

KATEGORİSEL VERİ ANALİZİ

(χ^2 testi)

Prof. Dr. Fazıl GÖKGÖZ

1

Giriş

- Değişkenleri nitel ve nicel değişkenler olarak iki kısımda inceleyebiliriz.
- Bu konuya kadar nicel değişkenler için hesaplamalar ve testler yapıldı.
- Ancak, günlük hayatta çoğunlukla nitel veriler üzerinden hareketle araştırmalar yapılabilmekte ve bunların sonuçları da sayısal olarak sunulabilmektedir.

Prof. Dr. Fazıl GÖKGÖZ

2

Örnek

- Çalışanların eğitim seviyesi ile meslekleri arasındaki bağıntıyı incelemek istersek bunları sayısal olarak ifade edemeyiz.
- Eğitim seviyesini, ilk öğretim, orta öğretim, lisans ve lisansüstü şeklinde, meslekleri ise isimleri ile ifade etmek mümkündür.
- Bu bağlamda, araştırma için veriyi χ^2 testine göre değerlendirmek uygun bir yöntemdir.

χ^2 TESTİ

χ^2 (*ki-kare*) testi, sayısal olmayan değişkenler arasındaki ilişkinin var olmadığı başlangıç hipotezi (H_0) kurularak daha sonrasında bu hipotezin kabul/red durumunu belirlemek için yapılan istatistiksel test olarak bilinir.

χ^2 BAĞIMSIZLIK TESTİ

- Burada elimizde iki nitel değişken vardır. Bu değişkenler içinse farklı özellikler (*sınıflar*) söz konusudur.
- χ^2 testi ile bu değişkenler arasında ilişki olup olmadığı kontenjans tabloları yardımı ile bulunur.

Prof. Dr. Fazıl GÖKGÖZ

5

Kontenjans Tablosu

- Yatay (satır) ve dikey (sütun) eksenlerde değişkenlere ait gözlenen frekans değerleri yazılarak elde edilen bir tablodur. Genel olarak kontenjans tablosu şu şekildedir:

1. Değişkenin Şıkları	2. Değişkenin Şıkları						Toplam	
	1	2	3	j		c
1	n_{11}	n_{12}	n_{13}	n_{1j}	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	n_{2j}	n_{2c}	$n_{2.}$
3	n_{31}	n_{32}	n_{33}	n_{3j}	n_{3c}	$n_{3.}$
.....	-	-	-	-	-	-	-	-
.....	-	-	-	-	-	-	-	-
i	n_{i1}	n_{i2}	n_{i3}	n_{ij}	n_{ic}	$n_{i.}$
.....	-	-	-	-	-	-	-	-
.....	-	-	-	-	-	-	-	-
r	n_{r1}	n_{r2}	n_{r3}	n_{rj}	n_{rc}	$n_{r.}$
Toplam	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.j}$	$n_{.c}$	$n_{..} = n$

Prof. Dr. Fazıl GÖKGÖZ

6

.....

- Bu tabloda satır ve sütunun kesiştiği hücreye ilgili veri olarak gözlenen frekans değeri yazılır.
- Görüldüğü gibi matris formunda bir tablo oluşmaktadır.
- Burada örneğin n_{34} değeri birinci değişkenin 3. özelliği ile ikinci değişkenin 4. özelliğinin ikisine birden sahip olan frekans değeridir.

Prof. Dr. Fazıl GÖKGÖZ

7

χ^2 Testinin Aşamaları

- Eğer problem içerisinde hazır olarak verilmediyse frekans değerlerine göre kontenjans tablosu oluşturulur.
- Probleme göre sıfır hipotezi (H_0) ve alternatif hipotez (H_A) kurulur.
- Çözüm için gerekli olan anlamlılık düzeyi (α) belirlenir.
- r =tablodaki satır sayısı ve c =tablodaki sütun sayısı olmak üzere $df=(r-1)(c-1)$ serbestlik derecesi belirlenir.
- Tablodan α anlamlılık düzeyi ve df serbestlik derecesine ait χ^2 değeri (kritik değer) bulunur.

Prof. Dr. Fazıl GÖKGÖZ

8

.....

- B_{ij} beklenen frekansı her bir hücre için ayrı ayrı bulunur. ($i=1,2,\dots,r$)($j=1,2,\dots,c$)
 $B_{ij}=(i. \text{ Satır değerleri toplamı}) \times (j. \text{ Sütun değerleri toplamı})/\text{Genel Toplam}$
- **B_{ij} beklenen değerler** bulunduktan sonra ilk tablodaki **G_{ij} gözlenen değerler** de kullanılarak test istatistiği hesaplanır:

$$\chi^2 = \sum_{i,j} \frac{(G_{ij} - B_{ij})^2}{B_{ij}}$$

Prof. Dr. Fazıl GÖKGÖZ

9

.....

- En son olarak bulunan test istatistiği kritik değerden (tablo değeri) büyük ise sıfır hipotezi reddedilir, küçük ise kabul edilir.
- Şimdi tüm bu aşamaları uygulayarak örnek üzerinde χ^2 testini örnek üzerinde uygulayalım.

Prof. Dr. Fazıl GÖKGÖZ

10

ÖRNEK

- Televizyon izleyicilerinin öğrenim düzeyleri ve TV programlarından tercih ettikleri türler sorgulanarak, bu iki değişken arasında bir bağıntı bulunup bulunmadığını, başka bir anlatımla, iki değişkenin birbirinden bağımsız olup olmadığını ortaya koymaya çalışalım. Bu amaçla, 200 kişiyi kapsayan bir örneklem üzerinde yapılan gözlem sonuçları aşağıdaki tablo ile verilmiştir. Tercih edilen program türü ile öğrenim düzeyinin ilişkili olup olmadığını $\alpha=0,01$ anlamlılık düzeyi ile test ediniz.

Prof. Dr. Fazıl GÖKGÖZ

11

VERİ

TV Program Türü	Öğrenim Düzeyi			
	İlk	Orta	Yüksek	Toplam
Dizi	50	20	10	80
Eğlence	20	30	10	60
Magazin	20	10	30	60
Toplam	90	60	50	200

Prof. Dr. Fazıl GÖKGÖZ

12

ÇÖZÜM

- Bize verilen tablo 3 satır ve 3 sütunu olan bir kontenjans tablosudur. Buna göre gözlenen değerler örneğin; $G_{31}=20$, $G_{22}=30$, $G_{13}=10$, ...
- Hipotezlerimizi kuralım:
 H_0 = TV izleyicilerinin program tercihi ile öğrenim düzeyleri arasında bir ilişki yoktur, bunlar bağımsız değişkenlerdir.
 H_A =Bu iki değişken birbiri ile ilişkilidir.
- $\alpha=0.01$ anlamlılık düzeyi verilmiştir. Diğer yandan $r=3$ ve $c=3$ için serbestlik derecesi $df=(3-1)(3-1)=4$ olarak bulunur.
- $df=4$ serbestlik derecesi ve $\alpha=0,01$ anlamlılık düzeyine karşı gelen kritik değer tablodan $\chi^2 =13$ olarak bulunur.

Prof. Dr. Fazıl GÖKGÖZ

13

.....

- B_{ij} =beklenen frekansları hesaplayalım:
 $B_{11}=(\text{birinci satır toplamı} \times \text{birinci sütun toplamı})/\text{genel toplam}=(90 \times 80)/200=36$
ve benzer şekilde hesaplanırsa elde edilen değerler;

Program türü	Öğrenim düzeyi						toplam
	ilk		orta		yüksek		
	G	B	G	B	G	B	
Dizi	50	36	20	24	10	20	80
Eğlence	20	27	30	18	10	15	60
Magazin	20	27	10	18	30	15	60
toplam	90		60		50		200

Prof. Dr. Fazıl GÖKGÖZ

14

.....

- Bulunan bu verilere göre test istatistiğini hesaplırsak;
$$\chi^2 = (50-36)^2/(36) + (20-24)^2/(24) + (10-20)^2/(20) + (20-27)^2/(27) + (30-18)^2/(18) + (10-15)^2/(15) + (20-27)^2/(27) + (10-18)^2/(18) + (30-15)^2/(15) = 42.93$$

χ^2 test istatistiđi ile hesaplanan 42.93 deđeri kritik deđer olan 13den büyük olduđu için sıfır hipotezi (H_0) reddedilir. Bařka ifadeyle, bu iki deđiřken birbiri ile bađıntılıdır.

Kİ-KARE HOMOJENLİK TESTİ

- Bu testin kullanıldıđı ikinci bir alan ise birden fazla örneklemin aynı popölasyondan gelip gelmediđini belirleme iřlemidir.
- Homojenlik testi de bađımsızlık testi gibi aynı řekilde yapılmaktadır.
- Ancak, burada bađımsız örneklemlerin aynı popölasyondan olup olmadıđı arařtırıldıđı için hipotez testi bu řekilde kurulmalı ve benzer řekilde çözüm ařamaları izlenmelidir.

GÖZLENEN FREKANSLAR (VERİ)

Ürünler	Tüketici Görüşleri			Toplam
1. Ürün	60	30	10	100
2. Ürün	80	50	20	150
Toplam	140	80	30	250

Prof. Dr. Fazıl GÖKGÖZ

17

KONTENJANS KATSAYISI

- Ki-kare bağımsızlık testleri ile iki değişken arasındaki ilişki durumunu araştırıyorduk. Fakat bazı zamanlarda ilişkinin olup olmadığının tespiti yeterli değildir. Bunun yanı sıra aradaki ilişkinin derecesini (gücü) de bilmek gereklidir.
- Kontenjans katsayısı buradaki değişkenler arasındaki ilişkinin düzeyini ölçmeye yarayan bir katsayıdır.
- Kontenjans katsayısı ilişkinin olmadığı durumlarda "0", çok yüksek ilişki olduğunda ise "1"e çok yakın bir değer çıkar (ancak 1 değerini almaz).

Prof. Dr. Fazıl GÖKGÖZ

18

.....

- Test istatistiđi ile bulunan (formulasyondan) deđer χ^2 olmak üzere kontenjans katsayısı řu řekilde hesaplanır: (n=genel toplam veya gözlem sayısı)

$$k = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

- Bađımsızlık testi sırasında çözdüğümüz örnek için n=200 ve $\chi^2 = 42,93$ olduđu için aranan kontenjans katsayısı k=0,42 olarak bulunur. Buna göre program türü ve öğrenim derecesi arasında orta düzeyde bir ilişkinin bulunduđu ifade edilebilir.

Prof. Dr. Fazıl GÖKGÖZ

19

ÖDEV

Bölgesel satış yapan bir üretim işletmesi, 2 yeni ürün geliştirerek piyasaya sürmüştür. Tüketicilerin bu ürünlerle ilgili görüşlerini (beğendikleri, beğenmedikleri ya da ilgisiz kaldıkları) belirlemek amacıyla, birinci ve ikinci ürünlerle ilgili olarak iki rassal örneklem oluşturulmuştur. İlk ürünle ilgili birinci örneklemde 100 tüketiciyle, ikinci ürünle ilgili ikinci örneklemde de 150 tüketiciyle görüşülmüştür. Veri aşağıdaki tabloda belirtilmiştir. Seçilen örneklemelerin, aynı ana kütleyle ait olup olmadığını, %5 anlamlılık düzeyinde test ediniz.

(Tüketici görüşleri; fiyat, kalite, kolay ulaşabilme vb. gibi objektif ölçütlerle ve piyasadaki benzer ürünlerle mukayese sonucu oluşmuştur.)

Prof. Dr. Fazıl GÖKGÖZ

20