İstatistik Araştırma Dergisi Cilt 01 No 01 Sayfa 23-37 Nisan 2002

Üç Yönlü Tablolarda χ^2 İstatistiğinin Kullanılması

Mehmet AKYOL* Fikret GÜRBÜZ**

ÖZET

Hipotez testlerinde kullanılan olasılık dağılımları, değişkenlerin sürekli veya kesikli olmasına göre değişmektedir. Her zaman sürekli değişkenleri bulmak veya verileri bu şekilde ifade etmek mümkün olmayabilir. Böyle durumlarda özellikleri kategorilerine göre sınıflandırmak, her sınıf içine düşen birim sayısını (frekansı) hesaplamak suretiyle ki-kare testi uygulanır. Çalışmada sayarak elde edilen verilerde üç sınıflama kriteri olması durumunda, iki yönlü tablolarda kullanılan χ^2 analizindeki değişiklikler ele alınmıştır.

Uygulama olarak 3 yönlü gerçek bir veri seti dikkate alınmış ve χ^2 istatistiği kullanılarak analiz edilmiştir. Veri seti; cinsiyet (k=1,2), intihar şekli (i=1,2,...,6) ve intihar nedeni (j=1,2,...,8) değişkenlerinden oluşmaktadır. Analiz sonucunda, değişkenler arasındaki tüm birinci derece ilişkiler önemli bulunmuştur.

Anahtar Kelimeler: *İki yönlü tablolar, Üç yönlü tablolar, Çok yönlü tablolar, χ^2 analizi.*

1. GİRİŞ

Hipotez testlerinde kullanılan olasılık dağılımları, değişkenlerin sürekli veya kesikli olmasına göre değişmektedir.

Değişkenler sürekli olduğunda z, t ve F dağılımları kullanılır fakat sürekli değişkenleri bulmak veya değişkenleri bu şekilde ifade etmek her zaman mümkün olmayabilir. Mesela, boy ve beden ağırlığı ile ilgili bilgiler, hem rakamlarla ifade edilir, hem de hafif-ağır ve kısa boylu-uzun boylu olarak gruplandırılabilir. Çoğu sınıflandırmalar da açık renkli-koyu renkli, iyi-kötü, fertlerin tahsil durumları vb. gibidir. Babaların öğrenim durumları ile çocukların öğrenim durumları arasında bir ilişki arandığında, öğrenim durumlarını rakamlarla ifade etmek sözkonusu değildir. Bu sebeple, böyle durumlarda özellikleri kategorilerine göre sınıflandırmak ve bu kategorilerde gözlenen ve beklenen frekansları hesaplamak suretiyle ki-kare testi uygulanır.

* Dr., Devlet İstatistik Enstitüsü, Eğitim Merkezi Koordinatörü, Ankara, Türkiye, (Haberleşme adresi)

** Prof.Dr., Ankara Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik Anabilim Dalı, Ankara, Türkiye

1.1. Tek Yönlü Tablolar

Gerçekleşmiş sonuçlardan oluşan dağılımın beklenen frekanslardan oluşan dağılıma uyup uymadığının, diğer bir deyişle bu beklenen frekanslarla gözlenen frekanslar arasındaki farkın tesadüften ileri gelip gelmediği hipotez kontrolü yapılarak bulunabilir.

Örnekten elde edilen gözlenen frekanslar n_i (i=1,2,...,r) ve beklenen frekanslar E_i (i=1,2,...r) ile gösterilir. Uygunluk testinde serbestlik derecesi, dağılımın tahmin edilen parametresi yoksa kategori sayısının bir eksiğine eşittir. Testte kullanılacak istatistik aşağıdaki gibidir. Bu dağılımın χ^2'ye uygunluğu $E_i \geq 5$ olduğunda artar.

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - E_i)^2}{E_i}$$

Gözlem yoluyla elde edilen herhangi bir dağılımın hangi teorik dağılıma uyduğunun belirlenmesinde izlenilecek en basit yol, sözkonusu dağılımdan hesaplanan momentler ile seçilen teorik dağılımın momentlerinin karşılaştırılması şeklindedir. Beklenen frekanslar gözlenen frekanslara yeterince yakın bulunduğunda ise ilgilenilen teorik dağılıma uyduğu kabul edilir. Teorik dağılımın, örnek istatistiklerden tahmin edilen parametre sayısı ise serbestlik derecesinin belirlenmesinde önemli rol oynar. Mesela normal dağılım halinde, serbestlik derecesi (r-1)-2=r-3 (ortalama μ ve standart sapma σ parametreleri) , poisson dağılımında (r-1)-1=r-2 (ortalama λ parametresi) şeklindedir.

1.2. rxc (İki) Yönlü Tablolar

İki özelliğe sahip populasyondan çekilmiş örnek, rxc tabloda gösterilir. $n_{..}$ gözlem iki nicel değişkene göre sınıflandırılmıştır. Birinci değişken satırda r (row) sınıftan, ikinci değişken de sütunda c (column), sınıftan oluşur.

i. satır j. sütundaki gözlenmiş frekans n_{ij} olarak gösterilir. i. satırdaki gözlenmiş frekansların toplamı, $n_{i.}$, j. sütundaki gözlenmiş frekansların toplamı $n_{.j}$ ile gösterilir. Bunlar marjinal toplamlar olarak ifade edilir. Marjinal toplamları şu şekilde bulmak mümkündür:

$$n_{i.} = n_{i1} + n_{i2} + \ldots + n_{ic}$$
$$= \sum_{j=1}^{c} n_{ij},$$
$$n_{.j} = n_{1j} + n_{2j} + \ldots + n_{rj}$$
$$= \sum_{i=1}^{r} n_{ij}.$$

Benzer şekilde genel toplam;

$$n_{..} = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}$$

$$= \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{c} n_{.j}$$

şeklinde gösterilir.

İki değişken arasında bağımsızlık testi yapılırken populasyondaki frekans bilinmemektedir. Bundan dolayı gözlenmiş frekanslardan hareketle hesaplanan $p_{i.}$ ve $p_{.j}$ olasılıkları yerine $\hat{p}_{i.}$ ve $\hat{p}_{.j}$ olasılıkları kullanılabilir. Bu olasılıklar örnekteki gözlenen frekanslar yardımıyla şu şekilde hesaplanabilir.

$$\hat{p}_{i.} = \frac{n_{i.}}{n_{..}} \quad \text{ve} \quad \hat{p}_{.j} = \frac{n_{.j}}{n_{..}}$$

Ençok olabilirlilik tahmin yöntemine ij hücresindeki beklenen frekans;

$$E_{ij} = N\,\hat{p}_{i.}\hat{p}_{.j}$$

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

şeklinde hesaplanır.

Bağımsızlığın ileri sürüldüğü H_0 hipotezinin testi için Pearson tarafından ileri sürülen χ^2 istatistiği şu şekilde hesaplanır:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Bu dağılımın χ^2'ye uygunluğu $E_{ij} \geq 5$ olduğunda artar. Serbeslik derecesi; s.d=(r-1)(c-1) dır.

2. ÜÇ YÖNLÜ TABLOLARDA χ^2 İSTATİSTİĞİNİN KULLANILMASI

Üç yönlü tabloların analizi, iki yönlü tabloların analizi ile karşılaştırılarak verilebilir. rxc tabloları için kullanılan terminoloji rxcxl tabloları için de genişletilebilir. Genel olarak üç yönlü tablolar Tablo 1'deki gibi gösterilebilir.

Üç yönlü tablolarda r satırları, c sütunları ve l tabakaları gösterir. ijk gözündeki gözlenmiş frekans n_{ijk} olarak verilir. $i=1,2,\ldots,r$, $j=1,2,\ldots,c$ ve $k=1,2,\ldots,l$. Çeşitli alt indisler dikkate alınarak yapılan toplamlar marjinal toplamlar olarak adlandırılır. Örnek olarak i ve j'in tüm değerleri için yapılan toplam, k'ıncı tabakanın marjinal toplamını verir. Benzer şekilde her j ve k, her i ve k indislerine ait n_{ijk} değerlerinin toplanmasıyla i'inci satırın ve j'inci sütunun marjinal toplamları elde edilir. Bu toplamlar tek değişkenli marjinal toplamlar olarak bilinir. Bu toplamlar şu şekilde verilebilir:

$$n_{i..} = \sum_{j=1}^{c}\sum_{k=1}^{l} n_{ijk},$$

$$n_{.j.} = \sum_{i=1}^{r}\sum_{k=1}^{l} n_{ijk},$$

$$n_{..k} = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ijk}.$$

Tablo 1. Üç yönlü tabloların genel formu

Satır	Tabaka	Sütun (Değiken 2)			
(Değişken 1)	(Değişken 3)	1	2	...	c
1	1	n_{111}	n_{121}	...	n_{1c1}
	2	n_{112}	n_{122}	...	n_{1c2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{11l}	n_{12l}	...	n_{1cl}
2	1	n_{211}	n_{221}	...	n_{2c1}
	2	n_{212}	n_{222}	...	n_{2c2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{21l}	n_{22l}	...	n_{2cl}
⋮	1			...	
	2			...	
	⋮			...	
	l			...	
r	1	n_{r11}	n_{r21}	...	n_{rc1}
	2	n_{r12}	n_{r22}	...	n_{rc2}
	⋮	⋮	⋮	⋮	⋮
	l	n_{r1l}	n_{r2l}	...	n_{rcl}

Benzer şekilde iki değişkenli marjinal toplamlar da elde edilebilir.

$$n_{ij.} = \sum_{k=1}^{l} n_{ijk},$$

$$n_{i.k} = \sum_{j=1}^{c} n_{ijk},$$

$$n_{.jk} = \sum_{i=1}^{r} n_{ijk}.$$

Gözlenmiş frekansların genel toplamı ($n_{...}$) ise şu şekilde hesap edilir;

$$n_{...} = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{l} n_{ijk},$$

2.1. Karşılıklı Bağımsızlık Hipotezi

Üç yönlü tabloların incelenmesi iki yönlü tablolardaki iki değişkenin bağımsızlığı üzerinde kurulan hipotezle mümkün değildir. Üç yönlü tablolarda birden çok hipotez söz konusudur. Üç yönlü tablolar için en basit hipotez, değişkenlerin karşılıklı bağımsızlılığı hipotezidir. Üç yönlü tabloda değişkenlerin karşılıklı bağımsızlık hipotezi aşağıdaki gibi gösterilebilir. Bu hipoteze göre satırlar, sütunlar ve tabakalar birbirinden tamamen bağımsızdır.

$$H_0^{(0)}: p_{ijk} = p_{i..}\, p_{.j.}\, p_{..k}$$

p_{ijk}; ijk gözündeki gözlemlerinin gözlenme olasılığını, $p_{i..}$,$p_{.j.}$ ve $p_{..k}$ satır , sütun ve tabaka değişkenlerinin marjinal olasılıklarını gösterir. Verilen bu hipotez iki yönlü tablolar için ele alınan bağımsızlık hipotezinin üç yönlü tablolardaki karşılığıdır. Bu hipotezi test etmek için, iki yönlü tablolarda kullanılan yaklaşım kullanılır. İlk olarak $H_0^{(0)}$ hipotezinin geçerliliğinde, beklenen frekanslar hesaplanır. Daha sonra, bu frekanslar bilinen χ^2 istatistiğindeki gibi gözlenmiş frekanslarla eşleştirilir. Son olarak serbestlik derecesi ve önemlilik düzeyi belirlenir. Üç değişkenli karşılıklı bağımsızlık hipotezi durumunda örnek beklenen frekansların bulunmasında da iki yönlü tablolar için kullanılan yönteme benzer yöntem kullanılır.

$$E_{ijk} = n_{...}\hat{p}_{i..}\hat{p}_{.j.}\hat{p}_{..k},$$

eşitliğindeki $\hat{p}_{i..}, \hat{p}_{.j.}$ ve $\hat{p}_{..k}$ olasılıkları $p_{i..}$, $p_{.j.}$ ve $p_{..k}$ olasılıkların tahminleridir. Bu olasılık tahminleri, ilgili değişkenlerin marjinal toplamlarından hareketle hesap edilir.

$$\hat{p}_{i..} = \frac{n_{i..}}{n_{...}},\ \hat{p}_{.j.} = \frac{n_{.j.}}{n_{...}},\ \hat{p}_{..k} = \frac{n_{..k}}{n_{...}}$$

Bunlar ençok olabilirlilik tahminleridir. Marjinal olasılıklar $E_{ijk} = n_{...}\hat{p}_{i..}\hat{p}_{.j.}\hat{p}_{..k}$, eşitliğinde yerine konulduğunda,

$$E_{ijk} = n_{...}\frac{n_{i..}}{n_{...}}\frac{n_{.j.}}{n_{...}}\frac{n_{..k}}{n_{...}}$$

$$= \frac{n_{i..}n_{.j.}n_{..k}}{n_{...}^2}$$

herhangi bir gözün teorik olarak beklenen frekansı elde edilmiş olur. Bu eşitlikten yararlanarak hesaplanan beklenen frekanslarda sonra test istatistiği şu şekilde hesap edilir (Everitt, 1994).

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{l}\frac{(n_{ijk} - E_{ijk})^2}{E_{ijk}}$$

Bu dağılımın χ^2'ye uygunluğu $E_{ijk}\geq 5$ olduğunda artar. Üç yönlü tablolar için χ^2 istatistiğinin serbestlik derecesi aşağıdaki genel eşitlik yardımıyla hesap edilebilir.

Bu dağılımın χ^2'ye uygunluğu $E_{ijk}\geq 5$ olduğunda artar. Üç yönlü tablolar için χ^2 istatistiğinin serbestlik derecesi aşağıdaki genel eşitlik yardımıyla hesap edilebilir.

s.d.=(tablodaki göz sayısı-1) - (hipotez testi için verilerden yararlanılarak tahmin edilen olasılıkların sayısı)
=rcl-(r-1)-(c-1)-(l-1)-1
=rcl-r-c-l+2

Karşılıklı bağımsızlık hipotezi için tek değişkenli beklenen marjinal toplamların tek değişkenli gözlenen marjinal toplamlara ($E_{i..}=n_{i..}$, $E_{.j.}=n_{.j.}$ ve $E_{..k}=n_{..k}$) eşitliliği gibi kısıtlamalar sözkonusudur.

2.2. Kısmi Bağımsızlık Hipotezleri

Üç yönlü tablolarda bir değişkenin diğer iki değişkenden bağımsızlığını test etmeden önce karşılıklı bağımsızlık hipotezinin test edilmesine gerek vardır. Karşılıklı bağımsızlık hipotezi test edilip $H_0^{(0)}$ hipotezi red edilirse, değişkenler arasında bazı ilişkilerin varlığından söz edilir.

Üç değişkene ilişkin kurulan bağımsızlık hipotezinin red edilmesi durumunda değişkenlerin herhangi ikisi arasında ilişki olabilir. Bu durum kısmi (partial) bağımsızlıkla ilgilidir. Yani iki değişken diğer üçüncü değişkenden kısmi bağımsızlık hipotezleri şeklindedir. Bu hipotezler olasılıklarla ifade edilebilir. Üç yönlü tablolarda kısmi bağımsızlık hipotezleri aşağıdaki gibi verilebilir(Everitt, 1994).

1) $H_0^{(1)} : p_{ijk} = p_{i..}p_{.jk}$ (sütun ve tabaka sınıflaması satır sınıflamasından bağımsızdır)

2) $H_0^{(2)} : p_{ijk} = p_{.j.}p_{i.k}$ (satır ve tabaka sınıflaması sütun sınıflamasından bağımsızdır)

3) $H_0^{(3)} : p_{ijk} = p_{..k}p_{ij.}$ (satır ve sütun sınıflaması tabaka sınıflamasından bağımsızdır)

Bu hipotezlerin birincisinde ijk gözündeki gözlemin görünme olasılığı p_{ijk} , i'inci satır sınıflamasının olasılığı $p_{i..}$, ve j'inci sütun ve k'ıncı tabaka sınıflamasının olasılığı $p_{.jk}$ dır. Eğer bu hipotez geçerli ise sütun ve tabaka sınıflaması satır sınıflamasından bağımsızdır. Bu hipoteze göre uygun beklenen frekanslar şu şekilde hesap edilebilir.

$$E_{ijk} = n_{...}\hat{p}_{i..}\hat{p}_{.jk},$$

$p_{i..}$ ve $p_{.jk}$ olasılıklarının tahmin edicileri $\hat{p}_{i..}$ ve $\hat{p}_{.jk}$ olup, ilgili marjinal toplamlar cinsinden şu şekilde bulunabilirler.

$$\hat{p}_{i..} = \frac{n_{i..}}{n_{...}}, \quad \hat{p}_{.jk} = \frac{n_{.jk}}{n_{...}}$$

İki değişkenli marjinal toplam ($n_{.jk}$); i'inci değişken düzeyinde gözlenmiş frekansların toplanmasıyla bulunabilir. En çok olabilirlilik tahminleri aşağıdaki gibi kullanılarak beklenen frekanslar hesap edilebilir:

$$E_{ijk} = n_{...}\frac{n_{i..}}{n_{...}}\frac{n_{.jk}}{n_{...}}$$

$$= \frac{n_{i..}n_{.jk}}{n_{...}}$$

Test istatistiği daha önce üç yönlü tablolar için verilen χ^2 istatistiği ile aynı olup, serbestlik derecesi ise şu şekildedir;

$$s.d.=(r-1)(cl-1)$$
$$=rcl-r-cl+1$$

Burada $E_{i..}=n_{i..}$, $E_{.j.}=n_{.j.}$ ve $E_{..k}=n_{..k}$ marjinal kısıtlamaları yanında $E_{.jk}=n_{.jk}$ gibi marjinal kısıtlaması da söz konusudur.

$H_0^{(2)}$ hipotezinde verilen $p_{.j.}$ j'nci sütun sınıflamasının olasılığını ve $p_{i.k}$ i'nci satır ve k'ncı tabaka sınıflamasının olasılığını, $H_0^{(3)}$ hipotezinde verilen $p_{..k}$ k'nci tabaka sınıflamasının olasılığını ve $p_{ij.}$ i'nci satır ve j'nci sutun sınıflamasının olasılığını gösterir. Bu hipotezlere uygun beklenen frekansları, serbestlik dereceleri ve marjinal kısıtlamaları; $H_0^{(1)}$ hipotezindekine benzer şekilde beklenen frekanslar, serbestlik derecesi ve marjinal kısıtlar hesaplanabilir.

2.3. Koşullu Bağımsızlık Hipotezleri

Herhangi bir değişken verildiğinde diğer iki değişkenin bağımsızlığının araştırması söz konusu olabilir. Üç yönlü bir tabloda tabaka sınıflamasının verilmesinde, satır ve sütun sınıflamalarının bağımsızlığı, koşullu olasılığın tanımlanmasıyla elde edilebilir. Tabaka sınıflamasının verilmesinde, sütun ve satır sınıflamasının koşullu olasılığı (Christensen, 1990),

$$P(\text{satır}=i, \text{sütun}=j \mid \text{tabaka}=k)$$
$$=P(\text{satır}=i, \text{sütun}=j, \text{tabaka}=k)/P(\text{tabaka}=k)$$
$$=p_{ijk}/p_{..k}$$

şeklindedir. Tabaka sınıflamasının verilmesi durumunda satır ve sütun sınıflamasının bağımsızlığı ise;

$$P(\text{satır}=i, \text{sütun}=j \mid \text{tabaka}=k)$$
$$=P(\text{satır}=i \mid \text{tabaka}=k)P(\text{sütun}=j \mid \text{tabaka}=k)$$
$$=(p_{i.k}/p_{..k})\,(p_{.jk}/p_{..k})$$

dır. Buna göre koşullu modelin bağımsızlığı yukarıdaki eşitlikler dikkate alınarak yeniden yazıldığında,

$p_{ijk}=p_{i.k}\ p_{.jk}\ /\ p_{..k} \quad (p_{..k}>0)$

elde edilir. Bu koşullu bağımsızlık hipoteziyle satır ve sütunlar tabakalara bağımlı ya da bağımlı değildir sonucuna varılır. Misal olarak satır ve sütunlar tabakalar bağımsız ise her bir tabaka için

$P(\text{satır}=i, \text{sütun}=j \mid \text{tabaka}=k)=p_{i..}p_{.j.}$

olur ki buda satır ve sütunun tabakaya bağımlı olmadığını gösterir. Eğer satırlar sütun ve tabakalardan bağımsız ise $H_0^{(1)}$ modeli elde edilirki, koşullu olasılık olarak aşağıdaki gibi gösterilebilir:

$P(\text{satır}=i, \text{sütun}=j \mid \text{tabaka}=k)=p_{i..}(p_{.jk}/p_{..k})$

Sütun olasılığı tabakaya bağımlı olup satır olasılıkları bağımlı değildir. Benzer şekilde satır ve tabakalar sütunlardan bağımsız ise, $H_0^{(2)}$ modeli uygun olup koşullu olasılık olarak belirtilirse,

$P(\text{satır}=i, \text{sütun}=j \mid \text{tabaka}=k)=(p_{i.k}/p_{..k})p_{.j.}$

elde edilir. Üç yönlü tablolarda değişkenlerden 1'nin verildiğinde diğer iki değişkenin koşullu bağımsızlığında, üç koşullu bağımsızlık hipotezi sözkonusudur. Tabaka sınıflamasının verilmesi durumunda, satır ve sütun sınıflamasının koşullu bağımsızlığı ile ilgili hipotez;

$H_0^{(4)}: p_{ijk} = p_{i.k}p_{.jk}/p_{..k}$

olur. Sütun sınıflamasının verilmesi durumunda satır ve tabaka sınıflamasının koşullu bağımsızlığı ile ilgili hipotez;

$H_0^{(5)}: p_{ijk} = p_{ij.}p_{.jk}/p_{.j.}$

şeklindedir. Satır sınıflamasının verilmesi durumunda sütun ve tabaka sınıflamasının koşullu bağımsızlığına ait hipotez ise

$H_0^{(6)}: p_{ijk} = p_{ij.}p_{i.k}/p_{i..}$

dır. Daha önce verilen yöntemlere benzer şekilde bu hipotezler de test edilebilir. İşlemler $H_0^{(4)}$ hipotezi üzerinde açıklanırsa;

$$\begin{aligned} p_{ijk} &= \hat{p}_{i.k}\hat{p}_{.jk}/\hat{p}_{..k} \\ &=(n_{i.k}/n_{...})(n_{.jk}/n_{...})/(n_{..k}/n_{...}) \\ &=n_{i.k}n_{.jk}/n_{..k}n_{...} \end{aligned}$$

bulunur. Beklenen frekanslar ise,

$E_{ijk} = n_{..}p_{ijk}$

$$E_{ijk} = n_{...}\hat{p}_{i.k}\hat{p}_{.jk}/\hat{p}_{..k}$$
$$=n_{i.k}n_{.jk}/n_{..k}$$

şeklinde elde edilir. Buradan elde edilen beklenen frekanslar kullanılarak, daha önce verilen χ^2 istatistiği ile test edilir. Bu durumda serbestlik derecesi s.d.=(r-1)(c-1)l olur.

$H^{(4)}_0$ hipotezinin ilgili marjinal toplamlarıyla ilgili kısıtlamaları $E_{i.k}=n_{i.k}$, $E_{.jk}n_{.jk}$,$E_{i..}=n_{i..}$, $E_{.j.}=n_{.j.}$ ve $E_{..k}=n_{..k}$ şeklindedir.

2.4. İkinci Derece İlişkinin Olmadığı Hipotez

Üç yönlü tablolarda değişkenler arasında daha karmaşık ilişkilerin tesbitinden ziyade ilişkinin var olduğunu bilmek yeterlidir. Misal olarak; üç yönlü tablolarda değişkenlerin ikisi arasında ilişki olması durumunda, bu değişkenlerle üçüncü değişkenin farklı sınıfları arasındaki ilişkinin derecesinde farklılıklar olabilir. Sonuç olarak üç değişkenli ilişkinin işleme alınmasına gerek vardır. Böyle ilişkiler ikinci derece ilişkiler olarak tanımlanır ve değişken çiftleri arasındaki birinci derece ilişkilerden çok farklıdır. Misal olarak 2x2x2 yönlü tablolarda ikinci derece ilişkinin olmadığına dair hipotez aşığıdaki gibi gösterilebilir (Ku et. al., 1971)

$$H_o : \frac{p_{111}p_{221}}{p_{121}p_{211}} = \frac{p_{112}p_{222}}{p_{122}p_{212}},$$

Genel olarak p_{ijk} olasılıklarının tanımlandığı üç yönlü tablolardaki değişkenler arasında ikinci derece ilişkinin olmadığına ait hipotez şu şekilde verilebilir (Roy and Kastenbaum, 1956):

$$H_0^{(7)} : \frac{p_{rcl}p_{ijl}}{p_{icl}p_{rjl}} = \frac{p_{rck}p_{ijk}}{p_{ick}p_{rjk}},$$

i=1,K,r-1; j=1,K,c-1; k=1,K,l-1. Özellikle bu eşitliğin sol tarafı, üçüncü değişkenin l'inci sınıfında diğer iki değişken arasındaki ilişkinin ve eşitliğin sağ tarafı, üçüncü değişkenin k'ıncı sınıfındaki diğer iki değişken arasındaki aynı ilişkinin ölçülmesi olarak verilir. Bu ilişkinin ölçülmesinden sonra üçüncü değişkenin bütün sınıfları için aynı işlem yapılır.

Burada, bu hipotezin nasıl ve ne zaman test edileceği gibi zorluklar sözkonusudur. Çünkü bu hipotez altında beklenen frekansların elde edilmesi için daha öncede anlatıldığı gibi marjinal toplamlar direkt olarak bulunamaz.

İkinci derece ilişkinin olmadığı hipotezin testinde beklenen frekanslar gözlenmiş frekansların marjinal toplamlarından doğrudan doğruya hesaplanamayabilir. En çok olabilirlilik eşitlikleri de kesin çözüm değildir. Sonuç olarak beklenen frekanslar başka yollarla bulunmak zorundadır. Beklenen frekansların doğrudan doğruya hesaplanamadığı bu durumda tekrarlı beklenen frekanslar hesaplama metodu kullanılır. Bu durumda tek değişkenli beklenen marjinal toplamların gözlenmiş marjinal toplamlara eşitliği kısıtlamaları yanında $E_{ij.}=n_{ij.}$, $E_{i.k}=n_{i.k}$ ve $E_{.jk}=n_{.jk}$ kısıtlamaları da sözkonusudur. Tekrarlı beklenen frekansları hesaplama metoduna, E_{ijk} beklenen frekanslara başlangıç değerlerinin atanmasıyla başlanır. E_{ijk} 'nin herbiri için başlangıç değeri $E^{(0)}_{ijk}=1$'dir (Everitt, 1994).

Tekrar m (m=1,2,....);

Adım1;

$$E^{(1)}_{ijk} = \frac{E^{(0)}_{ijk}\, n_{ij.}}{E^{(0)}_{ij.}}.$$

Burada kısıtlama $E^{(1)}_{ij.}=n_{ij.}$ 'dir.

Adım 2;

$$E^{(2)}_{ijk} = \frac{E^{(1)}_{ijk}\, n_{i.k}}{E^{(1)}_{i.k}}.$$

Adım 2'deki kısıtlama $E^{(2)}_{i.k}=n_{i.k}$ şeklindedir.

Adım 3;

$$E^{(3)}_{ijk} = \frac{E^{(2)}_{ijk}\, n_{.jk}}{E^{(2)}_{.jk}}.$$

şeklinde olup ilgili kısıtlama $E^{(3)}_{.jk}=n_{.jk}$'dır.

Adım 3'de $E^{(0)}_{ijk}=E^{(3)}_{ijk}$ şeklinde alınarak, yukarıda algoritması verilen tekrarlı beklenen frekansları hesaplama metodunun **tekrar m=2; adım 1**'e geçilerek işlemlere devam edilir. Tekrar m ve m-1 sonucunda hesaplanan beklenen frekanslar arasındaki farklılık çok küçük bir değerden daha az olana kadar işlemler tekrarlanır.

Üç yönlü tablolarda ikinci derece ilişkinin olmadığı şeklindeki hipotezin test edilmesi için hesaplanması gereken parametre sayısı

$$1+(r-1)+(c-1)+(l-1)+(r-1)(c-1)+(r-1)(l-1)+(c-1)(l-1)$$

kadardır. Serbestlik derecesi ise

$$s.d.=rcl-\{1+(r-1)+(c-1)+(l-1)+(r-1)(c-1)+(r-1)(l-1)+(c-1)(l-1)\} \text{ dır.}$$

3. UYGULAMA

Bu çalışmada uygulama materyali olarak; Başbakanlık Devlet İstatistik Enstitüsü'nün (1997) derleyip yayınladığı, toplumun sosyal ve ekonomik yapısının göstergelerinden biri olan intihar istatistikleriyle ilgili veriler kullanılmıştır.

İntihar istatistikleriyle ilgili veri seti (Tablo 2); cinsiyet (C, k=1,2), intihar şekli (Ş, i=1,2,...,6) ve intihar nedeni (N, j=1,2,...,8) değişkenlerinden oluşmaktadır. 6x8x2 yönlü tablo'da oldukca küçük frekansların ve sıfırlı (boş) gözlerin bulunması nedeniyle ticari başarısızlık, diğer ve bilinmeyen sütunları tek bir sütun altında birleştirilerek oluşturulan 6x6x2 yönlü tablo Tablo.3'de verilmiştir.

Üç yönlü tablo için 1 karşılıklı bağımsızlık, 3 kısmi bağımsızlık, 3 koşullu bağımsızlık ve 1 ikinci derece ilişki hipotezleri test edilmiştir. Hipotezlerin testinde χ^2 istatistiği kullanılmış olup hipotezlere karşılık istatistik değerleri Tablo 4'de verilmiştir. Verilerin analizinde ilgili hipotezlere göre teorik olarak beklenen frekanslar ve χ^2 istatistiği, bu amaç için geliştirilen bir FORTRAN programı ile hesaplanmıştır (Akyol,1999). İkinci derece ilişkinin istatistik olarak önemli olmadığı hipotezinin geçerliliğinde beklenen frekansların hesabı, tekrarlı beklenen frekanslar algoritması yardımıyla yapılmıştır.

İntihar şekli, intihar nedeni ve cinsiyet birbirinden bağımsızdır şeklinde kurulan hipotezin ($H^{(0)}{}_0: p_{ijk}=p_{i..}p_{.j.}p_{..k}$), yani birinci ve ikinci derece ilişkilerin sıfır olduğu hipotezine göre hesaplanan teorik olarak beklenen frekanslar ve hipotez testi için gerekli χ^2 istatistiği hesaplanmıştır. Buna göre karşılıklı bağımsızlık ya da birinci ve ikinci derece ilişkilerin sıfır olduğuna ilişkin hipotez red edilmektedir ($p<0.01$). Bu durumda intihar şekli, intihar nedeni ve cinsiyet değişkenlerinden en az ikisi birbirine bağımlıdır. Yani değişkenler arasında birinci ya da birinci ve ikinci derece ilişkilerin varlığından söz edilebilir. Bu durumu ayrıntılı olarak irdelemek amacıyla kısmi bağımsızlık, koşullu bağımsızlık ve ikinci derece ilişkinin istatistik olarak önemli olmadığına ilişkin hipotez testlerinin yapılmasına gerek vardır.

İntihar nedeni ve cinsiyet sınıflamasının, intihar şekli sınıflamasından bağımsızlığı ile ilgili kısmi bağımsızlık hipotezinin ($H^{(1)}{}_0: p_{ijk}=p_{i..}\,p_{.jk}$) testi sonucunda, intihar nedeni ve cinsiyet sınıflamasının intihar şekli sınıflamasından bağımsız olmadığı hükmüne varılmıştır ($p<0.01$). Diğer bir deyişle $H_0{}^{(1)}$ hipotezinde sıfır olduğu ileri sürülen ŞxN, ŞxC ve ŞxNxC ilişki terimlerinden en az biri istatistiki olarak önemlidir. Kısmi bağımsızlık hipotezlerinden bir diğeri, intihar şekli ve cinsiyet sınıflamasının intihar nedeni sınıflamasından bağımsızlığı hipotezi ($H^{(2)}{}_0: p_{ijk}=p_{.j.}\,p_{i.k}$) olup, yapılan hipotez testi sonucunda $H^{(2)}{}_0$ hipotezi red edilmiştir ($p<0.01$). Böylelikle Ş ve C sınıflamasının N sınıflamasından bağımsız olmadığı hükmüne varılmıştır. Diğer bir deyişle H_0 hipotezinde sıfır olduğu ileri sürülen ŞxN, NxC ve ŞxNxC ilişki terimlerinden en az biri istatistiki olarak önemlidir. Kısmi bağımsızlık hipotezlerin sonuncusu olan Ş ve N sınıflamasının C sınıflamasından bağımsızlığı ile ilgili hipotez $H^{(3)}{}_0: p_{ijk}=p_{..k}\,p_{ij.}$ dir. Buna göre bu hipotez de red edilmiştir ($p<0.01$). Bu durumda Ş ve N sınıflaması C sınıflamasına bağımlıdır. Bu da ŞxC, NxC ve ŞxNxC ilişki terimlerinden en az biri istatistiki olarak önemlidir.

Tablo 2. Veri materyali

		İNTİHAR NEDENİ (j)								
İNTİHAR ŞEKLİ (i)	CİNSİYET (k)	Hastalık	Aile geçimsizliği	Geçim Zorluğu	Ticari başarısızlık	Hissi ilişkiler	Öğrenim başarısızlığı	Diğer	Bilinmeyen	Toplam
Kendini	Erkek	180	114	96	29	31	25	14	8	
Asarak	Kadın	98	74	25	0	34	6	14	1	749
Kimyevi madde	Erkek	16	16	9	3	6	1	0	0	
Kullanarak	Kadın	36	45	5	0	20	8	2	1	168
Kendini yüksekten	Erkek	68	19	15	1	2	1	4	1	
Atarak	Kadın	46	17	2	1	9	4	4	0	194
Kendini suya	Erkek	4	6	5	1	2	1	0	1	
Atarak	Kadın	7	3	2	0	3	1	1	0	37
Ateşli silah	Erkek	47	52	26	12	27	6	8	2	
Kullanarak	Kadın	11	24	3	0	17	8	4	0	247
Diğer	Erkek	17	15	8	1	4	2	1	0	
	Kadın	5	2	5	0	2	1	1	1	65
Toplam		535	387	201	48	157	64	53	15	1460

Tablo 3. Düzenlenmiş 6x6x2 yönlü veri tablosu

		İNTİHAR NEDENİ (j)						
İNTİHAR ŞEKLİ (i)	CİNSİYET (k)	Hastalık	Aile geçimsizliği	Geçim zorluğu	Hissi ilişkiler	Öğrenim Başarısızlığı	Diğer + Ticari başarısızlık + Bilinmeyen	Toplam
Kendini	Erkek	180	114	96	31	25	51	
Asarak	Kadın	98 .	74	25	34	6	15	749
Kimyevi madde	Erkek	16	16	9	6	1	3	
Kullanarak	Kadın	36	45	5	20	8	3	168
Kendini yüksekten	Erkek	68	19	15	2	1	6	
Atarak	Kadın	46	17	2	9	4	5	194
Kendini suya	Erkek	4	6	5	2	1	2	
Atarak	Kadın	7	3	2	3	1	1	37
Ateşli silah	Erkek	47	52	26	27	6	22	
Kullanarak	Kadın	11	24	3	17	8	4	247
Diğer	Erkek	17	15	8	4	2	2	
	Kadın	5	2	5	2	1	2	65
Toplam		535	387	201	157	64	116	1460

Tablo 4. Test sonuçları

Hipotez	χ^2	S.D.	Hüküm
$H^{(0)}_0: p_{ijk}=p_{i..}p_{.j.}p_{..k}$	284.83**	60	C, Ş ve N değişkenlerden en az ikisi birbirlerine bağımlıdır
$H^{(1)}_0: p_{ijk}=p_{i..}\ p_{.jk}$	225.64**	55	N ve C sınıflaması Ş sınıflamasına bağlıdır
$H^{(2)}_0: p_{ijk}=p_{.j.}\ p_{i.k}$	177.11**	55	Ş ve C sınıflaması N sınıflamasına bağlıdır
$H^{(3)}_0: p_{ijk}=p_{..k}\ p_{ij.}$	174.17**	35	N ve Ş sınıflaması C sınıflamasına bağlıdır
$H^{(4)}_0: p_{ijk}=p_{i.k}\ p_{.jk}/p_{..k}$	130.70**	50	C verildiğinde N ve Ş birbirine bağlıdır
$H^{(5)}_0: p_{ijk}=p_{ij.}\ p_{.jk}/p_{.j.}$	120.53**	30	N verildiğinde Ş ve C birbirine bağlıdır
$H^{(6)}_0: p_{ijk}=p_{ij.}\ p_{j.k}/p_{i..}$	82.91**	30	Ş verildiğinde C ve N birbirine bağlıdır
$H^{(7)}_0: \dfrac{p_{rcl}p_{ijl}}{p_{icl}p_{rjl}} = \dfrac{p_{rck}p_{ijk}}{p_{ick}p_{rjk}}$	36.83	25	İkinci derece ilişki söz konusu değildir

** ($p<0.01$)

Mümkün olan tüm kısmi hipotezler red edilmiştir. Bu durumda bir değişken verildiğinde, bu değişkenle diğer iki değişken bağımlı olsa dahi bu diğer iki değişkenin birbirinden bağımsızlığı ile ilgili koşullu bağımsızlık hipotez testlerinin yapılması gerekir. Bu durumda üç adet koşullu bağımsızlık hipotezi sözkonusudur. Bu hipotezlerin ilki, C verildiğinde Ş ve N sınıflamasının bağımsızlığıdır. Bu durumdaki hipotez $H^{(4)}_0: p_{ijk}=p_{i.k}\ p_{.jk}/p_{..k}$ dir. Bu durumda koşullu hipotez red edilmektedir ($p<0.01$). Yani ŞxN ve ŞxNxC ilişkilerinin en az birisi istatistik olarak önemlidir. Koşullu hipotezlerden bir diğeri olan İntihar nedeni verildiğinde, intihar şekli ve cinsiyetin bağımsızlığı ile ilgili koşullu bağımsızlık hipotezi $H^{(5)}_0: p_{ijk}=p_{ij.}\ p_{.jk}/p_{.j.}$ dir. Bu durumda da $H_0^{(5)}$ hipotezi de red edilmektedir ($p<0.01$). Yani ŞxC ve ŞxNxC ilişkilerinin en az birisi istatistik olarak önemlidir. Üçünücü ve son koşullu hipotez ise intihar şekli verildiğinde, intihar nedeni ve cinsiyet birbirinden bağımsızlığı ile ilgili hipotez $H^{(6)}_0: p_{ijk}=p_{ij.}\ p_{j.k}/p_{i..}$ dir. İlgili $H_0^{(6)}$ hipotezi red edilmektedir ($p<0.01$). Bu da NxC ve ŞxNxC ilişkilerinin en az birisi istatistik olarak önemli olduğu anlamındadır.

Yapılan hesaplamalar sonunda karşılıklı bağımsızlık, kısmi bağımsızlık ve koşullu bağımsızlık hipotezleri red edilmiştir. Bu durumda kurulabilecek hipotez ikinci derece ilişkinin istatistik olarak önemli olmadığına dair kurulan hipotezdir.

Bu hipotezin geçerliliğinde beklenen frekansların hesabı, tekrarlı bekelenen frekanslar algoritması yardımıyla yapılmıştır. Bu durumda ilgili $H_0^{(7)}$ hipotezi $\alpha=0.05$ anlamlılık düzeyinde kabul edilmektedir. Yani değişkenler arasında % 95 güvenle ikinci derece ilişki sözkonusu olmayıp, tüm birinci derece ilişkilerin önemli olduğu sonucu varılmıştır.

Tablo 3'e bakıldığında, kişiler sırasıyla hastalık, aile geçimsizliği, geçim zorluğu, hissi ilişkiler, diğer+ticari başarısızlık+bilinmeyen ve öğrenim başarızıslığı nedeniyle intihar etmektedirler.

İntihar edenlerden de ençoktan en aza doğru sırasıyla, kendini asarak, ateşli silah kullanarak, kendini yüksekten atarak, kimyevi madde kullanarak, kendini suya atarak ve diğer intihar şekillerini kullanarak intihar etmektedirler.

Cinsiyet dikkate alındığında ise, intihar edenlerinin %62'lik kısmı erkeklerden oluşmaktadır.

İntihar şekli ile intihar nedeni arasında bir ilişkinin varlığının sözkonusu olduğu bulunmuştu. İntihar nedenine göre kişilerin bazı intihar şekillerini tercih etmektedirler. Hastalık nedeniyle kişilerin çoğunluğu kendini asarak ve kendini yüksekten atarak intihar etmektedirler. Aile geçimsizliği ve geçim zorluğu içinde olan kişilerin büyük bir çoğunluğu kendini asarak intihar etmektedirler. Hissi ilişkiler, öğrenim başarısızlığı ve ticari başarısızlık+diğer+bilinmeyen şeklinde sorunları olanların büyük bir çoğunluğu kendini asarak ve ateşli silah kullanar intihar etmektedirler.

Yapılan istatistik analizi sonucunda intihar şekli ile cinsiyet arasında da bir ilişki bulunmuştur. Buna göre erkeklerde intihar şekilleri sırasıyla kendini asarak, Ateşli silah kullanarak, kendini yüksekten atarak, kimyevi madde kullanarak, diğer, kendini suya atarak şeklindedir. Kadınlar da ise kendini asarak, kimyevi madde kullarak, kendini yüksekten atarak, ateşli silah kullanarak, kendini suya atarak ve diğer şeklindedir. Her iki cinsiyet grubu ilk sırada kendini asarak intihar etmektedirler. İkinci sırada ise erkek ve kadınların özelliklerini yansıtır intihar şekilleri tercih edilmektedir. Erkekler ateşli silah kullanarak ve kadınlar ise kimyevi madde kulllanarak intihar etmektedirler.

İntihar nedeni ile cinsiyet arasında da bir ilişki bulunmuştur. Her iki cinsiyet grubu ilk iki sırada hastalık ve aile geçimsizliğ nedeniyle intihar etmektedirler. Daha sonraki sıralarda ise farklılık görülmektedir.

Hasta, aile geçimsizliği ve geçim zorluğu içinde olan kimseler kendilerini asma eğilimli olduklarından , bu kişilerin kendileri asmaları için etrafında ilgili materyallerin bulundurulmaması gerekir.

Bu çeşit irdemelerde toplum yararına sonuçlar çıkarılabilinmesi için konu ile ilgili psikolojik danışmanlık uzmanları, adli tıp uzmanları, sosyologlar ve din bilimcileriyle işbirliği yapılmalı, uygun yorumlamalar yapılıp, tedbirler tartışılmalıdır.

KAYNAKLAR

AKYOL, M. (1999), *Çok Yönlü Tablolarda İstatistiksel Analizler. Ankara Üniversitesi*, Fen Bilimleri Enstitüsü , Doktora Tezi. Ankara.

BAŞBAKANLİK DEVLET İSTATISTIK ENSTITÜSÜ (1997), *Türkiye İstatistik Yıllığı.* Yayın no: 2110 Ankara.

CHRİSTENSEN, R. (1990), *Log-Linear Models. Department of Mathematics and Statistics*, University of New Mexico, Albuquerque, NM 87131, USA.

EVERİTT, B.S. (1977), *The Analysis of Contingency Tables. Monographs Applied Probability and Statistics*, A halted Press Book John Wiley & Sons Inc., New York.

EVERİTT, B.S. (1994), *The Analysis of Contingency Tables. Monographs Applied Probability and Statistics 45*, Chapman & Hall, New York.

KU, H.H., RUTH, N. V. and KULLBACK, S. (1971), *Analysis of Multidimensional Contingency Tables,* Journal of the American Statistical Association, Volume:66, Pages:55-64.

ROY, S. N. and. KASTENBAUM, M.A. (1956), *On Hypothesis of no Interaction in a Multiway Contingency Table. Ann. Math.Statist,* Volume: 27, Pages: 749-751.

χ^2 Statistical Analysis Used of Three-Dimensional Table

ABSTRACT

Probability distributions used for hypothesis tests, changes in accordance with variables by being continuous or discontinuous. In every case, it could not be possible to find continuous variables or to express variables in continuous form. In such cases, chi-square test is applied by classifying characteristic according to their categories and calculating number of units in each cluster. In this study the differences of χ^2 analyze used for two-dimensional tables obtained for the case of having three characteristics.

For the application, three-dimensional data set is used and χ^2 analyze is applied. This data set is formed by variables such as, sex(k=1,2), type of suicide (i=1,2,...,6) and reason of suicide (j=1,2,...,8). At the end of the analysis, it is found out that all of the first degree relationships between variables are important.

Key words: *Two-dimensional tables, Three-dimensional tables, Multidimensional tables, The analysis of χ^2.*