

# Bilişimde İstatistiksel Analiz

## Ders 8

Ali Mertcan KOSE Msc.

`amertcankose@ticaret.edu.tr`

İstanbul Ticaret Üniversitesi



İSTANBUL TİCARET  
ÜNİVERSİTESİ

## Ki-Kare Testleri

- 1 Ki-kare testleri veri tipinin nitelik olduğu (kadın-erkek, iyileşti-iyileşmedi, hasta-sağlam, sosyo-ekonomik düzeyi; iyi-orta-kötü,... gibi) verilerde kullanılır.
- 2 Ayrıca sürekli ya da kesikli sayısal veri tipinde olduğu halde sonradan nitelik veri konumuna dönüştürülen veriler arasında fark olup olmadığının incelenmesinde kullanılır.
- 3 Veriler  $2 \times 2$ ,  $2 \times 3$ ,  $3 \times 3$ ,  $3 \times 4$ , ... Boyutlu çapraz tablo şeklinde olmalıdır.

## **2 × 2 ki-kare testleri:**

iki yüzde arasındaki farkın anlamlılık testinin uygulandığı durumlarda istenirse **2 × 2** ki-kare testinden de yararlanılabilir.

**2 × 2** ki-kare testinin avantajı, gruptaki gözlem sayılarının az olduğu durumlar için geliştirilmiş değişik ki-kare testlerinin olmasıdır. Gruplardaki gözlem sayısının az olması durumunda ki-kare testlerinden yararlanmak daha uygundur.

## Ki-kare için Genel Formül:

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

k: Toplam Göz Sayısı

# $\chi^2$ (Ki-Kare) Çözümlemesi

$2 \times 2$  (4 gözlü) ki-kare tablosu

	Sağlıktan yakınma var	Sağlıktan Yakınma Yok	Toplam
Sigara Kullanıyor			
Sigara Kullanmıyor			
Toplam			

# $\chi^2$ (Ki-Kare) Çözümlemesi

$2 \times 3$  (6 gözlü) ki-kare tablosu

Eğitim Düzeyi/Sağlık Bilgisi	İyi	Orta	Kötü	Toplam
Düşük				
Orta				
Toplam				

# $\chi^2$ (Ki-Kare) Çözömlemesi

**2 × 2** ya da 4 gözlü ki-kare düzenleri; her gözdeki gözlem sayısının ya da beklenen frekansların belli bir değerin altında olup olmaması durumuna göre değişik şekillerde ve değişik adlar altında uygulanır.

## ❶ Pearson Ki-kare

Gözlerdeki gözlem sayısının 25'in üzerinde olması durumunda uygulanır.

## ❷ Fisher kesin Ki-kare

herhangi bir gözdeki beklenen frekans değeri 5'in altında ise Fisher'in kesin ki-kare testinden yararlanılır.



# Örnek 1

Üniiversite öğrencilerinin cinsiyete göre şişmanlık oranları verilmiştir.

Cinsiyet/ Şişmanlık Durumu	Şişman	Şişman Değil	Toplam
Erkek	41	72	113
Kız	26	60	86
Toplam	67	132	199

Frekans 41 olan göz i.in beklenen frekans: Toplam 199 öğrenciden 67'si şıman ise 113 erkek öğrenciden kaç ı şişmandır?

# Örnek 1

Cinsiyet/ Şişmanlık Durumu	Şişman	Şişman Değil	Toplam
Erkek	41(38.05)	72(74.95)	113
Kız	26(28.95)	60(57.05)	86
Toplam	67	132	199

## Hipotezler

$H_0$  : Şişmanlık açısından kızlar ve erkekler arasında fark yoktur.

$H_1$  : Şişmanlık açısından kızlar ve erkekler arasında fark vardır.

## Test İstatistiğinin Hesaplanması

Gözlerde 25'in altında değer olmadığı için Pearson-ki-kare testi uygulanabilir.

$$\text{Erkek öğrenciler için } \chi_E^2 = \frac{(41-38.05)^2}{38.05} + \frac{(72-74.95)^2}{74.95} = 0.3448$$

$$\text{Kız öğrenciler için } \chi_K^2 = \frac{(26-28.95)^2}{28.95} + \frac{(60-57.05)^2}{57.05} = 0.4534$$

$$\text{Toplam Ki-kare } \chi_T^2 = 0.3448 + 0.4531 = 0.7979$$

# Örnek 1

$\alpha = 0.05$  anlamlılık düzeyi, tablo istatistiği serbestlik derecesi =  
(satır sayısı-1)  $\times$  (sütun sayısı-1)

$$= (2-1) \times (2-1) = 1$$

$$\chi^2_{tablo(sd=1, \alpha=0.05)} = 3.84$$

# Örnek 1

df	$\alpha$									
	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	-	-	-	-	0.02	2.71	3.84	5.02	6.64	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.12	0.22	0.35	0.58	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.54	20.09	21.96
9	1.74	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.58	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.08	4.66	5.63	6.57	7.79	21.06	23.69	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.27	7.02	8.23	9.39	10.87	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.51	71.42	76.15	79.49
60	35.53	37.49	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.22
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.30
100	67.33	70.07	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17

Figure 1: Ki-kare tablosu.

# Örnek 1

$\chi^2_{hesap} = 0.7979 < \chi^2_{tablo} = 3.841$   $p > 0.05$   $H_0$  kabul. Bu durumda Kız ve erkek öğrencilerin şişman olup olmama açısından aralarında istatistiksel olarak anlamlı bir farklılık yoktur. şişmanlık yüzdeleri: erkek öğrenciler için %36 (41/113), kız öğrenciler için %30.2 (26/86).

# Fisher Kesin Ki- Kare Testi

4 gözlü düzende gözlerden herhangi birisinde beklenen frekans 5'den küçükse ki-kare dağılımı çarpık ve kesikli olur. Bu durumda yukarıda anlatılan 4 gözlü düzende ki-kare testleri yerine Fisher kesin ki-kare testi uygulanır.

Sigara/Sağlıktan Yakınma	var	yok	Toplam
İçen	a	b	A
İçmeyen	c	d	B
Toplam	C	D	n



# Fisher Kesin Ki- Kare Testi

$$P = \sum_i^k \frac{A!B!C!D!}{a!b!c!d!n!}$$

# Fisher Kesin Ki- Kare Testi

- P istatistiği bir olasılık değeridir. İstatistiksel karar için; Eğer hipotez tek yönlü ise hesapla bulunan olasılık değeri saptanan yanılma olasılığından küçükse  $H_0$  hipotezi reddedilir, büyükse kabul edilir.
- Eğer hipotez çift yönlü ise hesapla bulunan olasılık değeri 2 ile çarpılır ve saptanan yanılma olasılığından küçükse  $H_0$  hipotezi reddedilir, büyükse kabul edilir.

# Fisher Kesin Ki- Kare Testi

Diyet türü/Kolesterol	Düşen	Düşmeyen	Toplam
A	8	4	12
B	1	12	13
Toplam	9	16	25

# Fisher Kesin Ki- Kare Testi

Diyet türü/Kolesterol	Düşen	Düşmeyen	Toplam
A	9	3	12
B	0	13	13
Toplam	9	16	25

# Fisher Kesin Ki- Kare Testi

$H_0$  : Kolesterol düşürme bakımından diyetler farksızdır.

$H_1$  : Kolesterol düşürme bakımından diyetler farklıdır.

$$p = \frac{12!13!9!16!}{8!4!1!12!25!} + \frac{12!13!9!16!}{9!3!0!13!25!} = 0.003261$$

$$\text{çift yönlü } p \text{ değeri} = 2 \times 0.003261 = 0.00652$$

$\alpha = 0.05$  alınmıştır. Bu durumda  $p=0.000652 < 0.05$  olduğu için  $H_0$  hipotezi reddedilir. Kolesterol düşürme bakımından diyetler arasında fark vardır ( $p < 0.05$ ).

A diyetinde bireylerin %66.7'sinin (8/12) kolesterolü düşerken, B diyetinde bireylerin %7.7'sinin (1/13) kolesterolü düşmektedir.

# Bağımlı grupta Ki- Kare Testi(McNemar) Testi

Niteliksel bir değişken yönünden, aynı bireylerden iki değişik zaman ya da iki değişik durumda elde edilen iki yüzden arasında fark olup olmadığının araştırılmasında kullanılır.

## ÖRNEK 2

Seminer öncesi bilgi düzeyi/Seminer sonrası bilgi düzeyi	Düşen	Düşmeyen	Toplam
Yeterli	25	15	40
Yetersiz	30	26	56
Toplam	55	41	96

## Hipotez Testi

$H_0$  = Bağımlı iki grup arasında farklılık yoktur.

$H_1$  = Bağımlı iki grup arasında farklılık vardır.

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

$$\chi^2 = \frac{(|(b-c)|-1)^2}{b+c}$$



## ÖRNEK 2

$$\chi^2 = \frac{(15-30)^2}{15+30} = 5$$

$\chi^2_{hesap}=5 > \chi^2_{tablo(sd=1, \alpha=0.05)}= 3.841$  Bu durumda Seminer öncesi ve sonrası bağımlı iki grup arasında istatistiksel olarak anlamlı derecede farklılık vardır.