X^2 OLASILIK DAĞILIM FONKSİYONU VE BİR UYGULAMA

Ass. Fethi ŞENİŞ

I— GİRİŞ

İstatistik Miminin incelediği çeşitli konulardan biri, olayların birbirinden bağımsız olarak meydana gelip gelmediğinin araştırılması ve bir ilişki sezinlenmişse bu ilişkinin derecesinin belirlenmesidir. Olaylar rakamsal olarak değerlendirilebiliyorsa aralarındaki ilişkinin incelenmesi ya da derecesinin belirlenmesi açısından büyük sorunlarla karşılaşılmadığı görülmektedir. Oysa, günlük yaşantıda karşılaşılan olayların bir çoğu niceliksel değerlendirilme olanağından yoksun bulunmakta ve bu tür olaylar arasındaki ilişkinin incelenmesi, bu olayların niteliklerine göre çeşitli sınıflara ayrılarak, sınıflara düşen birim sayılarının incelenmesiyle mümkün olabilmektedir.

Niteliklerine göre sınıflandırılan olayların incelenmesi kontenjans tabloları denilen bir yardımcı yöntemle mümkün olabilmektedir. Bu yöntem ilk kez 1875 yılında Helmert adlı bir Alman fizikçi tarafından kullanılmaya başlanmıştır. Geliştirilen ve x^2 (ki-kare) adı verilen olasılık dağılım fonksiyonu 1900 yılında Karl Pearson tarafından çeşitli testlerde kullanılmaya başlanmıştır (1).

(1) Rahman N.A., A Course in Theoretical Statistics, Griffin, London, 1968, S. 376.

Kontenjans tablolarında beklenen değerler ilk kez 1934 yılında Fisher tarafından düşünülmüş ve bu düşünceler Dr. Frank Yates tarafından x^2 testlerinin hassaslığı konusunda kullanılmaya başlanmıştır (2).

Niteliklerine göre sınıflandırılan olayların incelenmesinde kullanılan x^2 testinin uygulanabilmesi için x^2 dağılımının bilinmesi gereksinimi nedeniyle bu çalışmanın ilk bölümünde x^2 olasılık dağılım fonksiyonu tanıtılarak çeşitli özellikleri incelenmiştir. İkinci bölümde ise, bu dağılımın kullanılma yerleri araştırılmış, geliştirilmiş olan çeşitli x^2 :testleri tanıtılmıştır. Çalışmanın bütünlüğü düşünülerek son bölümde bir uygulamaya yer verilerek öğrencinin liseyi bitirme derecesi ile babanın tahsil durumu arasındaki ilişkinin araştırılması konusunda bir x^2 testi uygulanarak sonuçları incelenmiştir.

II— x^2 OLASILIK DAĞILIM FONKSİYONU

II-1 x^2 Olasılık Dağılım Fonksiyonunun Tanımı :

Olasılık dağılım fonksiyonu x^2 değişkeni cinsinden

$$f(\chi^2) = \frac{1}{2^{\frac{n}{2}}\,\gamma\left(\frac{n}{2}\right)} (\chi^2)^{\frac{n-2}{2}} e^{-\frac{\chi^2}{2}}$$

olarak tanımlanmaktadır. Bu fonksiyondaki n sabiti serbestlik derecesi (ş.d.) olarak isimlendirilmekte, γ ise

$$\gamma(k) = (1-k)!$$

veya

$$\gamma(k) = \int_0^\infty x^{k-1} e^{-x}\, dx$$

şeklinde tanımlanan gamma fonksiyonudur.

(2) Langley, R., Practical Statistics, Pan Books Ltd. London 1968, s. 285.

II-2 **Moment Belirleyen Fonksiyon Yaklaşımyla x^2 Olasılık Dağılım Fonksiyonu**

Bir olasılık dağılım fonksiyonu, bu dağılımın moment belirleyen fonksiyonu (MBF) tarafından bire bir olarak belirlenmektedir. Başka bir deyişle her olasılık dağılım fonksiyonunun kendine özgü bir MBF si vardır ve farklı MBF ler değişik olasılık dağılımlarına aittir.

MBF nin bu özelliği, bu fonksiyonların çeşitli dağılımlara ait momentlerin hesaplanmasının yanısıra, tesadüfi değişkenlerin olasılık dağılım fonksiyonları hakkında karar verme konusunda da etkin olmasını sağlamaktadır. Bir tesadüfi değişkenin olasılık dağılım fonksiyonu araştırılırken izlenecek yöntemlerden biri de şöyledir: Tesadüfi değişkenin MBF si hesaplanır ve çeşitli olasılık dağılım fonksiyonlarına ait MBF lerle karşılaştırılır. Bir MBF ile benzerlik bulunması durumunda bu MBF nin ait bulunduğu olasılık dağılım fonksiyonu, tesadüfi değişkenin dağılım fonksiyonu olarak belirlenir.

Bu çalışmada çeşitli tesadüfi değişkenlerin olasılık dağılım fonksiyonu yukarda anlatılan yöntemle araştırılmaktadır. Bu nedenle sıkça kullanılacak olan x^2; dağılımının MBF si aşağıdaki bölümde hesaplanmaktadır.

II-2.1 x^2 Dağılımının Moment Belirleyen Fonksiyonu

f (x) olasılık dağılımına sahip bulunan bir x tesadüfi değişkenin MBF si

$$M_x(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \, f(x) \, dx$$

olarak tanımlanmaktadır.

Bu tanımdan yararlanarak x^2 dağılımının MBF sinin belirlenişi aşağıdaki paragraflarda gösterilmiştir.

Simgeleme kolaylığı nedeniyle

$X^2 =$ v yazılıp MBF tanımı kullanılarak

$$M_v(\theta) = \int_{0}^{\infty} e^{\theta v} \, f(v) \, dv \quad \text{bulunur.}$$

f(v) yerine x^2 dağılım fonksiyonu yazılarak

$$M_v(\theta) = \frac{1}{2^{\frac{n}{2}}\,\gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{\theta v}\, v^{\frac{n-2}{2}}\, e^{-\frac{v}{2}}\, dv$$

ve;

$$= \frac{1}{2^{\frac{n}{2}}\,\gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{-\frac{v}{2}(1-2\theta)}\, v^{\frac{n}{2}-1}\, dv$$

elde edilir.

$$y = \frac{v}{2}(1-2\theta) \quad \text{dönüşümü yardımıyla}$$

$$dv = \frac{2}{1-2\theta}\, dy$$

$$M_v(\theta) = \frac{1}{2^{\frac{n}{2}}\,\gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{-y} \left(\frac{2y}{1-2\theta}\right)^{\frac{n}{2}-1} \frac{2}{1-2\theta}\, dy$$

$$= \frac{(1-2\theta)^{-\frac{n}{2}}}{\gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{-y}\, y^{\frac{n}{2}-1}\, dy$$

şekline dönüştürülmüş olur. Bu ifadedeki integral, gamma fonksiyonu tanımından dolayı $\gamma\left(\frac{n}{2}\right)$ ye eşit olduğundan ve

$$M_v(\theta) = M_{x^2}(\theta) = (1-2\theta)^{-\frac{n}{2}}$$

olması nedeniyle x^2 nin MBF si, n serbestlik derecesi için

$$M_{x^2}(\theta) = (1-2\theta)^{-\frac{n}{2}}$$

şeklinde bulunmaktadır.

II-2.2 x^2 Olasılık Dağılımının Momentleri

Bir olasılık dağılım fonksiyonunun bazı momentlerinin bilinmesi bu dağılımın karakterinin genel çizgileriyle bilinmesini sağlar. MBF nin momentlerin hesaplanmasında kullanılmasına örnek olarak bu bölümde x^2 dağılımının ilk dört momenti hesaplanarak bu momentlere ait özellikler incelenecektir.

Aritmetik ortalamaya göre momentler, sıfıra göre momentler cinsinden aşağıdaki eşitlik yardımıyla hesaplanmaktadır (3).

$$\mu_k = \sum_{j=o}^{k} (-1)^j \binom{k}{j} \mu^j \mu'_{k-j} \tag{2}$$

Böylece

$$\mu'_k = \left. \frac{\partial^k M_{x^2}(\theta)}{\partial\theta^k} \right|_{\theta = o}$$

(3) B. HARRIS, Theory of Probabilitiy, Addison-Wesley, Massachusetts, 1966, s. 101.

$$M_{X^2}(\theta) = (1-2\theta)^{-\frac{n}{2}}$$

ve

$$\binom{k}{j} \quad \frac{k!}{j!\,(k-j)!}$$

hatırlatmalarıyla

X^2 dağılımının momentlerinin hesaplanmasına geçilebilecektir.

(2) Numaralı eşitlik kullanılarak,

$$\mu_1 = \mu \qquad (3)$$

$$\mu_2 = \mu_2' - \mu^2 \qquad (4)$$

$$\mu_3 = \mu_3' - 3\mu\mu_2' + 2\mu^3 \qquad (5)$$

$$\mu_4 = \mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4 \qquad (6)$$

bulunmaktadır. Öte yandan MBF'nin türevleri yardımıyla sıfır etrafındaki momentler

$$\mu = \mu_1' = \mu_1 = \left.\frac{\partial M_{X^2}(\theta)}{\partial \theta}\right]_{\theta=o} = n\,(1-2\theta)^{-\frac{n+2}{2}}\Bigg]_{\theta=o} = n$$

$$\mu_2' = \frac{\partial^2 M_{X^2}(\theta)}{\partial \theta^2} = n(n+2)\,(1-2\theta)^{-\frac{n+4}{2}}\Bigg]_{\theta=o} = n\,(n+2)$$

$$\mu_3' = \frac{\partial^3 M_{X^2}(\theta)}{\partial \theta^2}\; n(n+2)\,(n+4)\,(1-2\theta)^{-\frac{n+6}{2}}\Bigg]_{\theta=o} = n(n+2)\,(n+4)$$

$$\mu'_4 = n(n+2)(n+4)(n+6)(1-2\theta)^{-\frac{n+8}{2}}\Bigg|_{\theta=0} = n(n+2)(n+4)(n+6)$$

olarak bulunmaktadır. Bu değerler sırasıyla (3), (4), (5) ve (6) No.lu eşitliklerde kullanılarak x^2 dağılımının ilk dört momenti

$$\mu = n$$
$$\mu_2 = \sigma^2 = 2n$$
$$\mu_3 = 8n$$
$$\mu_4 = 8n^2 + 48n = 8n(n+6)$$

olarak hesaplanmış olur.

X^2 dağılımının üçüncü ve dördüncü momentleri yardımıyla çarpıklık ve basıklık ölçüleri için standart momentler;

$$\text{çarpıklık ölçüsü, } \alpha_3 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{8n}{(2n)^{\frac{3}{2}}} = \sqrt{\frac{8}{n}},$$

$$\text{basıklık ölçüsü, } \alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{8n(n+6)}{(2n)^2} = 2 + \frac{12}{n}$$

şeklinde hesaplanmaktadır.

II-3 x^2 Olasılık Dağılım Fonksiyonunun Özellikleri

Tesadüfi olaylarda x^2 olasılık dağılım fonksiyonunun uygulanış alanı, bu dağılım fonksiyonunun ve bu dağılıma sahip tesadüfi değişkenlerin özellikleriyle belirlenmektedir. Bu nedenle aşağıda x^2 olasılık dağılımının ve sözkonusu dağılıma sahip tesadüfi değişkenlerin özelliklerine kısaca değinilecektir.

II-3.1 x^2 Olasılık Dağılımına Sahip Tesadüfi Değişkenlerin Toplanabilme Özelliği

X_1 ve X_2 tesadüfi değişkenleri, serbestlik dereceleri n_1 ve n_2 olan ve her biri ki-kare dağılımına sahip iki bağımsız değişken olsun. Bu iki değişken cinsinden $Y=X_1+X_2$ şeklinde tanımlanan Y değişkeni de ki-kare olasılık dağılımına sahiptir.

Bu özellik ki-kare dağılımının MBF si yardımıyla şöylece gösterilebilir.

MBF tanımından, $Y = X_1^2 \quad X_2^2$ tesadüfi değişkeninin MBF si

$$M_Y(\theta) = M_{X_1^2}(\theta).\ M_{X_2^2}(\theta)$$

şeklinde yazılabilmektedir (4).

X_1^2 ve X_2^2 tesadüfi değişkenleri kikare dağılımına sahip olduğundan,

(4) Bir x değişkeninin MBF si $(e^{\theta x})$ in beklenen değeri yani $E\left(e^{\theta x}\right)$ olarak tanımlanır. Buna göre,

$$M_Y(\theta) = M_{X_1^2+X_2^2}(\theta) \quad E\left[e^{\theta(X_1^2+X_2^2)}\right] = E\left[e^{\theta X_1^2} . e^{\theta X_2^2}\right]$$

ve tesadüfi değişkenlerinin bağımsızlığı nedeniyle

$$E\left[e^{\theta X_1^2}\ e^{\theta X_2^2}\right] = E\left[e^{\theta X_1^2}\right] . E\left[e^{\theta X_2^2}\right]$$ eşitliği geçerlidir.

$E\left[e^{\theta X_1^2}\right]$ ve $E\left[e^{\theta X_2^2}\right]$ ise tanıma göre sırasıyla X_1^2 ve X_2^2 de-

$$M_Y(\theta) = M_{X_1^2}(\theta)\ M_{X_2^2}(\theta)$$

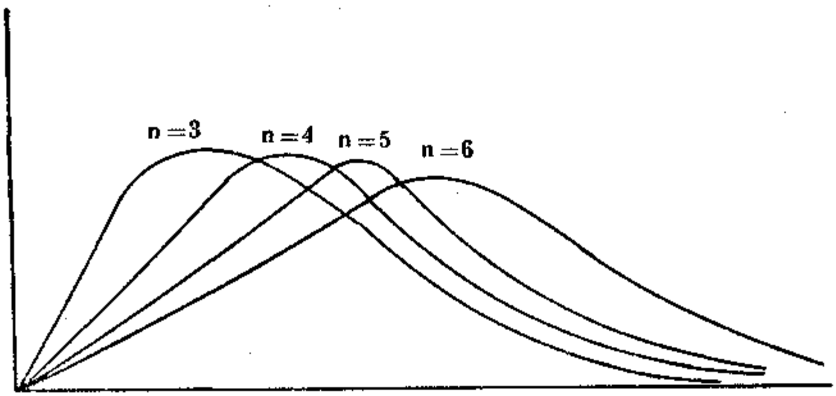
$$M_Y(\theta) = (1-2\theta)^{-\frac{n_1}{2}} (1-2\theta)^{-\frac{n_2}{2}}$$

$$= (1-2\theta)^{-\frac{n_1+n_2}{2}}$$

eşitliği bulunur. Bu ise serbestlik derecesi $n_1 + n_2$ olan bir ki-kare dağılım fonksiyonununu belirlemektedir.

II-3.2 Serbestlik Derecesindeki Değişmelere göre x^2 Dağılımının Özellikleri

X^2 Dağılımının dikkat çeken özelliklerinden biri de serbestlik derecesindeki (s.d.) değişikliklere karşı duyarlı olarak gösterdiği değişikliklerdir. Aşağıdaki şekilde, serbestlik derecesinin bazı değişik değerleri için x^2 eğrisinin alacağı şekiller gösterilmektedir.



Serbestlik derecesindeki değişmelerin x^2 eğrisinde oluşturacağı değişiklikler aşağıdaki şekilde özetlenebilir:

a) n serbestlik derecesi büyüdükçe eğri, standart sapması $\sqrt{2n}$ ve ortalaması n olan normal eğriye yaklaşmaktadır.

b) Serbestlik derecesinin büyümesi halinde $\sqrt{2X^2}$ değişkeninin dağılımı, ortalaması $\sqrt{2n-1}$ ve standart sapması 1 olan normal eğriye yaklaşmaktadır.

c) Yukarıda sayılan özellikler nedeniyle $n > 30$ olması halinde normal dağılım gerçek dağılıma yeterli bir yaklaşım sayılmaktadır (5).

III— x^2 DAĞILIMININ KULLANILMA YERLERİ

İstatistikte karşılaşılan çeşitli tesadüfi değişkenler yukarıda tanımı verilen ve çeşitli özellikleri incelenen x^2 dağılımına sahip bulunmaktadır. Bu bölümün ilk kesiminde sözkonusu dağılım fonksiyonuna sahip bazı tesadüfi değişkenler incelenmekte ikinci kesimde ise x^2 olasılık dağılımından yararlanarak geliştirilmiş bulunan çeşitli x^2 testleri tanıtılarak x^2 dağılımının bu testlerde kullanılış yöntemleri araştırılmaktadır.

III-1 x^2 Olasılık Dağılım Fonksiyonuna Sahip Bazı Tesadüfi Değişkenler

III-1.1 Kareler Toplamının Dağılımı

X tesadüfi değişkeni, ortalaması sıfır, varyansı bir olan ve normal dağılıma sahip bir tesadüfi değişken olsun, $X_1, X_2, \ldots, X_n$ değerleri bu dağılımdan tesadüfen seçilmiş değerler ise, bu durumda

$W = \sum_{i=1}^{n} X_i^2$ şeklinde tanımlanan bir W tesadüfi değişkeni x^2 dağılımına sahip olacaktır.

Bu özellik MBF yaklaşımıyla şu şekilde ispatlanabilmektedir:

$$W = x_1^2 + x_2^2 + \ldots + x_n^2$$

$$M_w(\theta) = M_{x_1^2 + x_2^2 + \ldots + x_n^2}(\theta)$$

(5) BURINGTON and MAY, Handbook of Probalitiy and Statistics, Handbook Publishers Inc., Sandusky-Ohio, 1958, s. 142.

olacaktır.

Diğer taraftan örneklemenin tesadüfi oluşu nedeniyle, X_i ler bağımsız olacak, böylece de

$M_w(\theta) = M_{X_1^2}(\theta) \,.\, M_{X_2^2}(\theta) \ldots M_{X_n^2}(\theta)$ elde edilecektir. (6)

X_i ler aynı dağılım fonksiyonuna sahip olduğundan,

$M_w(\theta) = M^n_{x^2}(\theta)$ eşitliği bulunur.

X değişkeni standart normal bir değişken olduğundan,

$$M_{x^2}(\theta) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{\infty} e^{\theta x^2} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(1-2\theta)} dx$$

bulunur.

$y = x\sqrt{1-2\theta}$

dönüşümü yardımıyla,

$$M_{x^2}(\theta) = (1-2\theta)^{-\frac{1}{2}} \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

$$= (1-2\theta)^{-\frac{1}{2}}$$

bulunacaktır.

Bu sonuç

$M_w(\theta) = M^n_{x^2}(\theta)$

(6) Bkz. s. 275.

eşitliğinde kullanılarak,

$$M_w(\theta) = (1-2\theta)^{-\frac{n}{2}}$$

elde edilir.

Bir olasılık dağılım fonksiyonu, moment belirleyen fonksiyonu tarafından bire bir olarak belirleneceğinden, bu sonuç w tesadüfi değişkeninin serbestlik derecesi n olan bir x^2 dağılım fonksiyonuna sahip olduğunu ortaya koymaktadır.

III-1.2 S 2 nin Dağılımı

Normal dağılıma sahip olduğu bilinen herhangi bir ana kütlenin ortalama ve varyansının bilinmemesi halinde, örnekleme yardimiyle ana kütlenin varyansı hakkında istenilen bir güven sınırı içinde tahmin yapmak S^2 dağılımının bir başka uygulama alanını oluşturmaktadır.

Normal dağılıma sahip olduğu ve varyansının o^2 olduğu bilinen bir ana kütleden seçilen m hacimli bir örneğin varyansının da S^2 olması, halinde, $\frac{mS^2}{\sigma^2}$ değişkeninin, serbestlik derecesi n-1 olan bir ki-kare dağılımına sahip olduğu gösterilmektedir (7).

Örneğin ana kütle varyansı için yapılacak bir tahminin %95 güven sınırları, yukarıda sözü edilen teoremden yararlanılarak aşağıdaki yöntemle belirlenebilir.

n-1 serbestlik derecesi için $X1^2>X2^2$ olması ihtimali %97.5 ve $X1^2<X2^2$ olması ihtimali % 2.5 olacak şekilde x^2 tablolarından bulunacak x_1^2 ve $x2^2$ değerleri

$$\chi_1^2 < \frac{mS^2}{\sigma^2} < \chi_2^2$$

olması ihtimalini 0.95 olarak belirler. Böylece de

(7) İspat için bkz. HOEL, Pul G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., London 1947, s. 136.

$$\frac{mS^2}{\chi_1^2} < \sigma^2 < \frac{mS^2}{\chi_2^2}$$

$$\frac{mS^2}{\chi_1^2} \quad \text{ve} \quad \frac{mS^2}{\chi_2^2}$$

güven sınırlarının bulunmasını sağlayacaktır (8).

III-2 x^2 Testi

II-2.1 Genel Bilgiler

X^2 Dağılımının en geniş kullanılma alanını, sözkonusu dağılımın özelliklerinden yararlanılarak geliştirilmiş bulunan çeşitli X^2 testleri oluşturmaktadır.

X^2 testleri genel olarak, yapılan tesadüfi bir örnekleme sonucu elde edilen verilerin belli bir varsayımın şartlarını sağlayıp sağlamamasmm sınanması problemlerinde kullanılmaktadır.

O_I = Gözlem sonucu elde edilen frekanslar

e_i = Belli bir varsayım altında beklenen frekanslar olmak

üzere

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i} \qquad (7)$$

şeklinde tanımlanan x^2 değişkeni, bütün ,ki-kare testlerinde kullanılmakta olan değişken olarak belirmektedir. Bu değişkenin ya-

(8) ÇÖMLEKÇİ. N., İSTATİSTİK, E.İ.T.İ.A. Yayınları No. 118, Kalite Matbaası, Ankara 1975, s. 250.

pılmış bulunan varsayıma bağlı olmaksızın ,x^2 dağılımına uygunluk göstermesi, ki-kare testlerinin çok geniş bir alanda kullanılmasını sağlar.

x^2 testlerinin değişik problemlerde kullanılışı ile ilgili olarak çeşitli yöntemler geliştirilmiş bulunmaktadır. Aşağıdaki bölümlerde bu yöntemler tanıtılarak x^2 testlerinin kullanılışı incelenecektir.

III-2.2 Kontenjans Tabloları ve x^2 Testi

Rakamsal olarak ölçülemeyen olaylar arasındaki ilişkinin araştırılması, olayların gruplara (ayrılıp) gruplarda gözlenen frekansların incelenmesiyle yapılmaktadır.

İlgilenilen olayların birden fazla vasıflarına ilişkin gözlenen frekansların satır ve sütunlar halinde gösterildiği tabloya «kontenjans tablosu» adı verilir.

Kontenjans tablolarının genel şekli aşağıda gösterildiği gibi olur.

	İkinci Olay					
Birinci Olay	Sınıflar	1	2		a	Toplam
	1	n_{11}	n_{12}		n_{1a}	n_{10}
	2	n_{21}	n_{22}		n_{2a}	n_{20}
	b	n_{b1}	n_{b2}		n_{ba}	n_{b0}
	Toplam	n_{01}	n_{02}		n_{0a}	m

Yukarıdaki b satır, a sütünlu tabloda (n_{ij}) , i'nci satırın j'nci sütununda gözlenen frekansı; (n_{i0}) i'nci satırın toplamını; (n_{oj}) j'nci sütunun toplamını; m ise genel toplamı vermektedir.

Olaylar arasında belli bir hipoteze göre bir ilişkinin olup olmadığı, beklenen frekanslarla gözlenen frekansların karşılaştırılması ile incelenmektedir. (7) no.lu eşitlikle tanımlanan ve $\%^2$ dağılımına sahip olduğu bilinen değişken, gözlenen ve beklenen frekansların farklılaşmasının bir ölçüsü olarak kullanılır.

Kontenjans tablolarında her satırın toplamı belli olduğuna göre, her satırdaki elemanlardan biri diğerleri ve satır toplamı yardımıyla hesaplanabilir. Buna göre her satırdaki a-1 frekansın bilinmesi diğerinin hesaplanabilmesi için yeterlidir. Diğer taraftan sütun toplamları da bilindiğine göre (b-1) frekans sonucunun hesaplanabilmesi için yeterlidir. Bu yorumla a sütunlu, b satirli bir kontenjans tablosunda (a-1). (b-1) adet frekansın diğerlerinin hesaplanmasına olanak bırakacak şekilde değişebilecek frekans sayısı olduğu görülecek ve n serbestlik derecesi n = (a-1). (b-1) olarak belirlenecektir.

Eğer teorik frekanslar ancak ana kütle parametrelerine dayanılarak tahmin edilebiliyorsa, M tahmin edilecek ana kütle parametre sayısı olarak, serbestlik derecesi,

n= (a-1). (b-1) — M olmaktadır (9).

Kontenjans tablolarında hesaplanan x^2 değişkeninin x^2 dağılım tablolarıyla karşılaştırılması, seçilen hipotezin sınanmasını sağlamakta olup bu yöntem x^2 testinin kontenjans tablolarındaki uygulanışı olmaktadır. Bu uygulama şu şekilde özetlenip genelleştirilebilir:

a) Gözlenen frekanslar yardımıyla kontenjans tabloları oluşturur. Frekansların toplamı N, gözlenen birim sayısına eşit olmalıdır.

b) Hipoteze göre grupların herbirinde beklenen değerler hesaplanır.

(9) Çömlekçi, N. a.g.e., s. 254.

c) $\chi^2 = \Sigma \dfrac{(0_{ij}-e_{ij})^2}{e_{ij}}$ değeri hesaplanır.

d) Serbestlik derecesi (a-1) (b-1) olarak belirlenir.

e) Güven sınır belirlenir.

f) x^2 tabloları yardımıyla (a-1) (b-1) serbestlik derecesi ve seçilen güven sınırı kullanılarak tablo değeri bulunur. Tablo değeri, hesaplanan x^2 değerinden büyükse gözlenen frekansların beklenen frekanslardan sapmasının fazla olmaması nedeniyle hipotez veya varsayım geçerlidir, aksi halde reddedilir.

III-2.3 Uygunluk Sınamalarında x^2 Testi

Uygunluk sınamaları x^2 dağılımının geniş bir uygulama alanını oluşturur. Genel olarak uygunluk sınamaları, gözlenen frekansların belli bir dağılıma uygun olup olmadığının araştırılması konusunda kullanılmaktadır. Kontenjans tablolarındaki uygulamaya benzer şekilde, iki veya daha çok grupta gözlenen birim sayısı ile belli bir H_0 hipotezine göre beklenen birim sayısı arasındaki farkların incelenmesi H_0 hipotezinin sınanmasına olanak vermektedir.

Sınıflar	1	2	. . .	k	Toplam
Birim Sayıları	o_1 (e_1)	o_2 (e_2)	. . .	o_k (e_k)	N

Uygunluk sınamalarında x^2 testi yukarıdaki türden tablolar kullanılarak uygulanmaktadır.

Bu tabloda (o_i) i'nci sınıfta gözlenen birim sayısının (e_i) H_0 hipotezine göre i'nci sınıfta beklenen birim sayısını göstermektedir.

(7) numaralı eşitlikte tanımlanan ve ki-kare dağılımına sahip x^2 tesadüfi değişkeni, gözlenen ve beklenen frekanslar' arasındaki farklılaşmanın bir ölçüsü olarak H_0 hipotezinin sınanmasını sağlar.

Uygunluk sınamalarında gözlenen birim sayıları belli olduğuna göre k-1 adet frekansın bilinmesi bütün frekansların bulunması için yeterli olacak ve serbestlik derecesi k-1 olacaktır.

Uygunluk sınamaları şu şekilde yapılmaktadır:

a) n adet gözlenen frekans k gruba toplanır.

b) Ho hipotezinde kabul edilen dağılıma göre beklenen frekanslar (e_i) hesaplanır.

c) $\chi^2 = \Sigma \frac{(0_i - e_i)^2}{e_i}$ değeri hesaplanır.

d) k-1 serbestlik derecesi belirlenir ve a güven sınırı seçilir.

e) k-1 serbestlik derecesi ve a güven sınırı kullanılarak tablo değeri bulunur. Tablo değeri hesaplanan x^2 değerinden büyükse H_0 kabul edilir, aksi halde gözlenen frekansların beklenen frekanslardan çok farklı oluşu nedeniyle hipotez reddedilir (10).

III-2.4 x^2 Testinin Uygulanabilme Koşulları

X^2 testlerinin kontenjans tabloları ve uygunluk sınamalarında kullanılışı ile ilgili olarak bazı koşulların sağlanması gerekir. Bu koşullar şu şekilde sıralanabilir:

a) x^2 olasılık dağılım fonksiyonu sürekli bir fonksiyondur. Oysa kontenjans tablolarında kullanılacak beklenen frekansların sınırlı olması nedeniyle, sürekli bir fonksiyona yeterli yaklaşım sağlanabilmesi için örnekteki birim sayısının büyük olması gerekir. Genellikle frekansların herbirinin 5 ten büyük olması halinde X^2 testlerinin yeterli olduğu kabul edilmektedir. Teorik frekansların çok küçük olması halinde x^2 değişkeni,

$$\chi^2 = \Sigma \frac{(\,|\,0_i - e_i\,|\, - 0.5)^2}{e_i}$$

(10) Uygunluk testlerinde kullanılan tablolar genel olarak tek satirli bir kontenjans tablosu niteliğindedir. Bu nedenle hipotezin sınanmasında kullanılan yöntemler benzer olmaktadır.

şeklinde değiştirilerek Yates düzeltmesi kullanılır. Teorik frekansların 5 ile 10 arasında olması halinde x^2 değişkeninin düzeltilmiş ve düzeltilmemiş değerleri karşılaştırılır. Farklı sonuçların elde edilmesi halinde örnek hacmi arttırılmalıdır.

b) Tablodaki frekanslar için kullanılacak kısıtlamalar (sınıflardaki frekans toplamlarının sınıftaki birim sayısına eşit olması gibi) birinci dereceden yani doğrusal olmalıdır.

Bu şartların sağlanması durumunda x^2 olasılık dağılım fonksiyonu kullanılarak kontenjans tabloları ve uygunluk sınamaları için yeterli sonuçlar elde edilebilmektedir.

IV— x^2 TESTİNE İLİŞKİN BİR UYGULAMA

— ÖĞRENCİNİN LİSEYİ BİTİRME DERECESİ İLE BABA TAHSİLİ ARASINDAKİ İLİŞKİNİN İNCELENMESİ —

IV-1 **Problemin Tanımı**

Ailedeki büyüklerin öğrenim durumlarının çocukların öğrenimlerindeki başarılarına etkili olup olmadığı sıkça tartışılan bir konudur. Bu tartışmalar çoğu kez öğrencinin başarısında, diğer etkilerin yanısıra, aile büyüklerinin öğrenim durumlarının büyük bir ağırlık taşıyacağı sonucuna ulaşmaktadır.

Bu bölümde, 1974 öğrenim yılında Eskişehir İ.T.İ. Akademisine kaydolan öğrenciler arasında Dr. İmdat Kara tarafından yapılan bir anketten sağlanan veriler kullanılarak yukarıdaki sorunun matematiksel bir yaklaşımı incelenmektedir.

Öğrencinin başarısının ölçümü olarak Liseden mezun oluş derecesi, büyüklerin öğrenim durumuna örnek olarak babanın öğrenimi ele alınmaktadır.

Liseden mezun oluş derecesinin öğrencinin başarısını tek başına ne denli ölçebileceği ayrıca tartışabilecek bir konu olabilir. Ancak, çalışmanın amacı öğrencinin başarısına etki eden faktörlerin araştırılması veya incelenmesi değildir. Uygulamamız bu tür araştırmalarda kullanılabilecek yöntemlerden ancak birini oluşturabilir, öte yandan rakamsal olarak değerlendirilmeyen olayların matematiksel yöntemlerle incelenmesine bir örnek olacağı düşünülerek çalışmanın kapsamına alınmasında bir sakınca görülmemiştir.

IV-2 x^2 Test Uygulanışı

1974 yılında E.İ.T.İ. Akademisine giren öğrenciler arasında yapılan anket incelenerek aşağıdaki tablo elde edilmiştir.

Liseden mezun oluş derecesi / baba öğrenim durumu	ORTA	İYİ	Toplam
İLKOKUL	42	18	60
ORTA OKUL	20	8	28
LİSE	21	9	30
YÜKSEK	16	6	22
Toplam	99	41	140

Ankete katılan öğrenciler arasında Liseyi pekiyi ile bitiren öğrenciye rastlanmamıştır.

Sınanmak istenen nitelik baba tahsil durumunun öğrencinin lise bitirme derecesinden bağımsız olup olmadığıdır. Bu nedenle,

H_0 hipotezi: Babanın tahsil durumu ile öğrencinin lise bitirme derecesi arasında bağımlılık yoktur.

şeklinde seçilir.

İncelenen 140 öğrenciden 99 tanesi liseyi orta derece ile bitirdiğinden topluluktan rasgele seçilen bir öğrencinin liseyi orta derece ile bitirmiş olma olasılığı

$$\frac{99}{140} = 0.707143$$ olacaktır.

Öte yandan, incelenen 140 öğrenciden 60 tanesinin babasının ilkokul öğrenimli olması nedeniyle rasgele seçilen bir öğrencinin babasının ilkokul öğrenimli olma olasılığı

$$\frac{60}{140} = 0.428571$$ olarak hesaplanacaktır.

H_0 hipotezine göre bu iki olay bağımsız olduğundan seçilen herhangi bir öğrencinin liseyi bitirme derecesinin orta; babasının tahsil durumunun ilkokul olması olasılığı, yukarıdaki olasılıkların çarpımı şeklinde

0.707143 x 0.428571 = 0.303061 olarak hesaplanacaktır.

Toplam olarak 140 öğrenci incelemeye katıldığından ve rasgele seçilen bir öğrenci için lise bitirme derecesinin orta, baba öğreniminin ilkokul olması olasılığı 0.303061 olduğundan bu şartları sağlayan öğrenci sayısının beklenen değeri

0.303061 x 140 = 42.4 olacaktır.

Yukarıda birinci satır ve birinci sütun için anlatılan işlemler diğer satır ve sütunlar için yapılarak aşağıdaki beklenen değerler tablosu elde edilir.

Lise bitirme derecesi / baba tahsil durumu	ORTA	İYİ
İLOKUL	42,4	17,5
ORTA OKUL	19,8	8,1
LİSE	21,2	9,4
YÜKSEK	15,5	6.4

Bu iki tablo yardımıyla

$$\chi^2 = \Sigma \frac{(0_i - e_i)^2}{e_i}$$ değeri

$X^2 = 0.08$ olarak hesaplanmaktadır.

Sıra sayısı, 4 sütun sayısı 2 olduğuna göre serbestlik derecesi n = (4-1) (2-1) = 3 olarak bulunur. Bu değerler kullanılarak ve % 99 güven derecesine göre ki-kare tablolarından x^2 tablo değeri 0.115 olarak bulunmaktadır.

IV—3 YORUM

Babanın öğrenim durumu ile öğrencinin lise bitirme derecesi arasındaki ilişkinin incelenmesinde uygulanan x^2 testi sonucunda; hesaplanan x^2 değerinin x^2 Tablo değerinden küçük olması nedeniyle H_0 hipotezi % 99 güven derecesinde kabul edilmektedir. Buna göre öğrencinin başarısı ile babanın öğrenim durumu arasında bir bağımlılık olmadığı % 99 güvenirlilikle söylenebilmektedir.

V— SONUÇ

X^2 Olasılık dağılım fonksiyonu rakamsal olarak değerlenemeyen olaylar arasındaki ilişkilerin incelenmesinde geniş bir uygulama alanına sahip bulunmaktadır. Bu dağılımdan yararlanarak geliştirilmiş bulunan çeşitli testler istatistikte değişik problemlerin çözümünde yarar sağlar.

Bu çalışmada x^2 olasılık dağılım fonksiyonu, özellikleri, ve geliştirilmiş çeşitli x^2 testleri incelenmiştir.

Çalışmanın son bölümünde bir uygulamaya yer verilmiş ve 1974 yılında Eskişehir İktisadi Ticari İlimler Akademisine kaydolan öğrenciler arasında yapılan bir anketten elde edilen veriler kullanılarak öğrencinin lise bitirme derecesi ile babasının öğrenim durumu arasındaki ilişkinin incelenmesinde ki-kare test uygulanışı ile elde edilen sonuç yorumlanmıştır.

YARARLANILAN KAYNAKLAR

BURINGTON and MAY : Handbook of Probability and Statististics, Hanbook Publishers Inc., Ohio, 1958.

CHOU. Y. Statistical Analysis, Holt, Rinehart and Winston, London, 1963.

ÇÖMLEKÇİ. N. İstatistik (İkinci Baskı), E.İ.T.I.A. Yayınları, Ankara, 1975.

FREUND. J. E. Mathematical Statistics, Prentice-Hall Inc. N.J., 1962.

HARRIS. B. Theory of Probability, Addison-Wesley, Massachusetts, 1966.

HOEL, P.G. Introduction to Mathematical Statistics, John Wiley and Sons Inc., London, 1947.

LANGLEY, R. Practical Statistics, Pan Books Ltd., London, 1968.

RAHMAN. N.A. A Course in Theoretical Statistics, Griffin, London, 1968.

SIEGEL. S. Nonparametric Satistics For the Behavioral Sciences, McGraw-Hill, New York, 1956.

YULE and KENDALL : An Introduction to Theory of Statistics, Griffin, London, 1957.