

Çok yönlü frekans tablolarının analizi üzerine bir çalışma

Hülya OLMUŞ*

Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 06500, Teknikokullar, Ankara

Özet

Çok yönlü frekans tablolarının analizi, iki veya daha fazla düzeyi olan üç veya daha fazla kesikli bağımsız değişkenler arasındaki ilişkileri değerlendirmek amacıyla kullanılan parametrik olmayan bir testtir. Bu değişkenler, sınıflama, nicel veya kategorik olabilir. Çok yönlü frekans analizi tablosunda, bağımlı değişken, bir ya da daha fazla kesikli değişkenler ve ilişkileri tarafından etkilenen hücre frekansıdır. Bu analiz, değişkenlerden biri bağımlı değişken olduğu zaman kullanılır. Bu durumda, temel etkiler ve etkileşimler test edilecektir. Bu analizde, ilk olarak etkiler tanımlanacak ve böylece tüm etkilerin her bir düzeyi için parametre tahminleri elde edilecektir. Eğer, kesikli ve sürekli değişkenlerin bir karışımı kullanılmak istenirse, genellikle Lojistik regresyon seçilir. Çok yönlü frekans analizi, Log-lineer analiz ve Lojit modellerin bir çeşididir. Bu üç analiz çeşidi, uyum iyiliği için Ki-karenin bir uzantısıdır. Elimizde var olan frekans verisi için hesaplama türleri, en fazla iki boyutlu olumsuzluk tabloları için uygundur. Böyle iki boyutlu tablolar, uyum iyiliği yaklaşımı ile analiz edilir. Çok yönlü frekans analizinde, Pearson Ki-kare yerine olabilirlik oran G^2 istatistiği kullanılacaktır çünkü, Pearson Ki-kare istatistiği toplamsal değil iken, iç-içe modeller için bu test toplamsaldır. Ayrıca, bu çalışmada kısmi birliktelik testi, modelde her bir bireysel etkinin anlamlılığını test etmek için kullanılır. Bu çalışmanın amacı, kesikli değişkenler arasında bir ilişki olup olmadığını araştırmak, en iyi modeli belirlemek ve uygun modele ilişkin parametre tahminlerini elde etmektir. Bunun için, Ankara ilinde devlet tiyatrolarının izleyici profilini çıkartmak amacıyla yapılmış bir anket çalışmasından yararlanılmıştır.

Anahtar kelimeler: Çok yönlü frekans analizi, olabilirlik oran G^2 istatistiği, kısmi birliktelik testi.

*Yazışmaların yapılacağı yazar: Hülya OLMUŞ. hulya@gazi.edu.tr; Tel: (312) 212 60 30.

Makale metni 01.06.2005 tarihinde dergiye ulaştı, 19.10.2006 tarihinde basım kararı alınmıştır. Makale ile ilgili tartışmalar 30.06.2007 tarihine kadar dergiye gönderilmelidir.

A study on analysis of multiway frequency tables

Extended abstract

Multiway frequency analysis is a nonparametric test that can be used to evaluate relations among three or more discrete independent variables with two or more levels. These variables may be nominal or categorical as qualitative. In a multiway frequency analysis table, cell frequency is the dependent variable that is influenced by one or more discrete variables and their associations. Multiway frequency analysis can be used when one of the variables is a dependent variable. In that case, main effects and interactions can be tested. In this analysis, effects are identified first and then parameter estimates are found for each level of all the effects. If the researcher wishes to use a mix of continuous and discrete variables, Logistic regression is usually the method of choice. Multiway frequency analysis is a version of Loglinear analysis and Logit models. These three analysis version are extensions of the chi-square for goodness-of-fit. The computational techniques for handling frequency data are appropriate for contingency tables of at most two dimensions. Such two-dimensional tables are analysed with the goodness-of-fit approach. In multiway frequency analysis, we use the Likelihood ratio G^2 statistic instead of the Pearson Chi-square because it is additive for nested models, whereas the Pearson statistic, in general, is not. The goal of this study is to discover whether there is an association among discrete variables, to determine the best model and to obtain parameter estimates of the derived model. Public survey which has been done for the evaluation of the profile of governmental theatre attendees in Ankara were used for this purpose.

Likelihood ratio G^2 statistic has the desirable property that it is additive, meaning the sum of the chi-square values for the individual effects in the model equals the chi-square for the total model. Therefore, if one considers the difference between two Likelihood chi-square statistics for related models, the result would be another Likelihood ratio G^2 statistic. This property enables one to make two important inferences:

- 1) nested models can be compared*
- 2) individual effects may be assessed*

Prior to proceeding with model selection, a special case of the Chi-square test, Likelihood ratio G^2 sta-

tistic is performed. With g_f as the observed frequency and B_f as the expected frequency for each of k cells, the Pearson chi-square is computed as:

$$\chi^2 = \sum_{f=1}^k \frac{(B_f - g_f)^2}{g_f}$$

Large differences yields large values of the statistics and more evidence that the model is inadequate. The Likelihood ratio G^2 statistic is computed by the following equation.

$$G^2 = 2 \sum g_f \ln(g_f / B_f)$$

G^2 was developed by Fisher (1924), based on his earlier work on maximum likelihood theory. As with X^2 , G^2 is distributed as X^2 for sufficiently large N .

In this study, before testing and selecting a model that best fits the observed data, screening is normally carried out to see if there are any significant effects to investigate. The researcher may be specifically interested in finding the best model is developed where an additive regression type equation is written for expected frequency as a function of the effects in the design. The procedure is similar to multiple regressions except that cell frequencies are predicted based on the combined effects of independent variables. The modelling process begins with all possible associations among the independent variables. That is, if there are three independent variable, it will start off with all of the one-, two-, and three way associations. This is called a full or saturated model, because it includes all possible effects. It has the same amount of cells in the frequency table as it does effects, so the expected cell frequencies will always exactly match the observed frequencies.

The statistic normally used to tell you how well the model of expected frequencies fits the observed frequencies is the Likelihood ratio G^2 statistic. In this study, partial associations test allow us to test the significance of each individual effect in the model. However, the likelihood ratio G^2 statistic can be used to compare a more complicated model with a simpler model which has one interaction or main effect dropped to assess the importance of that term.

Keywords: *Multiway frequency analysis, likelihood ratio G^2 statistic, partial association test.*

Giriş

İki kesikli değişken arasındaki bir ilişki, ki-kare (χ^2) testi ile araştırılır. Bu iki kesikli değişkene ek olarak üçüncü bir değişken ilave edilir ise, iki ve üç yönlü ilişkiler, çok yönlü frekans analizi ile incelenir. Sonuç olarak, üç veya daha fazla kesikli değişkenler (nicel, kategorik, sınıflama) arasındaki ilişkiler çok yönlü frekans analizi süresince araştırılacaktır. Çok yönlü frekans analizini yapabilmek için, tablolar, tek yönlü, iki yönlü, üç yönlü... tabloları içerecek biçimde düzenlenmelidir. Bu analiz, olabilirlik oran test istatistiği G^2 'yi kullanır. G^2 , Pearson ki-kare istatistiğine (χ^2) karşı bir alternatiftir. G^2 , yeteri kadar büyük N'ler için χ^2 olarak dağılır. Bu nedenle, değişkenler arasındaki ilişkilerin anlamlılığını değerlendirmek için, χ^2 tabloları kullanılır. Ayrıca, G^2 , etkilerin toplanabilirlik özelliğine sahiptir. Örneğin, A ve B değişkenlerini içeren bir iki-yönlü analizde,

$$G^2_{\text{toplam}} = G^2_A + G^2_B + G^2_{AB} \quad (1)$$

dır. G^2_{toplam} (G^2), iki yönlü tablodaki tüm ilişkilerin testini, G^2_A ve G^2_B , temel etkilerin uyum iyiliği testlerini ve G^2_{AB} ise A ve B arasındaki birlikteliğin testini gösterir. G^2 olabilirlik oran istatistiği aşağıdaki gibi tanımlanacaktır.

$$G^2 = 2 \sum (g_f) \ln(g_f/B_f) \quad (2)$$

(2) eşitliğinde, g_f , tablonun her bir hücresinde yer alan gözlenmiş frekansları; B_f , her bir hücredeki beklenen frekansları gösterir (Freeman, 1987; Vokey, 2002).

Çok yönlü frekans analizinin gösterimi için bir uygulama

Bu çalışmada, ele alınan veri Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü araştırma projesi öğrencilerinin 627 tiyatro izleyicisi arasında yapmış olduğu anket çalışmasından alınmıştır. Bu araştırmanın amacı, Ankara ilindeki devlet tiyatrolarının izleyici profilinin çı-

kartılmasıdır. Bu çalışmada, sadece üç kesikli değişken ele alınmıştır. Bu değişkenler ve değişkenlerin aldığı düzeyler şöyledir:

Medeni durum (M) : bekar, evli
Cinsiyet (C) : erkek, kadın
Tiyatroya gidenler arasında sinemayı tercih etme durumu (S) : 1. sırada, 2. sırada, ..., 5. sırada tercih etme durumu

Ankara ilindeki devlet tiyatrolarının izleyici profiline ait olan veri tablosu aşağıda verilmiştir (Tablo 1).

Tablo 1. Ankara ilindeki devlet tiyatrolarının izleyici profiline ait gözlenmiş frekanslar

Medeni durum	Cinsiyet	Tiyatroya gidenler arasında sinemayı tercih etme durumu					Toplam
		1	2	3	4	5	
Bekar	Erkek	75	37	8	6	7	133
	Kadın	103	93	31	12	7	246
	Toplam	178	130	39	18	14	379
Evli	Erkek	38	43	8	8	5	102
	Kadın	52	56	23	6	9	146
	Toplam	90	99	31	14	14	248

Çok yönlü frekans tablolarının analizi üç adımdan oluşur. Bu adımlar şunlardır.

- Eleme
- Uygun modelin seçimi ve testi
- Seçilen modelin değerlendirilmesi ve yorumlanması

Tablo 1'deki veriler ele alınarak, bu adımlar incelenmiştir.

Etkilerin elenme aşaması

Gözlenmiş verinin en iyi uyduğu modelin seçimi ve test edilme aşamasına geçilmeden önce, hangi etkilerin önemli olup olmadıkları araştırılmalıdır. Model, tüm mümkün etkileri içeriyor ise, modele "tam model" denir. Eğer, modelde bazı etkiler yer alıp, diğer etkiler yer almıyorsa, böyle modele "tamamlanmamış model" denir. Eğer, araştırmacı belli ilişkilerin öncelikli olarak modelden çıkarılmasını gerekli görmüyorsa, tam bir modelle başlar ve bu model üzerinde modele

katkısı en az olan etkileri sonradan eleme yolu ile çıkarır (O'Leary vd., 2005).

Toplam etkinin testi

Hiçbir etkinin var olmadığı kabul edilirse, beklenen frekanslar (B_f) her bir hücre için aynı olur. Toplam etkiyi (G^2) test etmek için gerekli olan beklenen frekanslar, örneklem büyüklüğünü tabloda yer alan hücre sayısına bölerek elde edilir. Buna göre,

$$B_f = n/mcs \quad (3)$$

dır. Eşitlik 3'te yer alan terimler aşağıda verilmiştir.

- n : Örneklem büyüklüğü
 m : Medeni durumun düzey sayısı
 c : Cinsiyet değişkeninin düzey sayısı
 s : Tiyatroya gidenler arasında sinemayı tercih etme durumunun düzey sayısı

Tablo 1'de verilen veriler için,

$$B_f = 627/(2)(2)(5) = 31.35 \quad (4)$$

olarak elde edilir. Toplam etkiyi test etmek için, Eşitlik 2 ve Tablo 1'deki veriler ve hücrelerin her biri ele alınarak, G^2 ve serbestlik derecesi (sd) aşağıdaki gibi elde edilmiştir.

$$G^2 = 2 \left[75 \ln \left(\frac{75}{31.35} \right) + \dots + 9 \ln \left(\frac{9}{31.35} \right) \right] \quad (5)$$

$$sd = (2)(2)(5) - 1 = 19$$

$\alpha = 0.05$ anlamlılık düzeyinde, 19 serbestlik dereceli χ^2 değeri 30.1435'tir. Bu değer, elde edilen G^2 ile karşılaştırıldığında, cinsiyet, medeni durum ve tiyatroya gidenler arasında sinemayı tercih edenlerin oluşturduğu toplam 20 hücre için eşit beklenen frekanslardan ayrılışlarının istatistiksel olarak önemli olduğuna karar verilir.

Temel etkilerin testi

Çalışmada yer alan üç kesikli değişkenin her birinin anlamlı olup olmadıklarını belirlemek için uyum iyiliği testi yapılmıştır.

Medeni durumun iki düzeyi için (bekar, evli), marjinal toplamalar ele alınarak, gözlenmiş frekanslar verilmiştir. Beklenen frekansları elde etmek için, örneklem büyüklüğü, ilgili hücrenin sayısına bölünerek elde edilir. Buna göre, medeni durumun düzey sayısı $m=2$ olduğundan,

$$B_f = 627/2 = 313.5 \quad (6)$$

elde edilmiştir. Böylece, elde edilen gözlenen ve beklenen frekanslar aşağıda verilmiştir.

g _r		B _f	
Bekar	Evli	Bekar	Evli
379	248	313.5	313.5

Medeni durum değişkeni için, uyum iyiliği testi ve serbestlik derecesi,

$$G_M^2 = 2 \left[379 \ln \left(\frac{379}{313.5} \right) + 248 \ln \left(\frac{248}{313.5} \right) \right] = 27.573 \quad (7)$$

$$sd = 2 - 1 = 1$$

ve cinsiyet değişkeni için,

$$G_C^2 = 39.734 \quad sd = 2 - 1 = 1 \quad (8)$$

elde edilir. $\alpha = 0.05$ anlamlılık düzeyinde, 1 serbestlik dereceli χ^2 değeri 3.84'dir. Sonuç olarak, tiyatro izleyicisi olan bekar (379) ve evli (248) olanların sayıları arasında ve tiyatro izleyicisi olan erkek (235) ve kadınların (392) sayıları arasında istatistiksel olarak anlamlı farklılık vardır. Sinemayı tercih etme durumu değişkeni için de,

$$G_S^2 = 429.902 \quad sd = 5 - 1 = 4 \quad (9)$$

elde edilmiştir. $\alpha = 0.05$ anlamlılık düzeyinde 4 serbestlik dereceli χ^2 değeri 9.48'dir. Sonuç olarak, tiyatroya gidenler arasında sinemayı tercih etme durumu değişkenine ait 1. düzeyi (268), 2. düzeyi (229), ..., 5. düzeyi (28) sayıları arasında da istatistiksel olarak anlamlı farklılık vardır.

Kısmi birlikteliklerin testi

Kısmi birliktelik testleri, tüm marjinal toplamların (test edilecek etki hariç), gözlenmiş marjinal frekanslara eşleştirildiği durumda, beklenen frekansların tam bir setini elde etmek için kullanılan bir ardışık hesaplama yöntemidir (Tabachnick ve Fidell, 2001; Wrigley, 2002).

Kısmi birliktelik testini açıklamak için, çalışmada sadece S×C birlikteliğinin hesapları açık olarak verilmektedir. Cinsiyet ve tiyatroya gidenler arasında sinemayı tercih etme durumunun meydana getirdiği her bir hücre, diğer değişken olan medeni durum (M) üzerinden toplanarak, gözlenmiş frekanslar aşağıdaki gibi elde edilir (Tablo 2).

Tablo 2. S×C birlikteliğinin testi için gözlenmiş frekanslar

Tiyatroya gidenler arasında sinemayı tercih etme durumu						
Cinsiyet	1	2	3	4	5	Toplam
Erkek	113	80	16	14	12	235
Kadın	155	149	54	18	16	392
Toplam	268	229	70	32	28	627

Beklenen frekanslar ise aşağıdaki tabloda verilmiştir (Tablo 3). Örneğin, ilk hücre için (sinemayı birinci sırada tercih eden erkekler) beklenen frekans,

$$B_f = (268)(235)/627 = 100.4466 \quad (10)$$

elde edilir.

Tablo 3 elde edildikten sonra, beklenen frekanslar diğer değişken olan medeni durumun her bir düzeyi için aynen yazılarak aşağıdaki tablo elde edilir (Tablo 4).

İki-yönlü tablo iki kere tekrarlandığından dolayı, örneklem büyüklüğü 627, 1254 olarak veya cinsiyeti erkek olanların sayısı da 235 yerine 470 olarak artmıştır. S×C birlikteliğinin testi için, başka bir iki-yönlü ilişki ele alınarak, Tablo 4'teki değerler üzerinde uyarlamalar yapılmış ve ikinci ardışık tahminleri elde edilmiştir. İkinci ardışık tahminleri elde edebilmek için, diğer bir ilişki olan S×M birlikteliği ele alınmıştır. S×M için gözlenmiş frekanslar ve marjinal toplamalar aşağıdaki tabloda verilmiştir (Tablo 5).

Tablo 3. S×C birlikteliğinin testi için beklenen frekanslar

Tiyatroya gidenler arasında sinemayı tercih etme durumu						
	1	2	3	4	5	Toplam
Erkek	100.4466	85.8293	26.2361	11.9936	10.4944	235
Kadın	167.5534	143.1707	43.7639	20.0064	17.5056	392
Toplam	268	229	70	32	28	627

Tablo 4. S×C birlikteliğinin kısmi testi için beklenen frekansların ilk ardışık tahminleri

Tiyatroya gidenler arasında sinemayı tercih etme durumu							
Medeni durum	Cinsiyet	1	2	3	4	5	Toplam
Bekar	Erkek	100.4466	85.8293	26.2361	11.9936	10.4944	235
	Kadın	167.5534	143.1707	43.7639	20.0064	17.5056	392
	Toplam	268	229	70	32	28	627
Evli	Erkek	100.4466	85.8293	26.2361	11.9936	10.4944	235
	Kadın	167.5534	143.1707	43.7639	20.0064	17.5056	392
	Toplam	268	229	70	32	28	627

Tablo 5. $S \times M$ için gözlenmiş frekanslar ve marjinal toplamlar

Medeni durum	Tiyatroya gidenler arasında sinemayı tercih etme durumu					Toplam
	1	2	3	4	5	
Bekar	178	130	39	18	14	379
Evli	90	99	31	14	14	248
Toplam	268	229	70	32	28	627

$S \times C$ birlikteliğinin kısmi testi için, ikinci ardışık tahminlerinin elde edilmesinde uygulanan işlem, sadece, sinemayı birinci sırada tercih eden bekar erkek ve kadın olanlar için gösterilmiştir. Burada:

- g_f : $S \times M$ için gözlenmiş frekansları
 M_T : $S \times M$ için marjinal toplamları
 $B_{f(\text{erkek})}^{\#1}$: Sinemayı birinci sırada tercih eden bekar erkekler için birinci ardışık tahminleri
 $B_{f(\text{kadın})}^{\#1}$: Sinemayı birinci sırada tercih eden bekar kadınlar için birinci ardışık tahminleri
 $B_{f(\text{erkek})}^{\#2}$: Sinemayı birinci sırada tercih eden bekar erkekler için ikinci ardışık tahminleri
 $B_{f(\text{kadın})}^{\#2}$: Sinemayı birinci sırada tercih eden bekar kadınlar için ikinci ardışık tahminleri
 g_f/M_T : Sinemayı birinci sırada tercih eden bekar erkek ve kadınlar için bir oranı

göstermek üzere,

$$\begin{aligned}
 g_f/M_T &= 178/268 = 0.664179 \\
 B_{f(\text{erkek})}^{\#2} &= B_{f(\text{erkek})}^{\#1} (0.664179) \\
 &= (100.4466)(0.664179) \\
 &= 66.7145
 \end{aligned} \tag{11}$$

ve

$$\begin{aligned}
 B_{f(\text{kadın})}^{\#2} &= B_{f(\text{kadın})}^{\#1} (0.664179) \\
 &= (167.5534)(0.664179) \\
 &= 111.2855
 \end{aligned} \tag{12}$$

elde edilmiştir. Tüm hücelere bu prosedür uygulanarak, aşağıdaki tablo elde edilir (Tablo 6).

Tablo 6 incelendiğinde, C , S ve $S \times M$ için gerçek toplamlar elde edilirken, $C \times M$ için gerçek toplamlar elde edilememiştir. Yani, C , S ve $S \times M$ 'nin her bir hücre değeri ile gözlenmiş frekanslar aynı iken, $C \times M$ 'nin her bir hücre değeri ile gözlenmiş frekanslar aynı değildir. Bu nedenle, $C \times M$ 'nin gözlenmiş değerleri ele alınarak, üçüncü ardışık tahminleri elde edilecektir. $C \times M$ için gözlenmiş ve beklenen frekanslar aşağıda verilmiştir (Tablo 7).

Tablo 6. $S \times C$ birlikteliğinin kısmi testi için beklenen frekansların ikinci ardışık tahminleri

Medeni durum	Cinsiyet	Tiyatroya gidenler arasında sinemayı tercih etme Durumu					Toplam
		1	2	3	4	5	
Bekar	Erkek	66.7145	48.7240	14.6173	6.7464	5.2472	142.0494
	Kadın	111.2855	81.2760	24.3827	11.2536	8.7528	236.9506
	Toplam	178	130	39	18	14	379
Evli	Erkek	33.7321	37.1052	11.6188	5.2472	5.2472	92.9505
	Kadın	56.2679	61.8948	19.3812	8.7528	8.7528	155.0495
	Toplam	90	99	31	14	14	248

Tablo 7. $C \times M$ için gözlenmiş ve beklenen frekanslar

	g_f		B_f	
	Erkek	Kadın	Erkek	Kadın
Bekar	133	246	Bekar 142.0494	236.9506
Evli	102	146	Evli 92.9506	155.0494
	235	392		

$S \times C$ birlikteliğinin kısmi testi için üçüncü ardışık tahminlerinin elde edilmesinde uygulanan işlem sadece, sinemayı birinci sırada tercih eden bekar erkekler için gösterilmiştir.

Tüm hücelere bu işlem uygulanarak, aşağıdaki tablo elde edilir (Tablo 8).

$$g_f/B_f=133/142.0494=0.936294$$

$$\begin{aligned} B_{f(\text{erkek})}^{\#3} &= B_{f(\text{erkek})}^{\#2} (0.936294) \\ &= (66.7145)(0.936294) \\ &= 62.4644 \end{aligned} \quad (13)$$

Tablo 8 incelendiğinde, test edilecek etki S×C hariç, gözlenmiş tüm marjinal frekansların, tüm beklenen marjinal frekanslara eşit olduğu görülmüştür. G_{SC}^2 'yi hesaplayabilmek için gerekli olan B_f değerleri elde edilmiştir ve G_{SC}^2 değeri,

$$\begin{aligned} G_{SC}^2 &= 2 \left[75 \ln \left(\frac{75}{62.4644} \right) + \dots + 9 \ln \left(\frac{9}{8.2419} \right) \right] \\ &= 16.7370 \end{aligned} \quad (14)$$

sd=4

dır. Aynı işlem, C×M ve S×M birliktelikleri içinde uygulandığında,

$$G_{CM}^2=8.209 \quad \text{sd}=1 \quad (15)$$

ve

$$G_{SM}^2=13.379 \quad \text{sd}=4 \quad (16)$$

olarak elde edilir. Sonuç olarak, $\alpha = 0.05$ anlamlılık düzeyinde 4 serbestlik derecesi için χ^2 değeri 9.48 ve 1 serbestlik derecesi için χ^2 değeri 3.84 olduğundan dolayı, S×C, S×M ve C×M birliktelikleri istatistiksel olarak anlamlıdır.

Üç yönlü etkilerin incelenmesi

Üç yönlü (S×C×M) ilişkiyi test etmek için, uzun bir süreç vardır. 20 hücrede uygun B_f 'yi hesaplayabilmek için 10 tane ardışık işlem gereklidir. S×C×M ilişkisi için,

$$G_{SCM}^2=5.130 \quad \text{sd}=4 \quad (17)$$

olarak elde edilmiştir. $\alpha = 0.05$ anlamlılık düzeyinde, 4 serbestlik derecesi için χ^2 değeri 9.48 olduğundan dolayı, üç-yönlü ilişki istatistiksel olarak anlamsızdır.

Üç yönlü ilişki için, $G_{SCM}^2=5.130$ elde edildiğinden, Cinsiyet ve Sinemayı tercih etme değişkenleri arasında birliktelik için kısmi olabilirlik oran istatistiği,

$$\begin{aligned} G_{SC(\text{kısmi})}^2 &= G_{SC}^2 - G_{SCM}^2 \\ &= 16.7370 - 5.1300 \quad \text{sd}=4 \\ &= 11.6070 \end{aligned} \quad (18)$$

olarak elde edilir. Ayrıca, C×M ve S×M birliktelikleri için kısmi olabilirlik oran istatistikleri,

$$G_{CM(\text{kısmi})}^2=3.079 \quad \text{sd}=1 \quad (19)$$

$$G_{SM(\text{kısmi})}^2=8.249 \quad \text{sd}=4 \quad (20)$$

dır.

Sonuç olarak, $\alpha = 0.05$ anlamlılık düzeyinde 4 serbestlik derecesi için χ^2 değeri 9.48 ve 1 ser-

Tablo 8. S×C birlikteliğinin kısmi testi için beklenen frekansların üçüncü ardışık tahminleri

Medeni durum	Cinsiyet	Tiyatroya gidenler arasında sinemayı tercih etme durumu					Toplam
		1	2	3	4	5	
Bekar	Erkek	62.4644	45.6200	13.6860	6.3166	4.9129	133
	Kadın	115.5356	84.3800	25.3140	11.6834	9.0871	246
	Toplam	178	130	39	18	14	379
Evli	Erkek	37.0162	40.7177	12.7499	5.7581	5.7581	102
	Kadın	52.9838	58.2823	18.2501	8.2419	8.2419	146
	Toplam	90	99	31	14	14	248

bestlik derecesi için χ^2 değeri 3.84 olduğundan dolayı, C×M ve S×M birliktelikleri istatistiksel olarak anlamsız, sadece S×C birlikteliği istatistiksel olarak anlamlı bulunmuştur.

Çok yönlü frekans analizinde, Ankara ilindeki devlet tiyatrolarının izleyici profiline ait veriler için eleme testlerinin sonuçları aşağıdaki tabloda verilmiştir (Tablo 9).

Tablo 9. Ankara ilindeki devlet tiyatrolarının izleyici profiline ait veriler için eleme testlerinin sonuçları

Etki	sd	G ²	P değeri
Toplam	19	523.762	<0.05
Sinemayı tercih etme (S)	4	429.902	<0.05
Cinsiyet (C)	1	39.734	<0.05
Medeni durum (M)	1	27.573	<0.05
S×C	4	11.607	<0.05
S×M	4	8.249	>0.05
C×M	1	3.079	>0.05
S×C×M	4	5.130	>0.05

Uygun modelin seçimi ve testi

Çok yönlü frekans analizinde, eleme sonuçları kullanılarak araştırmacı için yeterli bilgi elde edilir. Amaç, her bir hücrede beklenen frekansları gerçeğe en uygun şekilde tahmin edebilecek modeli oluşturmaktır. Aynı zamanda böyle bir model, en az sayıda etkiyi kullanacak olan modeldir. Örneğin, A, B ve C değişkenlerini içeren üç yönlü bir düzende tam model,

$$\ln B_{f_{ijk}} = \theta + \lambda_{A_i} + \lambda_{B_j} + \lambda_{C_k} + \lambda_{AB_{ij}} + \lambda_{AC_{ik}} + \lambda_{BC_{jk}} + \lambda_{ABC_{ijk}} \quad (21)$$

dır. Bu eşitlikte, $\ln B_{f_{ijk}}$, her bir hücre için beklenen frekanstır ve bu eşitlik, λ etki parametreleri ve θ sabitinin toplamına eşittir.

Seçilen modelin test işlemini yapmak için, artık frekanslara ihtiyaç vardır. Artık frekanslar, toplam G^2 değerinden, etkilerin her biri için elde edilen G^2 değeri çıkartılarak elde edilir.

Eğer, artık frekanslar anlamsız ise, seçilen model yeterli bir modeldir denir (O'Leary vd., 2005).

Çalışmada, (SC,M) en açık şekilde görülen modeldir. Tablo 9 incelendiğinde, P değeri <0.05 olan etkiler tiyatroya gidenler arasında sinemayı tercih etme (S), cinsiyet (C), medeni durum (M) ve sinemayı tercih etme ve cinsiyet (S×C) etkileridir. Bu modelde, medeni duruma ek olarak sadece S×C birlikteliği yer alır. Çünkü, S×C birlikteliği, S ve C ana etkilerini de içerir. (SC,M) modeli için G^2 değeri, Tablo 9'da verilen eleme sonuçlarından elde edilmiştir ve iki yönlü etki için kısmi testlerin G^2 değerleri kullanılmıştır. (SC,M) modeli için,

$$\begin{aligned} G_{(SC,M)}^2 &= G_T^2 - G_{SC}^2 - G_C^2 - G_S^2 - G_M^2 \\ &= 523.762 - 11.607 - 39.734 \\ &\quad - 429.902 - 27.573 \\ &= 14.946 \end{aligned} \quad (22)$$

$$sd = 19 - 4 - 1 - 4 - 1 = 9$$

elde edilir. $\alpha = 0.05$ anlamlılık düzeyinde 9 serbestlik derecesi için χ^2 değeri 16.92 olduğundan, artık frekans anlamsızdır. Sonuç olarak, seçilen model, yeterli bir modeldir ve bu model

$$\ln B_f = \theta + \lambda_M + \lambda_C + \lambda_S + \lambda_{SC} \quad (23)$$

şeklinde yazılır.

Seçilen modelin değerlendirilmesi ve yorumlanması

En iyi model seçildikten sonra, her bir hücre için beklenen frekanslar elde edilecektir. (SC,M) modeli için beklenen frekanslar aşağıdaki tabloda verilmiştir (Tablo 10).

Çok yönlü frekans analizinde, Eşitlik (21)'de kullanılan λ parametre tahminlerinin elde edilmesi için, ilk olarak sapma değerlerine ihtiyaç vardır (Tabachnick ve Fidell, 2001). Sapmalar,

$$\ln(P_{ijk}) = \ln(B_{f_{ijk}}/n) \quad (24)$$

Çok yönlü frekans tablolarının analizi

elde edilir. Örneğin, sinemayı birinci sırada tercih eden bekar erkekler için,

$$\begin{aligned} \ln(P_{111}) &= \ln(B_{111}/627) \\ &= \ln(68.3046/627) \\ &= -2.2170 \end{aligned} \quad (25)$$

elde edilir ve diğer geriye kalan tüm hücreler için $\ln(P_{ijk})$ 'lar hesaplanır ise, aşağıdaki tablo elde edilir (Tablo 11).

Tablo 11'de elde edilen $\ln(P_{ijk})$ 'lar yardımıyla, parametre tahminlerinin elde edilmesi üç adımdan oluşur.

1. Seçilen modelde yer alan etkilerin her birinin her düzeyi için ortalama ve genel ortalama hesaplanır. Genel ortalama,

$$\begin{aligned} \bar{x}_{...} &= (1/scm) \sum_{ijk} \ln(P_{ijk}) \\ &= (1/20)[(-2.2170) + \dots + (-4.5959)] \\ &= -3.48589 \end{aligned} \quad (26)$$

dır ve örneğin, tiyatroya gidenler arasında sinemayı birinci sırada tercih edenler için ortalama,

$$\begin{aligned} \bar{x}_{1..} &= (1/cm) \sum_{jk} \ln(P_{ijk}) \\ &= (1/4)[(-2.2170) + (-1.9009) \\ &\quad + (-2.6411) + (-2.3250)] \\ &= -2.2710 \end{aligned} \quad (27)$$

ve cinsiyeti erkek olanlar için ortalama,

$$\begin{aligned} \bar{x}_{1.} &= (1/sm) \sum_{ik} \ln(P_{ijk}) \\ &= (1/10)[(-2.2170) + (-2.5623) \\ &\quad + \dots + (-4.8836)] \\ &= -3.75523 \end{aligned} \quad (28)$$

elde edilir. Ayrıca, cinsiyeti erkek olup tiyatroya gidenler arasında sinemayı birinci sırada tercih edenler için ortalama,

$$\begin{aligned} \bar{x}_{11.} &= (1/m) \sum_k \ln(P_{ijk}) \\ &= (1/2)[(-2.2170) + (-2.6411)] \\ &= -2.42905 \end{aligned} \quad (29)$$

dır. Ayrıca, etkilerin her birinin her bir düzeyi için ayrı ayrı ortalamalar hesaplanmıştır.

Tablo 10. (SC,M) modeli için beklenen frekanslar

Medeni durum	Cinsiyet	Tiyatroya gidenler arasında sinemayı tercih etme Durumu					Toplam
		1	2	3	4	5	
Bekar	Erkek	68.3046	48.3572	9.6715	8.4625	7.2536	142.0494
	Kadın	93.6922	90.0654	32.6411	10.8804	9.6715	236.9506
Evli	Erkek	44.6954	31.6428	6.3285	5.5375	4.7464	92.9506
	Kadın	61.3078	58.9346	21.3589	7.1196	6.3285	155.0494
	Toplam	268	229	70	32	28	627

Tablo 11. (SC,M) modeli için Beklenen $\ln(P_{ijk})$ değerleri

Medeni durum	Cinsiyet	Tiyatroya gidenler arasında sinemayı tercih etme durumu				
		1	2	3	4	5
Bekar	Erkek	-2.2170	-2.5623	-4.1718	-4.3053	-4.4595
	Kadın	-1.9009	-1.9404	-2.9554	-4.0540	-4.1718
Evli	Erkek	-2.6411	-2.9864	-4.5959	-4.7294	-4.8836
	Kadın	-2.3250	-2.3645	-3.3795	-4.4781	-4.5959

2. Parametre tahminleri (λ), her etkinin her bir düzeyinin genel ortalamadan farkı alınarak elde edilir. Örneğin; sinemayı birinci sırada tercih edenler için,

$$\begin{aligned}\lambda_{S_1} &= \bar{x}_{1..} - \bar{x}_{...} \\ &= -2.2710 - (-3.48589) \\ &= 1.21489\end{aligned}\quad (30)$$

dır. Ayrıca, erkek olup sinemayı birinci sırada tercih edenler için,

$$\begin{aligned}\lambda_{SC_{11}} &= \bar{x}_{11..} - \bar{x}_{.1.} - \bar{x}_{1..} + \bar{x}_{...} \\ &= -2.42905 - (-3.75523) - \\ &\quad (-2.2710) + (-3.48589) \\ &= 0.11129\end{aligned}\quad (31)$$

elde edilir.

3. Son adımda ise, elde edilen λ değerleri standart hatalarına (SH) bölünerek, bir etkinin bir hücreye olan katılımı değerlendirilecektir. λ /SH oranına, standart normal sapma adı verilir (Tabachnick, 2001).

$$\begin{aligned}SH^2 &= \sum (1/s)^2 / B_f \\ &= (1/5)^2 / 268 + (1/5)^2 / 229 + \dots + \\ &\quad (1/5)^2 / 70 + (1/5)^2 / 32 + (1/5)^2 / 28 \\ &= 0.00357392622\end{aligned}\quad (32)$$

$$SH = 0.059782323 \quad (33)$$

elde edilir.

Tiyatroya gidenler arasında sinemayı tercih etme değişkeninin birinci düzeyine ilişkin standart normal sapma değeri,

$$\begin{aligned}\lambda / SH &= 1.21489 / 0.059782323 \\ &= 20.32189\end{aligned}\quad (34)$$

dır. $\alpha = 0.05$ anlamlılık düzeyinde, z değeri 1.64'dür. Standart normal sapma değeri (20.32189), z değeri ile karşılaştırıldığında, tiyatroya gidenler arasında sinemayı tercih etme değişkeninin birinci düzeyinin önemli bir etkisi vardır. Tüm λ ve λ /SH değerleri seçilen modelde yer alan etkilerin her bir düzeyi için hesaplanmış ve aşağıda verilen tablo elde edilmiştir (Tablo 12).

Tablo 12. (SC,M) modeli için tüm λ ve λ /SH değerleri

Etki	Düzye	λ	λ /SH
Cinsiyet (C)	Erkek	-0.26934	-6.52941
	Kadın	0.26934	6.52941
Medeni durum (M)	Bekar	0.21205	5.19254
	Evli	-0.21205	-5.19254
Sinemayı tercih etme durumu (S)	1	1.21489	20.32189
	2	1.02249	17.10355
	3	-0.28976	-4.84692
	4	-0.90581	-15.15180
	5	-1.04181	-17.42672
Sinemayı tercih etme durumu ile cinsiyet (S×C)	Sinemayı tercih etme 1. düzeyi ile erkek	0.11129	1.81168
	Sinemayı tercih etme 1. düzeyi ile kadın	-0.11129	-1.81168
	Sinemayı tercih etme 2. düzeyi ile erkek	-0.0416	-0.66755
	Sinemayı tercih etme 2. düzeyi ile kadın	0.0416	0.66755
	Sinemayı tercih etme 3. düzeyi ile erkek	-0.33886	-5.43762
	Sinemayı tercih etme 3. düzeyi ile kadın	0.33886	5.43762
	Sinemayı tercih etme 4. düzeyi ile erkek	0.14369	2.30577
	Sinemayı tercih etme 4. düzeyi ile kadın	-0.14369	-2.30577
	Sinemayı tercih etme 5. düzeyi ile erkek	0.12549	2.01372
	Sinemayı tercih etme 5. düzeyi ile kadın	-0.12549	-2.01372

Çok yönlü frekans tablolarının analizinde, değişkenin iki düzeyli olması durumunda kullanılan (+) ve (-) işaretler frekansın büyüklüğüne göre seçilir. Her bir düzey için elde edilen λ parametre tahminleri ayrı ayrı hesaplanmaz ve tek bir parametre tahmini elde edilir.

Sonuçlar

Bu çalışmada, çok yönlü frekans analizi ile ele alınan verideki üç kesikli bağımsız değişken arasındaki ilişkiler değerlendirilmiştir. Bu değişkenler arasındaki ilişkilerin anlamlı olup olmadıklarının belirlenmesinde, olabilirlik oran test istatistiği G^2 kullanılmıştır. İki yönlü birliktelikler için kısmi G^2 değerleri elde edilmiştir. Ayrıca, en iyi model belirlenmiş ve bu model için parametre tahminleri ve standart normal sapmalar elde edilmiştir. Tablo 12 incelendiğinde, aşağıdaki sonuçlara ulaşılabılır:

Tiyatroya gidenler arasında sinemayı tercih etme durumu ile cinsiyet arasındaki birlikteliklerin düzeyleri incelendiğinde, sadece kadın olup sinemayı 2. sırada tercih edenler arasındaki birlikteliğin hücre frekansı üzerine önemli bir katkısı bulunmamıştır. Çünkü, standart normal sapma değeri 0.66755, $\alpha = 0.05$ anlamlılık düzeyinde $z=1.64$ değeri ile karşılaştırıldığında, bu birlikteliğin önemsiz olduğu görülmektedir. Tiyatroya gidenler arasında sinemayı tercih etme durumu ile cinsiyet arasındaki birlikteliklerden, sadece sinemayı 3. sırada tercih etme ve cinsiyet arasındaki birliktelik hücre frekansı üzerinde en önemli katkıyı sağlar.

Sonuç olarak, örneğin, tiyatroya gidenler arasında sinemayı 2. sırada tercih eden evli kadınlar hücresi ele alınsın. Bu hücrede yer alan parametreler için standart normal sapmalar,

Sinemayı 2.sırada tercih etme	: 17.10355
Evli olma	: -5.19254
Kadın olma	: 6.52941
Kadın olup sinemayı 2. sırada tercih etme	: 0.66755

dır. $\alpha = 0.05$ anlamlılık düzeyinde $z=1.64$ değeri ile bu değerler karşılaştırılırsa, ele alınan hücre frekansı üzerindeki en önemli etkiler sırası

ile, sinemayı 2. sırada tercih etme, cinsiyeti kadın olanlar ve medeni durumu evli olanlar sağlar. Sinemayı 2. sırada tercih etme ve kadın arasındaki birlikteliğin hücre frekansı üzerinde önemli bir katkısı yoktur.

Benzer şekilde, tiyatroya gidenler arasında sinemayı 1. sırada tercih eden bekar kadınlar hücresi ele alınsın. Bu hücrede yer alan parametreler için standart normal sapmalar,

Sinemayı 1.sırada tercih etme	: 20.32189
Bekar olma	: 5.19254
Kadın olma	: 6.52941
Kadın olup sinemayı 1. sırada tercih etme	: -1.81168

dır. $\alpha = 0.05$ anlamlılık düzeyinde $z=1.64$ değeri ile bu değerler karşılaştırılırsa, ele alınan hücre frekansı üzerindeki en önemli etkiler sırası ile, sinemayı 1. sırada tercih etme, cinsiyeti kadın, medeni durumu bekar ve sinemayı 1. sırada tercih etme ve kadın arasındaki birlikteliğidir.

Bu çalışmada amaç, çok yönlü frekans analizinde yer alan temel işlemlerin nasıl kullanıldığını göstermek ve yorumlamaktır.

Kaynaklar

- Fisher, R.A., (1924). The conditions under with X^2 measures the discrepancy between observed observation and hypothesis, *Journal of Royal Statistical Society*, **87**, 442-450.
- Freeman, D. H., (1987). *Applied Categorical Data Analysis*, Marcel Dekker, NewYork.
- Tabachnick, B.G. ve Fidell L.S., (2001). *Using Multivariate Statistics*, Allyn&Bacon, USA.
- Wrigley, N., (2002). *Categorical Data Analysis for Geographers and Enviromental Scientists*, The Blackburn Press, USA.

-
- Vokey, J.H., (2002). *Multway Frequency Analysis for Experimental Psychologists*, [http://people.uleth.ca/~vokey/pdf/vokeydoc .pdf](http://people.uleth.ca/~vokey/pdf/vokeydoc.pdf), (18.04.2005).
- O'Leary, C., Baulch, J. ve Ling, L.Y., eds. (2005). *Multway Frequency Analysis*, <http://psych.sci.usq.edu.au/Teaching/Honours/Documents/MFA.pdf>, (18.04.2005).