XII. BÖLÜM

ÇAPRAZ TABLOLARDA Kİ-KARE TESTLERİ

XII.1 Giriş

Genel olarak Ki-kare testleri olarak da bilinen testler; bağımsızlık, ilişki ve uyum iyiliği testlerini içerir. Özellikle sayımla belirlenen kitlelerde nitel değişkenler ile ilgili hipotezleri test etmek için kullanılır.

XII.2 Ki-Kare Testlerinin Bazı Özellikleri

1) Gözlenen frekanslar ile beklenen frekanslar toplamı gözlem sayısı (yani örneklem hacmi) n'e eşittir.

2) Örneklemdeki gözlem sayısı dağılımı etkilemez. Ancak gözlenen frekansların oluşturduğu tabloda herhangi bir satır veya sütun üzeride beşten az beklenen frekans değeri olmamalıdır. Eğer var ise beşten küçük gözlemin bulunduğu hücre kendine en uygun satır veya sütun ya da hem satır ve hem de sütun ile birleştirilmelidir. Bu duruma bir diğer yaklaşım (özellikle çapraz tablolarda) beklenen frekansı beşten küçük olan hücre sayısının toplam hücre sayısının % 20'sinden fazla olmaması istenir. Eğer % 20'sini aşıyorsa uygun sınıflar arasında birleştirmeler yapılabilir.

3) (2×2) düzenindeki çapraz tablolarda herhangi bir hücrede beşten küçük beklenen frekans değeri var ise birleştirme yerine Fisher'in Kesin Ki-Kare test istatistiği yaklaşımı kullanılması uygundur.

4) Gözlem sayısı 50'den az olan örneklemlerde Ki-Kare test istatistiği için Yates düzeltmesi şarttır. Bunun nedeni kesikli verilere süreklilik kazandırmaktır. Yates düzeltmesi gözlenen frekanslarla beklenen frekanslar arasındaki farkın mutlak değerinden 0.5 çıkarılarak yapılır.

XII.3 Ki-Kare Testlerinin Kullanıldığı Konular

Ki-kare test istatistiğini kullanıldığı konulardan bazıları takipte verildiği gibidir.

1) Sonuçların ortaya çıkması eşit olasılıklı olan tesadüfî deneylerde gözlenen frekanslar ile beklenen frekanslar arasındaki farkın önemlilik kontrolü.
2) Sayımla belirlenen kitlelerde gruplar arası farkın önemlilik kontrolü.
3) Sayımla belirlenen kitlelerde homojenlik kontrolü.
4) Sayımla belirlenen kitlelerde bağımlılık ve ilişki kontrolü.
5) Örneklem dağılımları ile bilinen teorik olasılık dağılımları arasında uyumluluk kontrolü.

XII.4 Ki-Kare Testlerinin İşlem Akışı

a) **Hipotezler:** Bu aşamada araştırmanın amacına yönelik hipotezler belirlenir. H_0 ve H_A olmak üzere iki tür olarak belirlemek gerekir. Bir önceki bölümde verilen problemler için sıfır ve araştırma hipotezleri takipte verildiği gibi belirlenebilir.

1) H_0: " Gözlenen frekanslarla beklenen frekanslar arasındaki fark önemsizdir" sıfır hipotezine karşılık araştırma hipotezi, H_A: "Gözlenen frekanslarla beklenen frekanslar arasındaki fark önemlidir" şeklinde kurulur.

2) H_0: " Gruplar arası fark önemsizdir" sıfır hipotezine karşılık araştırma hipotezi, H_A: "Gruplar arası fark önemlidir" şeklinde kurulur.

3) H_0: " Gruplar arasındaki homojenlik yapısı önemsizdir" sıfır hipotezine karşılık araştırma hipotezi, H_A: "Gruplar arasındaki homojenlik önemlidir" şeklinde kurulur.

4) H_0: " Değişkenler arasındaki ilişki önemsizdir" sıfır hipotezine karşılık araştırma hipotezi, H_A: "Değişkenler arasındaki ilişki önemlidir" şeklinde kurulur.

5) H_0: " Örneklem dağılımı teorik dağılım ile uyumludur" sıfır hipotezine karşılık araştırma hipotezi, H_A: "Örneklem dağılımı teorik dağılım ile uyumlu değildir" şeklinde kurulur.

b) **Test İstatistiği:** İncelenen konuya göre değişiklik gösterir. Öyle olsa da öncelikle uygulama verisinin ürettiği gözlenen frekanslar tablosu yardımı ile ve H_0 hipotezinin doğruluğu altında beklenen frekanslar hesaplanır. Daha sonra test istatistiği belirlenir. Ki-kare testlerinde test istatistiği gözlenen frekanslar ile beklenen frekanslar arasındaki farkın bir ölçüsü olarak tanımlanan $\chi_H^2 \sim \chi_{(r)}^2$ istatistiğidir. Burada r s.d. olup yine uygulanan konuya göre değerlendirilir.

c) **Karar:** α önem seviyesi ve r s.d. olmak üzere $\chi_T^2 = \chi_{r;\,\alpha}^2$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. Eğer $\chi_H^2 \leq \chi_T^2 => H_0$ hipotezi kabul ve eğer $\chi_H^2 > \chi_T^2 => H_0$ hipotezi ret edilir.

d) **Yorum:** Karara göre yorum yapılır.

XII.4.1 Eşit Olasılıklı Tesadüfî Deneylerde Ki-Kare Testi

Eşit olasılıklı deney sonuçlarından meydana gelen çok sayıdaki gözlemin oluşturduğu bir örneklem düşünülsün ve j farklı sonuçtan oluşsun. Her farklı sonucu grup olarak düşündüğümüzde her bir grupta bir olayın ortaya çıkması eşit olasılıklı olduğundan düzenlenen örneklem gruplarında da olayların eşit olasılıklı olarak ortaya çıkması gayet tabiidir. Bununla birlikte genellikle gözlem sayısının az olduğu örneklemlerde tesadüfî farklılıklar ortaya çıkabilir. Böylece örneklemden elde edilen gözlenen frekanslar fg_j ile beklenen fb_j arasında fark görülebilir. Bu durumda araştırmacını amacı söz konusu farkın önemli olup olmadığını incelemektir. Çünkü bu fark tesadüfen ortaya çıkabileceği gibi deneysel bir hata sonucunda da ortaya çıkmış olabilir.

Şimdi bir tesadüfî deneyin sonucunda mümkün olan olaylar $O_1, \cdots O_j$ olmak üzere her bir olayın gerçekleşme olasılığı eşit olsun. Şöyle ki $P(O_1), = \cdots = P(O_j)$ yazılır. Deneyin n defa tekrarlandığını ve oluşturulan örnekleme göre söz konusu olaylar grubunun sırası ile $fg_1, \cdots, fg_J$ frekansları ile gözlemlendiğini düşünelim. Bu durumda veri düzeni takipte verilen tabladaki gibi oluşur.

Tablo: $(1 \times J)$ Düzenindeki Eşit Olasılıklı Deneylerde Veri Düzeni.

Olaylar	O_1	O_2	$\cdots$	O_J	Toplam
Gözlenen Frekans (fg_j)	fg_1	fg_2	$\cdots$	fg_J	n
Beklenen Frekans fb_j	fb_1	fb_2	$\cdots$	fb_J	n

Belli bir olasılık kuralına göre beklenen frekanslar $fb_j = (O_j) \times n,\ j = 1, \cdots, J$ eşitliği ile hesaplanır. Böylece eşit olasılıklı deneyler için test işlem akışı şu şekilde uyarlanabilir.

a) **Hipotezler:** H_0: " Gözlenen frekanslarla beklenen frekanslar arasındaki fark önemsizdir" ve araştırma hipotezi ise H_A: "Gözlenen frekanslarla beklenen frekanslar arasındaki fark önemlidir" şeklinde kurulur.

b) **Test istatistiği:** $\chi_H^2 = \sum_{j=1}^{J} \frac{(fg_j - fb_j)^2}{fb_j} \sim \chi_r^2$ olup $r = J - 1$ s.d.'ne sahiptir.

c) **Karar:** α önem seviyesi ve r s.d. olmak üzere $\chi_T^2 = \chi_{r;\,\alpha}^2$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. Eğer $\chi_H^2 \leq \chi_T^2 => H_0$ hipotezi kabul ve eğer $\chi_H^2 > \chi_T^2 => H_0$ hipotezi ret edilir.

d) **Yorum:** Karara göre yorum yapılır.

♦ **Örnek:** Bir madeni paranın 200 defa atılması tesadüfî deney sonucunda 115 kez yazı 85 kez de tura geldiği gözlemlenmiştir. % 5 önem seviyesinde paranın hilesiz olduğunu söylemek mümkün müdür?

♦ **Çözüm:**

a) Hipotezler: H_0: " Para Hilesizdir" ve araştırma hipotezi ise H_A: "Para Hilelidir" şeklinde kurulur.

b) Test istatistiği: $\chi_H^2 = \sum_{j=1}^{J} \frac{(fg_j - fb_j)^2}{fb_j} \sim \chi_r^2$ olup $r = 2 - 1 = 1$ s.d.'ne sahiptir. Deneye ilişkin beklenen frekanslar, H_0 hipotezinin doğruluğu altında ve paranın iki durumu olup eşit olasılıklı olduğundan $\frac{1}{2}$ olasılığına sahiptir. Buradan $fb_j = n \times P(O_j) = 200 \times \frac{1}{2} = 100$ $j = 1, 2$ olup takipteki tabloda verildiği gibi veri düzenine sahip olur.

Tablo: Madeni Para Atma Deneyinde Veri Düzeni.

Olaylar	Yazı	Tura	Toplam
Gözlenen Frekans (fg_j)	115	85	200
Beklenen Frekans fb_j	100	100	200

$\chi_H^2 = \sum_{j=1}^{2} \frac{(fg_j - fb_j)^2}{fb_j} = \frac{(115 - 100_j)^2}{100} + \frac{(85 - 100_j)^2}{100} = 4.5$ bulunur.

c) Karar: $\alpha = 0.05$ önem seviyesi ve 1 s.d. olmak üzere $\chi_T^2 = \chi_{1;\,0.05}^2 = 3.841$ bulunur (ki-kare dağılım tablosundan). Böylece $\chi_H^2 > \chi_T^2$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: % 95 güvenilirlikle ilgili paranın hileli olduğu söylenebilir.

♦ **Örnek:** Altı yüzlü bir oyun zarı 500 defa atılıyor. Deney sonuçları takipteki gibi elde edilmiştir.

Zarın Üste Gelen Yüzü	1	2	3	4	5	6	Toplam
Gözlenen Frekans (fg_j)	86	77	77	90	88	82	500

% 5 önem seviyesinde bu oyun zarının hilesiz olduğunu söylemek mümkün müdür?

♦ **Çözüm:**

a) Hipotezler: H_0: " Zar Hilesizdir" ve araştırma hipotezi ise H_A: "Zar Hilelidir" şeklinde kurulur.

b) Test İstatistiği: $\chi_H^2 = \sum_{j=1}^{J} \frac{(fg_j - fb_j)^2}{fb_j} \sim \chi_r^2$ olup $r = 6 - 1 = 5$ s.d.'ne sahiptir. Deneye ilişkin beklenen frekanslar, H_0 hipotezinin doğruluğu altında ve zarın altı noktalı yüzü olup eşit olasılıklı olduğundan $\frac{1}{6}$ olasılığına sahiptir. Buradan $fb_j = n \times P(O_j) = 500 \times \frac{1}{6} = 83.33$ $j = 1, 2, 3, 4$ $5, 6$ olur. Böylece test istatistiğinin değeri ise $\chi_H^2 = \sum_{j=1}^{6} \frac{(fg_j - fb_j)^2}{fb_j} \cong 1.864$ bulunur.

c) Karar: : $\alpha = 0.05$ önem seviyesi ve 5 s.d. olmak üzere $\chi_T^2 = \chi_{5;\,0.95}^2 = 11.070$ bulunur (ki-kare dağılım tablosundan). Böylece $\chi_H^2 \leq \chi_T^2$ olduğundan H_0 hipotezi kabul edilir.

d) Yorum: % 95 güvenilirlikle ilgili oyun zarının hilesiz olduğu söylenebilir.

XII.4.2 (2×2) Düzenindeki Verilerde Ki-Kare Testi

(2×2) düzenindeki çapraz tablolar bir birinden farklı herhangi iki niteliğin her birinin iki farklı özelliğine göre birimlerin aldığı değerleri içeren tablolardır. Böylesi bir veri düzeni takipteki bir tablo ile verilebilir.

Tablo: (2×2) Düzenindeki Deneylerde Veri Düzeni.

	II. NİTELİK			
I. NİTELİK	XXXXX	A*	B*	TOPLAM
	A	fg_{11} / fb_{11}	fg_{12} / fb_{12}	$T_{1.}$
	B	fg_{21} / fb_{21}	fg_{22} / fb_{22}	$T_{2.}$
	TOPLAM	$T_{.1}$	$T_{.2}$	$T_{..} = n$

Tablodan da görüldüğü gibi n adet gözlem birimi iki ayrı niteliğe göre dört hücreye dağılmaktadır. Tabloda; $T_{..} = n$: örneklem hacmini, fg_{ij}: Satır değişkeninin $i.$ ve sütun değişkeninin $j.$ seviyesindeki gözlenen birim sayısını (frekansını), fb_{ij}: Satır değişkeninin $i.$ ve sütun değişkeninin $j.$ seviyesindeki beklenen birim sayısını (frekansını) göstermektedir. Ayrıca $i.$ satır toplamı, $T_{i.}$ ve $j.$ sütun toplamı, $T_{.j}$ ile gösterilir ise $T_{i.} = \sum_{j=1}^{2} fg_{ij}$, $T_{.j} = \sum_{i=1}^{2} fg_{ij}$ve $T_{..} = n = \sum_{i=1}^{2} T_{i.} = \sum_{j=1}^{2} T_{.j}$ eşitlikleri yazılabilir. İki nitelik bağımsız iken (yani; H_0 hipotezi geçerli iken) ortak olasılık, marjinal olasılıklar çarpımına eşit olduğundan beklenen frekaslar, $fb_{ij} = \frac{T_{i.} \times T_{.j}}{T_{..}}$, $i = 1, 2;\ j = 1, 2$ eşitliği ile hesaplanır. Bu bilgiler ışığı altında test akışı aşağıdaki gibi verilir.

a) **Hipotezler:** H_0: " Birinci nitelik ile ikinci nitelik arasında ilişki yoktur" ve araştırma hipotezi ise H_A: "Birinci nitelik ile ikinci nitelik arasında ilişki vardır" şeklinde kurulur.

b) **Test istatistiği:** $\chi_H^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(fg_{ij} - fb_{ij})^2}{fb_{ij}} \sim \chi_r^2$ olup $r = (2-1) \times (2-1) = 1$ s.d.'ne sahiptir.

c) **Karar:** α önem seviyesi ve r s.d. olmak üzere $\chi_T^2 = \chi_{r;\,\alpha}^2$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. Eğer $\chi_H^2 \leq \chi_T^2 \Rightarrow H_0$ hipotezi kabul ve eğer $\chi_H^2 > \chi_T^2 \Rightarrow H_0$ hipotezi ret edilir.

d) **Yorum:** Karara göre yorum yapılır.

♦ **Örnek:** Bir üniversitede okuyan kız öğrenciler arasında yapılan bir araştırmada bir spor faaliyetine katılıp katılmadıkları (birinci nitelik) ve abilerinin olup olmadığı (ikinci nitelik) soruluş ve bu iki özelliğe ilişkin sınıflandırılmışlardır. Elde edilen sonuçlar aşağıdaki tabloda verildiği gibi elde edilmiştir. % 5 önem seviyesinde sınıflandırmanın iki özelliği arasında ilişki olduğu söylenebilir mi?

Tablo: Spor-Abi Değişkenleri Araştırması İçin Veriler

I. NİTELİK: Sporla İlgilenme	**II. NİTELİK: Abi**			
	XXXXX	Var	Yok	TOPLAM
	İlgileniyor	12 / 8.33	8 / 11.67	20
	İlgilenmiyor	13 / 16.67	27 / 23.33	40
	TOPLAM	25	35	60

♦ **Çözüm:**

a) Hipotezler: H_0: " Sporla ilgilenme ile abi arasında ilişki yoktur" ve araştırma hipotezi ise H_A: " Sporla ilgilenme ile abi arasında ilişki vardır" şeklinde kurulur.

b) Test İstatistiği: $\chi_H^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(fg_{ij}-fb_{ij}\right)^2}{fb_{ij}} \sim \chi_1^2$'dir. H_0 hipotezinin doğruluğu altında beklenen frekanslar: $fb_{11} = \frac{20\times25}{60} \cong 8.33$ $fb_{12} = \frac{20\times35}{60} \cong 11.67$ $fb_{21} = \frac{40\times25}{60} \cong 16.67$ ve $fb_{22} = \frac{40\times35}{60} \cong 23.33$ bulunur. Böylece test istatistiğinin değeri ise $\chi_H^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(fg_{ij}-fb_{ij}\right)^2}{fb_{ij}} \cong 4.16$ bulunur.

c) Karar: : $\alpha = 0.05$ önem seviyesi ve 1 s.d. olmak üzere $\chi_T^2 = \chi_{1;\,0.95}^2 = 3.841$ bulunur (ki-kare dağılım tablosundan). Böylece $\chi_H^2 > \chi_T^2$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: % 95 güvenilirlikle kız öğrencilerinin sporla ilgilenme durumu ile abilerinin olup olmaması arasında ilişki olduğu söylenebilir.

♦ **Örnek:** Erkek ve kadınlarda yapılan renk körlüğü testi sonuçları aşağıdaki tabloda verildiği gibi elde edilmiştir. % 5 önem seviyesinde sınıflandırmanın iki özelliği arasında ilişki olduğu söylenebilir mi?

Tablo: Cinsiyet-Renk Körlülüğü Değişkenleri Araştırması İçin Veriler

I. NİTELİK: Cinsiyet	**II. NİTELİK: Test Sonucu**			
	XXXXX	Normal	Renk Körü	TOPLAM
	Erkek	30 / 22.5	30 / 37.5	60
	Kadın	120 / 127.5	220 / 212.5	340
	TOPLAM	150	250	400

♦ **Çözüm:**

a) Hipotezler: H_0: " Sporla ilgilenme ile abi arasında ilişki yoktur" ve araştırma hipotezi ise H_A: " Sporla ilgilenme ile abi arasında ilişki vardır" şeklinde kurulur.

b) Test İstatistiği: $\chi_H^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(fg_{ij}-fb_{ij}\right)^2}{fb_{ij}} \sim \chi_1^2$'dir. H_0 hipotezinin doğruluğu altında beklenen frekanslar: $fb_{11} = \frac{60\times150}{400} = 22.5$ $fb_{12} = \frac{60\times250}{400} = 37.5$ $fb_{21} = \frac{340\times150}{400} = 127.5$ ve

$fb_{22} = \frac{340\times250}{400} = 212.5$ bulunur. Böylece test istatistiğinin değeri ise $\chi_H^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(fg_{ij}-fb_{ij})^2}{fb_{ij}} \cong 4.706$ bulunur.

c) Karar: : $\alpha = 0.05$ önem seviyesi ve 1 s.d. olmak üzere $\chi_T^2 = \chi_{1;\,0.95}^2 = 3.841$ bulunur (ki-kare dağılım tablosundan). Böylece $\chi_H^2 > \chi_T^2$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: % 95 güvenilirlikle cinsiyet itibari ile renk körlğü arasında bir ilişki olduğu söylenebilir.

XII.4.3 (2×2) Düzenindeki Verilerde Fisher'in Kesin Ki-Kare Testi

(2×2) düzenindeki çapraz tabloları için gözlenen frekanslardan herhangi biri beşten küçük olduğunda birleştirme yapılamadığı için Fisher'in kesin-ki kare test istatistiği kullanılır. Test işlem akışında sadece test istatistiği Pearson'un Ki-kare testi yerine Fisher'in kesin ki-kare test istatistiği hesaplanır. Karar bölümünde ise test sonucu önem seviyesi olasılığı ile karşılaştırılır.

Bu durumda izlenecek yol şu şekilde ifade edilir. Eğer $fg_{ij} < 5$ ise satır ve sütun toplamları değişmeksizin 5'ten küçük olan frekans her seferinde bir birim azaltılarak sıfır oluncaya kadar ilave tablolar düzenlenir. Bu şekilde düzenlenen tablolar yardımı ile Fisher'in kesin ki-kare test istatistiği hesaplanır. Şöyle ki $F\chi_H^2 = \sum_{Tablo\ Sayısı}\frac{T_{1.}!T_{2.}!T_{.1}!T_{.2}!}{fg_{11}!fg_{12}!fg_{21}!fg_{22}!T_{..}!}$ eşitliği ile hesaplanır. Bu sonuç $0 \le F\chi_H^2 \le 1$ arasında değerler alır. Karar kriteri olarak ise eğer $F\chi_H^2 \ge \alpha$ ise H_0 hipotezi kabul. $F\chi_H^2 < \alpha$ ise ret edilir.

♦ **Örnek:** Bir hastalıkta *A* ve *B* ilaçları denenmiş ve tedavi sonucunda iki durum gözlemlenmiştir. Sonuçlar aşağıdaki tabloda verildiği gibi elde edildiğine göre % 5 önem seviyesinde ilaç türü ile tedavi sonucu arasında bir ilişki olduğu söylenebilir mi?

I. NİTELİK: İlaç Türü	**II. NİTELİK: Tedavi Sonucu**			
	XXXXX	İyileşen	İyileşmeyen	TOPLAM
	A	2	14	16
	B	5	20	25
	TOPLAM	7	34	41

♦ **Çözüm:**

a) Hipotezler: H_0: " İlaç türü ile tedavi sonucu arasında ilişki yoktur" ve araştırma hipotezi ise H_A: "İlaç türü ile tedavi sonucu arasında ilişki vardır" şeklinde kurulur.

b) Test İstatsitiği: $fg_{11} > 5$ olduğundan Fisher'in kesin ki-kare istatistiği kullanılır. $fg_{11} = 2$ olduğundan iki ilave tablo düzenlenmelidir.

İlaç Türü	Tablo.1		Tablo.2	
A	1	15	0	16
B	6	19	7	18

$F\chi_H^2 = \sum_3 \frac{T_{1.}!T_{2.}!T_{.1}!T_{.2}!}{fg_{11}!fg_{12}!fg_{21}!fg_{22}!T_{..}!} = \frac{16!25!7!34!}{2!14!5!20!41!} + \frac{16!25!7!34!}{1!15!6!19!41!} + \frac{16!25!7!34!}{0!16!7!18!41!} \cong 0.284 + 0.126$ $+0.021 \cong 0.431$ bulunur.

c) Karar: : $\alpha = 0.05$ önem seviyesi olmak üzere $F\chi_T^2 = \alpha = 0.05$ yazılır. Böylece $F\chi_H^2 > F\chi_T^2$ olduğundan H_0 hipotezi kabul edilir.

d) Yorum: % 95 güvenilirlikle ilaçların tedavi sonucu arasında bir ilişki olduğu söylenemez.

XII.4.4 $(\mathrm{I} \times \mathrm{J})$ Düzenindeki Verilerde Ki-Kare Testi

Verlerin $(\mathrm{I} \times \mathrm{J})$ düzeninde olması bir çeşit (2×2) düzeninin genişletilmiş hali gibi düşünülebilir. $I = J$ olabileceği gibi $I \neq J$ de olabilir. Bu durumda n adet birim iki farklı nitelikten satır değişkeninin I ve sütun değişkeninin de J tane seviyesine göre $(I \times J)$ adet hücreye dağıtılır. Dağıtım tablosu aşağıda verildiği gibi olur.

Tablo: $(I \times J)$ Düzenindeki Deneylerde Veri Düzeni.

		II. NİTELİK				
	XXXXX	B_1	B_2	$\cdots$	B_J	TOPLAM
I. NİTELİK	A_1	$fg_{11}\,/\,fb_{11}$	$fg_{12}\,/\,fb_{12}$	$\cdots$	$fg_{1J}\,/\,fb_{1J}$	$T_{1.}$
	A_2	$fg_{21}\,/\,fb_{21}$	$fg_{22}\,/\,fb_{22}$	$\cdots$	$fg_{2J}\,/\,fb_{2J}$	$T_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_I	$fg_{I1}\,/\,fb_{I1}$	$fg_{I2}\,/\,fb_{I2}$	$\cdots$	$fg_{IJ}\,/\,fb_{IJ}$	$T_{I.}$
	TOPLAM	$T_{.1}$	$T_{.2}$	$\cdots$	$T_{.J}$	$T_{..} = n$

Tablodan da görüldüğü gibi n adet gözlem birimi iki ayrı niteliğe göre $(I \times J)$ adet hücreye dağılmaktadır. Tabloda; $T_{..} = n$: örneklem hacmini, fg_{ij}: Satır değişkeninin $i.$ ve sütun değişkeninin $j.$ seviyesindeki gözlenen birim sayısını (frekansını), fb_{ij}: Satır değişkeninin $i.$ ve sütun değişkeninin $j.$ seviyesindeki beklenen birim sayısını (frekansını) göstermektedir. Ayrıca $i.$ satır toplamı, $T_{i.}$ ve $j.$ sütun toplamı, $T_{.j}$ ile gösterilir ise $T_{i.} = \sum_{j=1}^{J} fg_{ij}$, $T_{.j} = \sum_{i=1}^{I} fg_{ij}$ve $T_{..} = n = \sum_{i=1}^{I} T_{i.} = \sum_{j=1}^{J} T_{.j}$ eşitlikleri yazılabilir. İki nitelik bağımsız iken (yani; H_0 hipotezi geçerli iken) ortak olasılık, marjinal olasılıklar çarpımına eşit olduğundan beklenen frekaslar, $fb_{ij} = \frac{T_{i.} \times T_{.j}}{T_{..}}$, $i = 1, 2 \cdots, I; \ j = 1, 2, \cdots, J$ eşitliği ile hesaplanır. Bu bilgiler ışığı altında test akışı (2×2) veri düzeninde verildiği gibidir. Burada sadece serbestlik derecesi $r = (I-1) \times (J-1)$ eşitliği ile verilir.

♦ **Örnek:** Bir mobilya işletmecisinin mobilyaları üç farklı işçi ekibi tarafından üretilmektedir. Ürünlerin üretiminde bazı hatalar görülmüştür. Bu amaçla yapılan bir çalışmada hatalar da üç krup altında toplamış ve 270 kusurlu mobilya; hata türü ve üretim ekibine göre sınıflandırmış ve elde edilen sonuçlar tabloda verildiği gibi elde edilmiştir.

Tablo: Ekip-Hata Deneyinde Veriler.

	II. NİTELİK: Hata Türü				
I. NİTELİK: Ekip grubu	XXXXX	H_1	H_2	H_3	TOPLAM
	E_1	15 / 21.926	45 / 37.926	20 / 20.148	80
	E_2	26 / 24.941	34 / 43.141	31 / 22.918	91
	E_3	33 / 27.133	49 / 46.933	17 / 24.934	99
	TOPLAM	74	128	68	270

$\alpha = 0.05$ önem seviyesinde ekipler ile hata türü arasında bir ilişki olduğu söylenebilir mi?

♦ **Çözüm:**

a) Hipotezler: H_0: " Ekip grubu ile hata türü arasında ilişki yoktur" ve araştırma hipotezi ise H_A: "Ekip grubu ile hata türü arasında ilişki vardır" şeklinde kurulur.

b) Test İstatistiği: $\chi_H^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(fg_{ij}-fb_{ij})^2}{fb_{ij}} \sim \chi_4^2$'dir. H_0 hipotezinin doğruluğu altında beklenen frekanslar: $fb_{11} = \frac{880\times 74}{270} = 22.926$, $fb_{12} = \frac{80\times 128}{270} = 37.926$, $fb_{13} = \frac{80\times 68}{270} = 20.148$, $fb_{21} = \frac{91\times 74}{270} = 24.941$, $fb_{22} = \frac{91\times 128}{270} = 43.141$ $fb_{23} = \frac{91\times 68}{270} = 22.918$, $fb_{31} = \frac{99\times 74}{270} = 27.133$, $fb_{32} = \frac{99\times 128}{270} = 46.933$ ve $fb_{33} = \frac{99\times 68}{270} = 24.934$ bulunur. Böylece test istatistiğinin değeri ise $\chi_H^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(fg_{ij}-fb_{ij})^2}{fb_{ij}} \cong 12.223$ bulunur.

c) Karar: : $\alpha = 0.05$ önem seviyesi ve 4 s.d. olmak üzere $\chi_T^2 = \chi_{4;\,0.95}^2 = 9.488$ bulunur (ki-kare dağılım tablosundan). Böylece $\chi_H^2 > \chi_T^2$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: % 95 güvenilirlikle ekip grupları ile hata türü arasında bir ilişki olduğu söylenebilir.

XII.4.5 Ki-Kare Testine Dayalı İlişki Katsayıları

Gerek (2×2) ve gerekse de $(\mathrm{i} \times \mathrm{j})$ veri düzeninde olsun H_0 hipotezinin ret edilmesi sonucunda ilişkinin derecesi belinmek istenebilir. Her iki değişkeni de sınıflama düzeyinde olan böylesi değişkenler için değişkenler arası ilişkiyi veren ölçüm katsayılarının çoğu χ_H^2 değerine bağlı olarak tanımlanır. Bu tür ilişki ölçülerinden bazıları: Cramer'in nusü (υ), Birliktelik katsayısı (ϕ), Uyumluluk katsayısı (c) ve değişkenler arası korelasyon (r). İle verilir.

XII.4.5.1 Cramer'in nu (υ) ilişki katsayısı

$\upsilon = \sqrt{\frac{\chi_H^2}{n\times enk(I-1,J-1)}}$, $0 \leq \upsilon \leq 1$ eşitliği ile verilir. $\upsilon \to 0$ => zayıf ilişki, $\upsilon \to 1$ => kuvvetli ilişki, $\upsilon = 0$ => ilişki yok ve $\upsilon = 1$ => Tam ilişki anlamındadır.

XII.4.5.2 Birliktelik katsayısı (ϕ)

$\phi = \sqrt{\frac{\chi_H^2}{n}}$, $0 \leq \phi \leq 1$ eşitliği ile verilir. $\phi \to 0$ => iki değişkene ilişkin özelliklerin birimlerde birlikte bulunması zayıflar, $\phi \to 1$ => iki değişkene ilişkin özelliklerin birimlerde birlikte bulunması kuvvetlenir,$\phi = 0$ => iki değişkene ilişkin özelliklerin birimlerde birlikte bulunması mümkün değil ve $\phi = 1$ => iki değişkene ilişkin özelliklerin birimlerde birlikte bulunması kesindir.

XII.4.5.3 Uyumluluk katsayısı (c)

$c = \sqrt{\frac{\chi_H^2}{n+\chi_H^2}}$, $0 \leq c \leq 1$ eşitliği ile verilir. $c \to 0$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma uygunluğu zayıflar, $c \to 1$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma uygunluğu kuvvetlenir, $c = 0$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma uygunluğu mümkün değil ve $c = 1$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma uygunluğu tamdır.

XII.4.5.4 Değişkenler arası korelasyon (r)

$I = J$ durumu söz konusu olduğunda iki nitel değişken arasındaki korelasyonu verir. $r = \sqrt{\frac{\chi_H^2}{n\times(I-1)}}$, $0 \leq r \leq 1$ eşitliği ile verilir. $r \to 0$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma ilişkisi zayıflar, $r \to 1$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma ilişkisi kuvvetlenir, $r = 0$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma ilişkisi yok ve $c = 1$ => Birimlerin iki değişkene ilişkin özelliklere sahip olma ilişkisi tamdır.

♦ **Örnek:** Cinsiyet-renk körlüğü ve ekip grubu-hata türü arsanıdaki problemlerde ilişki tespit edilmişti. Bu ilişki miktarlarını belirleyiniz.

♦ **Çözüm:** Hemen yukarıda tanımlanan ilişki katsayıları yardımı ile sözü edilen iki problem için ilişki miktarları takipteki tabloda sergilenmiştir.

Problem	n	χ_H^2	υ	ϕ	c	r
Cinsiyet-Renk körlülüğü	400	4.706	%10.85	%10.85	%10.78	%10.85
Ekip-Hata türü	270	12.223	%15.05	%21.28	%20.81	%15.05

XII.4.6 İkiden Fazla Grubun Farklılık Kontrolünde Ki-Kare Testi

Nitel değişkenli durumlarda ikiden fazla grup sözkonusu olduğu zaman bu gruplar arasındaki ilişkinin (farkın) önemliliği ki-kare testi ile incelenebilir. Bu durumda test işlem adımları şu şekilde ifade edilir.

a) Hipotezler: H_0: " Gruplar arasında fark yoktur" ve araştırma hipotezi ise H_A: "Gruplar arasında fark vardır" şeklinde kurulur.

b) Test istatistiği: Test işlemi ise $(\text{İ} \times J)$ düzenindeki veriler için izlenen tets istatistiği ile aynıdır. Şöyle ki $\chi_H^2 = \sum_{i=1}^{\text{İ}} \sum_{j=1}^{J} \frac{(fg_{ij}-fb_{ij})^2}{fb_{ij}} \sim \chi_r^2$ olup $r = (\text{İ} - 1) \times (J - 1)$ s.d.'ne sahiptir. Burda fb_{ij}beklenen frekas olup $fb_{ij} = \frac{T_{i.}\times T_{.j}}{T_{..}}$, $i = 1, 2, \cdots, \text{İ}$; $j = 1, 2, \cdots, J$ eşitliği ile hesaplanır.

c) Karar: α önem seviyesi ve r s.d. olmak üzere $\chi_T^2 = \chi_{r;\,\alpha}^2$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. Eğer $\chi_H^2 \leq \chi_T^2 => H_0$ hipotezi kabul ve eğer $\chi_H^2 > \chi_T^2 => H_0$ hipotezi ret edilir.

d) Yorum: Karara göre yorum yapılır.

H_0 hipotezinin kabul edilmesi durumunda grupların farsız olduğu söylenir. H_0 hipotezinin ret edilmesi durumunda gruplardan en az birini farklı olduğu; ancak hangi grup yada grupların farklı olduğu bilinemez. Farklı olan grup ya da grupları belirlemek için χ_H^2'ı oluşturan toplam terimine bakılır. Öncelikle χ_H^2'ye en fazla katkıyı yapan grup belirlenerek bu grup işlemden çıkarılır. Çünkü bu aşamada H_0 hipotezinin ret edilmesine sebep olan grup öncelikle bu gruptur. Bu işleme geriye kalan gruplar üzerinde benzer biçimde devam edilir. Tekrarda kullanılan hipotezler ise: H_0: " Geriye kalan gruplar arasındaki fark önemsizdir" ve araştırma hipotezi ise H_A: "Geriye kalan gruplar arasındaki fark önemlidir" şeklinde kurulur.

Bu işleme H_0 hipotezi kabul edilinceye kadar devam edilir. Eğer herhangi bir aşamada kabul edilmiş ise kabul edilen aşamada geriye kalan gruplar arasındaki farkın önemsiz olduğu ve işlemden çıkarılan grup ya da grupların arasındaki farklılık olduğu söylenir. Eğer herhangi bir aşamada H_0 hipotezi kabul edilmemişse tüm grupların birbirinden farklı olduğu söylenir.

Örnek: Bir şehirde deprem sigortasına karşı dört ayrı bölgenin duyarlılıkları karşılaştırılmak istenmektedir. Bu amaçla yapılan bir çalışnada aşağıdaki tablodaki ilk dört sütun ve ilk beş satır değerleri elde edilmiştir. $\alpha = 0.05$ önem seviyesinde bölgelerin duyarlılık oranlarının farklı olduğu söylenebilir mi?

Bölgler	Deprem Siortası (Gözlenen)			Deprem Siortası (Beklenen)		χ_H^2'a Katkısı
	Var	Yok	Toplam	Var	Yok	
A	117	404	521	169,49	351,51	24,090
B	222	334	556	180,87	375,13	13,862
C	133	204	337	109,63	227,37	7,385
D	109	263	372	121,01	250,99	1,768
TOPLAM	581	1205	1786	581	1205	47,105

Birinci İşlem:

a) Hipotezler: H_0: " Bölgeler arasında fark yoktur" ve araştırma hipotezi ise H_A: "Bölgeler arasında fark vardır" şeklinde kurulur.

b) Test istatistiği: $\chi_H^2 \cong 47.105$ bulunur. Bu sonuç talonun son saütun ve son satırındaki hüçrede verilmiştir. Diğer hesaplamalar ise beşinci ve altıncı sütunlarda verilmiştir.

c) Karar**:** $\alpha = 0.05$ önem seviyesi ve $r = (4-1) \times (2-1) = 3$ s.d. olmak üzere $\chi_T^2 = \chi_{3;\,0.05}^2 = 7.815$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. $\chi_H^2 > \chi_T^2$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: %95 güvenilirlikle en az bir bölge diğerlerinden farklıdır. Bu durmda ilk farklı olan bölge χ_H^2'a Katkısı en yüksek olan bölge A bölgesidir. Bu bölge işlenden çıkarılarak işlem yeniden tekrarlandığında takipte verilen tablodaki sonuçlar elde edilir.

Bölgler	Deprem Siortası (Gözlenen)			Deprem Siortası (Beklenen)		χ^2_H'a Katkısı
	Var	Yok	Toplam	Var	Yok	
B	222	334	556	203,94	352,06	2,526
C	133	204	337	123,61	213,39	1,126
D	109	263	372	136,45	235,55	8,720
TOPLAM	464	801	1265	464	801	12,372

İkinci İşlem:

a) Hipotezler: H_0: " Geriye kalan bölgeler arasında fark yoktur" ve araştırma hipotezi ise H_A: "Geriye kalan bölgeler arasında fark vardır" şeklinde kurulur.

b) Test istatistiği: $\chi^2_H \cong 12.372$ bulunur. Bu sonuç talonun son saütun ve son satırındaki hüçrede verilmiştir. Diğer hesaplamalar ise beşinci ve altıncı sütunlarda verilmiştir.

c) Karar**:** $\alpha = 0.05$ önem seviyesi ve $r = (3-1) \times (2-1) = 2$ s.d. olmak üzere $\chi^2_T = \chi^2_{3;\,0.05} = 5.991$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. $\chi^2_H > \chi^2_T$ olduğundan H_0 hipotezi ret edilir.

d) Yorum: %95 güvenilirlikle en az bir bölge diğerlerinden farklıdır. Bu durmda farklı olan bölge χ^2_H'a Katkısı en yüksek olan bölge D bölgesidir. Bu bölge işlenden çıkarılarak işlem yeniden tekrarlandığında takipte verilen tablodaki sonuçlar elde edilir.

Bölgler	Deprem Siortası (Gözlenen)			Deprem Siortası (Beklenen)		χ^2_H'a Katkısı
	Var	Yok	Toplam	Var	Yok	
B	222	334	556	221,03	334,97	0,007
C	133	204	337	133,97	203,03	0,012
TOPLAM	355	538	893	355	538	0,019

Üçüncü İşlem:

a) Hipotezler: H_0: " Geriye kalan bölgeler arasında fark yoktur" ve araştırma hipotezi ise H_A: "Geriye kalan bölgeler arasında fark vardır" şeklinde kurulur.

b) Test istatistiği: $\chi^2_H \cong 0.019$ bulunur. Bu sonuç talonun son saütun ve son satırındaki hüçrede verilmiştir. Diğer hesaplamalar ise beşinci ve altıncı sütunlarda verilmiştir.

c) Karar**:** $\alpha = 0.05$ önem seviyesi ve $r = (3-1) \times (2-1) = 2$ s.d. olmak üzere $\chi^2_T = \chi^2_{3;\,0.05} = 3.841$ tablo değeri (ki-kare dağılım tablosundan) belirlenir. $\chi^2_H < \chi^2_T$ olduğundan H_0 hipotezi kabul edilir ve işleme son verilir.

d) Yorum: %95 güvenilirlikle son iki bölgenin oranları arasında fark yoktur.Genel Yorum: Bu analiz sonucunda üç farklı grup oluşmuştur. Birinci grupta A bölgesi, ikinci grupta D bölgesi ve üçüncü grupta ise B ve C bölgeleri bulunmaktadır. Deprem sigortası yaptıranları oranı A bölesinde % 22.46 ile en düşük, D bölgesinde % 29.30 ila ikinci en düşük, % 39.47 C ve % 39.92 oranı ile B bölgesi en yüksek değere sahiptir.