



Kİ KARE ANALİZİ

Doç. Dr. Mehmet AKSARAYLI
www.mehmetaksarayli

Ki-Kare Analizleri

OLAY 1: Genelde bir istatistik sınıfında, öğrencilerin %60'ının devamlı, %30'unun bazen, %10'unun ise çok az derse geldikleri düşünülmektedir. Sizin sınıfınızdan 30 kişilik bir örnek alınmış ve %65'inin devamlı, %20'sinin bazen, ve %15'inin çok az derse geldiği bulunmuştur. Bu sınıfın devam durumunun genel devam durumuna uyduğu söylenebilir mi?

OLAY 2: Aynı öğretim üyesinin istatistik dersi verdiği 3 farklı sınıfta devam durumlarının aynı olup olmadığı araştırılmaktadır. Her üç sınıftan alınan ayrı örneklerin sonucu aşağıdaki tabloda görüldüğü gibidir.

	devamlı	bazen	çok az
sınıf1	15	8	3
sınıf2	14	6	4
sınıf3	6	7	7

OLAY 3: Bu fakülte'deki devam durumu ile geçme notu arasında bir ilişki olup olmadığı araştırılmak istenmektedir. Fakülte öğrencilerinden alınan 100 kişilik bir örneğe devam durumları ile geçme notları sorulmuş ve aşağıdaki tablo elde edilmiştir:

	<50	50-<70	70-<90	>=90
devamlı	15	20	10	1
bazen	12	19	5	3
çok az	2	5	5	3

OLAY 1: Genelde bir istatistik sınıfında, öğrencilerin %60'ının devamlı, %30'unun bazen, %10'unun ise çok az derse geldikleri düşünülmektedir. Sizin sınıfınızdan 30 kişilik bir örnek alınmış ve %65'inin devamlı, %20'sinin bazen, ve %15'inin çok az derse geldiği bulunmuştur. Bu sınıfın devam durumunun genel devam durumuna uyduğu söylenebilir mi?

OLAY 2: Aynı öğretim üyesinin istatistik dersi verdiği 3 farklı sınıfta devam durumlarının aynı olup olmadığı araştırılmaktadır. Her üç sınıftan alınan ayrı örneklerin sonucu aşağıdaki tabloda görüldüğü gibidir.

	devamlı	bazen	çok az
sınıf1	15	8	3
sınıf2	14	6	4
sınıf3	6	7	7

OLAY 3: Bu fakültedeki devam durumu ile geçme notu arasında bir ilişki olup olmadığı araştırılmak istenmektedir. Fakülte öğrencilerinden alınan 100 kişilik bir örneğe devam durumları ile geçme notları sorulmuş ve aşağıdaki tablo elde edilmiştir:

	<50	50-<70	70-<90	>=90
devamlı	15	20	10	1
bazen	12	19	5	3
çok az	2	5	5	3

Ki-Kare Testleri

İyi Uyum Testi

Varsayımları:

1. Çokdeğerli kategorik veri
2. Tek bir örnek, birden çok ilgilenilen oran
3. Örnek oranları hakkında varsayılan oranların test edilmesi

Homojenlik Testi

Varsayımları:

1. Tek bir kategorik değişken
2. Her gruptan alınan ayrı birer örnek
3. Varsayılan oran yok
4. Bir grup için oranın diğer grup veya gruplar için de aynı olup olmadığının test edilmesi

Bağımsızlık Testi

Varsayımları:

1. İki kategorik değişken
2. Sadece tek örnek, oranlarla ilgili varsayım yok
3. Bir kategorik değişkenin diğeriyle ilişkili olup olmadığının test edilmesi

Kİ-KARE TESTLERİ

A) Kİ-KARE DAĞILIMI VE ÖZELLİKLERİ

Örnekleme yoluyla elde edilen rakamların, anakütle rakamlarına uygun olup olmadığı; bir başka ifadeyle gözlenen değerlerin teorik (beklenen) değerlere uygunluk gösterip göstermediği ki-kare testi ile tespit edilir.

Standart normal değişken Z_i değerlerinin;

$$Z_i^2 = \left(\frac{X_i - \mu}{\sigma} \right)^2$$

şeklinde karesi alındığında, Z_i^2 değerlerinin dağılımı **ki-kare dağılımına** dönüşür.

5

X tesadüfi değişken dağılımından bir değer seçilip standart hale dönüştürülür ve karesi alınırsa X değişkeninin dağılımı ki-kare dağılımına dönüşür. Normal dağılım gösteren bir X tesadüfi değişken dağılımından tesadüfi ve birbirinden bağımsız olarak iki değer seçilsin. Seçilen değerleri;

$$Z_1 = \frac{X_1 - \mu}{\sigma} \quad Z_2 = \frac{X_2 - \mu}{\sigma}$$

şeklinde standart hale dönüştürülsün. Bu Z değerlerinin kareleri alınıp toplanırsa elde edilen $Z_1^2 + Z_2^2$ değişkeni ki-kare dağılımı gösterir. N adet örnek için bu prosedür takip edildiğinde;

$$Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_n^2$$

değişkeni de ki-kare dağılımı gösterir.

6

Bir ki-kare dağılımı diğer ki-kare dağılımlarından serbestlik derecelerine göre ayrılır. Kareleri alınıp toplandığında , ki-kare dağılımı gösteren bağımsız standart normal değer sayısına **serbestlik derecesi** denir. Bir ki-kare dağılımının **ortalaması**, dağılımın serbestlik derecesine ve **varyansı**, serbestlik derecesinin iki katına eşittir. Mesela, serbestlik derecesi 10 olan bir ki-kare dağılımının ortalaması 10 ve varyansı 20'dir.

Ki-kare değişkeni, χ^2 sembolü ile gösterilir. Bir dağılımı diğerinden ayırmak için , bu sembole serbestlik derecesini gösteren bir indis eklenebilir. Böylece 1, 2 ve n serbestlik dereceleriyle ki-kare dağılımı gösteren değişkenler, χ^2_1 , χ^2_2 ve χ^2_n şeklinde gösterilebilir.

Ki-kare dağılımı sağa çarpıktır ve normalden daha diktir. N büyüdükçe merkezi limit teoreminin bir sonucu olarak diklik ve asimetri azalır ve dağılım normale yaklaşır. Ki-kare değerleri 0 ile $+\infty$ arasında değişir.

Ki-Kare Bağımsızlık Testi

Varsayımları:

1. İki kategorik değişken
2. Sadece tek örnek, oranlarla ilgili varsayım yok
3. Bir kategorik değişkenin diğeriyle İLİŞKİLİ olup olmadığının test edilmesi

HANGİ OLAY BU TEST İLE İLGİLİDİR?

Ki-Kare Bağımsızlık Testi

Sık sık karşılaşılan araştırma sorularından biri de iki değişkenin birbiri ile ilişkili olup olmadığıdır. Mesela, bir sosyolog okul vasıtasıyla kazanılan eğitim seviyesinin gelirle ilişkili olup olmadığını araştırmak isteyebilir. Bir okulda çalışan beslenme uzmanı öğrencilerin beslenme uzmanı öğrencilerin beslenme seviyelerinin akademik performanslarıyla ilişkili olup olmadığını bilmek isteyebilir.

İki değişken arasında ilişki yoksa, birisinin dağılımı hiçbir şekilde diğerinin dağılımına bağımlı değilse bu iki değişken birbirinden bağımsızdır denir. İki değişken arasında ilişki yoksa, belirli bir değişkenin değerini bilmek, diğer değişkenin değerini tespite imkan sağlamaz.

Varsayımları:

1. İki kategorik değişken
2. Sadece tek örnek, oranlarla ilgili varsayım yok
3. Bir kategorik değişkenin diğeriyle İLİŞKİLİ olup olmadığının test edilmesi

HANGİ OLAY BU TEST İLE İLGİLİDİR?

9

Bu fakülte'deki devam durumu ile geçme notu arasında bir ilişki olup olmadığı araştırılmak istenmektedir. Fakülte öğrencilerinden alınan 100 kişilik bir örneğe devam durumları ile geçme notları sorulmuş ve aşağıdaki tablo elde edilmiştir:

	<50	50-<70	70-<90	>=90
devamli	15	20	10	1
bazen	12	19	5	3
çok az	2	5	5	3

H_0 : Devam yapısı, geçme notu ile **ilişkili değildir**.

H_1 : Devam yapısı, geçme notu ile **ilişkilidir**.

Bu hipotez testini uygulamak için yine gözlenen ve beklenen arasındaki farkların büyüklüğü ile ilgili Ki-Kare test istatistiği kullanılır.

Bu test istatistiği, $(k-1)(r-1)$ serbestlik dereceli tablo değeriyle karşılaştırılır.

$$\chi^2_{\text{tablo}} = \chi^2_{(k-1)(r-1), \alpha}$$

Burada k: sütun sayısı, r:satır sayısı

10

H_0 : Devam yapısı, geçme notu ile ilişkili değildir.

H_1 : Devam yapısı, geçme notu ile ilişkilidir.

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Yine veriler bir tabloda yer aldığı için test istatistiğinin hesaplamaları ayrı tablolarla yapılır.

Gözlenen	<50	50-<70	70-<90	>=90	satır toplamı
devamli	15	20	10	1	46
bazen	12	19	5	3	39
çok az	2	5	5	3	15
sütun toplamı	29	44	20	7	100

beklenen = $\frac{\text{satır toplamı} \times \text{sütun toplamı}}{\text{genel toplam}}$

Beklenen	<50	50-<70	70-<90	>=90
devamli	13.3	20.2	9.2	3.2
bazen	11.3	17.2	7.8	2.7
çok az	4.4	6.6	3.0	1.1

11

Gözlenen	<50	50-<70	70-<90	>=90	satır toplamı
devamli	15	20	10	1	46
bazen	12	19	5	3	39
çok az	2	5	5	3	15
sütun toplamı	29	44	20	7	100

Beklenen	<50	50-<70	70-<90	>=90
devamli	13.3	20.2	9.2	3.2
bazen	11.3	17.2	7.8	2.7
çok az	4.4	6.6	3.0	1.1

O - E	<50	50-<70	70-<90	>=90
devamli	1.7	-0.2	0.8	-2.2
bazen	0.7	1.8	-2.8	0.3
çok az	-2.4	-1.6	2.0	2.0

(O - E) ² /E	<50	50-<70	70-<90	>=90	
devamli	0.21	0.003	0.07	1.53	
bazen	0.04	0.20	1.01	0.03	
çok az	1.27	0.39	1.33	3.62	
					9.69

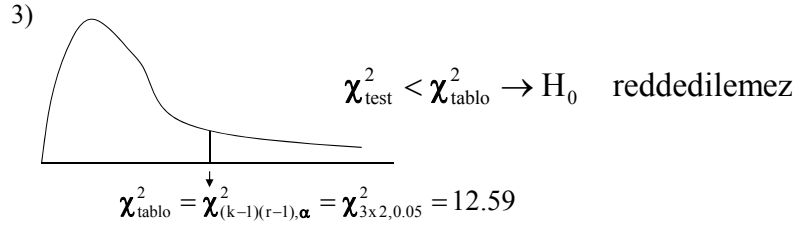
χ^2_{test}

12

1) H_0 : Devam yapısı, geçme notu ile ilişkili değildir.

H_1 : Devam yapısı, geçme notu ile ilişkilidir.

2)
$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i} = 9.69$$

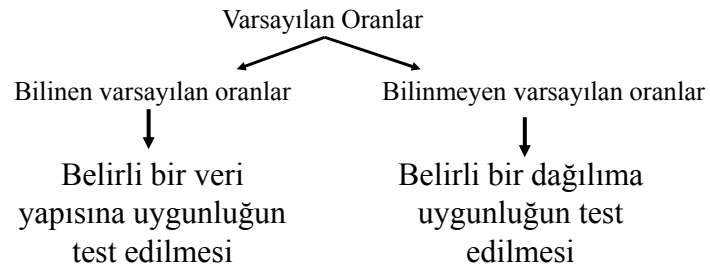


4) Yorum: %5 hata olasılığı ile, devam yapı ile geçme notu değişkenlerinin ilişkili olduğuna dair yeterli kanıt bulunamamıştır.

Ki-Kare İyi Uyum Testi

Varsayımları:

1. Çok değerli kategorik veri (2'den fazla kategorisi olan kategorik değişken)
2. Tek bir örnek, birden çok ilgilenilen oran
3. Örnek oranları hakkında varsayılan oranların test edilmesi



HANGİ OLAY BU TEST İLE İLGİLİDİR?

Genelde bir istatistik sınıfında, öğrencilerin %60'ının devamlı, %30'unun bazen, %10'unun ise çok az derse geldikleri düşünülmektedir. Sizin sınıfınızdan 100 kişilik bir örnek alınmış ve %65'inin devamlı, %20'sinin bazen, ve %15'inin çok az derse geldiği bulunmuştur. Bu sınıfın devam durumunun genel devam durumuna uyduğu söylenebilir mi?

	varsayılan oranlar	gözlenen oranlar
devamlı	0.60	0.65
bazen	0.30	0.20
çok az	0.10	0.15

$$H_0: \Pi_{\text{devamlı}}=0.60, \Pi_{\text{bazen}}=0.30, \Pi_{\text{çok az}}=0.10$$

H_1 : En az bir oran varsayılandan farklıdır.

15

$$H_0: \Pi_{\text{devamlı}}=0.60, \Pi_{\text{bazen}}=0.30, \Pi_{\text{çok az}}=0.10$$

H_1 : En az bir oran varsayılandan farklıdır.

Bu hipotez testini uygulamak için gözlenen ve beklenen arasındaki farkların büyüklüğüne bakmak gerekir. Bu farkların mutlak değeri ne kadar büyükse, sıfır hipotezi hakkında o kadar kuşkuya düşeriz. Sıfır hipotezi doğruyken ve örnek orta büyüklükteyken (beklenen değerlerin (E_i) herbiri en az 5 ise) , bu hipotez testi için aşağıdaki Ki-Kare test istatistiği kullanılır:

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Gözlenen değer
Beklenen değer

Bu test istatistiği, k-1 serbestlik dereceli tablo değeriyle karşılaştırılır.

Burada k: kategori sayısıdır.

$$\chi^2_{\text{tablo}} = \chi^2_{k-1, \alpha}$$

16

$H_0: \Pi_{\text{devamlı}}=0.60, \Pi_{\text{bazen}}=0.30, \Pi_{\text{çok az}}=0.10$

H_1 : En az bir oran varsayılandan farklıdır.

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Beklenen değer = örnek hacmi x beklenen olasılık $\rightarrow E_i = n \times p_i$

Gözlenen değer = örnek hacmi x gözlenen olasılık $\rightarrow O_i = n \times r_i$

	varsayılan oranlar	gözlenen oranlar	beklenen değerler (E _i)	gözlenen değerler (O _i)	O _i - E _i	(O _i - E _i) ²	((O _i - E _i) ²)/E _i
devamlı	0.60	0.65	60	65	5	25	0.42
bazen	0.30	0.20	30	20	-10	100	3.33
çok az	0.10	0.15	10	15	5	25	2.50
Ki-kare test =							6.25

17

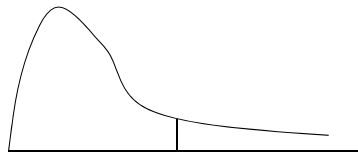
1) $H_0: \Pi_{\text{devamlı}}=0.60, \Pi_{\text{bazen}}=0.30, \Pi_{\text{çok az}}=0.10$

H_1 : En az bir oran varsayılandan farklıdır.

2)

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i} = 6.25$$

3)



$$\chi^2_{\text{test}} > \chi^2_{\text{tablo}} \rightarrow H_0 \text{ ret!}$$

$$\chi^2_{\text{tablo}} = \chi^2_{k-1, \alpha} = \chi^2_{2, 0.05} = 5.99$$

4) Yorum: %5 hata olasılığı ile, en az bir oran diğerlerinden farklı olduğu için, bu sınıfın devam durumu, genel devam durumu yapısına uymamaktadır.

18

Ki-Kare Homojenlik Testi

Homojenlik testinde her bir anakütleden belirli hacimlerde örnekler çekilir ve bu örneklerin belirli bir karakteristiği taşımaları bakımından benzer değişim gösterip göstermediklerine karar verilir.

Varsayımları:

1. Tek bir kategorik değişken
2. Her gruptan alınan ayrı birer örnek
3. Varsayılan oran yok
4. Bir grup için oranın diğer grup veya gruplar için de AYNI (HOMOJEN) olup olmadığının test edilmesi

HANGİ OLAY BU TEST İLE İLGİLİDİR?

Aynı öğretim üyesinin istatistik dersi verdiği 3 farklı sınıfta devam durumlarının aynı olup olmadığı araştırılmaktadır. Her üç sınıftan alınan ayrı örneklerin sonucu aşağıdaki tabloda görüldüğü gibidir.

	devamli	bazen	çok az
sınıf1	15	8	3
sınıf2	14	6	4
sınıf3	6	7	7

H_0 : sınıf1'in devam yapısı = sınıf2'nin devam yapısı = sınıf3'ün devam yapısı

H_1 : En az bir sınıfın devam yapısı diğerlerinden farklıdır.

Bu hipotez testini uygulamak için yine gözlenen ve beklenen arasındaki farkların büyüklüğü ile ilgili Ki-Kare test istatistiği kullanılır.

Bu test istatistiği, $(k-1)(r-1)$ serbestlik dereceli tablo değeriyle karşılaştırılır.

$$\chi^2_{\text{tablo}} = \chi^2_{(k-1)(r-1), \alpha}$$

Burada k: kategori sayısı (sütun sayısı),
r: grup sayısı (satır sayısı)

H₀: sınıf1'in devam yapısı = sınıf2'nin devam yapısı = sınıf3'ün devam yapısı

H₁: En az bir sınıfın devam yapısı diğerlerinden farklıdır.

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Fakat bu sefer veriler bir tabloda yer aldığı için test istatistiğinin hesaplamaları ayrı tablolarla yapılır.

Gözlenen	devamli	bazen	çok az	satır toplamı
sınıf1	15	8	3	26
sınıf2	14	6	4	24
sınıf3	6	7	7	20
sütun toplamı	35	21	14	70

beklenen = $\frac{\text{satır toplamı} \times \text{sütun toplamı}}{\text{genel toplam}}$

Beklenen	devamli	bazen	çok az
sınıf1	13	7.8	5.2
sınıf2	12	7.2	4.8
sınıf3	10	6	4

21

Gözlenen	devamli	bazen	çok az	satır toplamı
sınıf1	15	8	3	26
sınıf2	14	6	4	24
sınıf3	6	7	7	20
sütun toplamı	35	21	14	70

Beklenen	devamli	bazen	çok az
sınıf1	13	7.8	5.2
sınıf2	12	7.2	4.8
sınıf3	10	6	4

$$15 - 13 = 2$$

O - E	devamli	bazen	çok az
sınıf1	2	0.2	-2.2
sınıf2	2	-1.2	-0.8
sınıf3	-4	1	3

(O - E) ² /E	devamli	bazen	çok az
sınıf1	0.31	0.01	0.93
sınıf2	0.33	0.20	0.13
sınıf3	1.60	0.17	2.25
			5.93

χ^2_{test}

22

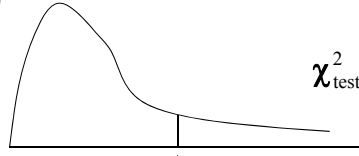
1) H_0 : sınıf1'in devam yapısı = sınıf2'nin devam yapısı = sınıf3'ün devam yapısı

H_1 : En az bir sınıfın devam yapısı diğerlerinden farklıdır.

2)

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i} = 5.93$$

3)



$$\chi^2_{\text{test}} < \chi^2_{\text{tablo}} \rightarrow H_0 \text{ reddedilemez}$$

$$\chi^2_{\text{tablo}} = \chi^2_{(k-1)(r-1), \alpha} = \chi^2_{2 \times 2, 0.05} = 9.49$$

4) Yorum: %5 hata olasılığı ile, sınıfların devam yapısının birbirinden farklı olduğu söylenemez.