olarak adlandırılır. Çünkü tüm tekli ve ikili etkileşimleri içermektedir. Doymuş bir modelde beklenen frekanslar ile gözlenen frekanslar, serbestlik derecesi kalmayacak biçimde tam bir eşleşme gösterir. Ayrıca bu modelde yer alan parametre sayısı ilgili kontenjans tablosunun hücre sayısı ile aynı olacaktır.

A değişkeninin B değişkeni üzerinde etkisi olmaması durumunda (veya B değişkeninin A değişkeni üzerinde etkisi olmaması durumunda) λ_{ij}^{AB} parametresi 0 değerini alarak modelden çıkarılacaktır:

$$\ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B \quad (2)$$

(2) nolu eşitlikte verilen model, doymamış bir modeldir ve literatürde bağımsızlık modeli olarak isimlendirilir. Bağımsızlık modeli olarak adlandırılmasının nedeni A ve B arasındaki etkileşim parametresini (λ_{ij}^{AB}) içermemesidir. (2) nolu eşitlikle gösterilen bağımsızlık modelinin testi, ki-kare bağımsızlık hipotezinin testine benzemektedir.

Bir adım ileri giderek üç yönlü kontenjans tablolarının modellenmesi ile ilgilenilmiş olsun. 2x2x2 boyutlu dolayısıyla 2 düzeyli 3 değişken içeren bir kontenjans tablosu için doymuş model aşağıdaki şekilde yazılabilir:

$$\ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad (3)$$

Doymuş logaritmik doğrusal model, orijinal frekans tablosuna tam uyan modeldir. Logaritmik doğrusal modellemeye en çok kullanılan yapı hiyerarşik yapıdır [5]. Doymuş modelin dışındaki logaritmik doğrusal modeller, doymamış modeller (unsaturated) olarak isimlendirilirler ve genellikle **hiyerarşik logaritmik doğrusal modeller** olarak bilinirler [4]. Hiyerarşik logaritmik doğrusal modeller veya kısa söyleyişle hiyerarşik modeller, aynı zamanda daha düşük düzeydeki birliktelikleri de modellemektedirler [6]. Basit bir deyişle daha az komplike modeller daha yüksek dereceli modellerde yer almaktadır [7]. Hiyerarşik olmayan modellerde ise bu tarz bir kısıtlama bulunmamaktadır. Örneğin

$$\ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B \quad (4)$$

modeli hiyerarşik olmayan bir modeldir. Çünkü λ_i^B ana etkileşim terimi modelde yer almamaktadır.

A, B ve C gibi üç değişken içeren bir kontenjans tablosu için oluşturulabilecek hiyerarşik logaritmik doğrusal modeller aşağıda gösterildiği şekildedir [8]:

$$1. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C$$

$$2. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$$

$$3. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC}$$

$$4. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$$

$$5. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC}$$

$$6. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$$

$$7. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

$$8. \ln(F_{ijk}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

Yukarıda verilen 8 modelin tümü, hiyerarşik modeldir. Çünkü daha yüksek düzeydeki birliktelikleri içeren modeller, aynı zamanda daha düşük düzeydeki birliktelikleri de modelde içermektedir. Hiyerarşik bir modelde bir λ parametresi sıfır bulunduğu, daha yüksek dereceli etkileşimleri gösteren tüm λ parametreleri sıfıra eşit olur. Benzer biçimde bir λ parametresi sıfırdan

farklı bulunduğu da, daha düşük düzeydeki etkileşimleri içeren tüm parametreler de sıfırdan farklı bulunacaktır [9]. Örneğin modelde λ_{ijk}^{ABC} terimi yer almakta ise, aynı zamanda λ_{ij}^{AB} , λ_{ik}^{AC} , λ_{jk}^{BC} , λ_i^A , λ_j^B , λ_k^C ve genel ortalama olan μ 'de modelde yer almak durumundadır [8].

İstatistiksel paket programları hiyerarşik olmayan modelleri test edememektedir. Çünkü hiyerarşik olmayan modellerde potansiyel modeller arasından seçim için bir istatistiksel süreç sağlanamamaktadır. Bunun sebebi ise bir etkileşim teriminin anlamlı olmadığı durumlarda, daha yüksek dereceli etkileşimlerin anlamlı olup olmadığına yorumunun anlam taşımamasındandır.

Etkileşimlerin test edilmesinde iki farklı yaklaşım izlenebilir:

1. En yüksek düzeyde etkileşim içeren modelden başlanılarak modelin uygunluğu kabul edilebilir düzeye gelinceye değin yüksek dereceden etkileşim terimlerini modelden çıkarmak,
2. En basit model olan bağımsızlık modelinden başlayarak, etkileşim terimlerini sırasıyla kabul edilebilir düzeye ulaşılan değin modele almak.

III. EN İYİ MODELİN SEÇİLMESİ

Bir modelin uyum iyiliği, her bir model için gözlenen ve beklenen frekansların karşılaştırılmasıyla sağlanır. Pearson χ^2 istatistiği veya olabilirlik oran (G^2) ki-kare istatistiği modelin uygunluğunun testi için yaygınlıkla kullanılan iki test istatistiğidir. Pearson χ^2 istatistiği aşağıdaki formül yardımıyla elde edilmektedir [10]:

$$Q_p = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (5)$$

Formülde yer alan e_{ij} 'ler beklenen frekansları göstermektedir.

Olabilirlik oran χ^2 istatistiği ise aşağıdaki formül yardımıyla bulunmaktadır:

$$G^2 = 2 \sum_i \sum_j n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right) \quad (6)$$

Bu iki test istatistiği modelin verilere uygun olup olmadığının belirlenmesinde kullanılabilmesine karşılık,

çok boyutlu tabloların analizinde daha çok olabilirlik oran χ^2 istatistiği (G^2) kullanılmaktadır [11].

$r \times c \times l$ boyutlu kontenjans tablosu için etkiler, sıfır hipotezleri, beklenen frekans tahminleri ve serbestlik derecesi formülleri şu şekilde özetlenebilir [12]:

1. Genel Etki

Sıfır hipotezi:

$$p_{111} = \dots = p_{rcl} = \frac{1}{rcl}$$

Beklenen frekans tahmini:

$$n_{111} = \dots = n_{rcl} = \frac{N}{rcl}$$

Serbestlik derecesi:

$$rcl - 1$$

2. Tek Faktör Etkileri

Sıfır hipotezleri:

$$p_{1..} = \dots = p_{r..} = \frac{1}{r}$$

$$p_{.1.} = \dots = p_{.c.} = \frac{1}{c}$$

$$p_{..1} = \dots = p_{..l} = \frac{1}{l}$$

Beklenen frekans tahminleri:

$$n_{1..} = \dots = n_{r..} = \frac{N}{r}$$

$$n_{.1.} = \dots = n_{.c.} = \frac{N}{c}$$

$$n_{..1} = \dots = n_{..l} = \frac{N}{l}$$

Serbestlik dereceleri:

$$r - 1$$

$$c - 1$$

$$l - 1$$

3.İki Faktör Marjinal Etkileşim Etkileri

Sıfır hipotezleri:

$$p_{ij.} = p_{i..} p_{.j.}$$

$$p_{i.k} = p_{i..} p_{..k}$$

$$p_{.jk} = p_{.j.} p_{..k}$$

Beklenen frekans tahminleri:

$$\frac{n_{i..} n_{.j.}}{N}$$

$$\frac{n_{i..} n_{..k}}{N}$$

$$\frac{n_{.j.} n_{..k}}{N}$$

Serbestlik dereceleri:

$$(r-1)(c-1)$$

$$(r-1)(l-1)$$

$$(c-1)(l-1)$$

$r \times c \times l$ boyutlu kontenjans tablosu için etkilere karşılık gelen G^2 etki istatistikleri ise aşağıda belirtildiği şekildedir [12]:

1.Genel Etki İçin

$$G_g^2 = 2 \sum^{rcl} n_{ijk} \log \left(n_{ijk} \left[\frac{rcl}{N} \right] \right)$$

2.Tek Faktör Etkileri İçin

$$G_x^2 = 2 \sum^r n_{i..} \log(Nn_{i..} / r)$$

$$G_y^2 = 2 \sum^c n_{.j.} \log(Nn_{.j.} / c)$$

$$G_z^2 = 2 \sum^l n_{..k} \log(Nn_{..k} / l)$$

3.İki Faktör Etkileşim Etkileri İçin

$$G_{xy}^2 = 2 \sum^r \sum^c n_{ij.} \log \left[\frac{Nn_{ij.}}{n_{i..} n_{.j.}} \right]$$

$$G_{xz}^2 = 2 \sum^r \sum^l n_{i.k} \log \left[\frac{Nn_{i.k}}{n_{i..} n_{..k}} \right]$$

$$G_{yz}^2 = 2 \sum^c \sum^l n_{.jk} \log \left[\frac{Nn_{.jk}}{n_{.j.} n_{..k}} \right]$$

4.Üç Faktör Etkileşim Etkileri İçin

$$G_{xyz}^2 = 2 \sum^r \sum^c \sum^l n_{ijk} \log \left[\frac{n_{ijk}}{\hat{m}_{ijk}} \right]$$

Bu G^2 etki istatistiklerinden hareketle bulunan G^2 uyum iyiliği istatistikleri ise aşağıda gösterilmiştir:

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2) = G_{(x,y,z)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xy}^2) = G_{(xy,z)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{yz}^2) = G_{(xz,y)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xz}^2) = G_{(yz,x)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xy}^2 + G_{xz}^2) = G_{(xy,xz)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xy}^2 + G_{yz}^2) = G_{(xy,yz)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xz}^2 + G_{yz}^2) = G_{(xz,yz)}^2$$

$$G_g^2 - (G_x^2 + G_y^2 + G_z^2 + G_{xy}^2 + G_{xz}^2 + G_{yz}^2) = G_{(xy,xz,yz)}^2$$

Elde edilen bu nihai olabilirlik oran ki-kare istatistikleri, p anlamlılık düzeyi veya tablo ki-kare değeri ile kıyaslanmak suretiyle ve diğer mümkün modellerden elde edilen G^2 istatistikleri incelenerek modele ilişkin bir karara ulaşmak mümkündür.

Çok yönlü tablolarda tablonun boyutu arttıkça yüksek dereceden etkileşimlerin modele alınıp alınmamasına karar vermek için hiyerarşik yapıdaki logaritmik doğrusal modeller incelenir [13]. En iyi

modelin seçilmesi amacıyla (verilere en uygun model) tüm etkiler ile etkileşimler için G^2 istatistikleri hesaplanır. Anlamli bulunan etkiler ve etkileşimler verilere en uygun bulunacak modelde yer alması muhtemel olan etkiler ve etkileşimler olacaktır. Etkiler ve etkileşimlerin anlamlılığının test edilmesinin ardından anlamlı bulunan etkiler ve etkileşimleri içeren hiyerarşik logaritmik doğrusal modellerin uygunluğu test edilecektir. Örneğin İki model anlamlı bulunduysa daha az terim içeren model tercih edilecektir [12]. Diğer bir yaklaşıma göre ise hiyerarşik bir yapıda olan iki modelden daha az terim içeren model seçilerek, diğer anlamlı model ile aralarındaki G^2 farklılığı hesaplanır. Eğer farklılık anlamlı ise daha az terim içeren model seçilir. Eğer aralarındaki farklılık anlamlı değil ise ikinci model dolayısıyla daha fazla terimi içeren model anlamlı olacaktır [14].

Logaritmik doğrusal modellerde test edilen değişken ve oluşturulan hipotez sayısı çok fazladır. Bu nedenle örneklem hacminin küçük seçilmesi sonuç elde etmeyi güçleştireceğinden büyük hacimli örneklerle çalışılması gerekmektedir. Bu doğrultuda örneklem hacminin tablodaki hücre sayısının 5 katı olması önerilmektedir [15].

Uygulama bölümünde 2002 III. Dönemi için açıklanan hanehalkı işgücü anket sonuçlarından yararlanılarak her biri üç değişken içeren üç farklı kontenjans tablosu için en iyi model araştırılmıştır.

IV. UYGULAMA

Uygulama bölümünde 2002 III. Dönemi için uygulanan hanehalkı işgücü anket sonuçlarından yararlanılmıştır. 24487 hanehalkı ve 72620 kişi için uygulanan anket değişkenlerinden yaş grubu, hanehalkı reisine yakınlık, medeni durum, işteki durum, en son bitirilen okul, meslek ve istihdamda olup olunmadığını gösteren iş değişkenleri ele alınmıştır.

Bu değişkenlerin analizde ele alınan kategorileri aşağıdaki biçimdedir:

Yaş Grubu

0. 0-14
1. 15-19
2. 20-24
3. 25-29
4. 30-34
5. 35-39
6. 40-44
7. 45-49
8. 50-54
9. 55-59
10. 60-64
11. 65+

Hanehalkı reisine yakınlık

1. Hanehalkı reisi
2. Eşi
3. Çocuğu
4. Gelini veya damadı
5. Torunu
6. Ebeveyni
7. Diğer akrabası
8. Akraba olmayan

İş: Referans haftası içinde çalışıp/çalışmadığı veya işi ile bağlantısının devam edip etmediği

1. Evet (İstihdamda)
2. Hayır (İstihdamda değil)

Medeni Durum

1. Hiç evlenmedi
2. Evli
3. Boşandı
4. Eşi öldü

İşteki Durum

1. Düzenli ücretli (Maaşlı)
2. Yevmiyeli (Mevsimlik, arızı, geçici)
3. İşveren
4. Kendi hesabına

En son bitirilen okul

1. İlkokul
3. İlköğretim
4. Ortaokul
5. Lise
6. Mesleki lise
7. 2 yıllık önlisans
8. 4 yıllık lisans

Meslek

1. Profesyonel meslek mensupları
2. Yardımcı profesyonel meslek grupları
3. Büro ve müşteri hizmetlerinde çalışan elemanlar
4. Hizmet ve satış elemanları
5. Nitelikli tarım, hayvancılık, avcılık, ormancılık ve su ürünlerinde çalışanlar
6. Sanatkarlar ve ilgili işlerde çalışanlar

Bu değişkenlerin tümünün ele alınarak bir kontenjans tablosunun oluşturulması boş hücrelerin çokluğundan dolayı mümkün olamamıştır. Dolayısıyla asıl inceleme konumuz olan iş değişkeni ile diğer değişkenler üç yönlü bir kontenjans tablosu biçiminde ele alınarak üç farklı model grubu arasında en iyi modelin seçilmesine karar verilmiştir. Öncelikle bu değişkenlerden yaş grubu (1), hanehalkı reisine yakınlık (2) ve iş (3) değişkenleri ele alınarak oluşturulan kontenjans tablosu

için Statistica paket programında etkiler ve etkileşimlerin anlamlılığı test edilmiştir. Test edilen etkiler ve etkileşimler için elde edilen olabilirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri aşağıda özetlenmiştir:

Ana Etki: (1)

Olabilirlik Oran χ^2 istatistiği: 132727,9

S.d:150 p=0,000

Pearson χ^2 istatistiği: 201905,6

S.d:150 p=0,000

Ana Etki: (2)

Olabilirlik Oran χ^2 istatistiği: 65863,14

S.d:152 p=0,000

Pearson χ^2 istatistiği: 82905,82

S.d:152 p=0,000

Ana Etki: (3)

Olabilirlik Oran χ^2 istatistiği: 143289,1

S.d:158 p=0,000

Pearson χ^2 istatistiği: 226643,9

S.d:158 p=0,000

İkili Etkileşim: (12)

Olabilirlik Oran χ^2 istatistiği: 19388,25

S.d:80 p=0,000

Pearson χ^2 istatistiği: 17532,42

S.d:80 p=0,000

İkili etkileşim: (13)

Olabilirlik Oran χ^2 istatistiği: 122948,7

S.d:140 p=0,000

Pearson χ^2 istatistiği: 164642,6

S.d:140 p=0,000

İkili Etkileşim: (23)

Olabilirlik Oran χ^2 istatistiği: 55568,44

S.d:144 p=0,000

Pearson χ^2 istatistiği: 62135,51

S.d:144 p=0,000

Ana etkiler ve ikili etkileşimlerin tümü için gözlenen anlamlılık düzeyleri $p < 0,05$ bulunduğundan tüm ana etki ve etkileşimler anlamlı bulunmuştur. Dolayısıyla ana etkiler ve ikili etkileşimler için tüm mümkün modeller oluşturularak bu modellerin verilere uygun olup olmadığı test edilebilir. Aşağıda bu ikinci değişken grubuna ilişkin ana etkiler ve etkileşimleri içeren tüm mümkün modeller için olabilirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri özetlenmiştir:

1.Model: (1,2)

Olabilirlik Oran χ^2 istatistiği: 52515,92

S.d:143 p=0,000

Pearson χ^2 istatistiği: 58723,21

S.d:143 p=0,000

2.Model: (1,3)

Olabilirlik Oran χ^2 istatistiği: 129941,9

S.d:149 p=0,000

Pearson χ^2 istatistiği: 193085,1

S.d:149 p=0,000

3.Model: (2,3)

Olabilirlik Oran χ^2 istatistiği: 63077,07

S.d:151 p=0,000

Pearson χ^2 istatistiği: 72465,47

S.d:151 p=0,000

4.Model: (1,2,3) (bağımsızlık modeli)

Olabilirlik Oran χ^2 istatistiği: 49729,82

S.d:142 p=0,000

Pearson χ^2 istatistiği: 52934,64

S.d:142 p=0,000

5.Model: (12,3)

Olabilirlik Oran χ^2 istatistiği: 16602,18

S.d:79 p=0,000

Pearson χ^2 istatistiği: 15524,69

S.d:79 p=0,000

6.Model: (13,2)

Olabilirlik Oran χ^2 istatistiği: 42736,58

S.d:133 p=0,000

Pearson χ^2 istatistiği: 41647,39

S.d:133 p=0,000

7.Model: (23,1)

Olabilirlik Oran χ^2 istatistiği: 42221,26

S.d:135 p=0,000

Pearson χ^2 istatistiği: 40816,72

S.d:135 p=0,000

8.Model: (12,13)

Olabilirlik Oran χ^2 istatistiği: 9608,907

S.d:70 p=0,000

Pearson χ^2 istatistiği: 8971,502

S.d:70 p=0,000

9.Model: (12,23)

Olabilirlik Oran χ^2 istatistiği: 9093,581

S.d:72 p=0,000

Pearson χ^2 istatistiği: 8619,141

S.d:72 p=0,000

10.Model: (13,23)

Olabilirlik Oran χ^2 istatistiği: 35227,98

S.d:126 p=0,000

Pearson χ^2 istatistiği: 31977,70

S.d:126 p=0,000

11.Model: (12,13;23)

Olabilirlik Oran χ^2 istatistiği: 1428,757

S.d:63 p=0,000

Pearson χ^2 istatistiği: 1674,454

S.d:63 p=0,000

Bu modellere bakıldığında modellerin tümünün $p<0,05$ bulunduğundan verilere uygun olmadığı görülebilir. Dolayısıyla yaş grubu, hanehalkı reisine yakınlık ve iş değişkenleri için oluşturulacak tüm

modeller $\alpha=0,05$ anlamlılık düzeyinde anlamsız bulunmuştur.

İkinci bir model grubu olarak **medeni durum (1)**, **iş (2)** ve **işteki durum (3)** değişkenleri ele alınarak oluşturulan kontenjans tablosu için etkiler ve etkileşimlerin anlamlılığı test edilmiştir. Test edilen etkiler ve etkileşimler için elde edilen olabilirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri aşağıda özetlenmiştir:

Ana Etki: (1)

Olabilirlik Oran χ^2 istatistiği: 35338,08

S.d:28 p=0,000

Pearson χ^2 istatistiği: 40616,06

S.d:28 p=0,000

Ana Etki: (2)

Olabilirlik Oran χ^2 istatistiği: 37689,37

S.d:30 p=0,000

Pearson χ^2 istatistiği: 50986,82

S.d:30 p=0,000

Ana Etki: (3)

Olabilirlik Oran χ^2 istatistiği: 53089,71

S.d:28 p=0,000

Pearson χ^2 istatistiği: 76845,28

S.d:28 p=0,000

İkili Etkileşim: (12)

Olabilirlik Oran χ^2 istatistiği: 11596,14

S.d:32 p=0,000

Pearson χ^2 istatistiği: 11862,98

S.d:32 p=0,000

İkili Etkileşim: (13)

Olabilirlik Oran χ^2 istatistiği: 30507,36

S.d:20 p=0,000

Pearson χ^2 istatistiği: 22068,43

S.d:20 p=0,000

İkili Etkileşim: (2,3)

Olabilirlik Oran χ^2 istatistiği: 33962,24

S.d:30 p=0,000

Pearson χ^2 istatistiği: 33846,22

S.d:30 p=0,000

Ana etkiler ve ikili etkileşimlerin tümü için gözlenen anlamlılık düzeyleri $p < 0,05$ bulunduğundan tüm ana etki ve etkileşimler anlamlı bulunmuştur. Dolayısıyla bu ikinci değişken grubuna ilişkin ana etkiler ve ikili etkileşimler için tüm mümkün modeller oluşturularak bu modellerin verilere uygun olup olmadığı test edilebilir. Aşağıda ana etkiler ve etkileşimleri içeren tüm mümkün modeller için olabilirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri özetlenmiştir:

1.Model: (1,2)

Olabilirlik Oran χ^2 istatistiği: 11620,85

S.d:35 p=0,000

Pearson χ^2 istatistiği: 11954

S.d:35 p=0,000

2.Model: (1,3)

Olabilirlik Oran χ^2 istatistiği: 32949,16

S.d:32 p=0,000

Pearson χ^2 istatistiği: 26345,34

S.d:32 p=0,000

3.Model: (2,3)

Olabilirlik Oran χ^2 istatistiği: 33966,72

S.d:34 p=0,000

Pearson χ^2 istatistiği: 33850,51

S.d:34 p=0,000

4.Model: (1,2,3) (bağımsızlık modeli)

Olabilirlik Oran χ^2 istatistiği: 2474,980

S.d:31 p=0,000

Pearson χ^2 istatistiği: 2290,428

S.d:31 p=0,000

5.Model: (12,3)

Olabilirlik Oran χ^2 istatistiği: 2450,277

S.d:28 p=0,000

Pearson χ^2 istatistiği: 2147,161

S.d:28 p=0,000

6.Model: (13,2)

Olabilirlik Oran χ^2 istatistiği: 33,1801

S.d:19 p=0,022966

Pearson χ^2 istatistiği: 227,2043

S.d:19 p=0,000

7.Model: (23,1)

Olabilirlik Oran χ^2 istatistiği: 2470,518

S.d:27 p=0,000

Pearson χ^2 istatistiği: 2233,300

S.d:27 p=0,000

8.Model: (12,13)

Olabilirlik Oran χ^2 istatistiği: 8,46669

S.d:16 p=0,933718

Pearson χ^2 istatistiği: 14,33257

S.d:16 p=0,573958

9.Model: (12,23)

Olabilirlik Oran χ^2 istatistiği: 2445,804

S.d:24 p=0,000

Pearson χ^2 istatistiği: 2141,798

S.d:24 p=0,000

10.Model: (13,23)

Olabilirlik Oran χ^2 istatistiği: 28,7054

S.d:15 p=0,017585

Pearson χ^2 istatistiği: 168,4319

S.d:15 p=0,000

11.Model: (12,13,23)

Olabilirlik Oran χ^2 istatistiği: 3,811476

S.d:12 p=0,986597

Pearson χ^2 istatistiği: 5,671220

S.d:12 p=0,931723

Medeni durum (1), iş (2) ve işteki durum (3) değişkenleri için test edilen model değerlerine bakıldığında $\alpha=0,05$ anlamlılık düzeyi için verilere uygun modeller 14. (12,13) ve 17. (12,13,23) modellerdir. 14. modelin s.d.=16 olabirlik oran χ^2 istatistiği=8,46669, 17. modelin s.d.=12 olabirlik oran χ^2 istatistiği=3,811476 olduğundan fark alındığında s.d.=4 ve olabirlik oran χ^2 istatistiği=4,655214 bulunmaktadır. $\alpha=0,05$ anlamlılık düzeyinde s.d.=4 için ki-kare tablo değeri 9,49 olduğundan iki model arasındaki fark anlamlı bulunamamıştır. Bu sebeple daha az terim içeren 14. model (12,13) modeli verilerimize en uygun model olarak kabul edilmiştir.

Son bir model grubu olarak **en son bitirilen okul (1), iş (2) ve meslek (3)** değişkenleri ele alınarak oluşturulan kontenjans tablosu için etkiler ve etkileşimlerin anlamlılığı test edilmiştir. Test edilen etkiler ve etkileşimler için elde edilen olabirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri aşağıda özetlenmiştir:

Ana Etki: (1)

Olabilirlik Oran χ^2 istatistiği: 35379,83

S.d:91 p=0,000

Pearson χ^2 istatistiği: 44146,42

S.d:91 p=0,000

Ana Etki: (2)

Olabilirlik Oran χ^2 istatistiği: 30783,83

S.d:96 p=0,000

Pearson χ^2 istatistiği: 59649,43

S.d:96 p=0,000

Ana Etki: (3)

Olabilirlik Oran χ^2 istatistiği: 48560,80

S.d:91 p=0,000

Pearson χ^2 istatistiği: 92909,93

S.d:91 p=0,000

İkili Etkileşim: (12)

Olabilirlik Oran χ^2 istatistiği: 14065,59

S.d:84 p=0,000

Pearson χ^2 istatistiği: 14335,08

S.d:84 p=0,000

İkili Etkileşim: (13)

Olabilirlik Oran χ^2 istatistiği: 21343,40

S.d:49 p=0,000

Pearson χ^2 istatistiği: 15522,72

S.d:49 p=0,000

İkili Etkileşim: (23)

Olabilirlik Oran χ^2 istatistiği: 27268,01

S.d:84 p=0,000

Pearson χ^2 istatistiği: 38744,67

S.d:84 p=0,000

Ana etkiler ve ikili etkileşimlerin tümü için gözlenen anlamlılık düzeyleri $p<0,05$ olduğundan tüm ana etki ve etkileşimler anlamlı bulunmuştur. Dolayısıyla üçüncü değişken grubuna ilişkin ana etkiler ve ikili etkileşimler için tüm mümkün modeller oluşturularak bu modellerin verilere uygun olup olmadığı test edilebilir. Aşağıda ana etkiler ve etkileşimleri içeren tüm mümkün modeller için olabirlik oran ki-kare istatistiği, Pearson ki-kare istatistiği, gözlenen anlamlılık düzeyleri ve serbestlik dereceleri özetlenmiştir:

1.Model: (1,2)

Olabilirlik Oran χ^2 istatistiği: 14092,67

S.d:90 p=0,000

Pearson χ^2 istatistiği: 14379,41

S.d:90 p=0,000

2.Model: (1,3)

Olabilirlik Oran χ^2 istatistiği: 31869,61

S.d:85 p=0,000

Pearson χ^2 istatistiği: 42865,59

S.d:85 p=0,000

3.Model: (2,3)

Olabilirlik Oran χ^2 istatistiği: 27273,63

S.d:90 p=0,000

Pearson χ^2 istatistiği: 38761,24

S.d:90 p=0,000

4.Model: (1,2,3) (bağımsızlık modeli)

Olabilirlik Oran χ^2 istatistiği: 10582,47

S.d:84 p=0,000

Pearson χ^2 istatistiği: 13755,03

S.d:84 p=0,000

5.Model: (12,3)

Olabilirlik Oran χ^2 istatistiği: 10555,38

S.d:78 p=0,000

Pearson χ^2 istatistiği: 13713,53

S.d:78 p=0,000

6.Model: (13,2)

Olabilirlik Oran χ^2 istatistiği: 56,2462

S.d:48 p=0,193763

Pearson χ^2 istatistiği: 250,8701

S.d:48 p=0,000

7.Model: (23,1)

Olabilirlik Oran χ^2 istatistiği: 10576,85

S.d:78 p=0,000

Pearson χ^2 istatistiği: 13750,46

S.d:78 p=0,000

8.Model: (12,13)

Olabilirlik Oran χ^2 istatistiği: 29,16570

S.d:42 p=0,933182

Pearson χ^2 istatistiği: 99,15006

S.d:42 p=0,000002

9.Model: (12,23)

Olabilirlik Oran χ^2 istatistiği: 10549,76

S.d:72 p=0,000

Pearson χ^2 istatistiği: 13720

S.d:72 p=0,000

10.Model: (13,23)

Olabilirlik Oran χ^2 istatistiği: 50,6280

S.d:42 p=0,169905

Pearson χ^2 istatistiği: 212,2511

S.d:42 p=0,000

11.Model: (12,13,23)

Olabilirlik Oran χ^2 istatistiği: 28,46584

S.d:36 p=0,810098

Pearson χ^2 istatistiği: 98,78715

S.d:36 p=0,000

En son bitirilen okul (1), iş (2) ve meslek (3) değişkenleri için oluşturulan modellere baktığımızda, $\alpha=0,05$ anlamlılık düzeyi için anlamlı olan modeller 12. (13,2), 14. (12,13), 16. (13,23) ve 17. (12,13,23) modellerdir. En az terimi içeren ilk anlamlı model olan 12. (13,2) model için s.d.=48 ve olabilirlik oran χ^2 istatistiği= 56,2462 ve ikinci anlamlı bulunan model olan 14. (12,13) model için s.d.=42 ve olabilirlik oran χ^2 istatistiği=29,16570 bulunmuştur. Bu iki model arasındaki fark için s.d =6 ve olabilirlik oran χ^2 istatistiği=27,08050 olarak bulunmaktadır. $\alpha=0,05$ anlamlılık düzeyi ve s.d.=6 için χ^2 istatistiği=12,59 olduğundan aradaki fark anlamlıdır. Dolayısıyla 12. (13,2) modelden ziyade 14. (12,13) model verilerimiz için daha uygun görünmektedir. Anlamlı bulunan diğer modeller olan 16. (13,23) model ile 17. (12,13,23) modeller için serbestlik dereceleri ve olabilirlik oran χ^2 istatistikleri sırasıyla 42 ile 36 ve 50,6280 ile 28,46584 olarak bulunmuştu. Bu iki modelin serbestlik dereceleri arasındaki fark 6 ve olabilirlik oran χ^2 istatistikleri arasındaki fark ise 22,16216 olarak bulunmaktadır. $\alpha=0,05$ anlamlılık düzeyi ve s.d.=6 için χ^2 istatistiği=12,59 olduğundan aradaki fark anlamlıdır. Bu sebeple 17. (12,13,23) olan modelin verilerimize uygun olduğu söylenebilir. Verilerimiz için uygun olan 14. (12,13) ile 17. (12,13,23) model kıyaslanırsa iki modelin serbestlik dereceleri arasındaki fark 6 ve olabilirlik oran χ^2 istatistikleri arasındaki fark ise 0,69986 olarak bulunmuştur. $\alpha=0,05$ anlamlılık düzeyi ve s.d.=6 için χ^2 istatistiği=12,59 olduğundan aradaki fark anlamlı değildir. Sonuç olarak üçüncü model grubu için en uygun model 14. (12,13) modeldir.

V. SONUÇ

Çapraz tablolar diğer bir deyişle kontenjans tabloları sosyal bilimcilerin sıkça kullandıkları bir araçtır. Bu tabloların analizinde, tablonun içerdiği değişken sayısı ve değişkenlerin kategorilerinin fazla oluşu klasik χ^2 analizinin kullanımını imkansız hale getirmektedir. Logaritmik doğrusal modeller çok sayıda değişken ve kategorinin söz konusu olduğu kontenjans tabloları için sıkça kullanılmaktadır. Ele alınan uygulama çalışmasında 2002 III. Dönemi için uygulanan hanehalkı işgücü anket sonuçlarından hareketle, her biri üç değişken içeren üç

farklı model grubunun içinden en iyi modelin seçilmesi amaçlanmıştır. İlk model grubunu oluşturan yaş grubu (1), hanehalkı reisine yakınlık (2) ve iş (3) değişkenleri ele alınarak oluşturulan kontenjans tablosu için Statistica paket programında öncelikle tüm ana etki ve ikili etkileşimlerin anlamlı bulunmasının ardından tüm mümkün modellerin testine geçilmiştir. Bu modellere bakıldığında modellerin tümünün anlamsız olduğu verilere uygun olmadığı görülmüştür. Dolayısıyla yaş grubu, hanehalkı reisine yakınlık ve iş değişkenleri için oluşturulacak tüm modeller $\alpha=0,05$ anlamlılık düzeyinde verilere uygun bulunmamıştır.

İkinci model grubu ise medeni durum (1), iş (2) ve işteki durum (3) değişkenleri ele alınarak oluşturulmuştur. Bu değişkenler ele alınarak tüm ana etki ve ikili etkileşimlerin anlamlı bulunmasının ardından, oluşturulan modeller için kıyaslamalar yapıldığında 14. model (12,13) bu değişken grubu için verilerimize en uygun model olarak kabul edilmiştir.

Son model grubu olarak da en son bitirilen okul (1), iş (2) ve meslek (3) değişkenleri ele alınarak tüm ana etki ve ikili etkileşimlerin anlamlı bulunmasının ardından hiyerarşik logaritmik doğrusal modellerin testi gerçekleştirilmiştir. Bu değişken grubu için oluşturulan modellere bakıldığında $\alpha=0,05$ anlamlılık düzeyinde en uygun modelin 14. (12,13) olduğuna karar verilmiştir.

Sonuç olarak ele alınan değişkenler için anlamlı bulunan modeller aşağıdaki modellerdir.

1. (medeni durum \times iş)(medeni durum \times işteki durum)
2. (en son bitirilen okul \times iş)(en son bitirilen okul \times meslek)

Son bir cümle olarak medeni durum ile iş değişkeninin etkileşimi ve en son bitirilen okul ile iş değişkeninin etkileşiminin verilere uygun modellerde yer alması, kişilerin istihdamda olup olmasını etkileyen en önemli faktörlerin kişilerin medeni durumu ile en son bitirilen okul olduğu sonucuna ulaşılabilir.

YARARLANILAN KAYNAKLAR

- [1] POWERS, D.A.; XIE, Y., **Statistical Methods For Categorical Data Analysis**, Academic Press, USA, 2000, s.87.
- [2] Loglinear Models For Multinomial Counts, <http://www.stat.psu.edu/~jls/stat544/2001/lec21.pdf>.
- [3] Log-Linear, Logit and Probit Models, <http://www2.chass.ncsu.edu/garson/pa/65/logit.htm>
- [4] LAKHAN, V.C.; LAVALLE, P.D., "Use Of Loglinear Models To Assess Factors Influencing Concern For The

Natural Environment", **Environmental Management**, Vol.30, No:1, Springer Verlag, New York, 2002, s.80.

- [5] LEMAY, P., Hierarchial, Non-hierarchial and Conditional Models, <http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/node239.html>.
- [6] AGRESTI, A., **An Introduction to Categorical Data Analysis**, John Wiley Sons, USA, 1996, s.168.
- [7] JEANSONNE, A., Loglinear Models, <http://www.sfsu.edu/~efc/classes/biol710/loglinear/Log%20Linear%20Models.htm>, 2002.
- [8] Open And Distance Learning, <http://www.mathpsyc.uni-bonn.de/doc/cristant/node32.html>.
- [9] LE, C.T., **Applied Categorical Data Analysis**, John Wiley&Sons., New York, 1998, ss.15-102.
- [10] Chi-Square Tests and Statistics, <http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect19.htm>.
- [11] Hierarchical Log-linear Models, <http://faculty.ucr.edu/~hanneman/soc203a/loglin.htm>.
- [12] Open And Distance Learning, <http://www.mathpsyc.uni-bonn.de/doc/cristant/node34.html>.
- [13] ÖZDAMAR, K., **Paket Programlar ile İstatistiksel Veri Analizi-1**, Kaan Kitabevi, Eskişehir, 1999, s.450.
- [14] EVERITT B.S., **The Analysis Of Contingency Tables**, London Chapman And Hall, Great Britain, 1977.
- [15] ACAR F., "Çapraz Tabloların Çözümlemesinde Logaritmik Doğrusal Modellerin Kullanımı", IV. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, 14-16 Mayıs 1999, Antalya, s.790.



Ayşe OĞUZLAR

Uludağ University,
Department of Econometrics,
Görükle - BURSA

Tel: +90 (224) 442 89 41 - 41158
avseog@uludag.edu.tr

Ayşe OĞUZLAR has Ph.D. of Statistics-at Uludağ University Institute of Social Sciences. She is Assistant Professor since October 2000 in Uludağ University, Department of Econometrics. Her scientific interests are multivariate statistical analysis, categorical data analysis and data mining.