HACETTEPE ÜNİVERSİTESİ
İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ DERGİSİ
CİLT 7 / SAYI 1-2 / 1989

LOG-LİNEAR MODELLER

SOSYAL ARAŞTIRMALARDA KULLANILAN YENİ ÇÖZÜMLEME TEKNİĞİ

Dr. Hamza UYGUN (*)

1. Giriş

Sosyal bilimlerde nitel değişkenlerle çalışmanın vazgeçilmezliği ve bu alanda kullanılabilecek istatistiksel çözümleme tekniklerinin sınırlılığı toplanan verilerden genellemelere ulaşılmasının kesinlik düzeyinde olumsuz etkiler yaratmaktadır. Bu zayıflık sosyal araştırmaların sonucunda elde edilen bilgi konusunda çok dikkatli olmamızı zorunlu kılar. Toplanan veriler ile bu verilere dayalı genellemeler arasında ilişkiyi daha duyarlı hale getirebilmek için temelde iki yola başvurulur. Bunlar en genelde, 1. Kullanılan değişkenlerin ölçüm düzeylerini daha duyarlı hale getirmek ya da 2. Kullanılan değişkenlerin ölçüm düzeylerine dokunmaksızın daha duyarlı çözümlemeler yapabilmeyi sağlayacak teknikler geliştirmek olarak belirlenebilir. Log-linear modeller ikinci türden çabaların bir ürünüdür. Andersen (1980), Bishop, Finnberg ve Holland (1975), Goodman (1970), (1984), Haberman (1974), (1978), (1979) ile Knoke ve Burke (1980)'nin çalışmaları bu tür ürünlerdendir.

Nominal ve ordinal düzeyde ölçüm yapılan nitel değişkenler arasındaki ilişkiye test etmek için yaygın kullanılan teknik Ki-Kare'dir. Bu teknik yalnızca iki değişken arasında ilişkinin olup olmadığı konusunda bilgi verir. Bu konuda karara varabilmek için araştırmacı beklenen değerler ile gözlenen değerler arasındaki farkı dikkate alır. Beklenen değerler hesaplanırken değişkenlerin birbirlerinden bağımsız olduğu varsayılır. Buna göre hücrelere düşen frekanslar bulunur. Örneğin D1 ve D2 değişkenleri arasında bir ilişkinin olup olmadığını test edelim. Bu iki değişken birbirinden bağımsız ise $N_{ij} = (N_{i.}\, N_{.j})/N..$ olacaktır.

(*) Hacettepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Kamu Yönetimi Bölümünde Uzman

	D1			
	N_{11}	N_{12}	N_{13}	$N_{1.}$
D2	N_{21}	N_{22}	N_{23}	N_{24}
	N_{31}	N_{32}	N_{33}	N_{34}
	$N_{.1}$	N_{2}	N_{3}	$N_{..}$

N_{ij} , i. sıra j. kolondaki hücrenin frekansını,

$N_{i.}$, i. sıranın toplam frekansını,

$N_{.j}$, j. kolonun toplam frekansını,

$N_{..}$, ise genel toplam frekansı göstermektedir.

İlişkiyi test etmek için her hücrenin beklenen frekansı (Nij) ile gözlenen frekansı (n_{ij}) karşılaştırılır. Bu değerler arasında farkın olması iki değişkenin arasında ilişkinin varolduğunu gösterir. Ki-Kare tekniği bu farkı saptamak ve bulunulan farkın anlamlı olup olmadığını saptamak için kullanılır.

Son yıllarda ise yapılan sosyal araştırmaların çözümlenmelerinde Log-Linear modellerin kullanıldığına sıkca rastlanılmaktadır. Bu modellerin çapraz tabloların çözümlenmesinde kullanılması çözümlemeye incelik, duyarlılık getirmiştir (Andersen, 1980; 162). Nominal ve ordinal düzeyde ölçüm yapılmış verilerle çok değişkenli çözümleme yanında regrasyon çözümlemesi olanağı da sağlanmıştır (Knoke ve Burke, 1980; 7).

Log-Linear modellerin kullanımının yaygınlaşmasında, bu tekniğin değişkenler arasındaki ilişkiyi belirlemedeki duyarlılığı yanında bu modellerle yapılan uzun ve sıkıcı hesaplamaların bilgisayar programları yoluyla çözülmüş olmasının da rolü büyüktür.

İstatistiksel çözümleme, toplanan verilerin dağılımının seçilen kuramsal modele uygun kuramsal dağılım ile kıyaslanmasına dayalıdır. Bu nedenle bu yazıda önce kuramsal dağılımın bulunması daha sonra da bu dağılımla elimizdeki verilerin dağılımının karşılaştırıması ve karara ulaşılması ele alınacaktır.

2. Modelin Seçimi ve Kuramsal Dağılımın Bulunması

Model en geniş anlamda ilişkilendirilecek öğelerin saptanması ve bu öğeler arasındaki ilişkilerin nasıl kurulacağına ilişkin genellemelerdir. Örneğin öğrencinin başarısını, gelir düzeyi, sosyo-ekonomik düzeyi ve çevreye bağlı olarak açıklamaya çalışalım. Olayı formülleştirirsek, Başarı = f (gelir, sosyo-ekonomik düzey, çevre) şeklinde gösterebiliriz. Burada gelir, sosyo-ekonomik düzey, çevre ve başarı ilişkilendirilecek öğeleri oluşturmaktadırlar.

Başarının diğer öğelere bağımlı olduğu ise ilişkilendirme biçimini gösterir.

Kuramsal dağılımın bulunması açısından da model yukarıda açıklanan tanıma uygun özellikler gösterir. Yalnız burada kuramsal dağılımdaki frekansları, yani çapraz tablonun hangi öğelerden (değişkenlerden) oluşacağı ve bu öğeler arasındaki ilişkilere bağlı olarak çapraz tablonun herbir hücresine düşen sayıyı bulmak amaçlanacaktır. İlişkilendirilecek değişken sayısına ve ilişkinin biçimine göre kurulacak model değişecektir. Model seçim işlemi ise araştımanın dayalı olduğu kuramsal çerçeve tarafından belirlenecektir. Bu nedenle araştırmalarda veri toplamadan önceki aşama olarak kuramsal çerçevenin belirlenmesi önemli olmaktadır. Bu konular Hanushek ve Jaçkson (1977)'de daha ayrıntılı olarak işlenmiştir.

Çapraz tabloların log-linear model aracılığıyla çözümlenmesinde iki yaklaşım vardır. Birinci yaklaşım yalnız değişkenler arasında ilişki olup olmadığını saptamaya yöneliktir. Bağımlı ve bağımsız değişken ayırımı yapılmaz. İkinci yaklaşım ise değişkenlerden birini bağımlı değişken olarak ele alır. Birinci model Genel Log-Linear model olarak isimlendirilirken ikinci yaklaşım ise Logit model olarak isimlendirilir (Knoke ve Burke, 1980; 11-12).

Bu yazıda birinci yaklaşım üzerinde durulacaktır. Ayrıca hesaplamaların elle yapılabilmesi nedeniyle yalnızca 2x2 boyutlu tablolar ele alınacaktır. Böylece log-linear modellerde kuramsal dağılımın oluşturulması kolaylıkla gözlenebilecektir. Çözümlemede kullanılan tablonun daha büyük boyutlarda olması halinde hesaplamaların elle yapılması olanaksızlaşacağı için kuramsal dağılımın oluşturulması bilgisayar programları aracılığıyla olacaktır. Bu durum ise bu modellerin ve bu modellerle yapılan işlemlerin somut olarak aktarılmasını zorlaştıracaktır.

2.1. Kuramsal dağılımın (Beklenen Frekansların) Bulunması

Beklenen frekansların bulunmasında seçilen model önemli bir rol oynamaktadır. Eğer modelimiz ilişkilendirdiğimiz değişkenler arasındaki olası her ilişkiyi içeriyorsa, bu modele doymuş model diyoruz. Örneğin 2x2 boyutlu bir çapraz tabloda her hücreye düşen beklenen değer (N_{ij}) aşağıdaki formül ile elde ediliyor ise

$$N_{ij} = G d_i^{(2)} d_j^{(1)} d_{ij}^{(12)}$$

bu modele doymuş model diyoruz. Çünkü burada her hücreye düşen frekansın hesap edilmesinde olası bütün ilişkileri ve bu ilişkilerin hücre frekanslarına etkilerini dikkate alıyoruz. Yani hiçbir ilişki ve etki model dışında bırakılmıyor.

Tablo 2

		1. Değişken 1. Seçenek	1. Değişken 2. Seçenek
2.	1. Seçenek	$N_{11} = G\, d_1^{(2)}\, d_1^{(1)}\, d_{11}^{(12)}$	$N_{12} = G\, d_1^{(2)}\, d_2^{(1)}\, d_{12}^{(12)}$
Değişken	2. Seçenek	$N_{21} = G\, d_2^{(2)}\, d_1^{(1)}\, d_{21}^{(12)}$	$N_{22} = G\, d_2^{(2)}\, d_2^{(1)}\, d_{22}^{(12)}$

Burada,

N_{ij}, i. satır ve j. sutundaki frekansı

G, tablodaki hüclerin beklenen frekanslarının geometrik ortalamasını

$d_i^{(n)}$ n. değişkenin i. satır frekansına etkisini,

$d_j^{(n)}$ n. değişkenin j. sutun frekansına etkisini,

$d_{ij}^{(nm)}$ n. ve m. değişkenlerin i. satır ve j. sutundaki hücre frekansına birlikte etkilerini göstermektedir.

Değişkenin seçeneklerinin etkilerinin çarpımı 1' eşittir. Yani 2. değişkenin 1. satıra etkinin aynı değişkenin 2. satıra etisi ile çarpımı ve 1. değişkenin 1. sutun ve 2. sutuna etkilerinin çarpımı 1'dir. Bu ifadeyi formülleştirirsek,

$$d_1^{(1)} \cdot d_2^{(1)} = 1$$

$d_1^{(2)} \cdot d_2^{(2)} = 1$'yi elde ederiz.

Bu formüllerden yola çıkarak,

$$d^{(1)} = d_1^{(1)} = 1/d_2^{(1)}$$

$d^{(2)} = d_1^{(2)} = 1/d_2^{(2)}$ formülleri elde edilir.

Bir değişken dağılım üzerinde etkili değilse onun etki değeri 1 olacaktır. Örneğin birinci değişkenin dağılım üzerinde etkisi yoksa $d_1^{(1)}$ ve $d_2^{(1)}$'nin değeri 1 olacaktır.

Doymuş bir modelde beklenen frekanslar (N_{ij}) ile gözlenen frekanslar (n_{ij}) aynı değerleri alacağından parametre tahminlerini elde etmek kolay olacaktır. Çünkü parametrelerin tahmin edilmesinde gerekli olan beklenen frekansların yerine eşitleri olan gözlenen frekanslar kullanılır.

Parametrelerin tahmin edicilerini bulabilmek için önce odds ve odds oranı kavramlarının açıklanması gerekmektedir.

Odds: Herhangi bir değişkenin bir seçeneğinde bulunan frekansın aynı değişkenin bitişik diğer seçeneğindeki frekansına oranı.

$$\text{Odds} = N_{i,j} / N_{i+1,j} \quad \text{(2. Değişken için)}$$
$$= N_{i,j} / N_{i,j+1} \quad \text{(1. Değişken için)}$$

Odds Oranı: bitişik iki sıra ya da sutunun oddslarının oranı.

$$\Omega = \frac{N_{i,j} / N_{i+1,j}}{N_{i,j+1} / N_{i+1,j+1}} = \frac{N_{i,j} / N_{i,j+1}}{N_{i+1,j} / N_{i+1,j+1}}$$

$$\Omega = \frac{N_{i,j} \cdot N_{i+1,j+1}}{N_{i+1,j} \cdot N_{i,j+1}}$$

Tablo 2 için odds oranı ise

$$\Omega = \frac{N_{11} N_{22}}{N_{21} N_{12}}$$

Yukarıdaki formüllerden de anlaşılacağı gibi log-linear modellerde odds oranı önemli bir yer tutmaktadır. Bu oranın ve parametre tahminlerinin elde edilebilmesi için ise beklenen frekansların bilinmesi gerekmektedir. Bu değerlerin bulunmasını 2x2 boyutundaki bir tablo üzerinde ele alalım. Knoke ve Burke (1980) nin bir örgüte üye olup olmama ile seçimde oy kullanma değişkenlerinden oluşan tablosu (Tablo 3) aşağıdadır.

Tablo 3

Örgüte Üyelik ile Oy Verme Arasındaki İlişki

		Üyelik Durumu		
		Üye	Üye Değil	Toplam
Oy Kullanma	Kullanmış	689	298	987
Durumu	Kullanmamış	232	254	486
Toplam		921	552	1473

Bu tablodaki değerler gözlenen frekanslardır. Burada i. sıra ve j. sutuna düşen frekansı n_{ij}, i. sıranın toplamını $n_{i.}$, j. sutun toplamını $n_{.j}$, ve genel toplamı $n_{..}$ ile gösterelim. Bu durumda

$n_{11} = 689$ $n_{12} = 298$ $n_{1.} = 987$

$n_{21} = 232$ $n_{22} = 254$ $n_{2.} = 486$

$n_{.1} = 921$ $n_{.2} = 552$ $n_{..} = 1473$

olacaktır.

Beklenen frekanslar seçilen modele göre değişiklikler gösterecektir. i. sıra ve j. sutunun beklenen frekansının N_{ij}, i. sıra toplamını $N_{i.}$, j. sutun toplamını $N_{.j}$ ve genel toplamı $N_{..}$,ile gösterelim.

$N_{ij} = Gd_i^{(2)} d_j^{(1)} d_{ij}^{(12)}$ biçiminde formule edilen bir

dağılımda formüldeki parametrelerin tahmin edicilerini bulalım.

G tablonun hücrelerine düşen beklenen değerlerin geometrik ortalaması olduğundan bu parametrenin tahmin edicisi

$$G = (N_{11} N_{12} N_{21} N_{22})^{1/4}$$

olacaktır.

$d^{(12)}$ parametresinin tahmin edicisini odds oranından yararlanarak bulabiliriz. Odds oranı tanımından

1 $$\frac{N_{11} N_{22}}{N_{21} N_{12}} = \frac{(G\, d_1^{(2)} d_1^{(1)} d_{11}^{(12)})\ (G\, d_1^{(2)} d_2^{(1)} d_{12}^{(12)})}{(G\, d_2^{(2)} d_1^{(1)} d_{21}^{(12)})\ (G\, d_2^{(2)} d_2^{(1)} d_{22}^{(12)})}$$

eşitliğini yazarız. Buradan gerekli kısaltmaları yaparak

2 $$\frac{N_{11} N_{22}}{N_{21} N_{12}} = \frac{d_{11}^{(12)} d_{22}^{(12)}}{d_{21}^{(12)} d_{22}^{(12)}}$$

elde ederiz.

3 $$d^{(12)} = d_{11}^{(12)} = d_{22}^{(12)} = 1/d_{12}^{(12)} = 1/d_{21}^{(12)}$$

eşitliğini dikkate alarak formülü yeniden

4 $$\frac{N_{11} N_{22}}{N_{21} N_{12}} = (d^{(12)})^4$$

şeklinde yazabiliriz. Burada $d^{(12)}$ yi kuvvetten kurtararak

5 $$d^{(12)} = \sqrt[4]{\frac{N_{11} N_{22}}{N_{21} N_{12}}}$$

sonucunu elde ederiz.

$d^{(2)}$ ve $d^{(1)}$ parametrelerinin tahmin edicilerini bulmak için de sırasıyla (N_{11}/N_{21}) (N_{12}/N_{22}) ve (N_{11}/N_{12}) (N_{21}/N_{22}) çarpımlarından yararlanırız. $d^{(2)}$ 'yi bulmak için

6 $$\left(\frac{N_{11}}{N_{21}}\right) \cdot \left(\frac{N_{12}}{N_{22}}\right) = \left(\frac{G\, d_1^{(2)}\, d_1^{(1)}\, d_{11}^{(12)}}{G\, d_2^{(2)}\, d_1^{(1)}\, d_{21}^{(12)}}\right) \cdot \left(\frac{G\, d_1^{(2)}\, d_2^{(1)}\, d_{12}^{(12)}}{G\, d_2^{(2)}\, d_2^{(1)}\, d_{22}^{(12)}}\right)$$

eşitliğinde gerekli sadeleştirmeleri yapmamız gerekecektir. Gerekli kısaltmalar yapılarak

7 $$\left(\frac{N_{11}}{N_{21}}\right) \cdot \left(\frac{N_{12}}{N_{22}}\right) = \left(\frac{d_1^{(2)}}{d_2^{(2)}}\right)^2$$

elde edilir.

$d^{(2)} = d_1^{(2)} = 1/d_2^{(2)}$ olduğundan

$d_1^{(2)} / d_2^{(2)} = (d^{(2)})^2$ olacak ve buradan

8 $$\frac{N_{11} N_{12}}{N_{21} N_{22}} = \left(d^{(2)}\right)^2$$

ve

9 $$d^{(2)} = \sqrt[4]{\frac{N_{11} N_{12}}{N_{21} N_{22}}}$$

bulunacaktır.

Benzer şekilde (N_{11}/N_{12}) (N_{21}/N_{22}) çarpımından da

10 $$d^{(1)} = \sqrt[4]{\frac{N_{11} N_{21}}{N_{12} N_{22}}}$$

bulunacaktır.

Bu formüllerden yararlanarak modellerdeki parametrelerin tahminlerini bulalım.

$N_{ij} = G d_i^{(2)}\, d_j^{(1)}\, d_{ij}^{(12)}$ şeklinde formule edilen doymuş

modelde beklenen frekans ile gözlenen frekans eşit olduğundan formüldeki N_{ij}'lerin yerine n_{ij} değerleri kullanılacaktır. Bu durumda parametrelerin tahminleri,

$$G = (689 * 298 * 232 * 254)^{1/4}$$
$$= 331.66$$
$$d^{(2)} = (689 * 296/232*254)^{1/4}$$
$$= 1.37$$
$$d^{(1)} = (689 * 232 / 298 * 254)^{1/4}$$
$$= 1.21$$
$$d^{(12)} = (689 * 254 / 232 * 298)^{1/4}$$
$$= 1.26$$

olarak bulunacaktır. Bu tahminleri formülde yerlerine koyduğumuzda tablodaki değerler elde edilir. d parametreleri değişkenlerin dağılımı etkileme durumunu verirler. Eğer bir değişkene ait d parametresinin değeri 1 ise bu, o değişkenin dağılım üzerinde etkili olmadığını gösterir.

Doymamış modellerde hücrelerin gözlenen frekansı ile beklenen frekansları farklı olacağından, parametrelerin tahmin işlemlerinden önce beklenen frekansları bulmak gerekmektedir. Tablo 3 ile ele alınabilecek 4 olası modeli dikkate alarak beklenen frekansları hesap edelim. Bu dört model,

$$N_{ij} = Gd_i^{(2)} d_j^{(1)},$$
$$N_{ij} = Gd_i^{(2)},$$
$$N_{ij} = Gd_j^{(1)},$$
$$N_{ij} = G \text{ 'dir,}$$

Önce $N_{ij} = Gd_i^{(2)} d_j^{(1)}$ modeline göre beklenen frekansları bulalım. Modelde her iki değişken de ayrı ayrı bulunduğuna göre, $n_{i.} = N_{i.}$, $n_{.j} = N_{.j}$, ve $n_{..} = N_{..}$ koşulları sağlanmalı. Bu koşula göre boş tablomuz Tablo 4a şeklinde olacaktır.

Tablo 4 a

	Üye	Üye Değil	Toplam
Oy Kullanmamış	$N_{11} = ?$	$N_{12} = ?$	$N_{1.} = 987$
Oy Kullanmamış	$N_{21} = ?$	$N_{22} = ?$	$N_{2.} = 486$
Toplam	$N_{.1} = 921$	$N_{.2} = 552$	$N_{..} = 1473$

Hücrelere yerleştireceğimiz değerler,

$N_{11}/N_{21} = N_{1.} / N_{2.}$

$N_{12}/N_{22} = N_{1.} / N_{2.}$

$N_{11}+N_{21} = N_{.1}$

$N_{12}+N_{22} = N_{.2}$

ve

$N_{11}/N_{12} = N_{.1} / N_{.2}$

$N_{21}/N_{22} = N_{.1} / N_{.2}$

$N_{11}+N_{12} = N_{1.}$

$N_{21}+N_{22} = N_{2.}$

koşullarını sağlamalıdır. Eşitliklerde bilinen değerler yerlerine koyarak işlemleri yapalım.

$N_{11}/N_{21} = N_{1.} / N_{2.}$

$= 987 / 486$

$= 2.03$

$N_{11} = 2.03 * N_{21}$

$N_{11}+N_{21} = N_{.1}$

$N_{11}+N_{21} = 921$

$2.03 * N_{21}+N_{21} = 921$

$(2.03 + 1) * N_{21} = 921$

$3.03 * N_{21} = 921$

$N_{21} = 921 / 3.03$

$= 303.96$

$N_{11} = 2.03 * N_{21}$

$= 2.03 * 303.96$

$= 617.04$

Aynı yolla N_{12} ve N_{22}'nin değerlerini sırasıyla 369.82 ve 182.18 buluruz. Bulunan bu değerler tabloda yerlerine konulduğunda Tablo 4b elde edilir.

Tablo 4 b

	Üye	Üye Değil	Toplam
Oy Kullanmış	617.04	369.82	987
Oy Kullanmamış	303.96	182.18	486
Toplam	921	552	1473

$N_{ij} = Gd_i^{(2)} d_j^{(1)} d_{ij}^{(12)}$ modeline göre kuramsal dağılım (beklenen frekanslar) elde edildikten sonra parametrelerin tahminleri bulunabilir.

$G = (617.04 * 369.82 * 303.96 * 182.18)^{1/4}$
$= 335.28$

$d^{(2)} = (617.04 * 369.82/303.96 * 182.18)^{1/4}$
$= 1.42$

$d^{(1)} = (617.04 * 303.96/369.82*182.18)^{1/4}$
$= 1.29$

$d^{(12)} = 1$

İkinci olarak $N_{ij} = Gd_i^{(2)}$ modeline göre oluşturulan kuramsal dağılımı bulalım. Modelde yalnız ikinci değişken yani oy kullanma davranışı bulunduğuna göre $n_{i.} = N_{i.}$ ve $n_{..} = N_{..}$ koşulları sağlanmalıdır. Bu durum Tablo 5a'da gösterilmiştir.

Tablo 5 a

	Üye	Üye Değil	Toplam
Oy Kullanmamış	$N_{11} = ?$	$N_{12} = ?$	$N_{1.} = 987$
Oy Kullanmamış	$N_{21} = ?$	$N_{22} = ?$	$N_{2.} = 486$
Toplam	$N_{.1} = ?$	$N_{.2} = ?$	$N_{..} = 1473$

Bu modelde hücrelere yerleştireceğimiz değerler,

$N_{11} / N_{21} = N_{1.} / N_{2.}$

$N_{12} / N_{22} = N_{1.} / N_{2.}$

$N_{.1} / N_{.2} = 1$

koşullarını sağlamalıdır.

$N_{.1} / N_{.2} = 1$

$N_{.1} + N_{.2} = N_{..}$

eşitliklerinden,

$N_{.1} = N_{.2}$

$2 * N_{.1} = N_{..}$

$N_{.1} = N_{..}/2$

$= 1473/2$

$= 736.5$

$N_{.2} = N_{.1}$

$= 736.5$ bulunur.

$N_{11} / N_{12} = 1$ ve $N_{11} + N_{12} = 987$ eşitlikleri birlikte çözüldüğünde,

$N_{11} = N_{12}$

$2 * N_{11} = 987$

$N_{11} = 493.5$ ve $N_{12} = 493.5$ değerleri bulunur. Benzer şekilde $N_{21}/N_{22} = 1$ ve $N_{21}+N_{22} = 486$ eşitlileri birlikte çözüldüğünde ise N_{21} ve N_{22}'in değerleri 243 olarak bulunur.

Bulunan bu değerleri yerlerine koyarak bu modeldeki kuramsal dağılım olan Tablo 5b elde edilir.

Tablo 5 b

	Üye	Üye Değil	Toplam
Oy Kullanmış	493.5	493.5	987
Oy Kullanmamış	243	243	486
Toplam	736.5	736.5	1473

Bu modeldeki parametrelerin tahminleri ise şöyledir.

$G = (493.5 * 493.5 * 243 * 243)^{1/4}$

$= 346.3$

$d^{(2)} = (493.5 * 493.5 / 243 * 243)^{1/4}$

$= 1.43$

$d^{(1)} = 1$

$d^{(12)} = 1$

Üçüncü modelimiz $N_{ij} = Gd_j^{(1)}$ ye göre kuramsal dağılımı bulalım. Modelde yalnızca üyelik durumunun dağılım üzerinde etkin olduğu düşünüldüğüne göre $n_{.j} = N_{.j}$ ve $n_{..} = N_{..}$ koşulları sağlanarak diğer frekanslar elde edilecektir. Bu durum aşağıda Tablo 6a da gösterilmiştir.

Tablo 6 a

	Üye	Üye Değil	Toplam
Oy Kullanmamış	$N_{11} = ?$	$N_{12} = ?$	$N_{1.} = ?$
Oy Kullanmamış	$N_{21} = ?$	$N_{22} = ?$	$N_{2.} = ?$
Toplam	$N_{.1} = 921$	$N_{.2} = 552$	$N_{..} = 1473$

Bilinmeyen değerleri bulabilmek için,

$N_{11} / N_{12} = N_{.1} / N_{.2}$

$N_{21} / N_{22} = N_{.1} / N_{.2}$

$N_{1.} / N_{2.} = 1$

koşulları sağlanmalıdır.

$N_{1.} / N_{2.} = 1$ ve $N_{1.} + N_{2.} = N_{..}$ eşitliklerinde bilinenleri yerlerine koyarsak,

$N_{1.} = 736.5$

$N_{2.} = 736.5$

değerlerini buluruz.

$N_{11} / N_{21} = 1$ ve $N_{11} / N_{21} = N_{.1}$ eşitlikleri birlikte çözülerek,

$N_{11} = 460.5$

$N_{21} = 460.5$

$N_{12} / N_{22} = 1$ ve $N_{12} + N_{22} = N_{.2}$ eşitlikleri de birlikte çözülerek,

$N_{12} = 276$

$N_{22} = 276$

bulunur.

Bulduğumuz bu değerleri yerlerine koyduğumuzda ise bu modelin kuramsal dağılımını gösteren Tablo 6b'yi elde ederiz.

Tablo 6 b

	Üye	Üye Değil	Toplam
Oy Kullanmış	460.5	276	736.5
Oy Kullanmamış	460.5	276	736.5
Toplam	921	552	1473

Bu kuramsal dağılımdan $N_{ij} = Gd_j^{(1)}$ modelindeki parametrelerin tahminleri ise aşağıda hesap edilmiştir.

$G = (460.5 * 460.5 * 276 * 276)^{1/4}$

$= 356.51$

$d^{(1)} = (460.5 * 460.5/276*276)^{1/4}$

$= 1.29$

$d^{(2)} = 1$

$d^{(12)} = 1$

Son modelimiz ise yalnızca G parametresini içeren $N_{ij} = G$ modelidir. Modelde üyelik ve oy kullanmanın dağılım üzerinde etkili olmadığı ileri sürülüyor. Buna göre yalnızca $n_{..} = N_{..}$ eşitliğinden yola çıkarak dağılımın frekansları hesap edilecektir. Başlangıçta bilinenlerle bilinmeyenler aşağıda Tablo 7a da gösterilmiştir.

Tablo 7 a

	Üye	Üye Değil	Toplam
Oy Kullanmamış	$N_{11} = ?$	$N_{12} = ?$	$N_{1.} = ?$
Oy Kullanmamış	$N_{21} = ?$	$N_{22} = ?$	$N_{2.} = ?$
Toplam	$N_{.1} = 921$	$N_{.2} = 552$	$N_{..} = 1473$

Modelde G dışındaki parametrelerin görülmemesi o parametre değerlerinin 1 olduğunu yani etkisiz oldukarını göstermektedir. Buradan,

$N_{1.} / N_{2.} = 1$

$N_{.1} / N_{.2} = 1$

$N_{11} / N_{12} = 1$

$N_{11} / N_{21} = 1$

$N_{12} / N_{22} = 1$

$N_{21} / N_{22} = 1$

eşitliklerini yazabiliriz. Ayrıca

$N_{1.} + N_{2.} = N_{..}$

$N_{.1} + N_{.2} = N_{..}$

$N_{11} + N_{12} = N_{1.}$

$N_{21} + N_{22} = N_{2.}$

$N_{11} + N_{21} = N_{.1}$

$N_{12} + N_{22} = N_{.2}$

eşitliklerini de gözönünde bulundurarak beklenen değerleri bulabiliriz. Önce satır ve sutunların toplamlarını bulalım.

$N_{1.} / N_{2.} = 1$

$N_{1.} + N_{2.} = N_{..}$

eşitliklerinde bilinen tek değer olan $N_{..}$ nin değerini yerine koyarak,

$N_{1.} = 736.5$

$N_{2.} = 736.5$

değerlerini buluruz.

$N_{.1} / N_{.2} = 1$ ve $N_{1.} + N_{.2} = N_{..}$ eşitliklerinin birlikte çözümlerinden de $N_{.1}$ ve $N_{.2}$ nin değerleri 736.5 olarak bulunur. Benzer şekilde N_{11}, N_{12}, N_{21}, N_{22} nin değerlerinin 368.25 olduğu hesap edilir. Bulduğumuz bu değerlerle de kuramsal dağılımı gösteren Tablo 7b yi oluşturabiliriz.

Tablo 7 b

	Üye	Üye Değil	Toplam
Oy Kullanmış	368.25	368.25	736.5
Oy Kullanmamış	368.25	368.25	736.5
Toplam	736.5	736.5	1473

Bu kuramsal dağılımın da parametre tahminleri

$G = (368.25^4)^{1/4}$

$= 368.25$

$d^{(2)} = 1$

$d^{(1)} = 1$

$d^{(12)} = 1$

olarak bulunur.

Seçilen modeller ve parametre tahmin değerleri Tablo 8 de toplu olarak verilmiştir.

Tablo 8

Model	G	$d^{(2)}$	$d^{(1)}$	$d^{(12)}$
$N_{ij}=Gd_i^{(2)}d_j^{(1)}d_{ij}^{(12)}$	331.66	1.37	1.21	1.26
$N_{ij}=Gd_i^{(2)}d_j^{(1)}$	335.28	1.42	1.29	1
$N_{ij}=Gd_i^{(2)}$	346.30	1.43	1	1
$N_{ij}=Gd_j^{(1)}$	356.51	1	1.29	1
$N_{ij}=G$	368.25	1	1	1

3. Değişkenler Arasındaki İlişkilerin Değerlendirilmesi

Seçilen modele göre kuramsal dağılımının bulunmasından sonraki aşama değerlendirme olmaktadır. Burada değerlendirme ölçütü beklenen değerler ile gözlenen değerler arasındaki farktır. Bu değerler arasında fark varsa dağılım modele uymamakta, fark yoksa seçtiğimiz model elimizdeki verilere uymaktadır.

Çapraz tablolarda yoğun olarak kullanılan hipotezden yola çıkarak hem log-linear model ile hipotezin nasıl test edileceğini hem de klasik Ki-Kare modeliyle ilişkisini ele alalım. Bu hipotez iki değişkenin, bizim örneğimizde oy davranışı ile örgüte üyelik değişkenleri, birbirinden bağımsız olduğunu ileri sürer.

Oluşturduğumuz çapraz tabloyu klasik ki-kare tekniği ile çözümlemek istediğimizde, değişkenlerin birbirlerinden bağımsız olduklarını kabul ederek bulduğumuz beklenen değerlerle gözlenen değerler arasındaki farka bakılır. Bu değerler arasında bir farkın olmaması, yani Ki-Kare değerinin sıfır olması hipotezi doğrularken, farkın olması ise hipotezi yanlışlar. Bu nedenle ilişkinin varlığını kabul edebilmemiz için Ki-Karenin değerinin büyük olması gerekir.

Aynı tablo Log-Linear model ile çözümlenmek istendiğinde ise değişik hipotezler oluşturulabilir. Ki-Kare tekniğinde olduğu gibi tek bir hipoteze bağlı kalınamaz. Log-Linear modelde Ki-Kare tekniğinde ele alınan hipotez, $N_{ij} = Gd_i^{(2)} d_j^{(1)}$ modelle ifade edilir. Bu modele göre bulunan beklenen değerlerle gözlenen değerler arasındaki fark, modelin verilere uyup uymadığı konusunda bilgi verir. Modelde ileri sürüldüğü gibi iki değişken arasında bir ilişki yoksa kuramsal dağılım ile gözlenen frekansların oluşturduğu dağılım benzerlik gösterirler. Yani gözlenen değerlerle beklenen değerler arasında fark görülmez.

Log-Linear modellerde gözlenen frekanslarla beklenen frekanslar arasındaki fark Ki-Kare istatistiği yerine Olabilirlik Oranı (Likelihood Ratio) istatistiği ile test edilir. Olabilirlik Oranının Ki-Kareye tercih edilmesinin iki temel nedeni: 1) Olabilirlik Oranının beklenen frekansların en çok olabilirlik (Maximum Likelihood) yöntemi ile tahmin edilmesi ve 2) Olabilirlik oranı istatistiğinin çok boyutlu tablolarda koşullu bağımsızlıkların testini de sağlayan biçimde parçalara ayrılabilmesidir (Knoke ve Burke, 1980; 30).

Olabilirlik Oranı İstatistiği,

$$L^2 = 2\,\Sigma\, n_{ij} \ln (n_{ij} / N_{ij})$$

formülü ile gösterilir.

Formülden de anlaşılacağı gibi gözlenen frekans ile beklenen frekans değerleri birbirlerine yaklaştığında n_{ij} / N_{ij} değeri de 1 e yaklaşacağından L^2 değeri de sıfıra yaklaşacaktır. Bu nedenle L^2 değerinin büyüklüğü seçilen modelin gerçeklere uymadığını göstermektedir.

Olabilirlik oranı istatistiğinde dağılım serbestlik derecesi, değeri 1 e eşit olan parametre sayısına eşittir. Bu durumu bu yazıda kullanılan modeller üzerinde görelim.

Model	G	$d^{(2)}$	$d^{(1)}$	$d^{(12)}$	L^2	sd
$N_{ij} = Gd_i^{(2)} d_j^{(1)} d_{ij}^{(12)}$	331.66	1.37	1.21	1.26	0.00	0
$N_{ij} = Gd_i^{(2)} d_j^{(1)}$	335.28	1.42	1.29	1	66.78	1
$N_{ij} = Gd_i^{(2)}$	346.30	1.43	1	1	160.22	2
$N_{ij} = Gd_j^{(1)}$	356.51	1	1.29	1	240.63	2
$N_{ij} = G$	368.25	1	1	1	334.07	3

Modellerin olabilirlik oranına bakıldığında 1. model dışında hiçbir modelin verilere uymadığı görülmektedir. O halde bu çözümlemeler ışığı altında iki değişken arasında ilişkisinin anlamlı olduğu kabul edilecektir.

4. Sonuç

Nitel değişkenlerle çalışma zorunluluğunda olan sosyal bilimcilerin Ki-Kare tekniği dışında bir teknikle de verilerini çözümleyebilmeleri ve bu yeni teknikle daha farklı hipotezlerin sınanabilmesi, sosyal bilimlerde elde edilen genellemelerin daha duyarlı hale gelmesini sağlayacaktır.

Bu amaca hizmet etmek için bu yazıda nitel değişkenlerin çözümlenmesinde kullanımı artan log-linear modeller tanıtılmaya çalışılmıştır.

KAYNAKLAR

ANDERSEN, E.B. (1980), **Discrete Statistical Models With Social Science Applications**, Amsderdam: Nort-Holland Publishing Company.

BISHOP, Y.M.M., FIENBERG, S.E., HOLLAND, P.W. (1975), **Discrete Multivariate Analysis: Theory and Practice**, Cambridge: MIT Press.

GOODMAN, L.A. (1970), "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications,", **Journal of American Statistical Association**, 65, 226-255.

GOODMAN, L.A. (1984), **The Analysis of Cross Classified Data Having Ordered Categories**, Harvard: Harvard Uni. Press.

HABERMAN, S.J. (1974), **The Analysis of Frequency Data**, Chicago: The University of Chicoga Press.

HABERMAN, S.J. (1978), **Analysis of Qualitativa Data: Volume 1**, New York: Academic Press.

HABERMAN, S.J. (1979), **Analysis of Qualitativa Data: Volume 2**, New York: Academic Press.

HANUSHEK, E.A., JACKSON, J.E. (1977), **Statistical Methods for Social Scientists**, New York: Academic Press.

KNOKE, D., BURKE, P.J. (1980), **Log-Linear Models**, Beverly Hills: Sage University Publications.