

Kİ-KARE (χ^2) DAĞILIMI VE BU DAĞILIM İLE İLGİLİ TESTLER

11.1. Ki-Kare (χ^2) Dağılımı

İstatistikte önemli diğer bir sürekli dağılım ki-kare dağılımıdır. Ki-kare değişkeni 0 ile sonsuz arasında dağılım gösteren tek taraflı bir değişkendir.

Ortalaması μ ve standart sapması σ olan normal dağılan bir popülasyondan çekilen her bir X değeri için $\frac{X-\mu}{\sigma}$ şeklinde hesaplanacak r tane bağımsız Z-değerinin karelerinin toplamının gösterdiği dağılım ki-kare (χ^2) dağılımı olarak adlandırılır (eşitlik 11.1).

$$Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + \dots + Z_r^2 = \sum_{i=1}^r Z_i^2 \quad \dots(11.1)$$

Bu dağılımın fonksiyonu aşağıdaki gibidir:

$$f(\chi^2) = \frac{1}{2^{\frac{r}{2}} \left(\frac{r}{2}\right)!} (\chi^2)^{\frac{(r-2)}{2}} e^{-\frac{\chi^2}{2}} \quad \dots(11.2)$$

(11.2) numaralı eşitlikte, r, serbestlik derecesidir yani χ^2 -değerini oluşturan bağımsız Z-değerlerinin sayısıdır.

Üzerinde durulan bir özelliğin değişik hallerini (kırmızı, beyaz, yeşil gibi) bilinen oranlarda barındıran bir popülasyondan geri iadeli olarak rastgele alınan n hacimli örneklerdeki sayıları (frekansları) ile bilinen oranlara göre bulunması gereken frekanslardan 11.3 numaralı eşitlikten hesaplanan değerler de χ^2 -dağılımı gösterirler. χ^2 -dağılımı istatistik testlerde daha çok bu özelliğinden yararlanarak kullanılır.

$$\sum_{i=1}^k \frac{(\text{Gözlenen frekans}(f) - \text{Beklenen frekans}(f'))^2}{\text{Beklenen frekans}(f)} = \chi^2 \quad \dots(11.3)$$

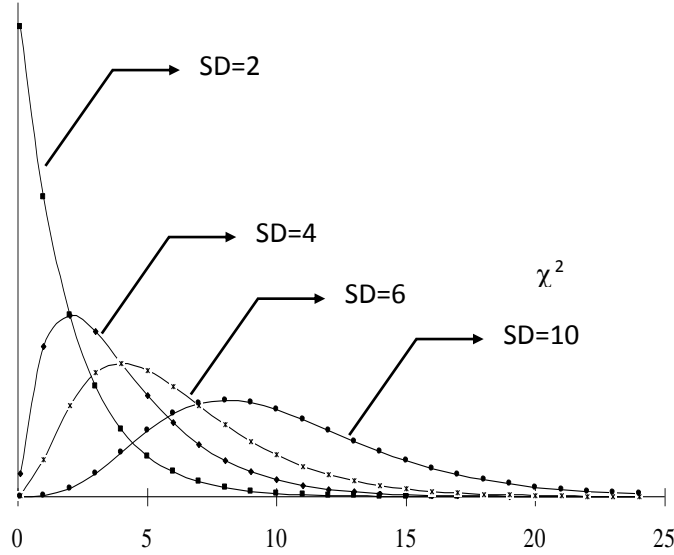
(11.3) numaralı eşitlikte, k, sınıf sayısı, f=her bir sınıfta gözlenen frekans, f', öne sürülen hipoteze göre her sınıfta olması beklenen frekanstır. (11.3) numaralı eşitliğe göre hesaplanan ki-kare değerinin serbestlik derecesi, ki-kare değerinin hesaplanması sırasında parametre yerine kullanılan istatistik sayısına bağlı olarak (k-1), (k-2), (k-3)...vs. olabilir. Bununla ilgili örnekler 11.3 numaralı bölümde verilecektir.

(11.3) numaralı eşitlik kullanılarak hesaplanan istatistiğin ki-kare dağılımı göstermesi için sınıf sayısının yeterli olması ve belirlenen hipoteze göre sınıflar için hesaplanacak beklenen frekansın 5'ten küçük olmaması gerekir. Serbestlik derecesi 1 olduğunda hesaplanan χ^2 -değerlerinin teorik χ^2 -değerine daha iyi yaklaşması için YATES düzeltmesi yapılarak (11.4) numaralı eşitlikten χ^2 -değerinin hesaplanması gerekir.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$\sum_{i=1}^k \frac{(|f - f'| - 0.5)^2}{(f')} = \chi^2 \quad \dots(11.4)$$

(11.2) numaralı fonksiyondan da görülebileceği gibi χ^2 -dağılımı serbestlik derecesine bağlı bir dağılımdır (Şekil 11.1).



ŞEKİL 11.1. Farklı serbestlik dereceli ki-kare dağılımları

Her serbestlik derecesi için farklı bir χ^2 -dağılımı vardır. Farklı serbestlik dereceli χ^2 -dağılımlarında belirli yüzdelere karşılık gelen χ^2 -değerleri Tablo D'de verilmiştir.

Şekil 11.1'de görüldüğü gibi 2 serbestlik dereceli ki-kare dağılımının tepe değeri yoktur. χ^2 -dağılımı sıfırdan başlar, sonsuza kadar devam eder ve sonsuzda χ^2 -eksenine asimptot olur. Bu dağılımlarda, (SD-2) değeri tepe değeridir. χ^2 -dağılımının ortalaması, serbestlik derecesine (SD'ye) ve varyansı serbestlik derecesinin iki katına (2SD'ye) eşittir.

11.2. Ki-Kare Kontrolleri

Bir örnekten elde edilen veriler belirtilen özelliklerine göre sınıflandırılabilir. Ve bir araştırmacı öne sürülen bir hipoteze göre her sınıfta olması beklenen frekans ile gözlenen frekanslar arasında uyum olup olmadığını kontrol etmek isteyebilir. Ki-kare kontrolleri 3 grup altında incelenebilir:

1. Homojenlik Kontrolleri
2. Dağılımlara Uyum Kontrolleri
3. Bağımsızlık Kontrolleri

Ki-kare kontrollerinde izlenecek adımlar Z- ve t-kontrollerinde anlatıldığı şekildedir. Araştırmacı ilk olarak hipotezlerini oluşturmalı, I. tip hata ihtimalini belirlemeli,

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

test istatistiğini hesaplamalı ve Tablo D'den örneği için geçerli olan serbestlik dereceli ki-kare dağılımında, I. tip hata ihtimaline karşılık gelen χ^2 -değeri ile hesapladığı χ^2 -değerini karşılaştırarak hangi hipotezi kabul edeceğine karar vermelidir.

11.2.1. Homojenlik Kontrolleri

ÖRNEK 1:

Bir eczaneye bir hafta içinde gelen müşterilerin günlere göre dağılımı aşağıdaki gibi bulunmuştur. Söz konusu eczaneye gelen müşterilerin sayısı günlere göre değişmekte midir?

Günler	Gözlenen frekans	Beklenen frekans	$(f-f')^2/f'$
1	22	26	0.165
2	29	26	0.346
3	31	26	0.962
4	21	26	0.962
5	27	26	0.038
6	26	26	0.000
Toplam	156	156	$\chi^2=2.923$

Araştırmacı yine Bölüm 7 ve 8'de açıklandığı gibi hipotezlerini kurmalıdır.

H₀: Eczaneye bir hafta boyunca gelen müşterilerin günlere dağılımı homojendir. Günler arasında eczaneye gelen müşteri sayısı bakımından fark önemli değildir.

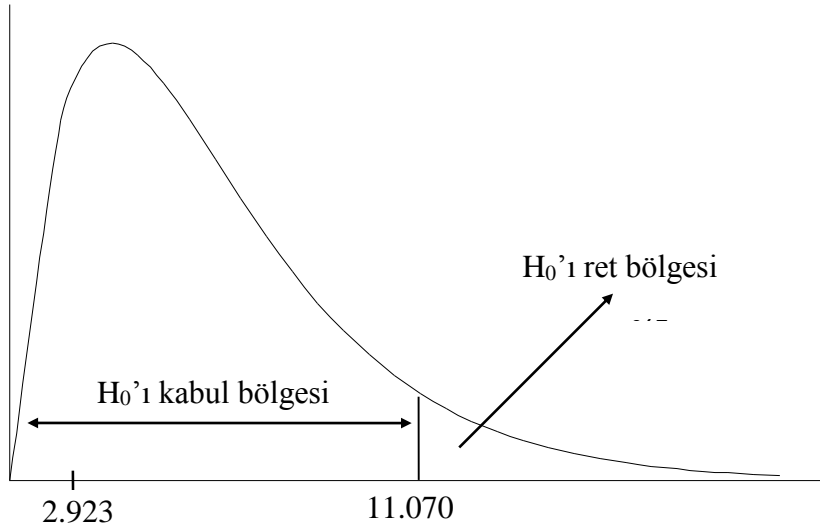
H₁: Eczaneye bir hafta boyunca gelen müşterilerin günlere dağılımı homojen değildir. Günler arasında eczaneye gelen müşteri sayısı bakımından fark önemlidir.

Burada hangi hipotezi kabul edeceğine karar vermek için araştırmacının χ^2 -değerini 11.3 numaralı eşitliği kullanarak hesaplaması gerekir. Bunun için de haftanın her günü için beklenen müşteri sayısının bulunması lazımdır. Kontrol hipotezi ile araştırmacı müşteri sayısı bakımından haftanın günlerinin homojen olduğunu öne sürdüğüne göre, toplam müşteri sayısı (156) haftanın günlerine eşit olarak dağılmalıdır. Yani her gün için beklenen frekans $156/6=26$ 'dır. Beklenen frekanslar bulunduktan sonra 11.3 numaralı eşitlik kullanılarak χ^2 -değeri aşağıdaki şekilde hesaplanır:

$$\begin{aligned}\chi^2 &= \frac{(22-26)^2}{26} + \frac{(29-26)^2}{26} + \frac{(31-26)^2}{26} + \\ &\frac{(21-26)^2}{26} + \frac{(27-26)^2}{26} + \frac{(26-26)^2}{26} = 0.165 + 0.346 + 0.962 + 0.962 + 0 \\ &= 2.923 \text{ olarak bulunur.}\end{aligned}$$

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Araştırmacı 1. tip hata ihtimalini %5 olarak kararlaştırmış olsun. Son olarak hangi hipotezin kabul edileceğine karar vermek gerekir. Bölüm 11.1’de açıklandığı gibi ki-kare dağılımı serbestlik derecesine bağlı bir dağılımdır. Bu örnekte serbestlik derecesi, $SD=6-1$ (gün sayısı-1)=5’tir. Tablo D’de, 5 serbestlik dereceli χ^2 -dağılımında %5’lik alanın başladığı χ^2 -değeri 11.070’tir (Şekil 11.2). Şekil 11.2’de görüldüğü gibi hesaplanan ki-kare değerinin 5 serbestlik dereceli ki-kare dağılımına dahil olma ihtimali %5’den büyüktür. Yani hesaplanan χ^2 -değeri kabul bölgesindedir. Böylece kontrol hipotezi reddedilemez, yani eczaneye bir hafta boyunca gelen müşterilerin günlere dağılımı homojendir ve günler arasında eczaneye gelen hasta sayısı bakımından fark önemli değildir.



ŞEKİL 11.2. 5 serbestlik dereceli ki-kare dağılımında H_0 hipotezini ret ve kabul bölgeleri

ÖRNEK 2:

Bir kliniğe bir yıl içinde gelen 240 hastanın aylara göre dağılımı aşağıdaki gibi bulunmuştur. Kliniğe gelen hastaların aylara göre dağılımı homojen midir?

H_0 : Hastaneye bir yıl boyunca gelen hastaların aylara göre dağılımı homojendir. Aylar arasında hastaneye gelen hasta sayısı bakımından fark önemli değildir.

H_1 : Hastaneye bir yıl boyunca gelen hastaların aylara göre dağılımı homojen değildir. Aylar arasında hastaneye gelen hasta sayısı bakımından fark önemlidir.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

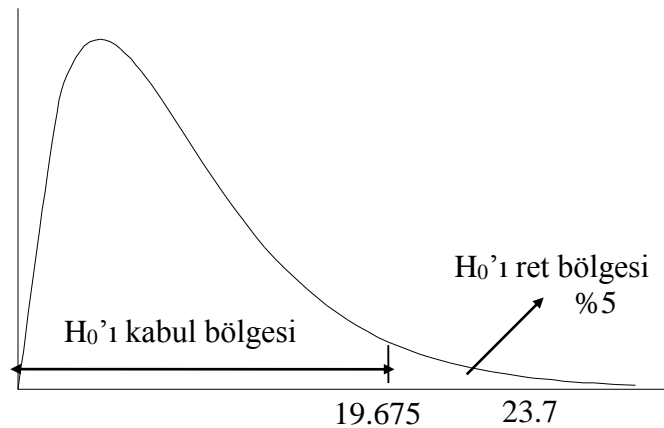
Aylar	Gözlenen hasta sayısı (f)	Beklenen hasta sayısı (f')	(f-f') ² /f'
1	32	20	7.2
2	30	20	5.0
3	24	20	0.8
4	22	20	0.2
5	19	20	0.05
6	14	20	1.8
7	12	20	3.2
8	12	20	3.2
9	14	20	1.8
10	18	20	0.2
11	22	20	0.2
12	21	20	0.05
Toplam	240	240	$\chi^2=23.7$

Burada hangi hipotezi kabul edeceğine karar vermek için araştırmacının χ^2 -değerini 11.3 numaralı eşitliği kullanarak hesaplaması gerekir. Bunun için de her ay beklenen hasta sayısının bulunması lazımdır. Kontrol hipotezi ile araştırmacı hasta sayısı bakımından ayların homojen olduğunu öne sürdüğüne göre, her ay için beklenen frekans $240/12=20$ 'dir. Beklenen frekanslar bulunduktan sonra 11.3 numaralı eşitlik kullanılarak χ^2 -değeri aşağıdaki şekilde hesaplanır:

$$\chi^2 = \frac{(32-20)^2}{20} + \frac{(30-20)^2}{20} + \dots + \frac{(21-20)^2}{20} = 7.2 + 5.0 + \dots + 0.05$$

$$= 23.7$$

Araştırmacı 1. tip hata ihtimalini %5 olarak kararlaştırmış olsun. Burada serbestlik derecesi, $SD=12-1=11$ 'dir. Tablo D'de, 11 serbestlik dereceli χ^2 -dağılımında %5'lik alanın başladığı χ^2 -değeri 19.675'tir (Şekil 11.3). Şekil 11.3'de görüldüğü gibi hesaplanan ki-kare değerinin 11 serbestlik dereceli ki-kare dağılımına dahil olma ihtimali %5'den küçüktür. Bu durumda kontrol hipotezi reddedilir.



ŞEKİL 11.3. 11 serbestlik dereceli ki-kare dağılımında H_0 hipotezini ret ve kabul bölgeleri

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

11.2.2. Uyum Kontrolleri

Bölüm III'te en çok rastlanan dağılımların binom, Poisson ve normal dağılım olduğu belirtilmişti. Araştırmacı üzerinde çalıştığı örneği oluşturan verilerin en çok rastlanan bu üç dağılımdan birine uyum gösterip göstermediğini kontrol etmek için de ki-kare dağılımını kullanabilir. Dağılımlara uyum kontrolleri Bölüm 3'te verilen örnekler için açıklanacaktır.

11.2.2.1. Binom Dağılımına Uyum Kontrolü

ÖRNEK 1:

Bir bölgeden seçilen 5 çocuklu 300 ailenin kız çocuk sayısı bakımından dağılımı aşağıdaki gibi bulunmuştur. 300 ailenin kız çocuk sayısı bakımından dağılımının $\pi=1/2$ olan binom dağılımına uygun olduğu söylenebilir mi?

Sınıflar (kız çocuk sayısı)	0	1	2	3	4	5
Aile sayısı	10	56	83	96	43	12

Örnekte gözlenen dağılımın binom dağılımına uygunluğunu kontrol etmek için beklenen frekansların bulunması gerekir. Bölüm 3.1'de açıklandığı şekilde 5 çocuktan hiçbirinin, 1'nin, 2'sinin, 3'ünün, 4'ünün ve hepsinin kız olma ihtimalleri bulunmalıdır. Bu ihtimaller binom ihtimal fonksiyonu kullanılarak veya $(p+q)^5$ binomu açılarak bulunabilir. $(p+q)^5$ binomu açıldığı zaman;

$$(\pi+(1-\pi))^5 = \pi^5 + 5\pi^4(1-\pi) + 10\pi^3(1-\pi)^2 + 10\pi^2(1-\pi)^3 + 5\pi(1-\pi)^4 + (1-\pi)^5$$

terimleri elde edilir. Birinci terim 5 çocuğun kız, 2. terim 4, 3. terim 3, 4. terim 2, 5. terim 1 ve 6. terim 0 kız çocuk ihtimalini verir. Bir aile için bulunan bu ihtimaller, örnekte 300 aile olduğu için 300 ile çarpılırsa her bir sınıf için beklenen frekanslar bulunur. Gözlenen frekanslar, olasılıklar ve beklenen frekanslar Tablo 11.1'de verilmiştir.

TABLO 11.1. Binom dağılımına uyum kontrolü ile ilgili hesaplamalar

Sınıflar	Olasılık	f	f'	$(f-f')^2 / f'$
0	0.03125	10	9.375	0.0417
1	0.15625	56	46.875	1.7763
2	0.31250	83	93.750	1.2327
3	0.31250	96	93.750	0.0540
4	0.15625	43	46.875	0.3203
5	0.03125	12	9.375	0.7350
Toplam	1.00	300	300	$\chi^2 = 4.160$

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

H₀: Üzerinde çalışılan örnekte, istenen olayın ihtimali 0.5 olan bir binom dağılımına uygun dağılmaktadır. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: Üzerinde çalışılan örnekte, istenen olayın ihtimali 0.5 olan bir binom dağılımına uygun dağılmamaktadır. Örneğin bu populasyondan tesadüfen alınmış bir örnek olduğu söylenemez. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Tablo 11.1'de verildiği gibi bu örnek için χ^2 -değeri 4.16 olarak bulunmuştur. Araştırmacı kontrolünü %1 seviyesinde yapmayı kararlaştırmış olsun. Bu örnekte serbestlik derecesi, bir önceki örnekten farklı olarak (sınıf sayısı-1)'dir. Çünkü bu örnekte populasyona ait olasılık ($\pi=1/2$) bilinmektedir ve sadece hesaplamalarda örnek genişliği (300) kullanılmıştır. Bunun için $SD=6-1=5$ 'tir. Tablo D'den 5 serbestlik dereceli ki-kare dağılımında %1'lik alan 15.086 değerinden başladığı bulunur. Hesaplanan χ^2 -değeri (4.16), 15.086 değerinden küçüktür ve kontrol hipotezini kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi reddedilemez, örneğin kız çocuk sayısı bakımından $\pi=1/2$ olan bir binom dağılımına uyum gösterdiği söylenir.

ÖRNEK 2:

Tablo 3.1'de verilen sigara içmeyen öğrencilerin gözlenen ve $\pi=0.75$ olan binom dağılımına göre beklenen sayıları verilmişti. Örneğin, üzerinde durulan olayın oluş ihtimali 0.75 olan binom dağılımına uyum gösterdiği söylenebilir mi?

Sigara içmeyen öğrenci sayısı	F	f'	$(f-f')^2 / f'$
0	2	0.8	
1	12	9.4	1.416
2	45	42.2	0.186
3	80	84.4	0.229
4	61	63.2	0.077
Toplam	200	200	$\chi^2=1.908$

H₀: Üzerinde çalışılan örnek, $\pi=0.75$ olan bir binom dağılımına uygun dağılmaktadır. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: Üzerinde çalışılan örnek, $\pi=0.75$ olan bir binom dağılımına uygun dağılmamaktadır. Örneğin bu populasyondan tesadüfen alınmış bir örnek olduğu söylenemez. Beklenen ve gözlenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Ki-kare değeri hesaplanırken birinci sınıf için beklenen frekans 5'ten küçük olduğu için birinci ve ikinci sınıf birleştirilmiştir. Çünkü hatırlanacağı üzere Bölüm 11.1'de ki-kare dağılımına uyum için beklenen frekansın 5'ten büyük olması gerektiği belirtilmişti. İlk iki sınıf birleştirildiği için toplam sınıf sayısı 4 olmuştur. Bu örnek için χ^2 -değeri 11.3 numaralı eşitlik kullanılarak;

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$\chi^2 = \frac{(14-10.2)^2}{10.2} + \frac{(45-42.2)^2}{42.2} + \frac{(80-84.4)^2}{84.4} + \frac{(61-63.2)^2}{63.2}$$

$$= 1.908 \text{ olarak bulunur.}$$

Araştırmacı I. tip hata ihtimalini %5 olarak belirlemiş olsun. Bu örnekte serbestlik derecesi $SD=4-1=3$ 'tür. Hesaplamalarda örnek genişliği (200) kullanılmıştır. Tablo D'den 3 serbestlik dereceli ki-kare dağılımında %5'lik alan 7.815 değerinden başladığı bulunur. Hesaplanan χ^2 -değeri (1.908), 7.815 değerinden küçüktür ve kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi yani, örneğin $\pi=0.75$ olan bir binom dağılımına uyum gösterdiği kabul edilir.

11.2.2.2. Poisson Dağılımına Uyum Kontrolü

ÖRNEK:

300 tablet bulunan 200 ambalaj için gözlenen ve ortalaması 1.23 olan Poisson dağılımına göre beklenen frekanslar Tablo 3.4'te verilmişti. 200 ambalajlık örneğin ortalaması 1.23 olan Poisson dağılımı gösterdiği söylenebilir mi?

Kırık tablet sayısı	f	f'	(f-f') ² / f'
0	56	58.46	0.104
1	77	71.91	0.360
2	40	44.22	0.403
3	20	18.13	0.193
4	6	5.58	0.0108
5	1	1.70	
Toplam	200	200	$\chi^2=1.07$

H₀: Ambalajlardaki kırık tablet sayısı ortalaması 1.23 olan Poisson popülasyona uygun bir dağılım göstermektedir.

H₁: Ambalaj kırık tablet sayısı ortalaması 1.23 olan Poisson popülasyona uygun bir dağılım göstermemektedir.

Ki-kare değeri hesaplanırken son sınıf için beklenen frekans 5'ten küçük olduğu için son iki sınıf birleştirilmiştir. Son iki sınıf birleştirildiği için toplam sınıf sayısı 5 olmuştur. Bu örnek için χ^2 -değeri 11.3 numaralı eşitlik kullanılarak;

$$\chi^2 = \frac{(56-58.46)^2}{58.46} + \dots + \frac{(7-7.28)^2}{7.28} = 1.07 \text{ olarak bulunur.}$$

Araştırmacı I. tip hata ihtimalini %5 olarak belirlemiş olsun. Bu örnekte serbestlik derecesi $SD=5-2=3$ 'tür. Çünkü bu örneğin alındığı popülasyon için ortalama kırık tablet sayısı örnekten tahmin edilmiştir. Ve ayrıca hesaplamalarda örnek genişliği (200)

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

kullanılmıştır. Tablo D'den 3 serbestlik dereceli ki-kare dağılımında %5'lik alan 7.815 değerinden başlamaktadır. Hesaplanan χ^2 -değeri (1.0708), 7.815 değerinden küçüktür ve kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi, örneğin kırık tablet sayısı 1.23 olan Poisson bir dağılıma uyum gösterdiği kabul edilir.

11.2.2.3. Normal Dağılıma Uyum Kontrolü

ÖRNEK:

Tablo 3.5'te 120 bebeğin doğum ağırlığı için gözlenen ve ortalaması 3.414 ve standart sapması 0.2152 olan normal dağılıma göre olması beklenen frekansları verilmişti. Bebeklerin doğum ağırlığının normal dağılıma uygun bir dağılım gösterdiği söylenebilir mi?

Sınıflar	F	f	(f-f') ² / f'
2.85-2.94	1	1.750	5.23
2.95-3.04	2	1.480	
3.05-3.14	9	7.440	
3.15-3.24	15	13.104	0.274
3.25-3.34	19	19.164	0.001
3.35-3.44	24	21.744	0.234
3.45-3.54	19	20.808	0.157
3.55-3.64	14	15.432	0.133
3.65-3.74	8	9.660	0.285
3.75-3.84	5	4.680	7.416
3.85-3.94	3	1.920	
3.95-4.04	1	0.816	
Toplam	120	119.998	$\chi^2=2.701$

H₀: Örnekteki bebeklerin doğum ağırlığı normal dağılım göstermektedir. Gözlenen ve beklenen frekanslar arasındaki fark tesadüften ileri gelmektedir.

H₁: Örnekteki bebeklerin doğum ağırlığı normal dağılım göstermemektedir. Gözlenen ve beklenen frekanslar arasındaki fark tesadüften ileri gelmemektedir.

Ki-kare değeri hesaplanırken beklenen frekans 5'ten küçük olduğu için ilk iki sınıf ve son 3 sınıf birleştirilmiştir. Birleştirme işlemi son 3 sınıfın birleştirilmesiyle toplam sınıf sayısı 9 olmuştur. Bu örnek için χ^2 -değeri 11.3 numaralı eşitlik kullanılarak;

$$\chi^2 = \frac{(3-5.23)^2}{5.23} + \frac{(9-7.44)^2}{7.44} + \dots + \frac{(9-7.416)^2}{7.416} = 2.701 \text{ olarak bulunur.}$$

Araştırmacı I. tip hata ihtimalini %5 olarak belirlemiş olsun. Bu örnekte serbestlik derecesi daha önce verilen örneklerden farklı olarak (sınıf sayısı-3)'tür. Serbestlik derecesi bulunurken 3 çıkarılmasının nedeni, hesaplamalar yapılırken;

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

1. Örnek genişliği kullanılmıştır.
 2. Populasyon ortalaması bilinmediği için örnekten hesaplanan (3.414) ortalama kullanılmıştır.
 3. Populasyona ait standart sapma bilinmediği için örnekten hesaplanan (0.2152) standart sapma kullanılmıştır.
- Bu sebeple de serbestlik derecesi hesaplanırken sınıf sayısından 3 çıkarılmıştır.

******Yapılan açıklamalardan anlaşılacağı üzere, eğer populasyona ait ortalama ve standart sapma biliniyorsa serbestlik derecesi hesaplanırken sınıf sayısından bir çıkarılır.*******

Söz konusu örnek için serbestlik derecesi $SD=9-3=6$ 'dır. Tablo D'den 6 serbestlik dereceli ki-kare dağılımında %5'lik alan 12.592 değerinden başlamaktadır. Hesaplanan χ^2 -değeri (2.701), 12.592 değerinden küçüktür ve kontrol hipotezini kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi, yani bebeklerin doğum ağırlığının normal dağılım göstermekte olduğu kararına varılır.

11.2.3. Bağımsızlık Kontrolleri

Yapılan bir araştırmada toplanan veriler iki özelliğin çeşitli hallerine göre sınıflandırılarak iki yanlı tablolar oluşturulabilir. Bu gibi durumlarda amaç, üzerinde çalışılan örnekten elde edilen verilerde bir özelliğin çeşitli hallerine göre dağılımın, diğer özelliğin bütün halleri için aynı olup olmadığını kontrol etmek, yani kısaca söz konusu iki özellik arasında bir bağımlılığın olup olmadığını araştırmak olabilir. Özelliklerin iki veya daha fazla hali olabilir. Özelliklerin hal sayısına bağlı olarak oluşturulan tablolar 2×2 , $2 \times C$ veya $R \times C$ tabloları olabilir (C, İngilizce Column (sütun), R, Row (sıra)). Bağımsızlık kontrollerinin nasıl yapıldığı örnekler ile açıklanacaktır.

ÖRNEK 1:

Bir hastalığı tedavi için geliştirilen bir ilacı denemek amacı ile yürütülen bir araştırma sonucunda elde edilen veriler, kullanılan ilaçlara ve hastaların iyileşip iyileşmemelerine göre sınıflandırılarak aşağıdaki şekilde verilmiştir. Hastaların iyileşip iyileşmemelerinin kullanılan ilaca göre değiştiği söylenebilir mi?

	İyileşen hastalar	İyileşemeyen hastalar	Toplam
Kullanılmakta olan ilaç	$f_{11}=40$	$f_{12}=15$	$f_{1.}=55$
Yeni geliştirilen ilaç	$f_{21}=58$	$f_{22}=7$	$f_{2.}=65$
Toplam	$f_{.1}=98$	$f_{.2}=22$	$f_{..}=120$

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

H₀: Hastaların iyileşip iyileşmemeleri kullanılan ilaca göre değişmemektedir. Hastalığın iyileşme durumu ile kullanılan ilaçlar birbirinden bağımsızdır. Yani $f-f'=0$ 'dır.

H₁: Hastaların iyileşip iyileşmemeleri kullanılan ilaca göre değişmektedir. Hastalığın iyileşme durumu ile kullanılan ilaçlar birbirine bağımlıdır. Yani $f-f' \neq 0$ 'dır.

Araştırmacı bu şekilde hipotezlerini oluşturduktan sonra 11.3 numaralı eşitliği kullanarak χ^2 -değerini hesaplamalıdır. Bunun için önce beklenen frekansların bulunması gerekir.

Yukarıda verildiği gibi araştırmada 120 hasta bulunmaktadır. Bu hastalardan 55 tanesi kullanılmakta olan ilaç ile, 65 tanesi yeni geliştirilen ilaç ile tedavi edilmiştir. Yine verilerden görüleceği gibi 120 hastanın 98 tanesi iyileşmiş, 22 tanesi iyileşmemiştir. Araştırmacının kontrol hipotezi geçerli ise iyileşme oranı kullanılan ilaca göre değişmemektedir. Yani 120 hastanın 98 tanesi iyileştigiğine göre iyileşme oranı $98/120 \approx 0.8167$ 'dir ve bu oran ilaçlara göre değişmemektedir. İyileşme oranı 0.8167 olarak bulunduğu göre iyileşmeme oranı $1-0.8167=0.1833$ veya iyileşmeyen hasta sayısı 22 olduğuna göre $22/120=0.1833$ 'tür. Bu durumda kullanılmakta olan ilaç ile tedavi edilen 55 hastanın %81.67'sinin iyileşmiş olması beklenir, yani bu örnekte tedavi edilen herhangi bir hastanın iyileşme olasılığı %81.67'dir.

Diğer taraftan hastaların 55'i kullanılmakta olan ilaç ile tedavi edilmiştir. Bunun toplam hastalara oranı $\frac{55}{120} = 0.4583$ 'tür. Diğer bir deyişle örnekte ele alınan herhangi bir hastanın kullanılmakta olan ilaç grubundan olma olasılığı 0.4583'tür. Kontrol hipotezine göre hastaların iyileşip iyileşmediği kullanılmakta olan ilaçlardan bağımsız olduğuna göre, herhangi bir hastanın hem kullanılmakta olan ilaç grubundan ve hem de iyileşen gruptan olma olasılığı $(0.8167)(0.4583)=0.3743$ 'tür. Çünkü bağımsız olayların birlikte olma olasılığı, bunların olasılıklarının çarpımına eşittir. Kontrol hipotezine göre kullanılmakta olan ilaç grubundan iyileşenlerin beklenen sayısı $(120)(0.3743)=44.92$ 'dir. Yani $f'_{11} = 44.92$ 'dir. Bu teorik frekans olasılıklar yerine frekansların oranları konarak da hesaplanabilir.

Yani;

$$f'_{11} = \left(\frac{98}{120} \frac{55}{120}\right)120$$

kısaltma yapılarak $f'_{11} = \frac{(98)(55)}{120} 120 = 44.92$ olarak bulunabilir. Genel olarak;

$$f'_{11} = \left(\frac{f_{1.}}{f_{..}}\right)\left(\frac{f_{.1}}{f_{..}}\right)f_{..} \text{ eşitliği } f'_{11} = \frac{f_{1.}f_{.1}}{f_{..}} \text{ şeklinde yazılabilir.}$$

Burada 98, beklenen frekansın hesaplanacağı gözün bulunduğu sütunun toplamı, 55, beklenen frekansın hesaplanacağı gözün bulunduğu sıranın toplamı, 120 ise genel toplamdır. O halde beklenen frekansı daha kolay hesaplamak için 11.5 numaralı eşitlik kullanılabilir:

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$\text{Beklenen frekans} = \frac{\text{Sıra toplamı} \times \text{Sütun toplamı}}{\text{Genel toplam}} \quad \dots(11.5)$$

2 x 2 tablolarında, beklenen frekanslar hesaplanırken farkında olunması gereken bir nokta şudur: Gözlerden herhangi biri için beklenen frekans hesaplandıktan sonra, diğer gözlerin beklenen frekansı satır ve sütun toplamlarından bunu çıkararak bulunabilir. Örneğin kullanılmakta olan ilaç ile tedavi edilen hastalardan 44.92 tanesinin iyileşmesi beklenmektedir. Bu ilaç ile tedavi edilen toplam hasta sayısı 55 olduğuna göre iyileşmemesi beklenen hasta sayısı $55 - 44.92 = 10.08$ 'dir. Aynı şekilde iyileşen hasta sayısı 98'dir. Bunlardan 44.92 tanesi kullanılmakta olan ilaç ile tedavi edilen ve iyileşmesi beklenen hasta sayısı olduğuna göre yeni ilaç ile tedavi edilen ve iyileşmesi beklenen hasta sayısı $98 - 44 = 53.08$ 'dir. Yeni ilaç ile tedavi edilip iyileşmemesi beklenen hasta sayısı, ya kullanılmakta olan ilaç ile tedavi edilip iyileşmemesi beklenen hasta sayısını 22'den çıkararak ya da yeni ilaç ile tedavi edilip iyileşmesi beklenen hasta sayısını 65'den çıkararak bulunabilir.

Beklenen frekanslar hesaplandıktan sonra χ^2 -değeri 11.3 numaralı eşitlik kullanılarak;

$$\begin{aligned} \chi^2 &= \frac{(40 - 44.92)^2}{44.92} + \frac{(15 - 10.08)^2}{10.08} + \frac{(58 - 53.08)^2}{53.08} + \frac{(7 - 11.92)^2}{11.92} \\ &= 0.539 + 2.401 + 0.456 + 2.031 = 5.427 \text{ olarak bulunur.} \end{aligned}$$

χ^2 -değeri hesaplandıktan sonra araştırmacının yapması gereken hangi hipotezi kabul edeceğine karar vermesidir. Bunun önce örneği için serbestlik derecesini bulmalıdır. 2x2 tablolarında serbestlik derecesi 1'e eşittir. Çünkü yukarıda açıklandığı şekilde araştırmacı bir göz için beklenen frekansı hesaplamış ise buna bağlı olarak diğerlerini bulabilir. Fakat 2xC ve RxC tablolarında (2x2 tabloları dahil) serbestlik derecesini daha kolay bulmak için 11.6 numaralı eşitlik kullanılabilir.

$$\text{Serbestlik derecesi} = (\text{Satır sayısı} - 1) \times (\text{Sütun sayısı} - 1) \quad \dots(11.6)$$

Örnek için 11.6 numaralı eşitlik kullanılacak olursa serbestlik derecesi yine $SD = (2 - 1) \times (2 - 1) = 1$ olarak bulunur. Tablo D'de 1 serbestlik dereceli ki-kare dağılımında %5'lik alan 3.841 değerinden başlamaktadır. Hesaplanan χ^2 -değeri (5.427), 3.841 değerinden büyüktür ve kontrol hipotezini ret bölgesine düşmektedir. Bu durumda kontrol hipotezi reddedilir, yani hastaların iyileşip iyileşmemeleri kullanılan ilaca göre değişmektedir. Araştırmacı, hastalığın iyileşme durumu ile kullanılan ilaçların birbirine bağımlı olduğu kararına varır.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

ÖRNEK 2:

Sigara kullanma alışkanlığının bölgelere göre değişip değişmediğini araştırmak üzere 5 bölgede uygulanan anket sonuçları aşağıdaki gibi bulunmuştur. Sigara kullanma alışkanlığı bölgeden bölgeye göre değişmekte midir?

BÖLGELER	Sigara kullanma alışkanlığı olanlar	Sigara kullanma alışkanlığı olmayanlar	Toplam
Bölge 1	$f_{11}=44$ $f'_{11} = 51.13$	$f_{12}=30$ $f'_{12} = 22.87$	$f_{1.}=74$
Bölge 2	$f_{21}=120$ $f'_{21} = 107.1$	$f_{22}=35$ $f'_{22} = 47.90$	$f_{2.}=155$
Bölge 3	$f_{31}=55$ $f'_{31} = 46.98$	$f_{32}=13$ $f'_{32} = 21.02$	$f_{3.}=68$
Bölge 4	$f_{41}=19$ $f'_{41} = 41.46$	$f_{42}=41$ $f'_{42} = 18.54$	$f_{4.}=60$
Bölge 5	$f_{51}=75$ $f'_{51} = 66.33$	$f_{52}=21$ $f'_{52} = 29.67$	$f_{5.}=96$
Toplam	$f_{.1}=313$ $f'_{.1}=313$	$f_{.2}=140$ $f'_{.2}=140$	453

H₀: Sigara içme alışkanlığı bölgelere göre değişmemektedir. Kişilerin sigara içme alışkanlığı oturduğu bölgelerden bağımsızdır.

H₁: Sigara içme alışkanlığı bölgelere göre değişmektedir. Kişilerin sigara içme alışkanlığı oturduğu bölgelerden bağımsız değildir.

Hipotezler kurulduktan sonra 11.3 numaralı eşitlik kullanılarak χ^2 -değerinin hesaplanması gerekmektedir. χ^2 -değerinin hesaplanması için gerekli olan beklenen frekanslar 11.5 numaralı eşitlik kullanılarak aşağıdaki şekilde bulunur.

$$f'_{11} = \frac{313 \times 74}{453} = 51.13 \quad f'_{12} = \frac{140 \times 74}{453} = 22.87$$

$$f'_{21} = \frac{313 \times 155}{453} = 107.1 \quad f'_{22} = \frac{140 \times 155}{453} = 47.90$$

$$f'_{31} = \frac{313 \times 68}{453} = 46.98 \quad f'_{32} = \frac{140 \times 68}{453} = 21.02$$

$$f'_{41} = \frac{313 \times 60}{453} = 41.46 \quad f'_{42} = \frac{140 \times 60}{453} = 18.54$$

$$f'_{51} = \frac{313 \times 96}{453} = 66.33 \quad f'_{52} = \frac{140 \times 96}{453} = 29.67$$

Beklenen frekanslar hesaplandıktan sonra χ^2 -değeri;

$$\chi^2 = \frac{(44 - 51.13)^2}{51.13} + \frac{(30 - 22.87)^2}{22.87} + \dots + \frac{(21 - 29.67)^2}{29.67} = 55.716 \text{ olarak bulunur.}$$

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Örnek için serbestlik derecesi 11.6 numaralı eşitlik kullanılarak, $SD=(5-1) \times (2-1)=4$ olarak bulunur. Bunun anlamı, eğer araştırmacı 5x2 tablosunda 10 gözden farklı sıralarda olan 4'üne ait beklenen frekansları bulmuş ise yukarıda verildiği gibi her göz için ayrı ayrı beklenen frekansları bulmak yerine 4 göz için hesaplanan beklenen frekansları satır ve sütun toplamlarından çıkararak diğer gözler için beklenen frekansları bulabilir. Tablo D'de 4 serbestlik dereceli ki-kare dağılımında %5'lik alan 9.488 değerinden başlamaktadır. Hesaplanan χ^2 -değeri (55.716), 9.488 değerinden büyüktür ve ret bölgesine düşmektedir. Bu durumda kontrol hipotezi reddedilir, yani sigara içme alışkanlığı bölgelere göre değişmektedir. Kişilerin sigara içme alışkanlığı oturduğu bölgelerden bağımsız değildir. Anket sonucunda elde edilen veriler incelenecek olursa, kontrol hipotezinin doğru olması durumunda 4. bölgeden ankete katılan 60 kişiden 41.46 kişinin sigara içmesi beklenirken 19 kişinin sigara içme alışkanlığı olduğu gözlenmiştir. Aynı şekilde bu bölgede 60 kişiden 18.54'ünün sigara içmemesi beklenirken anket sonucunda sigara içme alışkanlığı olmayan 41 kişi olduğu gözlenmiştir. Kontrol hipotezinin reddedilmesine sebep bu gibi farklılıklardır.

11.2.3.1. Bağımlılık (Contingency) Katsayısı

İki yanlı tablolarda χ^2 -kontrolü yapılarak H_0 hipotezi reddedilmiş ise üzerinde durulan iki özellik bağımlı demektir. İki yanlı tablolarda bağımlılığın derecesini belirtmek için en yaygın olarak kullanılan Pearson'un “**contingency**” katsayısıdır. Kısaca CC (coefficient of contingency) olarak gösterilir ve 11.7 numaralı eşitlikte verildiği gibi hesaplanır.

$$CC = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad \dots(11.7)$$

2 numaralı örnekte hesaplanan χ^2 -değerine göre sigara içme alışkanlığının bölgelere göre değiştiğine, yani kişilerin sigara içme alışkanlığı ile oturduğu bölgelerin bağımlı olduğuna karar verilmişti. Bu örnek için iki özellik arasındaki bağımlılığın derecesi;

$$CC = \sqrt{\frac{55.716}{453 + 55.716}} = 0.331, \text{ yani } \%33.1 \text{ olarak bulunur.}$$

Sigara içme alışkanlığı ile oturuş bölgesi arasında %33.1'lik bir bağımlılık vardır denir.

Contingency katsayısı sadece iki özellik arasındaki bağımlılığın derecesini verir. Fakat bu bağımlılığın yönü hakkında bir bilgi vermez. Üzerinde durulan iki özellik tam bağımlı olsa bile bu katsayısı, 11.7 numaralı eşitlikten görülebileceği gibi, 1 değerine ulaşmaz. Örneğin, 100 bireylik bir örnekten toplanan veriler 2'şer hali olan “a” ve “b” gibi iki özelliğe göre aşağıdaki gibi sınıflandırılmış olsun.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

	B1	B2	Toplam
A1	$f_1=0$	$f_3=50$	50
A2	$f_2=50$	$f_4=0$	50
Toplam	50	50	100

Bu verilerde bir özelliğin çeşitli hallerine göre dağılım, diğer özelliğin hallerine tamamen bağımlıdır. Her göz için beklenen frekans 25 ve bulunacak χ^2 -değeri 100'dür.

Bu örnek için contingency katsayısı;

$$CC = \sqrt{\frac{100}{100+100}} = \sqrt{\frac{1}{2}} = 0.707 \text{ olarak bulunur. Bu değer, } 2 \times 2 \text{ tablolarında iki}$$

özellik tam olarak bağımlı ise bulunacak maksimum değerdir. Bunun için de hesaplanan contingency katsayısının, RxC tablolarında contingency katsayısının alabileceği en büyük değere göre düzeltilmesi gerekir. 2x2 tablolarında contingency katsayısının alabileceği maksimum değer belirtildiği gibi;

$$CC_{\max} = \sqrt{\frac{1}{2}} = 0.707 \text{ 'dir.}$$

Bu değer hesaplandıktan sonra CC, maksimum contingency katsayısına bölünerek düzeltilmiş CC hesaplanır.

RxC tablolarında contingency katsayısının alabileceği en büyük değer ise 11.8 numaralı eşitlikte verildiği gibi hesaplanır.

$$CC_{\max} = \sqrt{\frac{R-1}{R}} \quad \text{eğer } R < C \text{ ise} \quad \dots(11.8)$$

$$CC_{\max} = \sqrt{\frac{C-1}{C}} \quad \text{eğer } C < R \text{ ise}$$

CC_{\max} hesaplandıktan sonra $CC_{\text{düz.}} = CC / CC_{\max}$ şeklinde hesaplanır.

İkinci örnek için $CC=0.331$ olarak bulunmuştur. Örnekteki 5x2 tablosu için hesaplanacak en büyük değer;

$$CC_{\max} = \sqrt{\frac{1}{2}} = 0.707 \text{ 'dir.}$$

Bu durumda $CC_{\text{düz.}} = 0.331 / 0.707 = 0.468$ olarak bulunur.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Birçok kitapta RxC tablolarında satır sayısı sütun sayısına eşit olmadığı zaman, küçük olanı kullanmak yerine her ikisine göre maksimum CC hesaplanmakta, bunların ortalaması CC_{max} olarak kullanılmaktadır. Yani, yukarıda verilen örnek için;

$$\text{Satır sayısına göre; } CC_{max} = \sqrt{\frac{5-1}{5}} = 0.894$$

$$\text{Sütun sayısına göre; } CC_{max} = \sqrt{\frac{2-1}{2}} = 0.707$$

Düzeltilmiş CC'nin hesaplanması için CC_{max} ;

$$CC_{max} = (0.894 + 0.707) / 2 = 0.801 \text{ olarak bulunur.}$$

Buna göre düzeltilmiş CC ise $CC_{düz.} = 0.331 / 0.801 = 0.413$ olarak hesaplanır.

11.3. Bilgisayar Uygulaması

ÖRNEK 1.

Bağımsızlık kontrollerinde Örnek 1'de, yeni geliştirilen bir ilacı denemek amacı ile yürütülen bir araştırma sonucunda elde edilen verilerin analizi açıklanmıştı. Bu örneğin MINITAB paket programı kullanılarak çözümlenmesi ise aşağıdaki gibidir.

Araştırmacı ilk olarak verilerini C1 ve C2 sütunlarına işler ve "İYİLESEN" ve "İYİLESMEYEN" olarak sütunları adlandırabilir. PRINT komutu kullanılarak işlenen veriler aşağıdaki gibi görüntülenir. 1. satır kullanılmakta olan ilaç ile, 2. satır ise yeni geliştirilen ilaç ile tedavi edilen hastalardır.

```
MTB > PRINT C1 C2
```

Data Display

Row	İYİLESEN	İYİLESMEYEN
1	40	15
2	58	7

Veriler C1 ve C2 sütununa işlendikten sonra ki-kare kontrolünün yapılması için kullanılacak komut CHISQUARE'dir. Veya MINITAB paket programından bir komutun ilk dört harfini vermek yeterli olduğu için "CHIS" komutu verilir. Komut verildikten sonra ise aşağıda görüldüğü gibi verilerin işlendiği sütun adları belirtilir. Aşağıdaki gibi komut verildiğinde MINITAB paket programı her göz için beklenen frekansları, $(f-f')^2 / f$ değerlerini, ki-kare test değerini ve serbestlik derecesini ve hesaplanan ki-kare değerinin söz konusu serbestlik dereceli ki-kare dağılımına dahil olma olasılığını verir. Araştırmacı Tablo D'yi kullanmak yerine bu olasılığı kullanarak karar verebilir.

Aşağıdaki ki-kare testi sonuçlarına bakıldığında zaman P-value=0.020 (%2.0) olduğu görülür. Bunun anlamı; hesaplanan ki-kare değerinin, 1 serbestlik dereceli ki-kare

14. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

dağılımına dahil olma olasılığı %5'ten küçüktür. Bu olasılık %5'ten küçük olduğu için kontrol hipotezi reddedilir ve hastaların iyileşme durumu ile kullanılan ilaçların birbirine bağımlı olduğu kabul edilir.

MTB > CHIS C1 C2

Chi-Square Test: İYİLESEN; İYİLESMEYEN

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	İYİLESEN	İYİLESMEYEN	Total
1	40 44,92 0,538	15 10,08 2,397	55
2	58 53,08 0,455	7 11,92 2,029	65
Total	98	22	120

Chi-Sq = 5,420; DF = 1; P-Value = 0,020

ÖRNEK 2.

Bağımsızlık kontrollerinde Örnek 2'de verilen sigara kullanma alışkanlığının bölgelere göre değişip değişmediğinin araştırıldığı anket sonuçlarının MINITAB paket programı kullanılarak değerlendirilmesi:

PRINT komutu ile MINITAB paket programına işlenen veriler aşağıdaki gibidir. Sütunlar sigara kullanan ve kullanmayan kişiler, satırlar ise bölgelerdir.

MTB > PRINT C3 C4

Data Display

Row	KULLANAN	KULLANMAYAN
1	44	30
2	120	35
3	55	13
4	19	41
5	75	21

MTB > CHIS C3 C4

14. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Chi-Square Test: KULLANAN; KULLANMAYAN

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	KULLANAN	KULLANMAYAN	Total
1	44 51,13 0,994	30 22,87 2,223	74
2	120 107,10 1,555	35 47,90 3,475	155
3	55 46,98 1,367	13 21,02 3,057	68
4	19 41,46 12,165	41 18,54 27,197	60
5	75 66,33 1,133	21 29,67 2,533	96
Total	313	140	453

Chi-Sq = 55,700; DF = 4; P-Value = 0,000

Sonuçları alan araştırmacı en önemli farklılığın 4. bölgeden kaynaklandığını görecektir. P-value değeri %1'den dahi küçüktür. Bu durumda sigara içme alışkanlığının bölgelere göre değiştiğine karar verilir.