

Üçgensel Olumsuzluk Tablolarında Yarı-Bağımsızlık Modeli için Güç Analizi

Serpil AKTAŞ*

Tülay SARAÇBAŞI

ÖZET

Üçgensel olumsuzluk tabloları tamamlanmamış olumsuzluk tablolarının özel bir durumudur. Üçgensel tablolarda ana köşegen elemanlarının alt ya da üst üçgen kısımlarında yapısal sıfırlı elemanlar bulunur. Bu tür tablolarda satır ve sütun değişkenleri arasındaki bağıntı yarı-bağımsızlık modeli ile araştırılır. Yarı-bağımsızlık modeli test edilirken ki-kare ya da olabilirlik oran istatistiklerinden yararlanılır. H_0 hipotezi doğru olmadığında bu test istatistikleri yaklaşık olarak merkezsiz olmayan ki-kare dağılımı göstermektedir. Bu çalışmada iki değişkenli normal dağılımdan yararlanarak türetilen 30.000 üçgensel tabloda yarı-bağımsızlık modeli test edilmiştir. Bu modele uyan ve uymayan tablo yüzdeleri verilmiştir. Yarı -bağımsızlık modelinden elde edilen olabilirlik oran istatistiklerinin testleri için güç değerleri hesaplanmıştır. Yarı-bağımsızlık modelinin reddedildiği tablo yüzdeleri ile güç değerleri karşılaştırılmıştır.

Anahtar Kelimeler : Olumsuzluk tablosu , Yarı-bağımsızlık modeli , Güç analizi

1.GİRİŞ

Tamamlanmamış olumsuzluk tablolarını ilk defa Bishop-Fienberg (1969) ve Mantel (1970) tanıtmışlardır. Üçgensel(triangular) olumsuzluk tabloları tamamlanmamış olumsuzluk tablolarının özel bir durumudur.Üçgensel olumsuzluk tablolarında sıklıklar ana köşegen elemanlarının alt ya da üst üçgen kısımlarında yer alır. Sarkar(1989) aşağıdaki koşulları sağlayan üst-sağ, üst-sol, alt-sağ ve alt-sol olmak üzere dört çeşit üçgensel olumsuzluk tablo tanımı yapmıştır.

Üst-sağ (sol) üçgensel olumsuzluk tablosu π_{ij} , i. satır ve j. kolon'a karşılık gelen gözlenen sıklık olmak üzere, $\pi_{ij} = 0 \quad i > j$, ve alt-sol(sağ) üçgensel olumsuzluk tablosu ise $\pi_{ij} = 0 \quad i < j$ olarak tanımlanır. Üst-sağ üçgensel olumsuzluk tablosu Tablo 1'de verilmiştir.

* Hacettepe Üniversitesi, Fen Fakültesi İstatistik Bölümü Beytepe-Ankara, e-mail: spxl@hacettepe.edu.tr

Tablo 1. Üst-Sağ Üçgensel Olumsuzluk Tablo Yapısı

i/j	1	2	3	...	R
1	π_{11}	π_{12}	π_{13}	...	π_{1R}
2		π_{22}	π_{23}	...	π_{2R}
3			π_{33}	...	π_{3R}
...			
R				...	π_{RR}

Üçgensel olumsuzluk tablolarında tablonun tamamlanmamış yapısı nedeniyle satır ve kolon arasındaki bağımsızlık ilk kez Goodman(1968) tarafından tanımlanan yarı-bağımsızlık(YB) modeli ile incelenir. YB modeli üst-sağ üçgensel olumsuzluk tablosu,

$$\begin{aligned} \pi_{ij} &= \alpha_i \beta_j & i \leq j \\ &= 0 & i > j \\ i &= 1, \dots, R, \text{ ve } j = 1, \dots, R \end{aligned} \quad (1)$$

olarak tanımlanmıştır.

Burada $\alpha_i > 0$ ve $\beta_j > 0$ pozitif sabitlerdir.

Sarkar(1989) yarı-bağımsızlık modelini sıralanabilir ilişki cinsinden uyumlu(concordant) C_π ve uyumsuz(discordant) D_π gözlem çiftleri arasındaki farka dayanan ve Eşitlik 2'de verilen η_π ile

$$\begin{aligned} C_\pi &= 2 \sum_{i < k} \sum_{j < l} \pi_{ij} \pi_{kl} \quad , \quad D_\pi = 2 \sum_{i > k} \sum_{j > l} \pi_{ij} \pi_{kl} \\ \eta_\pi &= 2 \sum_{i < k} \sum_{j < l} (\pi_{ij} \pi_{kl} - \pi_{il} \pi_{kj}) \end{aligned} \quad (2)$$

tanımlanmıştır.

Burada $\eta_\pi = 0$ ise yarı-bağımsızlık modelinin varlığından söz edilir. Dolayısıyla yarı-bağımsızlık modeli için $H_0: \eta_\pi = 0$ hipotezine karşılık $H_s: \eta_\pi > 0$ ($\eta_\pi < 0$) seçenek hipotezi ile test edilebilir. H_0 hipotezini H_s 'e karşı test etmek için uygun test istatistiği Eşitlik (3) ile,

$$z = \frac{\sqrt{n} \hat{\eta}_\pi}{\hat{\sigma}_\pi} \quad (3)$$

tanımlanır.

z 'nin dağılımı $N(0,1)$ 'dir (Sarkar, 1989). η_π 'nin dağılımı ile ilgili teorem:

Teorem: Çokterimli(multinomial) örneklem altında $\sqrt{n}(\hat{\eta}_{\pi} - \eta_{\pi})$ 'nin dağılımı $N(0, \sigma_{\pi}^2)$ dir.

$$\sigma_{\pi}^2 = \sum_{i \leq j} \phi_{\pi,ij}^2 \pi_{ij} - 4\eta_{\pi}^2 \quad (4)$$

Eşitlik (4) deki $\phi_{\pi,ij}$ $i \leq j$ için

$\phi_{\pi,ij} = 2[P(X < i, Y < j) + P(X > i, Y > j) - P(X < i, Y > j) - P(X > i, Y < j) - P(X > j) - P(Y < i)]$ biçiminde ifade edilir H_0 doğru iken σ_{π}^2 nin en çok olabilirlik tahmini $\hat{\sigma}_{\pi}^2$ 'dir Sarkar(1989).

2. OLUMSALLIK TABLOLARINDA GÜÇ ANALİZİ

Testin gücü bilindiği gibi seçenек hipotezi doğru olduğunda yokluk hipotezini reddetme olasılığıdır ve $Güç=1-P(II.Tip Hata)$ olarak tanımlanır. Yokluk hipotezi doğru olmadığında χ^2 ve G^2 gibi test istatistiklerinin dağılımı büyük örneklemelerde yaklaşık olarak merkezselsel olmayan ki-kare dağılımı gösterir. Merkezi olmayan ki-kare dağılımının olasılık yoğunluk fonksiyonu χ'^2 aşağıdaki şekilde tanımlanır.

X_1, X_2, \dots, X_v raslantı değişkenlerinin herbiri normal dağılım gösterebilir ve eğer a_1, a_2, \dots, a_v sabitleri $\lambda = \sum_{i=1}^v a_i^2$ ve $\chi'^2 = \sum_{i=1}^v (X_i - a_i)^2$ ise yoğunluk fonksiyonu

$$f(\chi'^2) = \frac{\exp\left\{-\frac{1}{2}(\chi'^2 + \lambda)\right\}}{2^{\frac{1}{2}v}} \sum_{j=0}^{\infty} \frac{(\chi'^2)^{\frac{1}{2}v+j-1} \lambda^j}{\Gamma\left(\frac{1}{2}v + j\right) 2^{2j} j!} \quad (5)$$

ile verilir. Burada v serbestlik derecesidir (Pearson and Hartley, 1976). Agresti (1990)'nin tanımladığı merkezselsel olmayış parametreleri λ , G^2 için Eşitlik (6)'da χ^2 için Eşitlik (7)'de vermiştir. Eşitlikler bağımsızlık modeli için tanımlanmıştır.

$$\lambda = 2n \sum_i \sum_j \pi_{ij} \log \frac{\pi_{ij}}{\pi_{ij(m)}} \quad (6)$$

$$\lambda = n \sum \frac{(\pi_{ij} - \pi_i \pi_j)^2}{\pi_i \pi_j} \quad (7)$$

Eşitlik (6)'da $\pi_{ij(m)}$, π_{ij} 'nin en çok olabilirlik tahminidir

$\lambda = 0$ olduğu durum merkezselsel ki-kare dağılımını verir. H_0 hipotezi doğru olduğunda $\pi_{ij} = \pi_i \pi_j$ olur.

Üçgensel olumsuzluk tablolarının çözümlenmesinde yararlanılan yarı-bağımsızlık modeli benzetim çalışması ile türetilen tablolarda çözümlenmiş ve yarı bağımsızlık testlerinin güçleri hesaplanmıştır.

3. BENZETİM ÇALIŞMASI

Yarı bağımsızlık modelini ilişki cinsinden incelemek ve güçlerini hesaplamak için benzetim çalışması yapılmıştır. Goodman (1981,1985) çalışmalarında ilişkisiz olumsuzluk tablolarını türetebilmek için korelasyon katsayısı sıfıra yakın iki değişkenli normal dağılımlı kitlelerden çekilen örneklemelerden yararlanmıştır. Üçgensel olumsuzluk tablolarında yarı bağımsızlık modeline uyumu belirleyen G^2 istatistiği ile Sarkar'ın önerdiği yöntemi karşılaştıran, test istatistiklerinin güç değerlerini hesaplayan benzetim çalışmasında bağımsız üçgensel tablolar oluşturabilmek için Goodman (1981), Saraçbaşı ve Muluk (1985) çalışmalarından yararlanılmıştır. Kitle korelasyon katsayısı (ρ), 0.0, 0.2, 0.4 olan iki değişkenli normal dağılımlı kitlelerden örneklem büyüklüğü (n), 100, 250, 500, 1000 ve tablo boyutu (R), 4,5,6,7 ve 8 olan üst-sağ üçgensel olumsuzluk tabloları türetilmiştir. Herbir kombinasyondan 500 tekrar yapılarak toplam 30.000 üst-sağ üçgensel olumsuzluk tablosu oluşturulmuştur (Aktaş, 1999). Türetilen 30.000 tablonun YB modeline uyumu G^2 ve Sarkar (1989)'ın önerdiği Z değerlerine göre incelenmiştir. Tabloların tümünde güç değerleri herbir örneklemde 500 tekrardan elde edilen veriler için S-Plus programından yararlanılarak hesaplanmıştır. Yarı-bağımsızlık modeli varsayımı altında olabilirlik oran istatistiği için güç değerleri Tablo 2'de verilmiştir. Güç değerlerinin benzetim parametrelerine göre değişimi ise Şekil 1-2 ve 3'de görülmektedir.

Tablo 3'de ise YB modeli için 500 tekrarda elde edilen reddedilen ve kabul edilen hipotezlerin oranları verilmiştir. Sarkar (1989) yöntemine göre hesaplanan $\hat{\eta}_\pi$ değerlerinin 500 tekrar için ortalama ve standart hata değerleri Tablo 4'de verilmiştir.

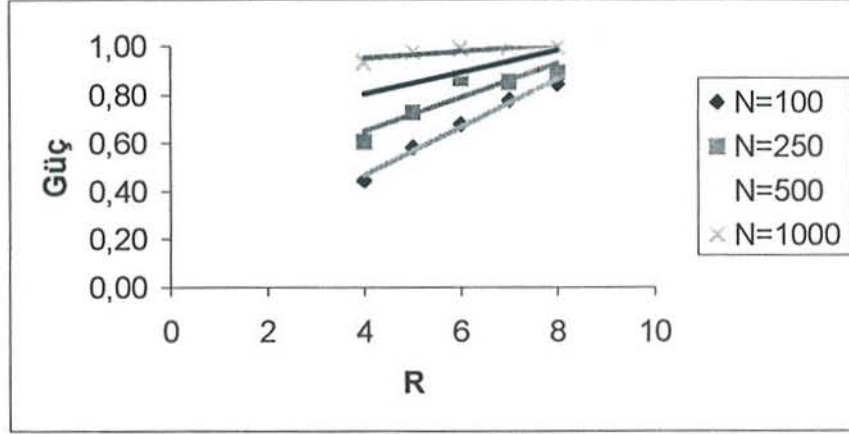
Tablo 2. YB Modeli İçin Olabilirlik Oran İstatistiğinin Güç Değerleri

n=100			
R	ρ		
	0.0	0.2	0.4
4	0.4464	0.4563	0.4359
5	0.5829	0.5830	0.5602
6	0.6797	0.6642	0.6720
7	0.7789	0.7679	0.7889
8	0.8454	0.8348	0.8368

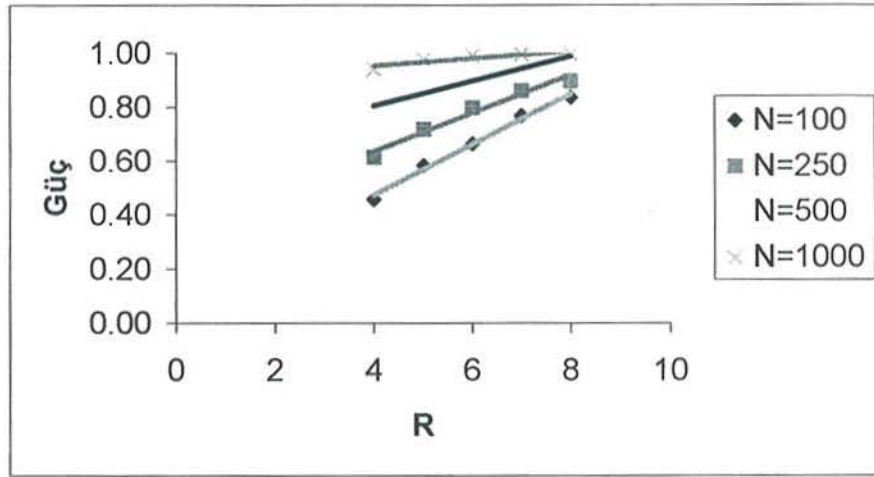
n=250			
R	ρ		
	0.0	0.2	0.4
4	0.6077	0.6131	0.6123
5	0.7292	0.7180	0.7277
6	0.8690	0.7967	0.7942
7	0.8533	0.8622	0.8503
8	0.8916	0.8969	0.8954

n=500			
R	ρ		
	0.0	0.2	0.4
4	0.7777	0.7693	0.7902
5	0.8731	0.8798	0.8960
6	0.9118	0.9171	0.9247
7	0.9398	0.9509	0.9366
8	0.9646	0.9630	0.9405

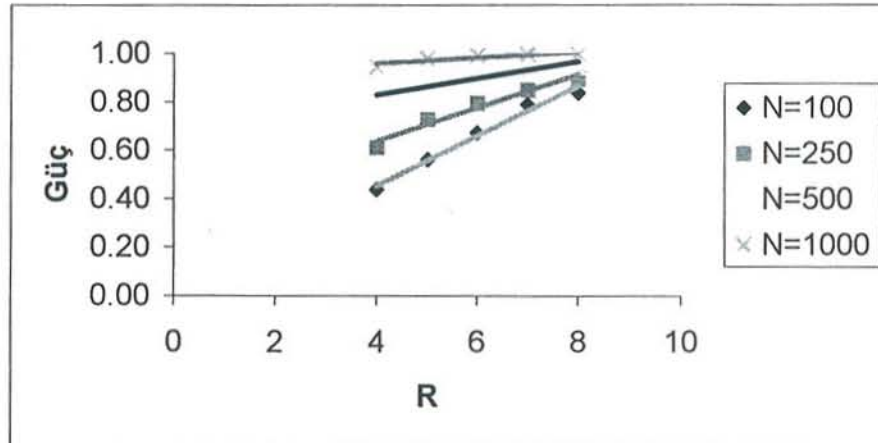
n=1000			
R	ρ		
	0.0	0.2	0.4
4	0.9337	0.9384	0.9453
5	0.9764	0.9776	0.9796
6	0.9912	0.9887	0.9914
7	0.9943	0.9961	0.9958
8	0.9964	0.9963	0.9954



Şekil 1. $\rho=0.0$ için Güç Değerleri



Şekil 2. $\rho=0.2$ için Güç Değerleri



Şekil 3. $\rho=0.4$ için Güç Değerleri

Tablo 2 incelendiğinde tablo boyutu ve örneklem büyüklüğü arttıkça güç değerlerinde de artış görülmektedir; ancak korelasyon katsayısı gücü etkilememektedir.

$n=100$, $\rho=0.4$ ve $R=4$ olduğu durumda güç minimum; $n=1000$, $\rho=0.0$ ve $R=8$ olduğu durumda ise güç maksimumdur. En yüksek güç değerleri $n=1000$ için elde edilmiştir. Bilindiği gibi gücün 1'e yakın olması beklenir. Bu değer, seçenek hipotezi doğru iken yokluk hipotezinin reddedilme olasılığını verir.

Şekil 1-2-3 güç değerlerinin boyut, örneklem büyüklüğü ve korelasyon katsayısına göre değişimini göstermektedir. Tüm şekillerde güç, örneklem büyüklüğü ve boyuta göre artış göstermektedir. Ancak, küçük örneklerde boyut büyüdükçe bu artışın daha hızlı olduğu gözlenmektedir. $n<250$ iken tablo boyutunun artırılması güçte hızlı bir artış yaratmaktadır.

Tablo 3. YB Modeli için 500 Tekrarda Reddedilen ve Kabul Edilen Hipotezlerin Oranları

<i>n</i>	ρ	<i>R</i>	<i>red</i>	<i>kabul</i>	<i>n</i>	ρ	<i>R</i>	<i>red</i>	<i>kabul</i>
100	0.0	4	0.21	0.79	500	0.0	4	0.76	0.24
100	0.0	5	0.26	0.74	500	0.0	5	0.80	0.20
100	0.0	6	0.21	0.79	500	0.0	6	0.77	0.23
100	0.0	7	0.19	0.81	500	0.0	7	0.71	0.29
100	0.0	8	0.13	0.87	500	0.0	8	0.70	0.30
100	0.2	4	0.23	0.77	500	0.2	4	0.75	0.25
100	0.2	5	0.24	0.76	500	0.2	5	0.82	0.18
100	0.2	6	0.18	0.82	500	0.2	6	0.78	0.22
100	0.2	7	0.17	0.83	500	0.2	7	0.78	0.22
100	0.2	8	0.09	0.91	500	0.2	8	0.66	0.34
100	0.4	4	0.21	0.79	500	0.4	4	0.78	0.22
100	0.4	5	0.18	0.82	500	0.4	5	0.84	0.16
100	0.4	6	0.18	0.82	500	0.4	6	0.81	0.19
100	0.4	7	0.19	0.81	500	0.4	7	0.70	0.30
100	0.4	8	0.09	0.91	500	0.4	8	0.48	0.52
250	0.0	4	0.47	0.53	1000	0.0	4	0.97	0.03
250	0.0	5	0.50	0.50	1000	0.0	5	0.98	0.02
250	0.0	6	0.42	0.58	1000	0.0	6	0.99	0.01
250	0.0	7	0.37	0.63	1000	0.0	7	0.99	0.01
250	0.0	8	0.30	0.70	1000	0.0	8	0.98	0.02
250	0.2	4	0.49	0.51	1000	0.2	4	0.97	0.03
250	0.2	5	0.47	0.53	1000	0.2	5	0.98	0.02
250	0.2	6	0.42	0.58	1000	0.2	6	0.99	0.01
250	0.2	7	0.40	0.60	1000	0.2	7	0.99	0.01
250	0.2	8	0.29	0.71	1000	0.2	8	0.98	0.02
250	0.4	4	0.48	0.52	1000	0.4	4	0.98	0.02
250	0.4	5	0.49	0.51	1000	0.4	5	0.99	0.01
250	0.4	6	0.44	0.56	1000	0.4	6	0.99	0.01
250	0.4	7	0.36	0.64	1000	0.4	7	0.99	0.01
250	0.4	8	0.27	0.73	1000	0.4	8	0.98	0.02

Tablo 3, 500 tekrarda YB modeli için reddedilen ve kabul edilen hipotez oranlarını vermektedir. Sonuçlara bakıldığında reddedilen hipotezlerin oranında örneklem

büyüklüğü arttıkça büyüme görülmektedir. Bu sonuç Tablo 2'deki güç değerlerinin verdiği sonuçlar ile çakışmaktadır. Örneklem büyüklüğü arttıkça kabul edilen hipotezlerin oranında azalmaktadır. Dolayısıyla YB için güç değerleri ve Tablo 3'deki sonuçlar birbirini desteklemektedir.

Tablo 4. Sarkar Yöntemine Göre Hesaplanan $\hat{\eta}_\pi$ Parametrelerinin 500 Tekrar İçin Ortalama ve Standart Hata Değerleri

n	ρ	R	$\hat{\eta}_\pi$	$S\hat{\eta}_\pi$	n	ρ	R	$\hat{\eta}_\pi$	$S\hat{\eta}_\pi$
100	0.0	4	0.0621	0.0018	500	0.0	4	0.0488	0.00057
100	0.0	5	0.0714	0.0022	500	0.0	5	0.0696	0.00078
100	0.0	6	0.0725	0.0022	500	0.0	6	0.0745	0.00074
100	0.0	7	0.0739	0.0024	500	0.0	7	0.0755	0.00078
100	0.0	8	0.0676	0.0021	500	0.0	8	0.0731	0.00093
100	0.2	4	0.0654	0.0018	500	0.2	4	0.0482	0.00062
100	0.2	5	0.1399	0.021	500	0.2	5	0.0692	0.0009
100	0.2	6	0.0729	0.020	500	0.2	6	0.0730	0.00084
100	0.2	7	0.0722	0.0021	500	0.2	7	0.0757	0.00092
100	0.2	8	0.0680	0.0022	500	0.2	8	0.0738	0.00086
100	0.4	4	0.0634	0.0018	500	0.4	4	0.0497	0.00068
100	0.4	5	0.0690	0.0021	500	0.4	5	0.0713	0.0009
100	0.4	6	0.0700	0.0021	500	0.4	6	0.0730	0.00087
100	0.4	7	0.0233	0.0024	500	0.4	7	0.0754	0.00073
100	0.4	8	0.0676	0.0020	500	0.4	8	0.0740	0.00073
250	0.0	4	0.0555	0.0011	1000	0.0	4	0.0425	0.00042
250	0.0	5	0.0698	0.0013	1000	0.0	5	0.0671	0.00063
250	0.0	6	0.0737	0.0014	1000	0.0	6	0.0725	0.00053
250	0.0	7	0.0724	0.0012	1000	0.0	7	0.0761	0.00051
250	0.0	8	0.0698	0.0014	1000	0.0	8	0.0763	0.00054
250	0.2	4	0.0551	0.0011	1000	0.2	4	0.0438	0.00042
250	0.2	5	0.0696	0.0013	1000	0.2	5	0.0675	0.00057
250	0.2	6	0.0749	0.0013	1000	0.2	6	0.0713	0.00054
250	0.2	7	0.0727	0.0014	1000	0.2	7	0.0749	0.00068
250	0.2	8	0.0730	0.0014	1000	0.2	8	0.0713	0.00059
250	0.4	4	0.0555	0.0011	1000	0.4	4	0.0435	0.00037
250	0.4	5	0.0702	0.0014	1000	0.4	5	0.0669	0.00053
250	0.4	6	0.0742	0.0013	1000	0.4	6	0.0714	0.00069
250	0.4	7	0.0748	0.0013	1000	0.4	7	0.0763	0.00069
250	0.4	8	0.0726	0.0012	1000	0.4	8	0.0743	0.00062

Tablo 4'de ortalama $\hat{\eta}_\pi$ değerleri verilen çözümlerinin tümü için Z değerlerinin güç değerleri tüm örneklerde "1" olarak bulunmuştur. Tablo 4' de yer alan $\hat{\eta}_\pi$ değerlerinin istatistiksel olarak anlamlı olması güç değerlerindeki desteklemektedir. Tamamlanmamış tablo yapısı gösteren üçgensel olumsuzluk tablolarının yarı bağımsızlık modeli ile çözümlenmesi iterasyona bağlı çözümlenmedir. Sarkar'ın önerdiği yöntem uygulaması daha kolay olan bir yöntemdir. Yapılan çalışmada Yarı- bağımsızlık ve

Sarkar Yöntemi'nin paralel sonuçlar verdiği görülmüştür. Bu nedenle üçgensel olumsuzluk tablo çözümlerinde Sarkar Yöntemi'nin uygulamadaki kolaylığı nedeniyle YB çözümlenmesi yerine kullanabileceği önerilebilir.

KAYNAKLAR

- AGRESTI, A., (1990), *Categorical Data Analysis*, John Wiley&Sons, New-York.
- AKTAŞ, S., (1999), *Üçgensel Olumsuzluk Tablolarında Bağımsızlık Çözümlemesi*, Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- BISHOP, Y.M.M.,FIENBERG, S.E., (1969), *Incomplete two-dimensional contingency tables*, *Biometrics* 28, 177-202.
- GOODMAN, L. A., (1968), *The analysis of cross-classified data: Independence, quasi-independence and interactions in contingency tables with or without missing entries*, *JASA* 63, 1091-1131.
- GOODMAN,L.A.,(1981), *Association models and the bivariate normal for contingency tables*, *Biometrics*, 35,651-655.
- GOODMAN, L.A., (1985), *The Analysis of cross-classified data having and/or unordered categories: Association models, correlation models, and asymmetry models or contingency tables with or without missing entries*, *The Annals of Statistics*, Vol.13, No.1, 10-69.
- MANTEL, N., (1970), *Incomplete contingency tables*, *Biometrics* 26, 291-304
- PEARSON, E.S. and HARTLEY H.O., (1976), *Biometrika Tables for Statisticians Volume II*, Cambridge University Press.
- SARAÇBAŞI, T. ve MULUK, F.Z.(1985) *İki değişkenli normal dağılımdan oluşturulan olumsuzluk çizelgelerinde ilişki katsayılarının incelenmesi*, H. Ü. Fen ve Mühendislik Bilimleri Dergisi, 10, 9-33.
- SARKAR, S. K., (1989), *Quasi-independence in ordinal triangular contingency tables.*, *JASA* 84,592-597.

Power Analysis of Quasi-Independence Model for Triangular Contingency Tables

ABSTRACT

Triangular contingency tables are the special class of incomplete contingency tables. The quasi-independence model investigates usual independence model between row and column classifications. It is known that when the alternative hypothesis is true, chi-squared or

likelihood ratio statistics is distributed as noncentral chi-squared distribution. In this study, the powers of tests for triangular contingency tables were calculated by generating 30.000 random tables from the bivariate normal distribution and quasi-independence model was tested. The proportions of fitted and not fitted to this model were given. Power values were calculated for the likelihood ratios tests. The proportions of rejected hypothesis for the quasi-independence and the power values were compared.

Key Words: *Contingency tables, Quasi-independence model, Power Analysis*