

Kİ-KARE (χ^2) DAĞILIMI ve Kİ-KARE TESTLERİ (HOMOJENLİK, UYUM, BAĞIMSIZLIK KONTROLLERİ)

10.1. Giriş

Yürütülen bir çalışmadan elde edilen veriler sayısal veriler olabileceği gibi nominal veya ordinal veriler olabilir. BÖLÜM 1’de açıklandığı gibi nominal “isim ile belirtilmiş” anlamında kullanılmaktadır. İsimlendirilmiş (nominal) veriler (erkek / kadın), (sarı / beyaz), (var / yok) vb şekilde elde edilmiş verilerdir. Bu tip veriler sadece isimle belirtilmiştir ve analiz aşamasında isim verilerek yeni kategoriler (sınıflar) oluşturabilir. Ordinal veriler ise sıralı veya sıralandırılmış verilerdir. Eğer üzerinde durulan değişkene ait değerler sıralanabilir kategorilerden oluşuyorsa bu tip veriler sıralı veya sıralandırılmış verilerdir. Örneğin bir sınav sonucunda öğrenciler A, B, C gibi notlar alabilir. Bu notlar öğrencilerin sınavdaki yeterliliklerine göre verilir ve A alan öğrenci B alan öğrenciden daha çalışkan olduğu verilen nottan anlaşılır. Diğer bir örnek olarak at yarışlarını verebiliriz. Gerek nominal gerekse ordinal verilerin elde edilmesinde bunların sayıları (frekansları) üzerinde durulur. Yani bu tür veriler sayılarak elde edilirler. Erkek sayısı, kadın sayısı, A notunu alanların sayısı, B notunu alanların sayısı,...vb. Ölçmek suretiyle elde edilen veriler de sonradan belirli bir kritere kategorik hale dönüştürülebilir. İnsanlarda tansiyon bilindiği üzere mmHg olarak ölçülmekte, ancak ifade edilirken düşük, normal, yüksek şeklinde kategorik hale dönüştürülmektedir. Bunun gibi fen, sosyal bilimlerde daha bir çok örnek bulmak mümkündür.

Kategorik değişkenlerden elde edilen veriler araştırmacı tarafından belirlenen kategoriler (sınıflar) içinde yer alır. İsimlendirilmiş veya sıralandırılmış veriler için araştırmacı kategoriler oluşturmuş ise bu şekilde elde edilen kategorik verilerin analizi ki-kare testleri kullanılarak yapılır. Ki-kare testleri, kategorik değişkenlerin dağılımlarının birbirinden farklı olup olmadığını kontrol etmek için kullanılır.

Ki-kare testleri, gözlenen frekansların teorik olarak beklenen frekanslardan farklı olup olmadığını kontrol etmek için kullanılır. Ki-kare testleri, sayarak elde edilmiş frekansların çeşitli kategorik sınıflara dağılımlarını incelemek amacıyla kullanılmaktadır. Dolayısıyla örneğin veya populasyonun ortalama ve varyansı ile ilgilenmez. Ki-kare testleri iki gruba ayrılır:

1. Uyum kontrolleri: Araştırmacının gözlediği frekanslar ile teorik olarak beklenen frekansları karşılaştırarak gözlenen frekansların beklenen frekanslar ile uyum içinde olup olmadığını kontrol eder. Bu kontroller homojenlik, belirtilen oranlara uyum veya belirli istatistik dağılımlara uyum kontrolleridir.
2. Bağımsızlık kontrolleri: Sayılarak elde edilen verilerin iki veya daha fazla sayıdaki kategorik faktöre göre olan dağılımlarının, söz konusu kategorik faktörlerden bağımsız olup olmadığını ele aldığı çalışmalardır.

Ki-kare testlerinde kontrol veya karşıt hipotezlerinden hangisinin kabul edileceğine karar verme aşamasında test dağılımı olarak ki-kare dağılımı kullanılır.

10.2. Ki-Kare (χ^2) Dağılımı

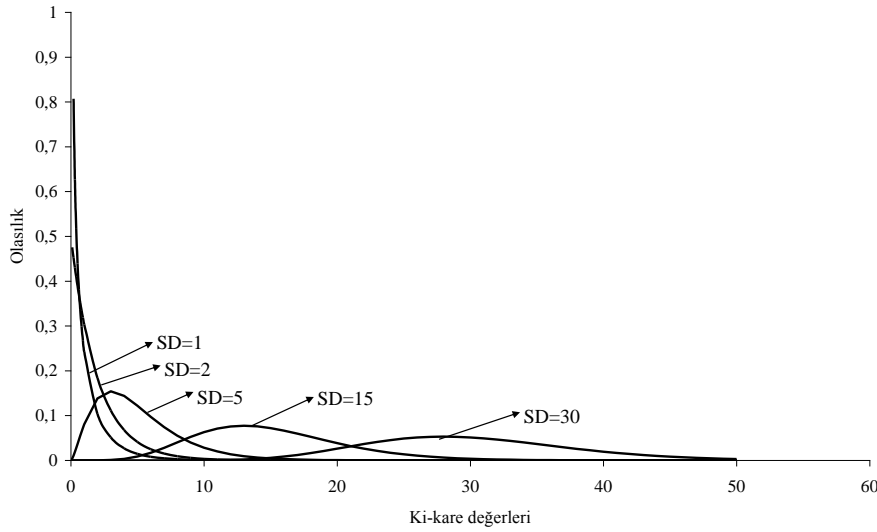
Standart normal dağılımdan tesadüfen v adet Z -değeri alınarak (10.1) numaralı eşitlikte görüldüğü gibi bunların teker teker kareleri alınıp toplansa bir tane χ^2 değeri elde edilir. Bu işlem mümkün olan sayıda tekrarlanırsa v serbestlik dereceli ki-kare (χ^2_v) dağılımı elde edilir.

$$\chi^2_v = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + \dots + Z_v^2 = \sum_{i=1}^v Z_i^2 \quad \dots(10.1)$$

(10.1) numaralı eşitlik kullanılarak hesaplanan ki-kare değerleri (10.2) numaralı eşitlikte verilen olasılık yoğunluk fonksiyonuna uygun dağılım gösterir.

$$f(\chi^2_v) = \frac{1}{2^{\frac{v}{2}} \left(\frac{v}{2}\right)!} (\chi^2)^{\frac{(v-2)}{2}} e^{-\frac{\chi^2}{2}} \quad \dots(10.2)$$

(10.2) numaralı eşitlikte, v , serbestlik derecesidir (SD). (10.2) numaralı eşitlikte verilen olasılık yoğunluk fonksiyonundan anlaşıldığı gibi ki-kare dağılımı serbestlik derecesine bağlı bir dağılımdır. Yani ki-kare dağılımının bir parametresi vardır ve bu da serbestlik derecesidir. Diğer bir deyişle sonsuz sayıdaki ki-kare dağılımları birbirlerinden serbestlik dereceleri ile ayrılırlar. Ki-kare dağılımının şekli parametresine (serbestlik derecesine) göre değişmektedir (Şekil 10.1)



ŞEKİL 10.1. Farklı serbestlik dereceli ki-kare dağılımları. Grafikte, SD: serbestlik derecesidir.

İsmlendirilmiş ve sıralandırılmış şekilde elde edilen veriler için kategoriler oluşturulduğu zaman araştırmacının her kategori için gözlediği frekansların yanı sıra belirli oranlara ve özelliklere göre beklediği frekanslar vardır. Her bir kategori için gözlenen ve beklenen frekanslardan (10.3) numaralı eşitlik kullanılarak hesaplanacak değerler de belirli varsayımlar altında ki-kare dağılımı gösterirler.

$$\chi^2 = \sum_{i=1}^k \frac{[\text{Gözlenen frekans}(f) - \text{Beklenen frekans}(f')]^2}{\text{Beklenen frekans}(f')} \quad \dots(10.3)$$

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

(10.3) numaralı eşitlikte, k: kategori (sınıf) sayısı, f: her bir sınıf için gözlenen frekans, f' : belirli oranlara ve özelliklere göre her bir sınıf için beklenen frekanslardır. (10.3) numaralı eşitlik kullanılarak hesaplanan ki-kare değerinin serbestlik derecesi yapılan kontrol sırasında parametre yerine kullanılan istatistik sayısına bağlı olarak (k-1), (k-2), (k-3)...vs. olabilir. (10.3) numaralı eşitlik kullanılarak hesaplanacak ki-kare değerinin ki-kare dağılımı gösterebilmesi için her sınıf için hesaplanacak beklenen frekansın 5'ten küçük olmaması gerekir.

Ki-kare testleri uygulanırken serbestlik derecesinin 1 olması durumunda hesaplanan ki-kare değerinin ki-kare dağılımına daha iyi yaklaşması için (10.4) numaralı eşitlikte verildiği gibi Yates düzeltmesinin yapılması gerekir.

$$\chi^2 = \sum_{i=1}^k \frac{[|f - f'| - 0.5]^2}{f'} \quad \dots(10.4)$$

Ki-kare dağılımı, serbestlik derecesine bağlı tek taraflı ve sürekli bir dağılımdır. Dağılımın şekli serbestlik derecesine bağlı olarak değişir. Şekil 10.1'de görüldüğü gibi 1 serbestlik dereceli ki-kare dağılımı X ve Y eksenlerine $+\infty$ 'da asimptot oluşturur. 2 serbestlik dereceli ki-kare dağılımı Y-eksenini keser ve X-eksenine $+\infty$ 'da asimptot oluşturur. 3 ve daha fazla serbestlik dereceli ki-kare dağılımları 0'dan başlar ve (SD-2) noktasına kadar artarak maksimum (tepe değeri) oluşturur. Bu noktadan itibaren azalarak X-eksenine $+\infty$ 'da asimptot oluşturur. Bütün Ki-kare dağılımlarının ortalaması serbestlik derecesine (SD), varyansı ise serbestlik derecesinin iki katına (2 SD) eşittir. Şekil 10.1'den görüldüğü gibi serbestlik derecesi arttıkça ki-kare dağılımı simetrikleşir ve normal dağılıma yaklaşır.

Ki-kare dağılımı serbestlik derecesine bağlı bir dağılım olduğu için sonsuz tane ki-kare dağılımı vardır. Farklı serbestlik dereceli ki-kare dağılımlarında farklı yüzdelik alanların başladığı ki-kare değerleri Tablo D'de verilmiştir.

10.3. Homojenlik Kontrolü

Homojenlik kontrolü, oluşturulan kategorilere (sınıflara) göre araştırmada dikkate alınan bireylerin dağılımının homojen olup olmadığını, yani oluşturulan sınıflar arasında, her bir sınıfta bulunan birey sayısı bakımından farklılığın önemli olup olmadığını kontrol eder.

ÖRNEK 1:

A, B, C, D ve E gibi 5 çeşit yem bitkisi tohumu eşit oranda belirlenen bir alana ekilmiştir. 2 yıl sonra bu alandan toplanan 200 bitkinin tohum çeşitlerine göre dağılımı Tablo 10.1'de verildiği gibi gözlenmiştir. Toplanan bitkilerin dağılımının homojen olduğu, yani çeşitlerin dağılımının tohumda olduğu gibi homojen kaldığı söylenebilir mi?

Tablo 10.1. Toplanan bitkilerin çeşitlere göre dağılımı

Bitki çeşidi	Gözlenen frekans (f)	Beklenen frekans (f')	$\frac{(f - f')^2}{f'}$
A	40	40	0.000
B	35	40	0.625
C	25	40	5.625
D	55	40	5.625
E	45	40	0.625
Toplam	200	200	$\chi^2 = 12.500$

Yapılan çalışmanın amacı, eşit oranda ekilen bitki tohumlarının eşit oranda kalıp kalmadığının, yani çıkan bitkilerin çeşitlere göre dağılımının homojen olup olmadığını araştırılmasıdır. Bunun için χ^2 -homojenlik kontrolünün uygulanması gerekir.

Daha önce yapılan hipotez kontrollerinde açıklandığı gibi ilk olarak kontrol ve karşıt hipotezlerin aşağıdaki şekilde kurulması gerekir.

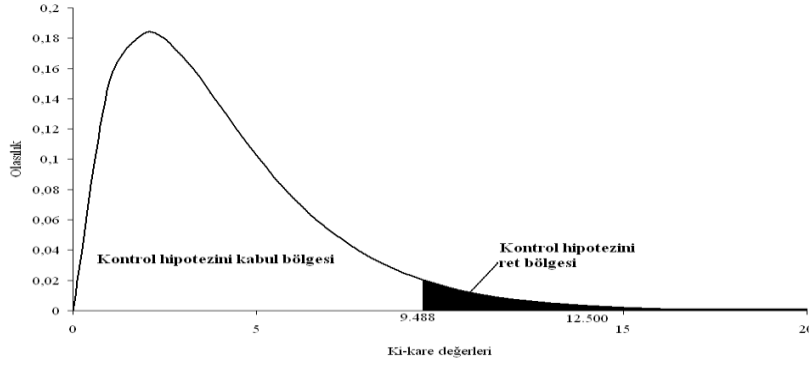
H₀: Toplanan bitkilerin çeşitlere göre dağılımı homojendir. Tohum çeşitleri arasında toplanan bitki sayısı bakımından fark tesadüften ileri gelmiştir. Kısaca, $(f - f') = 0$, yani bir tohum çeşidi için toplanan bitki sayısı ile toplanması beklenen bitki sayısı arasındaki fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir.

H₁: Toplanan bitkilerin çeşitlere göre dağılımı homojen değildir Tohum çeşitleri arasında toplanan bitki sayısı bakımından fark tesadüften ileri gelmemiştir. Kısaca, $(f - f') \neq 0$, yani bir tohum çeşidi için toplanan bitki sayısı ile toplanması beklenen bitki sayısı arasındaki fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez.

χ^2 -homojenlik kontrolünde χ^2 -istatistiği (10.3) numaralı eşitlik kullanılarak hesaplanır. Bunun için ilk olarak her bitki çeşidinden kaç bitkinin toplanması beklendiğinin hesaplanması gerekir. Çalışmada farklı çeşitlerden tohumların eşit oranda ekildiği ve kontrol hipotezinde de toplanan bitkilerin çeşitlere göre dağılımı homojen olduğu ileri sürüldüğüne göre toplanan bitkilerin eşit olarak bitki çeşitlerine dağılması gerekir, yani her bitki çeşidinden $200/5=40$ bitkinin toplanması beklenir. Bu her bir çeşit için beklenen frekanstır. χ^2 -kontrolleri yapılırken gözlenen ve beklenen frekansların toplamı her zaman birbirine eşittir. Her bir çeşit için beklenen frekanslar hesaplandıktan sonra χ^2 -değeri (10.3) numaralı eşitlik kullanılarak aşağıdaki şekilde hesaplanır ve:

$$\chi^2 = \frac{(40 - 40)^2 + (35 - 40)^2 + (25 - 40)^2 + (55 - 40)^2 + (45 - 40)^2}{40} = 12.5$$

χ^2 -değeri 12.5 olarak bulunur. Bu kontrolde sınıf sayısı 5 olduğundan serbestlik derecesi $(k-1)=4$ 'tür. Eğer yapılan kontrolde I. tip hata olasılığı %5 olarak kararlaştırılmışsa kritik χ^2 -değeri Tablo D'den 9.488 olarak bulunur. Şekil 10.2'de görüldüğü gibi Tablo D'den bulunan değer ki-kare dağılımını kontrol hipotezini kabul ve ret olmak üzere iki bölgeye ayırır.

ŞEKİL 10.2. 4 serbestlik dereceli ki-kare dağılımında H_0 hipotezini ret ve kabul bölgeleri

Şekil 10.2’de görüldüğü gibi hesaplanan ki-kare değerinin, 4 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %5’den küçüktür. Yani kontrol hipotezinin ret bölgesinde yer almaktadır. Bu sebeple kontrol hipotezi rey edilir. Yapılan homojenlik kontrolü sonucunda toplanan bitkilerin çeşitlere göre dağılımının homojen olmadığı, diğer bir deyişle tohum çeşitleri arasında toplanan bitki sayısı bakımından fark tesadüften ileri gelmediği kararına varılır.

ÖRNEK 2:

Bir bulvardan bir hafta boyunca geçen araçların günlere göre dağılımı Tablo 10.2’de verildiği gibi gözlenmiştir. Söz konusu bulvardan geçen araçların haftanın günlerine göre olan dağılımlarının homojen olduğu söylenebilir mi?

Tablo 10.2. Bir bulvardan geçen araçların haftanın günlerine göre dağılımı

Haftanın günleri	Gözlenen araç sayısı (f)	Beklenen araç sayısı (f')	$\frac{(f - f')^2}{f'}$
Pazartesi	425	369	8.499
Salı	355	369	0.531
Çarşamba	270	369	26.561
Perşembe	455	369	20.043
Cuma	550	369	88.783
Cumartesi	278	369	22.442
Pazar	250	369	38.377
Toplam	2583	2583	$\chi^2 = 205.236$

Yapılan çalışmada söz konusu bulvardan hafta boyunca geçen araçların haftanın günlerine göre dağılımının homojen olup olmadığı araştırılmaktadır. χ^2 -homojenlik testi uygulanarak kontrol edilecek hipotezler aşağıdaki şekilde kurulur.

- H₀:** Söz konusu bulvardan hafta boyunca geçen araçların günlere göre dağılımı homojendir. Günler arasında bulvardan geçen araç sayısı bakımından gözlenen fark tesadüften ileri gelmektedir. Kısaca, $(f-f')=0$ dir. Yani bulvardan bir günde geçen araç sayısı ile geçmesi beklenen araç sayısı arasındaki fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir.
- H₁:** Söz konusu bulvardan hafta boyunca geçen araçların günlere göre dağılımı homojen değildir. Günler arasında bulvardan geçen araç sayısı bakımından gözlenen fark tesadüften ileri gelmemektedir. Kısaca, $(f-f') \neq 0$ dir. Yani bulvardan bir günde geçen araç sayısı ile geçmesi beklenen araç sayısı arasındaki fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez.

Eğer kontrol hipotezi doğru ise bir hafta boyunca bulvardan geçen araçların günlere eşit olarak dağılması gerekir. Bu sebeple bulvardan haftanın her günü geçmesi beklenen araç sayısı $2583/7=369$ 'dur. Her bir gün için beklenen araç sayısı hesaplandıktan sonra χ^2 -değeri (10.3) numaralı eşitlik kullanılarak aşağıdaki şekilde hesaplanır.

$$\chi^2 = \frac{(425-369)^2 + (355-369)^2 + \dots + (250-369)^2}{369} = 205.236$$

Haftanın 7 günü olduğu için serbestlik derecesi $(7-1)=6$ 'dır. Eğer yapılan kontrolde I. tip hata olasılığı %1 olarak kararlaştırılmışsa Tablo D'den, 6 serbestlik dereceli ki-kare dağılımında %1'lik alanın 16.812'den başladığı bulunur. Hesaplanan ki-kare değerinin 7 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %1'den küçüktür. Bu sebeple kontrol hipotezi ret edilir. Yapılan homojenlik kontrolü sonucunda bulvardan geçen araçların haftanın günlerine göre olan dağılımlarının homojen olmadığı, yani haftanın günleri arasında bulvardan geçen araç sayısı bakımından farkın istatistik olarak önemli olduğu kararına varılır.

10.4. Uyum Kontrolleri

Yapılan bir araştırmada üzerinde durulan kategorilerde (sınıflarda) gözlenen frekansların her sınıf için bildirilen oranlarla uyum içinde olup olmadığı kontrol edilebileceği gibi toplanan verilerin belirli bir istatistik dağılıma uygun dağılıp dağılmadıkları da kontrol edilebilir.

10.4.1. Belirtilen Oranlara Uyum Kontrolü

ÖRNEK 1:

Bir ormandaki ağaçların %15'inin meşe, %20'sinin Ladin, %15'inin çam, %20'sinin köknar, %20'sinin ıhlamur ve %10'unun da kestane ağacı olduğu bildirilmiştir. Söz konusu ormandan tesadüfen seçilen 350 ağacın çeşitlere göre dağılımı Tablo 10.3'deki gibi gözlenmiştir. Bu ormandaki ağaçlar için bildirilen oranlar doğru mudur?

Bu çalışmanın amacı söz konusu ormanda bulunan ağaç çeşitleri için bildirilen oranların doğru olup olmadığını kontrol etmektir. Eğer ormandaki ağaç çeşitleri için bildirilen oranlar doğru ise her çeşit için gözlenen sayı ile beklenen sayılar arasındaki fark tesadüften ileri gelmelidir. Tesadüfen 350 seçildiğine göre 350 ağaçtan her çeşidin beklenen frekansı, her çeşit için bildirilen oranla seçilen ağaç sayısı çarpılarak bulunmuş ve Tablo 10.3'te verilmiştir.

Tablo 10.3. 350 ağacın çeşitlere göre dağılımı ve her çeşit için beklenen sayılar

Ağaç çeşitleri	Gözlenen ağaç sayısı (f)	Oran	Beklenen ağaç sayısı (f')	$\frac{(f - f')^2}{f'}$
Meşe	48	% 15	350(0.15)=52.5	0.386
Ladin	65	% 20	350(0.20)=70.0	0.357
Çam	56	% 15	350(0.15)=52.5	0.233
Kök nar	76	% 20	350(0.20)=70.0	0.514
Ihlamur	64	% 20	350(0.20)=70.0	0.514
Kestane	41	% 10	350(0.10)=35.0	1.029
Toplam	350	% 100	350	$\chi^2 = 3.033$

Her çeşit için beklenen frekanslar hesaplandıktan sonra χ^2 -değeri (10.3) numaralı eşitlik kullanılarak aşağıdaki şekilde bulunur.

$$\chi^2 = \frac{(48-52.5)^2}{52.5} + \frac{(65-70)^2}{70} + \dots + \frac{(64-70)^2}{70} + \frac{(41-35)^2}{35} = 3.033$$

Belirtilen ağaç çeşidi 6 olduğundan serbestlik derecesi 5'tir. Eğer yapılan kontrolde I. tip hata olasılığı %1 olarak kararlaştırılmışsa Tablo D'den 5 serbestlik dereceli ki-kare dağılımında %1'lik alanın 15.086 değerinden başladığı görülür. Hesaplanan ki-kare değerinin 6 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %1'den büyüktür. Dolayısıyla kontrol hipotezi kabul edilir. Yapılan uyum kontrolü sonucunda ormandaki ağaç çeşitleri için belirtilen oranların doğru olduğu, yani tesadüfen seçilen ağaların, ağaç çeşitlerine göre dağılımının belirtilen oranlar ile uyum içinde olduğuna karar verilir.

ÖRNEK 2:

Bir zar 120 kere atılmış ve her yüzün kaç kez geldiği Tablo 10.4'teki gibi gözlenmiştir.

Tablo 10.4. 120 zar atışının zarın yüzlerine göre dağılımı ve her yüz için beklenen atış sayısı

Zarın yüzleri	Gözlenen Zar yüzü sayısı (f)	Oran	Beklenen zar yüzü sayısı (f')	$\frac{(f - f')^2}{f'}$
1	15	1/6	120(1/6)=20	1.25
2	12	1/6	120(1/6)=20	3.20
3	22	1/6	120(1/6)=20	0.20
4	19	1/6	120(1/6)=20	0.05
5	24	1/6	120(1/6)=20	0.80
6	28	1/6	120(1/6)=20	3.20
Toplam	120	1.00	120	$\chi^2 = 8.70$

Bir zar hilesiz olarak atıldığı zaman her bir yüzün gelme olasılığı 1/6'dır. Zar 120 kere atılarak her yüzün gelme sayısı tablo 10.4'te verilmiştir. Eğer zar hilesiz olarak atılmış ise 120 atışın, zarın

yüzlerine 1/6 oranında dağılmış olması, yani Tablo 10.4'te görüldüğü gibi her bir yüzün $120(1/6)=20$ kere gelmiş olması beklenir. Zar, hilesiz olarak atıldığı zaman her bir yüz 20 kere gelmemiş olsa bile beklenen atış sayısı ile gözlenen atış sayıları arasındaki farklılığın tesadüften ileri geliyor olması gerekir. Gözlenen ve beklenen atış sayıları arasındaki farklılığın tesadüfi olup olmadığını kontrol etmek için gözlenen atış sayısının beklenen atış sayısı ile uyum için olup olmadığı (10.3) numaralı eşitlikten ki-kare değeri hesaplanarak aşağıdaki şekilde yapılır.

$$\chi^2 = \frac{(15-20)^2 + (12-20)^2 + \dots + (24-20)^2 + (28-20)^2}{20} = 8.70$$

Bir zarın 6 yüzü olduğu için serbestlik derecesi 5'tir. Yapılan kontrolde I. tip hata olasılığı %5 olarak kararlaştırılmışsa Tablo D'den 5 serbestlik dereceli ki-kare dağılımında %5'lik alanın 11.070 değerinden başladığı görülür. Hesaplanan ki-kare değerinin ($\chi^2= 8.70$) 5 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %5'den büyüktür. Dolayısıyla kontrol hipotezi kabul edilir. Yapılan uyum kontrolü sonucunda zarın hilesiz olarak atıldığı, yani 120 atışın zarın yüzlerine göre dağılımının belirtilen oranlara ile uyum içinde olduğu kararına varılır.

10.4.2. Dağılımlara Uyum Kontrolü

Üzerinde çalışılan her özellik dağılım fonksiyonu belirlenmiş bir dağılım gösterir. Çalışılan özelliklerin en yaygın olarak gösterdiği dağılımlar binomiyal, Poisson ve normal dağılımlardır. Bu dağılımlar BÖLÜM IV'te açıklanmıştır. Yapılan araştırmalarda üzerinde çalışılan özelliğe ait toplanan verilerin, BÖLÜM IV'te açıklanan dağılımlardan birine uyum gösterip göstermediği dağılımlara uyum kontrolü yardımıyla belirlenebilir.

10.4.2.1. Binom Dağılımına Uyum Kontrolü

Bölüm 4.1.2'de herhangi bir dersten öğrencilerin %60'nın başarılı olduğu saptanmıştır. Başarı oranı %60 olarak saptanan bu öğrencilerden 5'er öğrencilik 250 örnekte başarılı öğrenci sayısı bakımından gözlenen ve $\pi=0.60$ olan binomiyal dağılımına göre beklenen frekanslar Tablo 4.3'te verilmişti. 250 örnekte başarılı öğrenci sayısı bakımından gözlenen ve $\pi=0.60$ olan binomiyal dağılımına göre beklenen frekansları belirleyen araştırmacı, 250 örneğin başarılı öğrenci sayısı bakımından dağılımının $\pi=0.60$ olan binomiyal dağılıma uygun olup olmadığını kontrol edebilir.

Yapılan çalışmada oluşturulan 250 tesadüf örneğinde başarılı öğrenci sayısı bakımından dağılımın $\pi=0.60$ olan binomiyal dağılıma uygun olup olmadığını kontrol etmek için önce kontrol ve karşıt hipotezlerin kurulması gerekir.

H₀: Üzerinde çalışılan 250 örnekte başarılı öğrenci sayısı, $\pi=0.60$ ve $n=5$ olan binomiyal dağılıma uygun dağılmaktadır. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: Üzerinde çalışılan 250 örnekte başarılı öğrenci sayısı, $\pi=0.60$ ve $n=5$ olan binomiyal dağılıma uygun dağılmamaktadır. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Ki-kare testlerinde ki-kare değeri hesaplanırken beklenen frekansların 5'ten küçük olmaması gerekir. Tablo 10.5'te 1. sınıfın, yani 250 örnekte başarılı öğrenci olmaması durumu için beklenen frekans 2.56 olarak hesaplanmış olup 5'ten küçüktür. Bu sebeple 1. ve 2. sınıflar birleştirilerek ki-kare değeri hesaplanmıştır.

250 örneğin başarılı öğrenci sayısı bakımından dağılımının $\pi=0.60$ olan binomiyal dağılıma uygun olup olmadığını kontrol etmek için ki-kare değerinin hesaplanması Tablo 10.5'te gösterilmiştir.

TABLO 10.5. 5'er öğrencilik 250 örnekte başarılı öğrenci sayısı bakımından gözlenen ve $\pi=0.60$ olan binomiyal dağılımına göre beklenen frekanslar ve binomiyal dağılım uygunluk kontrolü için ki-kare değerinin hesaplanması

Başarılı öğrenci sayısı (r)	Gözlenen frekans (f)	Beklenen frekans (f')	$\frac{(f - f')^2}{f'}$
0	4	2.56	1.262
1	23 27	19.20 21.76	
2	54	57.60	0.225
3	91	86.40	0.245
4	61	64.80	0.223
5	17	19.44	0.306
Toplam	250	250	$\chi^2=2.261$

Araştırmacı I. tip hata olasılığını %5 olarak belirlemiş olsun. Beklenen frekanslar bulunurken popülasyona ait başarı oranı kullanılmıştır. Beklenen frekanslar 250 örnekten hesaplandığı için serbestlik derecesi $5-1=4$ 'tür. 4 serbestlik dereceli ki-kare dağılımında %5'lik alan Tablo D'den 9.488 olarak bulunur. Hesaplanan χ^2 -değerinin ($\chi^2=2.261$) 4 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %5'ten büyüktür. Yani kontrol hipotezinin kabul bölgesindedir. Bu durumda kontrol hipotezi kabul edilir. Üzerinde çalışılan 250 örnekte başarılı öğrenci sayısının, $\pi=0.60$ ve $n=5$ olan binomiyal dağılıma uygun dağıldığı kararına varılır.

Ki-kare testi kullanılarak binomiyal dağılıma uyum kontrolünde serbestlik derecesi hesaplanırken, sınıf sayısından 1 ve popülasyonun parametresi olan π yerine örnekten hesaplanan p kullanıldığı için de 1 olmak üzere sınıf sayısından iki çıkarılır. Eğer popülasyona ait istenen olayın oluş olasılığı biliniyor ise serbestlik derecesi hesaplanırken sınıf sayısından için 1 çıkarılır.

10.4.2.2. Poisson Dağılımına Uyum Kontrolü

Bölüm 4.2.3'te, sebze yetiştiricileri için hazırlanan ve içinde yaklaşık olarak 500 biber tohumu bulunan ambalajlardan 250 adet rastgele alınmış ve bu ambalajlardaki tohumlardan ölü veya canlı olanların sayısı belirlenmiş ve içlerinde yaklaşık olarak 500 biber tohumu bulunan 250 ambalaj için ölü tohum sayısı için gözlenen ve Poisson dağılımına göre beklenen frekanslar Tablo 4.6'da verilmiştir.

Bu çalışmada, 250 tohum ambalajındaki ölü tohumların ortalaması 2.48 olan Poisson dağılımına uygun bir dağılım gösterip göstermediği ki-kare testi kullanılarak kontrol edilecektir.

H₀: İçinde yaklaşık 500 adet biber tohumu bulunan 250 ambalajdaki ölü tohum sayısı ortalaması 2.48 olan Poisson dağılımına uygun bir dağılım göstermektedir. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: İçinde yaklaşık 500 adet biber tohumu bulunan 250 ambalajdaki ölü tohum sayısı ortalaması 2.48 olan Poisson dağılımına uygun bir dağılım göstermemektedir. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Tablo 10.6'da 7. ve 8. sınıfların beklenen frekansı 5'ten küçük olduğu için son üç sınıf birleştirilmiştir. 250 biber tohumu ambalajında ölü tohum sayısının dağılımının ortalaması 2.48 olan Poisson dağılımına uygun olup olmadığını kontrol etmek için ki-kare değerinin hesaplanması Tablo 10.6'da gösterilmiştir.

Araştırmacı I. tip hata olasılığını %5 olarak belirlemiş olsun. Bu kontrolde serbestlik derecesi (7-2)=5'tir. Çünkü 250 biber ambalajının alındığı tohum popülasyonu için ölü tohum sayısı örnekten tahmin edilmiş ve beklenen ölü tohum sayısı 250 ambalajlık örnekten hesaplanmıştır. 5 serbestlik dereceli ki-kare dağılımında %5'lik alan Tablo D'den 11.070 olarak bulunur.

Tablo 10.6. 250 ambalaj için gözlenen ve Poisson dağılımına göre beklenen frekanslar ve Poisson dağılımına uygunluk kontrolü için ki-kare değerinin hesaplanması

Ölü tohum sayısı	Gözlenen frekans (f)	Beklenen frekans (f')	$\frac{(f - f')^2}{f'}$
0	31	20.94	4.833
1	50	51.92	0.071
2	56	64.38	1.091
3	44	53.22	1.597
4	38	33.00	0.758
5	20	16.37	0.805
6	5	6.76	0.068
7	4	2.40	
8	2	1.01	
Toplam	250	250	$\chi^2=9.223$

Hesaplanan χ^2 -değerinin ($\chi^2=9.223$), 5 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %5'ten büyüktür. Yani kontrol hipotezinin kabul bölgesindedir. Bu durumda kontrol hipotezi kabul edilir. Dolayısıyla içinde yaklaşık 500 adet biber tohumu bulunan 250 ambalajdaki ölü tohum sayısı, ortalaması 2.48 olan Poisson dağılımına uygun bir dağılım göstermektedir kararına varılır.

Ki-kare testi kullanılarak Poisson dağılımına uyum kontrolünde serbestlik derecesi hesaplanırken, sınıf sayısından 1 ve popülasyonun parametresi olan μ yerine örnekten hesaplanan \bar{X} kullanıldığı için de 1 olmak üzere sınıf sayısından iki çıkarılır. Eğer popülasyona ait ortalama (μ) biliniyor ise serbestlik derecesi hesaplanırken sınıf sayısından 1 çıkarılır.

10.4.2.3. Normal Dağılıma Uyum Kontrolü

Bölüm 4.3.3'te 60 fasulye ağırlığının gözlenen ve ortalaması 0.661 ve standart sapması 0.0819 olan normal dağılıma göre beklenen frekansları Tablo 4.7'de verilmiştir. Araştırmacı, 60 fasulye ağırlığının, ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun bir dağılım gösterip göstermediğini kontrol etmek istemektedir.

Yapılan çalışmada, 60 fasulye ağırlığının ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun bir dağılım gösterip göstermediğini kontrol etmek için önce kontrol ve karşıt hipotezlerin kurulması gerekir.

H₀: 60 fasulye ağırlığı ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun dağılım göstermektedir. Gözlenen ve beklenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: 60 fasulye ağırlığı ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun dağılım göstermemektedir. Gözlenen ve beklenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Tablo 10.7'de, 1. sınıfın beklenen frekansı 5'ten küçük olduğu için 2. sınıfla ve 9. sınıfın beklenen frekansı 5'ten küçük olduğu için 8. sınıf ile birleştirilmiştir. 60 fasulye ağırlığının ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun olup olmadığını kontrol etmek için ki-kare değerinin hesaplanması Tablo 10.7'de gösterilmiştir.

Tablo 10.7. 60 fasulye tanesinin ağırlığına ait gözlenen ve ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma göre beklenen frekanslar ve normal dağılıma uygunluk kontrolü için ki-kare değerinin hesaplanması

Sınıflar	Gözlenen frekans (f)	Beklenen frekans (f')	$\frac{(f - f')^2}{f'}$
0.48 - 0.51	2	2.250	0.201
0.52 - 0.55	5 7	3.660 5.91	
0.56 - 0.59	6	6.630	0.060
0.60 - 0.63	9	9.930	0.088
0.64 - 0.67	13	11.580	0.174
0.68 - 0.71	8	10.674	0.670
0.72 - 0.75	9	7.770	0.195
0.76 - 0.79	5	4.476	0.033
0.80 - 0.83	3 8	3.030 7.506	
Toplam	60	60	$\chi^2=1.421$

Araştırmacı I. tip hata olasılığını %5 olarak belirlemiş olsun. Bu kontrolde serbestlik derecesi (sınıf sayısı-3)=(7-3)=4'tür. Çünkü 60 fasulye kullanılmıştır, yani örnek genişliği 60'tır, fasulye populasyonunun ağırlık ortalaması ve standart sapması örnekten hesaplanmıştır. Bu sebeple serbestlik derecesi bulunurken sınıf sayısından 3 çıkarılmıştır. 4 serbestlik dereceli ki-kare dağılımında %5'lik alan Tablo D'den 9.488 olarak bulunur. Hesaplanan χ^2 -değerinin ($\chi^2=1.421$) 4 serbestlik dereceli ki-kare dağılımına dahil olma olasılığı %5'ten büyüktür. Yani kontrol hipotezini kabul bölgesindedir. Bu

durumda kontrol hipotezi kabul edilir. Dolayısıyla 60 fasulye ağırlığı, ortalaması 0.661 ve standart sapması 0.082 olan normal dağılıma uygun dağılım göstermiştir kararına varılır.

Ki-kare testi kullanılarak Normal dağılıma uyum kontrolünde serbestlik derecesi hesaplanırken, sınıf sayısından 1 ve populasyonun parametreleri olan μ ve σ yerine örnekten hesaplanan \bar{X} ve S kullanıldığı için de her birisi için 1'er olmak üzere sınıf sayısından toplam olarak üç çıkarılır. Eğer populasyona ait ortalama ve standart sapma (μ ve σ) biliniyor ise serbestlik derecesi hesaplanırken sınıf sayısından 1 çıkarılır.

10.5. Bağımsızlık Kontrolleri

Toplanan veriler çeşitli faktör yada faktörlerin çeşitli hallerine göre sınıflandırılarak iki yanlı tablolar oluşturulabilir. Örneğin, herhangi bir dersten sınava girmiş öğrenciler cinsiyetlerine ve başarı durumlarına göre iki yanlı tablo oluşturulabilir veya yapılan bir zararlıyla mücadele çalışmasında böceklerin ölü, canlı veya felçli olma durumları kullanılan farklı ilaç çeşitlerine veya dozlarına göre sınıflandırılarak iki yanlı tablo oluşturulabilir.

Ele alınan faktörün hallerine göre sınıflandırılarak oluşturulan iki yanlı tablolar RxC tabloları olarak adlandırılır. Burada R, İngilizce row (sıra) ve C İngilizce column (sütun) kelimelerinin ilk harfleridir. RxC tabloları çalışılan özelliğin hallerine göre 2x3, 2x2, 3x4 vb. şeklinde düzenlenmiş olabilir. İki yanlı tabloda sıra ve sütunların kesiştiği noktalar hücre veya göz olarak adlandırılır.

Toplanan veriler iki yanlı tablo şeklinde düzenlendiği zaman üzerinde durulan faktörlerden birinin hallerine göre olan dağılımın diğer faktörün hallerine bağımlı (veya bağımsız) olup olmadığını kontrol edilmesi gerekebilir. Başka bir deyişle araştırmacı, iki özelliğin birbirinden bağımsız olup olmadığını kontrol etmek isteyebilir. Bu durumda bağımsızlık kontrolünün yapılması gerekir.

10.5.1. Ki-Kare Bağımsızlık Kontrolü

Bir araştırmada toplanan veriler, Tablo 10.7'de görüldüğü şekilde iki yanlı tablo olarak düzenlenmiş olabilir. İki yanlı tabloda her bir gözün frekansı a, b, c ve d ile gösterilmiştir. Bu durumda iki özelliğin birbirinden bağımsız olup olmadığı ki-kare bağımsızlık kontrolü uygulanarak kontrol edilebilir.

Tablo 10.7. İki özelliğin ikişer hali dikkate alınarak oluşturulmuş 2x2 tablosu

	A1	A2	Toplam
B1	a	b	a+b
B2	c	d	c+d
Toplam	a+c	b+d	N

Bağımsızlık kontrolü yapılırken de önce kontrol ve karşıt hipotezlerin aşağıdaki şekilde kurulması gerekir.

H₀: Gözlemlerin birinci faktörün hallerine göre dağılımı, ikinci faktörün hallerine göre değişmemektedir. Birinci faktörün hallerine göre olan dağılım ikinci faktörün hallerinden bağımsızdır. Kısaca, $(f - f') = 0$ 'dır.

H₁: Gözlemlerin birinci faktörün hallerine göre dağılımı, ikinci faktörün hallerine göre değişmektedir. Birinci faktörün hallerine göre olan dağılım ikinci faktörün hallerinden bağımsız değildir (bağımlıdır). Kısaca, $(f - f') \neq 0$ 'dır.

Hipotezler kurulduktan sonra hangi hipotezin kabul edileceğine karar vermek için (10.3) numaralı eşitlik kullanılarak χ^2 değeri hesaplanır. Ki-kare değerinin hesaplanabilmesi için önce iki yanlı tablodaki her göze ait beklenen frekansların hesaplanması gerekir. Beklenen frekanslar aşağıda açıklandığı şekillerde hesaplanabilir:

1. Eğer kontrol hipotezi doğru ise, yani gözlemlerin birinci faktörün hallerine göre dağılımı, ikinci faktörün hallerine göre değişmiyorsa A1 halinin B1 ve B2'de gözlenme olasılığı $\frac{a+c}{N}$ ve A2 halinin B1 ve B2'de gözlenme olasılığı $\frac{b+d}{N}$ 'dir. Bu durumda B1 halini gösteren $(a+b)$ tane bireyden

$$f' = (a+b) \times \left(\frac{a+c}{N}\right)$$

tanenin ve B2 halini gösteren $(c+d)$ tane bireyden de $f' = (c+d) \times \left(\frac{a+c}{N}\right)$

$$f' = (a+b) \times \left(\frac{b+d}{N}\right)$$

tanenin A2 halini göstermesi beklenir.

2. Bağımsız olayların birlikte gözlenme olasılığı, olayların olasılıklarının çarpımına eşittir. Eğer kontrol hipotezi doğru ise A ve B özellikleri birbirinden bağımsızdır. Bu durumda A1 ve B1 hallerinin birlikte gözlenme olasılığı $\left(\frac{a+b}{N}\right) \times \left(\frac{a+c}{N}\right)$ 'dir. Bu durumda N tane bireyden

$$f' = N \times \left(\frac{a+b}{N}\right) \times \left(\frac{a+c}{N}\right)$$

tanenin A1 ve B1 halini göstermesi beklenir. Görüldüğü gibi bu şekilde hesaplanan beklenen frekans bir önceki beklenen frekanstan başka bir şey değildir.

3. Beklenen frekanslar 1. veya 2. şekillerde hesaplanabilir. Fakat iki yanlı tablodaki satır ve sütun sayısı arttıkça 1. ve 2. şekilde açıklandığı gibi beklenen frekansların hesaplanmasında hata olasılığı artar. Her bir göz için beklenen frekansın daha kolay hesaplanabilmesi için (10.5) numaralı eşitlikte verilen eşitlik kullanılabilir.

$$\left[\begin{array}{c} \text{Herhangi bir gözün} \\ \text{beklenen frekansı} \end{array} \right] = \frac{\left[\begin{array}{c} \text{Beklenen frekans} \\ \text{hesaplanacak gözün} \\ \text{satır toplamı} \end{array} \right] \times \left[\begin{array}{c} \text{Beklenen frekans} \\ \text{hesaplanacak gözün} \\ \text{sütun toplamı} \end{array} \right]}{\left[\text{Genel toplam} \right]} \quad \dots(10.5)$$

İki yanlı tablodaki her bir göz için beklenen frekanslar hesaplandıktan sonra (10.3) numaralı eşitlik kullanılarak ki-kare değeri $\chi^2 = \sum_{i=1}^k \frac{(f - f')^2}{f'}$ şeklinde hesaplanır. Eşitlikte k: iki yanlı tablodaki toplam göz sayısıdır.

Daha önce de açıklandığı gibi ki-kare dağılımı serbestlik derecesine bağlı bir dağılımdır.

Ki-kare değerinin hesaplanabilmesi için önce her göz için beklenen frekansların hesaplanması gerekir.

Beklenen frekansların toplamı ile gözlenen frekansların toplamı birbirine eşit olması gerektiğinden iki satır ve sütundan oluşan iki yanlı tablolarda, yani 2x2 tablolarında gözlerden her hangi biri için beklenen frekanslar hesaplandıktan sonra diğer gözlerin beklenen frekansları gözlenen frekansların satır ve sütun toplamlarından söz konusu gözün beklenen frekansı çıkarılarak da bulunabilir. Başka bir deyişle iki yanlı tablolarda bir gözün beklenen frekansının hesaplanması için bir gözün beklenen frekansının hesaplanması yeterlidir. Bu sebeple de 2x2 tablolarında (10.3) numaralı eşitlik kullanılarak hesaplanan ki-kare değeri 1 serbestlik dereceli ki-kare dağılımı gösterir.

Oluşturulan iki yanlı tabloların satır ve sütun sayısı arttıkça yukarıda açıklandığı şekilde serbestlik derecesinin belirlenmesi zorlaşır. İki yanlı tablolarda bağımsızlık kontrolü yapılırken hesaplanan ki-kare istatistiğinin serbestlik derecesinin belirlenmesi için (10.6) numaralı eşitlik kullanılabilir.

$$\text{Serbestlik derecesi} = (\text{Satır sayısı} - 1) \times (\text{Sütun sayısı} - 1) \quad \dots(10.6)$$

Ki-kare değeri (10.3) numaralı eşitlik kullanılarak hesaplandıktan ve hesaplanan ki-kare değerinin serbestlik derecesi (10.6) numaralı eşitlik kullanılarak belirlendikten sonra, hesaplanan ki-kare değeri tablo değeri ile karşılaştırılarak hangi hipotezin kabul edileceğine karar verilir.

ÖRNEK 1:

Migren hastalığının cinsiyetten bağımsız olup olmadığını belirlemek üzere 400 bireyin cinsiyetlerine ve migren hastalığı olup olmamasına göre dağılımı Tablo 10.8'de verildiği gibi saptanmıştır. Bu sonuçlara göre Migren hastalığının cinsiyetten bağımsız olduğu söylenebilir mi?

Tablo 10.8. 400 bireyin cinsiyete ve migren hastalığına göre dağılımı ve her göz için beklenen frekanslar

	Migren Var	Migren Yok	Toplam
Kız	$f = 85$ $f' = \frac{120 \times 250}{400} = 75$	$f = 165$ $f' = 250 - 75 = 175$ veya $f' = \frac{280 \times 250}{400} = 175$	250
Erkek	$f = 35$ $f' = 120 - 75 = 45$ veya $f' = \frac{120 \times 150}{400} = 45$	$f = 115$ $f' = 280 - 175 = 105$ veya $f' = \frac{280 \times 150}{400} = 105$	150
Toplam	120	280	400

Migren hastası olup olmamanın cinsiyetten bağımsız olup olmadığını araştırılması için ki-kare bağımsızlık kontrolünün yapılması gerekir. Bunun için ilk olarak aşağıdaki şekilde hipotezler kurulur:

H₀: Migren hastası olup olmama cinsiyete göre değişmemektedir. Migren hastası olma durumu ile cinsiyet birbirinden bağımsızdır. Kısaca, $(f - f') = 0$ 'dır. Yani her göz için gözlenen ile beklenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

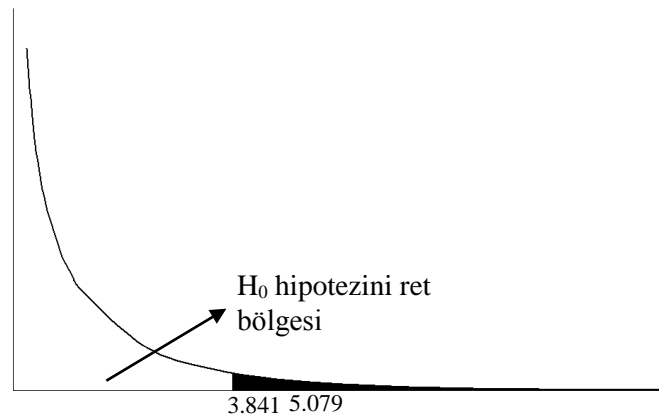
H₁: Migren hastası olup olmama cinsiyete göre değişmektedir. Migren hastası olma durumu ile cinsiyet birbirinden bağımsız değildir. Kısaca, $(f - f') \neq 0$ 'dır. Yani her göz için gözlenen ile beklenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Hipotezler kurulduktan sonra her göze ait beklenen frekanslar Tablo 10.8'de gösterildiği gibi hesaplanır. Beklenen frekanslar hesaplandıktan sonra (10.3) numaralı eşitlik kullanılarak ki-kare değeri;

$$\chi^2 = \frac{(85 - 75)^2}{75} + \frac{(165 - 175)^2}{175} + \frac{(35 - 45)^2}{45} + \frac{(115 - 105)^2}{105} = 5.079$$

olarak hesaplanır.

Ki-kare değeri hesaplandıktan sonra serbestlik derecesinin hesaplanması gerekir. Tablo 10.8'de görüldüğü gibi 2x2 tablolarında bir gözün beklenen frekansının hesaplanması yeterlidir. Diğer gözlerin beklenen frekansları, hesaplanan beklenen frekansın ilgili satır ve sütun toplamlarından çıkarılarak bulunabilir. Bu sebeple 2x2 tablolarında serbestlik derecesi 1'e eşittir. Eğer 10.6 numaralı eşitlik kullanılarak serbestlik derecesi hesaplanacak olursa $SD=(2-1) \times (2-1)=1$ olarak bulunur. Tablo D'de 1 serbestlik dereceli ki-kare dağılımında %5'lik alan 3.841 değerinden başlamaktadır. Şekil 10.3'te görüldüğü gibi hesaplanan ki-kare değeri 5.079 olup 3.841 değerinden büyüktür ve kontrol hipotezinin ret bölgesine düşmektedir.



ŞEKİL 10.3. 1 serbestlik dereceli ki-kare dağılımında H₀ hipotezini ret ve kabul bölgeleri

Bu durumda kontrol hipotezi ret edilir. Yani migren hastası olup olmama durumu cinsiyete göre değişmektedir. Araştırmacı, migren olup olmama durumu ile cinsiyetin birbirine bağımlı olduğu kararına varır.

ÖRNEK 2:

Bir bölgede tesadüfen seçilen 500 kişinin meslek grupları ve doğum yerlerine göre dağılımı Tablo 10.9'daki gibi gözlenmiştir. Gözlenen bu dağılım doğrultusunda bir meslek grubunu tercih etmenin doğum yerinden bağımsız olduğu ileri sürülebilir mi?

Tablo 10.9. 500 kişinin meslek grupları ve doğum yerlerine göre dağılımı ve her bir göz için beklenen frekanslar

	Köy	Kasaba	Şehir	Toplam
Öğretmen	$f = 55$ $f' = \frac{150 \times 145}{500}$ $f' = 43.5$	$f = 70$ $f' = \frac{150 \times 185}{500}$ $f' = 55.5$	$f = 25$ $f' = 51$	150
Avukat	$f = 25$ $f' = \frac{130 \times 145}{500}$ $f' = 37.7$	$f = 45$ $f' = \frac{130 \times 185}{500}$ $f' = 48.1$	$f = 60$ $f' = 53.45$	130
Mühendis	$f = 30$ $f' = \frac{105 \times 145}{500}$ $f' = 30.45$	$f = 25$ $f' = \frac{105 \times 185}{500}$ $f' = 38.85$	$f = 50$ $f' = 35.7$	105
Doktor	$f = 35$ $f' = 33.35$	$f = 45$ $f' = 42.55$	$f = 35$ $f' = 29.85$	115
Toplam	145 (145)	185 (185)	170 (170)	500

Meslek grubu tercih etmenin doğum yerinden bağımsız olup olmadığını kontrol etmek için ki-kare bağımsızlık kontrolünün yapılması gerekir. Bunun için ilk olarak aşağıdaki şekilde hipotezler kurulur:

H₀: Meslek grubu tercihi doğum yerine göre değişmemektedir. Seçilen meslek, doğum yerinden bağımsızdır. Kısaca, $(f - f') = 0$ 'dır. Yani her göz için gözlenen ile beklenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: Meslek grubu tercihi doğum yerine göre değişmektedir. Seçilen meslek, doğum yerine bağımlıdır. Kısaca, $(f - f') \neq 0$ 'dır. Yani her göz için gözlenen ile beklenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Hipotezler kurulduktan sonra her göze ait beklenen frekanslar Tablo 10.9'da görüldüğü gibi hesaplanır. Tablo 10.9'da görüldüğü gibi 6 gözün beklenen frekansları (10.5) numaralı eşitlik kullanılarak hesaplandıktan sonra taralı gözlerin beklenen frekansları her satırda hesaplanan beklenen frekans toplamı satır toplamlarından çıkarılarak hesaplanmıştır. Örneğin şehirde doğup öğretmen olan bireyler için beklenen frekans $150 - (43.5 - 55.5)$ olarak hesaplanmış ve $f' = 51$ olarak bulunmuştur. Diğer taralı gözlerin beklenen frekansları da benzer şekilde hesaplanmıştır.

Beklenen frekanslar hesaplandıktan sonra (10.3) numaralı eşitlik kullanılarak ki-kare değeri;

$$\chi^2 = \frac{(55 - 43.5)^2}{43.5} + \frac{(70 - 55.5)^2}{55.5} + \dots + \frac{(45 - 42.55)^2}{42.55} + \frac{(35 - 29.85)^2}{29.85}$$

$\chi^2 = 41.534$ olarak hesaplanır.

Ki-kare değeri hesaplandıktan sonra serbestlik derecesinin belirlenmesi gerekir. Altı gözün beklenen frekansı hesaplandıktan sonra diğer gözlerin beklenen frekanslarının satır ve sütun toplamları kullanılarak hesaplanması serbestlik derecesinin 6 olduğunu gösterir. Serbestlik derecesinin daha kolay belirlenmesi için (10.6) numaralı eşitlik kullanılarak serbestlik derecesi $SD = (4-1) \times (3-1) = 6$ olarak bulunur.

Tablo D'den 6 serbestlik dereceli ki-kare dağılımında %5'lik alan 12.592 değerinden başladığı görülür. Hesaplanan ki-kare değeri 41.534 olup tablo değerinden büyüktür ve kontrol hipotezinin ret bölgesine düşmektedir. Bu durumda kontrol hipotezi ret edilir ve meslek grubu tercih etmenin doğum yerinden bağımsız olmadığı, yani bireylerin meslek tercihlerinin doğum yerine göre değiştiği kararına varılır.

10.5.2. Bağımlılık (Contingency) Katsayısı

Düzenlenen iki yanlı tablolarda bağımsızlık kontrolü yapılarak kontrol hipotezi ret edilmiş ise bu iki faktör arasında bir bağımlılığın olduğunu gösterir. Böyle durumlarda bağımlılığın derecesinin belirlenmesi gerekir. Bu amaçla yaygın olarak kullanılan istatistiklerden biri Pearson'un bağımlılık katsayısı (**coefficient of contingency**) kullanılır. Bağımlılık katsayısı kısaca CC (coefficient of contingency) ile gösterilir ve (10.7) numaralı eşitlik kullanılarak hesaplanır.

$$CC = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad \dots(10.7)$$

Bağımsızlık kontrollerinde kontrol hipotezinin reddilmesi durumunda (10.7) numaralı eşitlik kullanılarak hesaplanan bağımlılık katsayısı iki özellik arasındaki bağımlılığın derecesini verir. Fakat bağımlılığın yönü hakkında bir bilgi vermez. İki özellik birbirine tam bağımlı olsa dahi (10.7) numaralı eşitlikten görüldüğü gibi bağımlılık katsayısı 1 (%100) değerine ulaşmaz. Bu sebeple hesaplanan bağımlılık katsayısının, RxC tablolarında alabileceği en büyük değere göre düzeltilmesi gerekir. Düzenlenen RxC tablolarında bağımlılık katsayısının alabileceği en büyük değer (10.8) numaralı eşitlik kullanılarak hesaplanır. İki yanlı tablolar düzenlenirken satır ve sütunların yeri değiştirilebileceğinden CC_{max} , CC_{max1} ile CC_{max2} 'nin ortalaması olarak da hesaplanabilir.

$$\begin{aligned} R < C \text{ ise } CC_{max1} &= \sqrt{\frac{R-1}{R}} \\ C < R \text{ ise } CC_{max2} &= \sqrt{\frac{C-1}{C}} \\ CC_{max} &= \frac{CC_{max1} + CC_{max2}}{2} \end{aligned} \quad \dots(10.8)$$

RxC tablolarında (10.8) numaralı eşitlik kullanılarak bağımlılık katsayısının alabileceği en büyük değer hesaplandıktan sonra düzeltilmiş bağımlılık katsayısı (10.9) numaralı eşitlik kullanılarak hesaplanır.

$$CC_{\text{düzeltilmiş}} = \frac{CC}{CC_{\text{max1}}} \text{ veya}$$
$$CC_{\text{düzeltilmiş}} = \frac{CC}{CC_{\text{max2}}} \text{ veya} \quad \dots(10.9)$$
$$CC_{\text{düzeltilmiş}} = \frac{CC}{CC_{\text{max}}}$$

Örneğin, Bölüm 10.5.1'de ÖRNEK 2'de, meslek grubu tercih etmenin doğum yerinden bağımsız olup olmadığını kontrol etmek için ki-kare değeri 41.534 olarak hesaplanarak meslek grubu tercih etmenin doğum yerinden bağımsız olmadığı, yani bireylerin meslek tercihlerinin doğum yerine göre değiştiği kararına varılmıştır. İki özellik arasındaki bağımlılığın derecesi (10.7) numaralı eşitlikten;

$$CC = \sqrt{\frac{41.534}{41.534 + 500}} = 0.277, \text{ yani } \%27.7 \text{ olarak bulunur.}$$

Hesaplanan bağımlılık katsayısı meslek grubu seçimi ile doğum yeri arasında %27.7'lik bir bağımlılığın olduğunu gösterir. Düzenlenen iki yanlı tabloda sütun sayısı satır sayısından az olduğu için (10.8) numaralı eşitlikten $CC_{\text{max2}} = \sqrt{\frac{3-1}{3}} = 0.816$ olarak bulunur. Eğer bu örnekte satır ve sütunların yerini değiştirilmiş olsaydı bu durumda satır sayısı sütun sayısından az olacak ve (10.8) numaralı eşitlikten $CC_{\text{max1}} = \sqrt{\frac{4-1}{4}} = 0.866$ olarak bulunacaktı. Bu durumda bağımlılık katsayısının alabileceği en büyük değer $CC_{\text{max}} = \frac{0.816 + 0.866}{2} = 0.841$ olarak hesaplanır. Düzeltilmiş bağımlılık katsayısı ise (10.9) numaralı eşitlik kullanılarak;

$CC_{\text{düzeltilmiş}} = \frac{0.277}{0.841} \cong 0.33$ olarak hesaplanır. Hesaplanan bağımlılık katsayısı meslek grubu seçimi ile doğum yeri arasında %33'lük bir bağımlılığın olduğu bulunmuş olur.

10.5.3. Fisher'in Kesin Olasılık (Fisher's Exact) Testi

Ki-kare bağımsızlık testi oluşturulan iki yanlı tablodaki gözlerin beklenen frekansının 5'ten az olmaması durumunda güvenilir sonuçlar verir. İki yanlı tablolar küçük örnekler için düzenlendiği zaman gözlerin beklenen frekansı 5'ten küçük olacağı için R.A. Fisher tarafından geliştirilen Fisher'in Kesin Olasılık testi kullanılır. Düzenlenen RxC tabloları için beklenen frekansın 5'ten küçük olması durumunda Fisher tarafından önerilen test de Fisher-Freeman-Halton testi olarak bilinir.

Küçük bir örnek için 2x2 tablo, Tablo 10.7’de görüldüğü ve aşağıda verildiği gibi düzenlenmiş olsun.

	A1	A2	Toplam
B1	a	b	a+b
B2	c	d	c+d
Toplam	a+c	b+d	N

İki yanlı tablodaki göz frekanslarının A ve B faktörleri arasında herhangi bir bağımlılığın olmadığı bir populasyondan çekilmiş olma olasılığı (10.10) numaralı eşitlik kullanılarak hesaplanır.

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!} \dots(10.10)$$

İlk olarak (10.10) numaralı eşitlik kullanılarak düzenlenen tablonun A ve B faktörleri bağımsız olan populasyondan alınmış olma olasılığı hesaplanır. Daha sonra her bir göz için gözlenen frekanslardan daha uç hallerin olasılıkları hesaplanır. Bu olasılıklar hesaplanırken düzenlenen tablodaki satır ve sütun toplamaları değişmeyecek şekilde daha uç tablolar hazırlanır. Gözlenen ve daha uç tablolar için hesaplanan olasılıklar toplanarak gözlenen ve daha uç hallerin meydana gelme olasılığı hesaplanır. Fisher’in kesin Olasılık Testinde hesaplanan olasılık, tek taraflı olasılıktır.

ÖRNEK:

Karadeniz bölgesinde tesadüfen seçilen 30 kişinin çay bahçesinde çalışma durumlarının cinsiyete göre dağılımı Tablo 10.10’da verildiği gibi gözlenmiştir.

Tablo 10.10. 30 kişinin çay bahçesinde çalışma durumlarının cinsiyete göre dağılımı

	Çay bahçesinde çalışan	Çay bahçesinde çalışmayan	Toplam
Kadın	f = 16	f = 2	18
Erkek	f = 8	f = 4	12
Toplam	24	6	30

Bu örnekte çay bahçesinde çalışıp çalışmamanın cinsiyetten bağımsız olup olmadığı kontrol edilmek istenmektedir. Düzenlenen iki yanlı tablo için ki-kare bağımsızlık testi uygulandığı zaman gözlerden biri için hesaplanan beklenen frekansın 5’ten küçük olduğu görülür. Elde edilen sonucun güvenilir olması için Fisher’in Kesin Olasılık testi uygulanır. Fisher’in Kesin Olasılık testi uygulanırken de ilk olarak hipotezler aşağıdaki şekilde kurulur.

H₀: Çay bahçesinde çalışıp çalışmama cinsiyetten bağımsızdır. Çay bahçesinde çalışıp çalışmama cinsiyete göre değişmemektedir.

H₁: Çay bahçesinde çalışıp çalışmama cinsiyetten bağımsız değildir. Çay bahçesinde çalışıp çalışmama cinsiyete göre değişmektedir.

Hipotezler kurulduktan sonra Fisher’in Kesin Olasılık est aşağıda açıklandığı şekilde uygulanır. İlk olarak gözlenen tablonun gözlenme olasılığı hesaplanır. Daha sonra gözlenen tablodan daha uç gözlenebilecek tabloların olasılıkları aşağıda görüldüğü gibi hesaplanarak bulunan olasılıklar toplanır. Bulunan olasılık gözlenen ve daha uç tabloların gözlenme olasılığıdır.

16	2	18	Gözlenen tablo için olasılık: $P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}$ $P = \frac{18!2!20!10!}{16!2!4!8!8!} = 0.002521$
4	8	12	
20	10	30	

17	1	18	Daha uç tablo için olasılık: $P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}$ $P = \frac{18!1!2!20!10!}{17!1!3!9!8!} = 0.0001318$
3	9	12	
20	10	30	

18	0	18	En uç tablo için olasılık: $P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}$ $P = \frac{18!0!2!20!10!}{18!0!2!1!20!} = 0.000002196$
2	10	12	
20	10	30	

Gözlenen ve daha uç tabloların gözlenme olasılıkları hesaplandıktan sonra bu olasılıklar toplanır ve toplam olasılık;

$P=0.002521+0.0001318+0.000002196 \approx 0.002655$ olarak bulunur. Bu olasılık sola doğru gözlenebilecek uç tabloların gözlenme olasılığıdır ve tek taraflı olasılıktır.

I. tip hata olasılığı %5 olarak kararlaştırılmış ise hesaplanan olasılık %5'ten küçük olduğu için kontrol hipotezi reddedilir ve çay bahçesinde çalışıp çalışmamanın cinsiyetten bağımsız olmadığı kararına varılır.

Hesaplanan olasılık tek taraflı olasılık olduğu için çift taraflı olasılık bu olasılığın iki katı alınarak bulunabileceği gibi sağ tarafa doğru gözlenebilecek uç tabloların olasılıklarının toplamı bulunarak da hesaplanabilir. Hesaplanan tek taraflı olasılığın iki katı alınarak bulunan çift taraflı olasılık ile diğer tarafa doğru uç tabloların olasılıklarının toplamının tek taraflı olasılığa eklenerek bulunacak çift taraflı olasılık arasındaki fark iki satır ve sütun toplamı arasındaki fark arttıkça artar.

Hesaplanan olasılık tek taraflı olasılığın iki katı alınarak çift taraflı olasılık $P=2(0.002655)=0.00531$ olarak bulunur.

10.5.4. G-testi

Düzenlenen iki yanlı tablolarda bağımsızlık kontrolü için yaygın olarak kullanılan yöntemlerden biri de G-testidir. G-testi olasılık oranı "likelihood-ratio" veya en yüksek olabilirlik "maximum likelihood" testi olarak da isimlendirilir. G-testi, ki-kare bağımsızlık testinin kullanıldığı bütün iki yanlı tablolara uygulanabilir. Ki-kare bağımsızlık testinin uygulanabileceği tablolarda G-testinin daha yaygın olarak kullanılmasının sebeplerinden biri beklenen frekansların 1 ve/veya 5'ten küçük olması durumlarında hesaplanan G istatistiğinin ki-kare dağılımına daha uygun olması, diğer bir deyişle G istatistiğinin ki-kare dağılımına uygunluğunun etkilenmemesidir.

Yürütülen bir çalışmada toplanan veriler ile iki yanlı tablo aşağıda verildiği gibi düzenlenmiş

	A1	A2	Toplam
B1	a	b	a+b
B2	c	d	c+d
Toplam	a+c	b+d	N

olsun. Bu şekilde düzenlenmiş bir iki yanlı tablo için G istatistiği (10.11) numaralı eşitlik kullanılarak hesaplanır.

$$G = a \ln a + b \ln b + c \ln c + d \ln d + N \ln N - (a+b) \ln (a+b) -$$

$$(c+d) \ln (c+d) - (a+c) \ln (a+c) - (b+d) \ln (b+d) \quad \dots(10.11)$$

(10.11) numaralı eşitlikte yapılan hesaplamalar aşağıdaki şekilde 3 adımda toplanır.

Adım 1: Her gözün frekansı için $\sum f \ln f$ değeri hesaplanır.

Adım 2: Satır ve sütun toplamaları için $\sum f \ln f$ değeri hesaplanır.

Adım 3: Toplam frekans için $N \ln N$ değeri hesaplanır.

3 adımda belirtilen hesaplamalar yapıldıktan sonra G istatistiği (10.12) numaralı eşitlikte verildiği şekilde düzenlenebilir.

$$G = 2[\text{Adım 1} - \text{Adım 2} + \text{Adım 3}] \quad \dots(10.12)$$

(10.11) veya (10.12) numaralı eşitlik kullanılarak hesaplanan G istatistiği serbestlik derecesi (satır sayısı-1)x(sütun sayısı-1) olan ki-kare dağılımına uygun bir dağılım gösterir.

ÖRNEK:

Yürütülen bir çalışmada, şehir ve köyde yaşayan 25 yaşındaki erkeklerin medeni durumlarına göre dağılımı Tablo 10.11'deki gibi gözlenmiştir. Yapılan çalışmada medeni durumun yaşanan yerden bağımsız olup olmadığını araştırılması amaçlanmıştır.

Tablo 10.11. Şehir ve köyde yaşayan 25 yaşındaki erkeklerin medeni durumlarına göre dağılımı

	Evli	Bekar	Toplam
Şehir	350	700	1050
Köy	150	100	250
	500	800	1300

Medeni durumun yaşanan yerden bağımsız olup olmadığı G-testi uygulanarak kontrol edilebilir. Bu durumda da ilk olarak hipotezlerin aşağıdaki gibi kurulması gerekir.

H₀: Medeni durum yaşanan yerden bağımsızdır. Yani evli veya bekar oluş şehir veya köyde yaşamaktan bağımsızdır.

H₁: Medeni durum yaşanan yere bağımlıdır. Yani evli veya bekar oluş şehir veya köyde yaşamaktan bağımsız değildir.

Hipotezler kurulduktan sonra G-istatistiğinin hesaplanması için Adım1-Adım 3'teki hesaplamalar yapılır.

Adım 1: $350 \ln 350 + 700 \ln 700 + 150 \ln 150 + 100 \ln 100 = 7848.145$

Adım 2: $1050 \ln 1050 + 250 \ln 250 + 500 \ln 500 + 800 \ln 800 = 17139.731$

Adım 3: $1300 \ln 1300 = 9321.155$

Gerekli hesaplamalar yapıldıktan sonra G istatistiği (10.12) numaralı eşitlikten $G = 2(7848.145 - 17139.731 + 9321.155) = 59.138$ olarak bulunur. Hesaplanan G istatistiği 1 serbestlik dereceli ki-kare dağılımına uygun dağılım gösterir. Eğer araştırmacı I. tip hata olasılığını %1 olarak belirlemiş ise Tablo D'den 1 serbestlik dereceli ki-kare dağılımında %1'lik alanının 6.635 değerinden başladığı görülür. Hesaplanan G-değeri (59.138) bu değerden çok büyük olup, kontrol hipotezini ret bölgesine düşer. Bu sebeple kontrol hipotezi ret edilerek medeni durumun yaşanılan yere göre değiştiğine, yani evli veya bekar oluşun şehir veya köyde yaşamaktan bağımsız olmadığına karar verilir.

10.6. Sorular

1. Ki-kare (χ^2) dağılımı nasıl elde edilir? Açıklayınız.
2. Ki-kare (χ^2) dağılımının özellikleri nelerdir? Açıklayınız.
3. Bir mer'adaki bitki kompozisyonunun %55'ini ayrık, %40'ını baklagil türü ve geri kalanını da zararlı otlardan oluştuğu bilinmektedir. 3 yıl otlatmadan sonra, söz konusu mer'adan tesadüfen seçilen 200 bitkiden; 115'inin ayrık, 75'inin baklagil türü ve 10'unun da zararlı ot olduğu tespit edilmiştir. Bu sonuçlara göre 3 yıllık otlatma sonunda, mer'adaki bitki kompozisyonu oranları değişmiş midir?
4. 5 farklı bölümü tercih eden 25 öğrencinin cinsiyetlere göre dağılımları tespit edildikten sonra, bölüm tercihinin cinsiyetten bağımsız olup olmadığını araştırmak amacıyla hesaplanan χ^2 değeri 14.425 olarak bulunmuştur. Buna göre bölüm tercihinin cinsiyetten bağımsız olduğu söylenebilir mi?
5. Bir araştırmada 100 kişiye diğ bakımına dikkat edip etmedikleri sorularak, cinsiyetlere göre aşağıdaki tablo hazırlanmıştır. Diğ bakımına dikkat etme durumu cinsiyete bağımlı mıdır?

Cinsiyet	Bakımlı	Bakımsız
Erkek	35	20
Kadın	25	20

6. Bir ilaç firması tarafından yeni bir ürün için yapılan tanıtım kampanyası sonucunda 7 bölgede bulunan eczanelerden bu ürünü satın alanların sayısı aşağıdaki gibi bulunmuştur. Yeni ürünü satın alanların sayısının bölgelere göre homojen dağılıp dağılmadığını kontrol ediniz.

BÖLGELER	1	2	3	4	5	6	7
ECZANELER	30	50	65	40	20	70	25

7. Bir hafta boyunca kütüphaneden alınan kitapların günlere göre sayısı aşağıdaki gibidir. Ödünç kitap almak haftanın günlerinden bağımsız mıdır?

Pazartesi	Salı	Çarşamba	Perşembe	Cuma	Cumartesi	Pazar
135	108	120	116	146	98	54

8. Dört ayrı bölgeden rastgele seçilen kişilere uygulanan test sonucu mide ülserinin en büyük nedeni olarak gösterilen *Helicobacter pylori* mikrobu taşıyıp taşımadıkları aşağıdaki gibi tespit edilmiştir. *Helicobacter pylori* mikrobu taşıyıp taşıyamama durumunun bölgelere bağlı olup olmadığını kontrol ediniz.

Bölgeler	A	B	C	D
Taşıyan	14	18	12	10
Taşımayan	36	32	48	90

9. Obezite hastası olan ve olmayan 200 bireyin cinsiyetlerine göre dağılımları aşağıdaki gibidir.

Cinsiyet	Obezite Var	Obezite Yok
Kız	25	75
Erkek	20	80

Obez olup olmamanın cinsiyetten bağımsız olduğu görüşüne katılır mısınız?

10. Migren hastası olan ve olmayan 200 bireyin cinsiyetlerine göre dağılımları aşağıdaki gibidir.

	Migren Var	Migren Yok
Kız	25	75
Erkek	20	80

Migren olup olmamanın cinsiyetten bağımsız olduğu görüşüne katılır mısınız?

11. Polen alerjisi olan ve olmayan bireylerin, yaş evrelerine göre dağılımı aşağıdaki gibidir. Polene karşı duyarlı olup olmamanın yaş evrelerinden bağımsız olduğu söylenebilir mi?

	Polen alerjisi var	Polen alerjisi yok
Çocuk	25	75
Genç	20	80
Yetişkin	15	65

12. Şehirde ve köyde yaşayan 20 yaşındaki erkeklerden evli ve bekar olanların sayıları aşağıdaki gibidir.

	Evli	Bekar
Şehir	407	726
Köy	266	220

Evli veya bekar oluş, şehirde veya köyde yaşamaktan bağımsız mıdır?

13. Trafik kazalarında yaralanan ve sağlam kalan sürücülerin, Emniyet kemeri kullanan ve kullanmayanlara göre dağılımı aşağıdaki gibidir. Trafik kazalarında yaralanmanın Emniyet kemeri kullanmaktan bağımsız olduğu söylenebilir mi? Eğer bağımsız değilse bağımlılık katsayısı (C.C) ne kadardır?

	Yaralı	Sağlam
E. kemeri var	407	726

E. kemeri yok	266	220
---------------	-----	-----

14. Test, yazılı ve sözlü sınav yöntemlerine tabi tutulan öğrencilerden başarılı ve başarısız olanların, sınav yöntemlerine göre dağılımı aşağıdaki gibidir.

	Test	Yazılı	Sözlü
Başarılı	50	47	56
Başarısız	5	14	8

Öğrencilerin başarılı olup olmamalarının, uygulanan sınav yönteminden bağımsız olduğu söylenebilir mi?

15. Belirli bir hastalığa karşı aşılanan ve aşılanmayan çocukların hastalığa tutulup tutulmama durumları aşağıdaki gibidir.

	Hasta olan	Hasta olmayan
Aşılanan	9	42
Aşılanmayan	17	22

Hasta olup olmama durumu aşı olup olmamaktan bağımsız mıdır?

16. Belirli bir bölgede meydana gelen Trafik kazaları sayısının mevsimlere göre dağılımı aşağıdaki gibidir.

İlkbahar	Yaz	Sonbahar	Kış
40	20	60	80

Bu bölgede meydana gelen Trafik kazası sayılarının mevsimlerden bağımsız olduğu söylenebilir mi?

17. Belirli bir bölgedeki bir yıl içerisinde Hırsızlık, Tecavüz ve Gasp suçlarını işleyenlerin yaş gruplarına göre dağılımları aşağıdaki gibidir.

Yaş grubu	Hırsızlık	Tecavüz	Gasp
20-30	25	55	20
31-40	35	25	15
41-50	40	10	10

Bu bölgede işlenen bu üç farklı suç türünün yaş gruplarından bağımsız olduğu söylenebilir mi? Eğer bağımsız değilse bağımlılık katsayısı (C.C) ne kadardır?

18. Bir bahçeden tesadüfen alınan 300 adet uzun sivri biberin yarısı acı yarısı ise tatlıdır. Tesadüfen alınan 400 adet kısa sivri biberin ise 3/4'ü acıdır. Sivri biberlerdeki acı ve tatlı oluşun uzun ve kısa oluştan bağımsız olduğu söylenebilir mi?
19. Öğretmen, avukat, mühendis ve doktor meslek gruplarını tercih edenlerin doğum yerlerine göre dağılımı aşağıdaki gibidir.

	Öğretmen	Avukat	Mühendis	Doktor
Köy	45	10	20	15
Kasaba	55	15	10	20
Şehir	15	55	45	10

Bu meslek gruplarını tercih etmek, doğum yerinden bağımsız mıdır? Eğer bağımsız değilse bağımlılık ne kadardır?