13.1. Khi-kare Dağılımı

 $(X_1, X_2, X_3, \ldots, X_n)$ ile gösterilen n adet gözlem değerinden oluşan bir X tesadüfi değişkeni tanımlayalım.

Bu tesadüfi değişkenin, beklenen değeri yani ortalaması,

$$E(X)=\mu$$

ve varyansı da,

$$V(X) = \sigma^2$$

olmaktadır.

Söz konusu tesadüfi değişkenin değerlerinin, standart değer dönüşümü ile standart hale getirildiğini varsayalım.

Standart değer dönüşümü,

$$z_i = \frac{X_i - \bar{X}}{\sigma}$$

formülüyle yapılmaktadır. Bu dönüşüm ile X tesadüfi değişkeninin her bir değerine karşılık bir z - standart değeri tanımlanmış olur.

z standart değerlerinin ortalaması sıfır ve varyansı bir olacak şekilde normal dağıldığı bilinmektedir.

(X_1 , X_2 , X_3 ,, X_n) değerlerine karşılık, standart değer dönüşümü ile hesapladığımız (z_1 , z_2 , z_3 ,, z_n) değerlerinin karelerini alarak tanımlayacağımız yeni z^2 tesadüfi değişkenin dağılımı Khi-kare dağılımına uyacaktır.

Khi-kare dağılımı, şekli serbestlik derecesine bağlı olarak değişmekte olan asimetrik bir olasılık dağılımıdır.

Khi-kare testleri olarak bilinen test grubunda, bağımsızlık, homojenlik, belirli bir dağılıma uygunluk testleri bulunmaktadır. Bu kitabın konusu içinde bizim ilgi alanımızı Khi-kare bağımsızlık testi oluşturmaktadır.

13.2. Khi-kare Bağımsızlık Testi

Khi-kare bağımsızlık testi, değişkenlerden en az birinin nominal ya da ordinal olması durumunda kullanılan ve iki değişken arasında ilişki olup olmadığını tespit etmeye yarayan bir testtir.

Söz gelimi bir TV programının izlenmesi ile cinsiyet arasında ilişki olup olmaması, bir ilçede yaşayan ve düşük-orta-yüksek eğitimli şeklinde ayrılan vatandaşların eğitim durumları ile belediye hizmetlerinden memnuniyetleri arasında ilişki bulunup bulunmaması, bebeklerin prematüre ve zamanında dünyaya gelmiş olması ile gelişim düzeyleri arasında bir ilişki bulunup bulunmaması araştırılmak istendiğinde Khi-kare bağımsızlık testinden yararlanılır.

Khi-kare bağımsızlık testinde toplanan veriler **kontenjans tablosu** adı verilen tablolar ile düzenlenir.

Kontenjans tabloları içerdikleri satır ve sütun sayılarına göre, 2x2'lik kontenjans tablosu, 2x3'lük kontenjans tablosu vb. şeklinde adlandırılır. Burada kontenjans tablosunun boyutlarını belirleyen kriter, değişkenlerin kaç kategoriye ayrılmış olduğudur. Sözgelimi cinsiyet ile belediye hizmetlerinden memnun olup olmama durumu için bir kontenjans tablosu oluşturulduğunda, cinsiyet kadın ve erkek şeklinde 2 kategoride değerlendirildiğinden, memnun olup olmama durumu da benzer şekilde memnun ve memnun değil şeklinde 2 kategoride değerlendirildiğinden 2x2'lik bir tablo söz konusu olmaktadır.

Memnun olup olmama durumu, memnun değil, orta derecede memnun ve çok memnun gibi 3 kategoride değerlendirilmiş olsaydı 2x3'lük bir kontenjans tablosu ile karşı karşıya olacaktık.

Aşağıda bir kontenjans tablosu örneği verilmiştir:

Bir banka, kendisinden internet bankacılığı hizmeti alan müşterilerinin memnuniyet derecelerini belirlemek istiyor ve müşterilerinden online olarak kısa bir anket formu doldurmalarını istiyor.

Müşterilere, internet bankacılığı hizmetinden memnun olup olmamalarının yanında, cinsiyet, yaş, eğitim, medeni durum gibi demografik sorular da soruluyor.

Müşterilerin cinsiyetine göre memnuniyet durumları aşağıdaki kontenjans tablosu ile düzenleniyor:

Tablo 13.1 : İnternet Bankacılığı Hizmeti Alan Müşterilerin Cinsiyete Göre Memnuniyet Durumu							
	Memnun		Memnun değil		Toplam		
Kadın	150	135,33	130	144,67	280		
Erkek	140	154,67	180	165,33	320		
Toplam	290		310		600		

Yukarıdaki kontenjans tablosu Kadın-Erkek şıkları (2 şık) ile Memnun-Memnun değil şıklarının (2 şık) çaprazlandığı 2x2 boyutlu bir tablodur.

Tabloda, cinsiyet ve memnuniyet değişkenlerinin şıklarının çaprazlanmasından oluşan 4 tane hücre bulunmaktadır:

Memnun olan kadın : 150 kişi

Memnun olan erkek : 140 kişi

Memnun olmayan Kadın : 130 kişi

Memnun olmayan erkek : 180 kişi

Değişkenlerin şıklarının çaprazlandığı hücrelerin içinde yer alan değerlere, **gözlenen frekanslar** denmektedir. Yukarıdaki kontenjans tablosunda gözlenen frekansları görüyoruz (siyah renkli olanlar).

Öte yandan araştırmaya katılanların cinsiyet dağılımına baktığımızda,

Kadın sayısı : 280 kişi

Erkek sayısı : 320 kişi

Memnuniyet durumunun dağılımı da,

Memnun olanlar : 290 kişi

Memnun olmayanlar : 310 kişi

şeklindedir.

Değişkenlerin şıklarının frekans toplamlarına marjinal toplamlar denmektedir. Tabloda yer alan şıklara ilişkin satır ve sütun toplamları marjinal toplamlardır ve yukarıda da marjinal toplamlar görülmektedir.

İki değişken arasında ilişki bulunmaması durumunda hücre freknaslarının ne olması gerektiğini gösteren frekanslara **beklenen frekanslar** denir.

Beklenen frekanslar, ilgili hücrenin satır ve sütun toplamlarının çarpılarak genel toplama bölünmesiyle hesaplanır.

Örneğin,

Memnun olan kadın hücresi için beklenen frekans:

$$\frac{290.280}{600} = 135,33 \text{ kişi}$$

Memnun olan erkek hücresi için beklenen frekans:

$$\frac{290.320}{600} = 154,67 \text{ kişi}$$

Memnun olmayan kadın hücresi için beklenen frekans:

$$\frac{310.280}{600} = 144,67 \text{ kişi}$$

Memnun olmyan erkek hücresi için beklenen frekans:

$$\frac{310.320}{600} = 165,33 \text{ kişi}$$

biçiminde hesaplanmaktadır.

Beklenen frekanslar kontenjans tablosunda kırmızı ile gösterilmiştir. Beklenen frekansların satır ve sütun toplamlarının gözlenen frekansların satır ve sütun toplamlarına eşit olduğuna dikkat ediniz.

Şimdi Khi-kare bağımsızlık testine uygun verilerin toplanarak kontenjans tablosu şeklinde düzenlenmiş olduğunu varsayarak Khi-kare bağımsızlık testinin nasıl uygulandığını aşama aşama öğrenelim:

1- Khi-kare bağımsızlık testinin test hipotezlerinin kurulması;

Khi-kare bağımsızlık testinin hipotezleri,

H₀: İki değişken birbirinden bağımsızdır ya da iki değişken arasında ilişki yoktur

H₁: İki değişken bağımsız değildir ya da iki değişken arasında ilişki vardır şeklinde kurulmaktadır.

2- Khi-kare bağımsızlık testinin test istatistiğinin hesaplanması:

Khi-kare bağımsızlık testinin test istatistiği,

$$\chi^2_{hes} = \sum_{i=1}^n \frac{(f_0 - f_i)^2}{f_i}$$

şeklinde hesaplanmaktadır. Formülde yer alan,

f₀: kontenjans tablosunda yer alan gözlenen frekansları,

 f_i : kontenjans tablosundaki her gözlenen frekansa karşılık olarak hesaplanan beklenen frekansları sembolize etmektedir.

Kontenjans tablosunda bulunan her hücre için bir beklenen frekans değeri hesaplanıp yukarıdaki formül kullanılarak her hücrenin khi-kare test istatistiğine yapacağı katkı miktarı belirlenmektedir.

Daha önce de ifade ettiğimiz gibi beklenen frekanslar, iki değişken arasında ilişki bulunmaması durumunda olması gereken frekansların neler olduğunu ifade eder. Dolayısıyla, gözlenen frekansların beklenen frekanslara yakın olması durumunda iki değişken arasında ilişki bulunmadığı anlaşılmaktadır. Gözlenen ve beklenen frekansların birbirine eşit olması, aralarındaki farkın sıfır olmasına sebep olacak bu durum da iki değişkenin tam bağımsız olduğu sonucunu doğuracaktır.

Benzer şekilde beklenen frekanslar ile gözlenen frekanslar arasında önemli farklılıklar gözlenmesi durumunda ise iki değişkenin ilişkili olduğu anlaşılmaktadır.

Khi-kare testinin bir zayıflığı, serbestlik derecesi arttıkça Khi-kare istatistiğinin değerinin artmasıdır. Bu durum, gerçekte ilişkili olmayan değişkenlerin aralarında ilişki olduğu şeklinde yanlış bir yoruma sebep olabilir.

3- Khi-kare tablo değerinin bulunması:

Khi-kare test istatistiğinin dağılımının khi-kare dağılımına uyduğunu daha önce ifade etmiştik.

Khi-kare bağımsızlık testinde tablo değeri α anlamlılık düzeyinde ve v= (n-1)(k-1) serbestlik derecesinde Khi-kare tablosundan elde edilir.

Serbestlik derecesi formülünde yer alan,

n : satır sayısını

k : sütun sayısını

ifade etmektedir. Satır ve sütun sayıları belirlenirken toplam satır ve sütunları dikkate alınmamaktadır.

4- Khi-kare testinde karar alma:

Khi-kare bağımsızlık testinde karar kriteri diğer tüm hipotezlerinde olduğu gibidir ve

$$\chi^2_{hes} > \chi^2_{tablo}$$
 ise H_0 Ret, khi-kare tablo $\chi^2_{hes} < \chi^2_{tablo}$ ise H_0 kabul

edilmektedir.

Khi-kare bağımsızlık testinin aşamalarını bu şekilde açıkladıktan sonra şimdi de konuyla ilgili örnek uygulamalar yapalım.

13.3. Khi-kare Bağımsızlık Testi Uygulamaları

Khi-kare bağımsızlık testini değişik boyutlu kontenjans tabloları üzerinde uygulayarak aşama çözüm yapalım:

Örnek:

Bir Televizyon programının izleyici kitlesi araştırılıyor. Bu amaçla 500 kişiyle görüşülerek, TV programını izleyip izlemedikleri soruluyor ve elde edilen bulgular aşağıdaki tablo ile düzenleniyor. Buna göre cinsiyetin programın izlenmesi üzerinde etkili bir değişken olup olmadığını araştıralım (α = 0,05).

	İzliyor		İzlemiyor		Toplam
Kadın	70	115	180	135	250
Erkek	160	115	90	135	250
Toplam	230		270		500

Öncelikle hipotezlerimizi belirleyelim:

H₀: Cinsiyet ile programın izlenmesi arasında ilişki yoktur.

H₁: Cinsiyet ile programın izlenmesi arasında ilişki vardır.

Şimdi de khi-kare test istatistiğini hesaplayalım:

$$\chi^2_{hes} = \sum_{i=1}^n \frac{(f_0 - f_i)^2}{f_i}$$

Yukarıdaki formül ile Khi-kare hesaplanmış değerini yani Khi-kare test istatistiğini bulabilmek için öncelikle her bir hücre için beklenen frekansların neler olduğunu hesaplamamız gerekiyor.

Kontenjans tablosunda kadın-erkek ve izliyor-izlemiyor şıklarının çaprazlandığı 4 adet hücre bulunmaktadır.

70 değerinin görüldüğü programı izleyen kadın sayısının yer aldığı hücrenin beklenen frekansı, bu hücrenin satır ve sütun toplamlarının çarpılıp, toplam gözlem sayısına bölünmesiyle elde edilir. Yani,

$$\frac{230.250}{500} = 115$$

şeklinde hesaplıyoruz.

Benzer şekilde diğer hücrelerin de beklenen frekanslarını da hesaplayalım:

$$\frac{270.250}{500} = 135$$
 $\frac{230.250}{500} = 115$ $\frac{270.250}{500} = 135$

Yukarıdaki gibi hesapladığımız beklenen frekans değerlerini, kontenjans tablosunun ilgili hücrelerinde kırmızı olarak gösterdik.

Gözlenen ve beklenen frekans toplamlarının eşit olduğuna dikkat çekmek isteriz (örneğin, 70+160=230 ve 115+115=230).

Beklenen frekanslar iki değişkenin birbirinden bağımsız olması durumunda gerçekleşmesi beklenen frekanslardır. Bağımsızlık olması durumunda, frekansların şıklara oransal olarak eşit dağılacağını varsayar. Dolayısıyla, gözlenen ve beklenen frekanslar arasındaki farklar ne denli fazla ise iki değişkenin ilişkili olma ihtimali o derece artmaktadır.

Şimdi de khi-kare test istatistiğini hesaplayalım:

$$\chi^{2}_{hes} = \sum_{i=1}^{n} \frac{(f_{0} - f_{i})^{2}}{f_{i}}$$

$$\chi^{2}_{hes} = \frac{(70 - 115)^{2}}{115} + \frac{(180 - 135)^{2}}{135} + \frac{(160 - 115)^{2}}{115} + \frac{(90 - 135)^{2}}{135}$$

Khi-kare test istatistiğinin hesaplanmasında kullandığımız yukarıdaki formülde, dört tane gözlenen ve beklenen frekansın farkları üzerinden hesaplanan katkı değerleri yer

almaktadır. Bu değerlerden her biri, o hücrenin khi-kare test istatistiğine yaptığı katkı miktarını göstermektedir.

$$\chi^{2}_{hes} = \frac{(70 - 115)^{2}}{115} + \frac{(180 - 135)^{2}}{135} + \frac{(160 - 115)^{2}}{115} + \frac{(90 - 135)^{2}}{135}$$

$$\chi^{2}_{hes} = \frac{(-45)^{2}}{115} + \frac{(45)^{2}}{135} + \frac{(45)^{2}}{115} + \frac{(-45)^{2}}{135}$$

$$\chi^{2}_{hes} = \frac{2025}{115} + \frac{2025}{135} + \frac{2025}{115} + \frac{2025}{135}$$

$$\chi^{2}_{hes} = 17,61 + 15 + 17,61 + 15$$

$$\chi^{2}_{hes} = 65,22$$

$$\alpha$$
 = 0,05 ve v= (n-1)(k-1) yani (2-1)(2-1)=1 serbestlik derecesinde tablo değerimiz;
$$\chi^2_{tablo} = 3,841$$

olmaktadır. Bu durumda,

 $\chi^2_{hes} > \chi^2_{tablo}$ olduğundan H_0 reddedilmekte ve alternatif hipotez kabul edilmektedir.

Sonuç olarak, TV programının izlenmesi ile cinsiyet arasında bir ilişki bulunduğunu % 95) olasılıkla (ya da % 5 hata olasılığıyla ($\alpha = 0.05$)) kabul etmiş oluyoruz.

Kontenjans tablosunda iki değişkenin şıklara dağılımına baktığımızda, programı izleyenlerin genel olarak erkekler olduğu anlaşılmaktadır. Dolayısıyla cinsiyet bu TV programının izlenmesi üzerinde etkili bir faktördür ve programı daha çok erkekler izlemektedir.

Örnek:

Bir hastanede dünyaya gelen bebeklerin annelerinin yaşı ile bebeklerin doğum ağırlıkları arasında ilişki bulunup bulunmadığı araştırılıyor. Bu amaçla 200 gebe kadın takibe alınarak aşağıdaki bulgulara ulaşılıyor:

	30 yaş altı		30-40 yaş ara	IS1	41 yaş ve üstü	Toplam
Düşük kilolu	30	48	40	44	50 28	120
Normal kilolu	90	72	70	66	20 42	180
Toplam	120		110		70	300

Öncelikle hipotezlerimizi belirleyelim:

H₀: Annelerin yaşı ile bebeklerin doğum ağırlığı arasında ilişki yoktur.

H₁: Annelerin yaşı ile bebeklerin doğum ağırlığı arasında ilişki vardır.

Khi-kare test istatistiğini hesaplayabilmemiz için öncelikle her bir hücre için beklenen frekansları hesaplamamız gerekiyor.

$$\frac{120.120}{300} = 48 \qquad \frac{110.120}{300} = 44 \qquad \frac{70.120}{300} = 28$$

$$\frac{120.180}{300} = 72 \qquad \frac{110.180}{300} = 66 \qquad \frac{70.180}{300} = 42$$

İki değişkenin bağımsız olması durumunda beklenen frekansların neler olması gerektiği yukarıdaki şekilde hesaplanarak kontenjans tablosunda ilgili hücrelerin içinde kırmızı olarak gösterilmiştir. Gözlenen ve beklenen frekansların satır ve sütun toplamlarının eşit olduğuna dikkat çekerek bu eşitliğin sağlanması ile herhangi bir işlem hatasının yapılmadığının da kontrolünün sağlanmış olacağını ifade etmek isteriz.

Şimdi de khi-kare test istatistiğini hesaplayalım:

$$\chi^{2}_{hes} = \sum_{i=1}^{n} \frac{(f_{0} - f_{i})^{2}}{f_{i}}$$

$$\chi^{2}_{hes} = \frac{(30 - 48)^{2}}{48} + \frac{(40 - 44)^{2}}{44} + \frac{(50 - 28)^{2}}{28} + \frac{(90 - 72)^{2}}{72} + \frac{(70 - 66)^{2}}{66} + \frac{(20 - 42)^{2}}{42}$$

$$\chi^{2}_{hes} = \frac{(-18)^{2}}{48} + \frac{(-4)^{2}}{44} + \frac{(22)^{2}}{28} + \frac{(18)^{2}}{72} + \frac{(-4)^{2}}{66} + \frac{(-22)^{2}}{42}$$
$$\chi^{2}_{hes} = \frac{324}{48} + \frac{16}{44} + \frac{484}{28} + \frac{324}{72} + \frac{16}{66} + \frac{484}{42}$$

$$\chi^2_{hes} = 6.75 + 0.3636 + 17.29 + 4.5 + 0.2424 + 11.52$$

$$\chi^2_{hes} = 40.66$$

Khi-kare test istatistiğini hesapladığımız yukarıdaki çözümün son işlem satırında, her bir hücrenin khi-kare istatistiğine yaptığı katkı görülmektedir. Bu katkı değerlerinin toplanması sonucunda Khi-kare test istatistiğine ulaşılmaktadır.

Hücrelerin Khi-kare test istatistiğine katkı değerleri ve toplam Khi-kare değeri:

$$6,75 + 0,3636 + 17,29 + 4,5 + 0,2424 + 11,52 = 40,66$$

hesaplanmıştır.

 $\alpha = 0.05$ ve v= (n-1)(k-1) yani (2-1)(3-1)=2 serbestlik derecesinde tablo değerimiz;

$$\chi^2_{tablo} = 5,991$$

olmaktadır. Bu durumda,

 $\chi^2_{hes} > \chi^2_{tablo}$ olduğundan H₀ reddedilmekte ve alternatif hipotez kabul edilmektedir.

Dolayısıyla, annelerin yaşı ile bebeklerin doğum ağırlığı arasında ilişki olduğu % 5 hata payıyla ve % 95 olasılıkla kabul edilmektedir.

Yukarıdaki kontenjans tablosunda, iki değişkenin şıklarının hücrelere nasıl dağıldığına baktığımızda, genç anneler daha çok normal kilolu bebek dünyaya getirirken; ileri yaştaki annelerin düşük ağırlıklı bebek dünyaya getirdiklerini görüyoruz.

Bu Bölümde Neler Öğrendik?

Khi-kare bağımsızlık testini incelediğimiz bu bölümde, öncelikle Khi-kare dağılımı üzerinde durarak, dağılımın temel özelliklerini ele aldık ve şekli serbestlik derecesine bağlı olarak değişen asimetrik bir dağılım olduğunu öğrendik.

Aralarında ilişki olup olmadığı araştırılan iki değişkenden en az birinin nominal ya da ordinal ölçekle ölçülmüş olması durumunda kullanılmakta olan Khi-kare bağımsızlık testinin nasıl dizayn edildiği üzerinde durduk.

Öte yandan, iki değişkenin birbirinden bağımsız olması durumunda, toplam frekansın hücrelere nasıl dağılması gerektiğini ifade eden beklenen frekansların nasıl hesaplanacağını öğrendik.

Gözlenen ve beklenen frekansların birbirine eşit olması durumunun iki değişkenin birbirinden tam olarak bağımsızlığını gösterdiğini ve gözlenen ve beklenen frekanslar arasındaki farklar büyüdükçe ise iki değişkenin birbiriyle ilişkili olma olasılığının arttığını gördük.

Khi-kare bağımsızlık testinin aşamalarını, test istatistiğinin ve tablo değerinin nasıl bulunduğunu, karar kuralını inceledik ve test sonucunda iki değişken arasında ilişki olduğu kabul edildiğinde, bu ilişkinin nasıl bir ilişki olduğunun kontenjans tablosundaki hücrelerin gözlenen frekanslarının yorumlanması ile nasıl belirlendiğini öğrendik.

Bölüm Soruları

Khi-kare dağılımı için aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

1-

	I-	Khi-kare dağılımı simetrik bir olasılık dağılımıdır.
	II-	Khi-kare dağılımı asimetrik bir olasılık dağılımıdır.
	III-	Khi-kare dağılımının şekli serbestlik derecesine göre değişir.
	IV-	z-standart değerlerinin dağılımı Khi-kare dağılımına uyar.
	V-	Khi-kare dağılımı ortalaması sıfır olan bir olasılık dağılımıdır.
	a-	I ve III
	b-	II ve III
	c-	I ve V
	d-	II ve V
	e-	II ve IV
	2-	Nominal ya da ordinal verilerin düzenlendiği tablolaradenir.
	a-	Kontenjans tablosu
	b-	Korelasyon tablosu
	c-	Olasılık tablosu
	d-	Khi-kare tablosu
	e-	Beklenen değer tablosu
• • • • •	3-	İki değişkenin ilişkisiz olması durumunda geçerli olması gereken frekanslaradenir.
	a-	Gözlenen frekans
	b-	Toplam frekans
	c-	Beklenen frekans
	d-	Marjinal frekans
	e-	Frekans
	4- de	Bir araştırma sonucunda verilerin düzenlenmesi ile oluşan frekanslara nir.
	a-	Gözlenen frekans
	b-	Toplam frekans
	c-	Beklenen frekans
	d-	Marjinal frekans
	e-	Frekans

- 5- Kontenjans tablosunda satır ve sütun toplamlarınadenir.
- a- Gözlenen frekans
- b- Toplam frekans
- c- Beklenen frekans
- d- Marjinal frekans
- e- Frekans
- 6- İki değişken arasında tam bağımsızlık olması durumunda khi-kare değeri olur.
 - a- Bir
 - b- Sıfır
 - c- Yüz
 - d- Sonsuz
 - e- Negatif

Aşağıda bir derginin satın alınması üzerinde cinsiyetin etkili bir faktör olup olmadığına yönelik olarak yapılan bir çalışmaya ilişkin kontenjans tablosu verilmektedir.

	Dergiyi düzenli okuyor	Dergiyi düzenli okumuyor	Toplam
Kadın	30	20	50
Erkek	15	35	50
Toplam	45	55	100

Aşağıdaki soruları yukarıdaki kontenjans tablosuna göre yanıtlayınız.

- 7- Kontenjans tablosunun serbestlik derecesi nedir?
- a- 2
- b- 3
- c- 0
- d- 1
- e- 4
- 8- Dergiyi düzenli okuyan kadınlar için beklenen değer nedir?
- a- 30
- b- 45
- c- 50
- d- 17,5
- e- 22,5

- 9- Dergiyi düzenli okuyan erkek hücresinin Khi-kare değerine katkısı nedir?
- a- 15
- b- 5
- c- 2,5
- d- 7,5
- e- 22,5
- 10- Dergiyi düzenli okumayan kadın hücresinin Khi-kare değerine katkısı nedir?
- a- 20
- b- 55
- c- 50
- d- 7,5
- e- 2,05

Cevaplar: