(1) Kikareye Dayalı Önemlilik Testleri

Gözlenen frekanslar $O_1, O_2, \dots, O_k$
Beklenen frekanslar $e_1, e_2, \dots, e_k$
olmak üzere gözlenen frekanslarla beklenen frekanslar arası uyumu ölçen istatistiğe **Kikare** adı denilir.

$$\boxed{\chi^2 = \frac{(o_1-e_1)^2}{e_1} + \frac{(o_2-e_2)^2}{e_2} + \dots + \frac{(o_k-e_k)^2}{e_k} = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}}$$

Gerekli koşullar

1. Herbir gözlem ya da frekans diğer gözlemlerin tümünden bağımsız olmalıdır.

2. Gerçek ve beklenen frekanslar arasındaki farkın normal dağılıma sahip olması amacıyla örneklem hacmi n yeter derecede büyük olmalıdır. $n \geq 50$

3. Gözelerden ~~[illegible]~~ beklenen değeri 5'in altında olmayacak.

Serbestlik Derecesi

a) $V = k-1$ b) $v = k-1-m$ m örneklem istatistiğinden tahmin edilen kitle parametresi

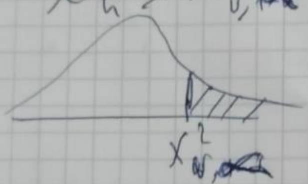
② Düzgün Dağılım ile Uyum

Örnek/ Bir zarın 120 kez atılmasıyla aşağıdaki sonuçlar elde edilmiştir.

Zarın Yüzü	1	2	3	4	5	6
Gözlenen frekans	25	15	12	26	15	27

a) 0,05 b) 0,01 anlam düzeyinde zarın düzgün bir zar olduğunu test ediniz.

$\chi^2_h > \chi^2_{v,\alpha}$ ise H_0 hipotezi reddedilir.



H_0: zar düzgün yüzlüdür.
H_1: zar düzgün yüzlü değildir.

$P(1) = \frac{1}{6} \cdot 120 = 20 \quad P(2) = \frac{1}{6} \cdot 120 = 20 \quad \ldots \quad P(6) = \frac{1}{6} \cdot 120 = 20$

Zarın yüzü	O	e	$(O-e)^2$	$(O-e)^2/e$
1	25	20	25	1,25
2	15	20	25	1,25
3	12	20	64	3,20
4	26	20	36	1,80
5	15	20	25	1,25
6	27	20	49	2,45

$\chi^2_h = 11,20$

$K = 6 - 1 = 5$

a) $\chi^2_{5;0,05} = 11,07$ b) $\chi^2_{5;0,01} = 15,09$

(3) Binom Dağılımı İle Uyum Testi

Örnek: 3 para 240 kez atılıyor. Paraların simetrik olup olmadığını test ediniz. $\alpha = 0,05$

	0 yazı	1 yazı	2 yazı	3 yazı
Gözlenen frekans	24	98	95	23
Beklenen frekans	30	90	90	30

H_0: Paralar simetriktir

H_1: Paralar simetrik değildir.

$s.d = k-1 = 4-1 = 3$

$P(x) = \binom{n}{x} p^x q^{n-x}$

$P(0) = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \frac{1}{8}$ $\quad n=240 \quad n.p = 240 \cdot \frac{1}{8} = 30$

$P(1) = \binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 = \frac{3}{8}$ $\quad n=240 \quad n.p = 240 \cdot \frac{3}{8} = 90$

$P(2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8}$ $\quad n=240 \quad n.p = 240 \cdot \frac{3}{8} = 90$

$P(3) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8}$ $\quad n=240 \quad n.p = 240 \cdot \frac{1}{8} = 30$

$$\chi^2_h = \sum\left(\frac{(o_i - e_i)^2}{e_i}\right) = \frac{(24-30)^2}{30} + \frac{(98-90)^2}{90} + \frac{(95-90)^2}{90} + \frac{(23-30)^2}{30}$$

$$\chi^2_h = \frac{36}{30} + \frac{64}{90} + \frac{25}{90} + \frac{49}{30} = 3,82$$

$\chi^2_{3,0.05} = 7,81$

$\chi^2_h = 3,82$

4) Poisson Dağılımı ile uyum

$\alpha = 0,05$

Bir sekreterin yazdığı 100 sayfalık bir yazıda her sayfadaki yaptığı hataların dağılımı aşağıda verilmiştir.

	0	1	2	3	4	5
Her sayfadaki hatalar	42	33	14	6	4	1

$$\bar{X} = \frac{\sum_{i=0} f_i x_i}{\sum f_i} = \frac{0.42 + 1.33 + 2.14 + 3.6 + 4.4 + 5.1}{100}$$

$$\bar{x} = \lambda = 1$$

H_0: veriler poisson dağılımı ile uyumludur.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1}.1^x}{x!} = \frac{e^{-1}}{x!} \quad x = 0,1,2,3,4,5$$

$$P(x) = \frac{e^{-1}}{x!}$$

$$P(0) = \frac{e^{-1}}{0!} = 0,368$$

$$P(1) = \frac{e^{-1}}{1!} = 0,368$$

$$P(2) = \frac{e^{-1}}{2!} = 0,184$$

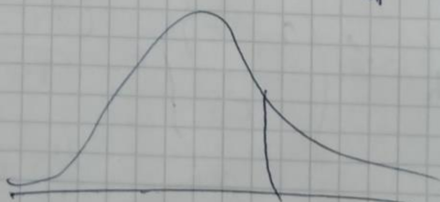
$$P(3) = \frac{e^{-1}}{3!} = 0,061$$

$$P(4) = \frac{e^{-1}}{4!} = 0,015$$

$$P(5) = \frac{e^{-1}}{5!} = 0,003$$

x	O_i	e_i	$\frac{(O_i - e_i)^2}{e_i}$
0	42	36,8	0,7348
1	33	36,8	0,3924
2	14	18,4	1,0521
3	6 } 11	6,1 } 8	1,1250
4	4	1,5	
5	1	0,3	

$\chi^2_h = 3,3043$



$\chi^2_{4-1-1,0,05} = 5,99$

$m = 1$ hesaplandı.

(5) Normal dağılıma uyum Testi

·Belli bir şikayet ile hastaneye başvuran rastgele seçilmiş 351 hastanın kolesterol düzeyleri tablodadır. $\alpha = 0{,}05$ anlamlılık düzeyinde normal dağılıma uyumu test ediniz.

Kolesterol	Hasta sayısı
$140 < x < 160$	14
$160 < x < 180$	62
$180 < x < 200$	80
$200 < x < 220$	65
$220 < x < 240$	50
$240 < x < 260$	30
$260 < x < 280$	27
$280 < x < 300$	23

H_0: Hastaların kolesterol düzeylerinin dağılımı normaldir.

H_1: Hastaların kolesterol düzeylerinin dağılımı normal değildir.

Anakütle parametreleri bilinmiyor

$S.d = k - m - 1 = k - 2 - 1 = 8 - 2 - 1 = 5$

$m \Rightarrow \mu, \sigma$ hesaplandığından 2

Kolesterol	x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
$140 < x < 160$	150	14	2100	53816
$160 < x < 180$	170	62	10540	109368
$180 < x < 200$	190	80	15200	38720
$200 < x < 220$	210	65	13650	260
$220 < x < 240$	230	50	11500	16200
$240 < x < 260$	250	30	7500	43320
$260 < x < 280$	270	27	7290	90828
$280 < x < 300$	290	23	6670	139932
			74450	492444

$\bar{X} = \dfrac{\sum x_i f_i}{\sum f_i}$

$\bar{x} = \dfrac{74450}{351} = 212$

$S_{\bar{x}}^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{n-1}$

$S_{\bar{x}}^2 = \dfrac{492444}{350} \cong 1407$ $\quad S_x = 37{,}5$

(6)

$140 < x < 160 \qquad z = \dfrac{x_i - \bar{x}}{S_x}$

$$\frac{140-212}{37{,}5} \le z \le \frac{160-212}{37{,}5}$$

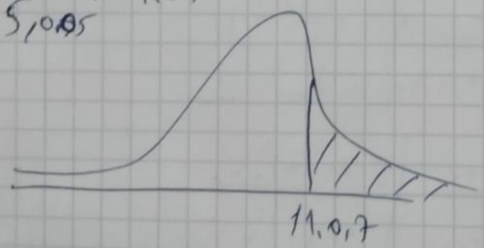
$-1{,}92 \le z \le -1{,}39 \rightarrow P = 0{,}0823 - 0{,}0274 = 0{,}0549$

$n \cdot p = 351(0{,}0549) = 19{,}26$

Kolesterol	Z	P	E	O	$\frac{(e_i-o_i)^2}{e_i}$
$140 < x < 160$	$-1{,}92 < z < -1{,}39$	0,0549	19,27	14	1,44
$160 \le x < 180$	$-1{,}39 < z < -0{,}85$	0,1154	40,51	62	11,41
$180 < x < 200$	$-0{,}85 < z < -0{,}32$	0,1768	62,06	80	5,19
$200 < x < 220$	$-0{,}32 < z < 0{,}21$	0,2087	73,25	65	0,93
$220 < x < 240$	$0{,}21 < z < 0{,}75$	0,1902	66,76	50	4,21
$240 < x < 260$	$0{,}75 < z < 1{,}28$	0,1263	44,33	30	4,63
$260 < x < 280$	$1{,}28 < z < 1{,}81$	0,0652	22,89	27	0,74
$280 < x < 300$	$1{,}81 < z < 2{,}35$	0,0257	9,02	23	21,66

$\chi^2_4 = 50{,}21$

$\chi^2_{5;0{,}05} = 11.07$



7 ~~İki~~ Ki Kare Bağımsızlık Testi

İlgilenilen 2 özellik arasında ilişki olup olmadığını test etmede kullanılır.

$S.d = (\text{satır sayısı} - 1) \times (\text{sütun sayısı} - 1)$

İki yönlü tablo — olumsallık tablosu — Kontenjans Tablosu

Örnek/ Bir ürünün pazar araştırmasında beğenilip beğenilmemesinin cinsiyete göre dağılımı aşağıdadır.
Cinsiyete bağımlılığı $\alpha = 0{,}05$ anlamlılık düzeyinde test ediniz.

	Kadın	Erkek	Toplam
Beğendim	90	40	130
Beğenmedim	35	110	145
Toplam	125	150	275

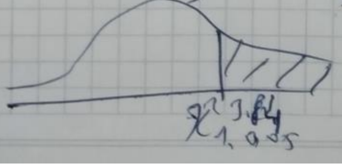
$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

$S.d = (2-1).(2-1) = 1$

Beğen Kadın Beklenen değeri = $(125 \times 130)/275 = 59.1$
Beğenen Erkek Beklenen değeri = $(150 \times 130)/275 = 70.9$
Beğenmeyen Kadın Beklenen değeri = $(125 \times 145)/275 = 65{,}9$
Beğenmeyen Erkek Beklenen değeri = $(150 \times 145)/275 = 79{,}1$

$$\chi^2_h = \frac{(90-59.1)^2}{59.1} + \frac{(40-70.9)^2}{70.9} + \frac{(35-65.9)^2}{65.9} + \frac{(110-79.1)^2}{79.1} = 56{,}19$$

$\chi^2_h = 56{,}19$



$\chi^2_{1,0{,}05} = 3{,}84$

8

Kikare Homojenlik Testi

İki veya daha fazla yığının dağılımlarının homojen (benzer) olup olmadığı test edilir.

Örnek! Kadın ve erkeklerin farklı yaş gruplarında son 2 yılda tiyatroya gitme sıklıkları arasında benzerlik olup olmadığını $\alpha = 0,025$ anlamlılık düzeyinde test ediniz.

	Kadın	Erkek	Toplam
20 yaş altı	38 e_{11}	30 e_{12}	68
20-50 yaş	46 e_{21}	52 e_{22}	98
50 yaş üstü	54 e_{31}	40 e_{32}	94
Toplam	138	122	260

$$\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \qquad s.d = (3-1)(2-1) = 2$$

$$e_{11} = \frac{138 \cdot 68}{260} = 36,09 \qquad e_{12} = \frac{122 \cdot 68}{260} = 31,9$$

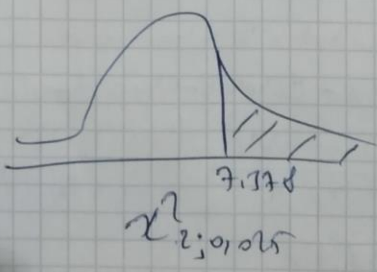
$$e_{21} = \frac{138 \cdot 98}{260} = 52,01 \qquad e_{22} = \frac{122 \cdot 98}{260} = 45,98$$

$$e_{31} = \frac{138 \cdot 94}{260} = 49,9 \qquad e_{32} = \frac{122 \cdot 94}{260} = 44,1$$

$$\chi^2_h = \frac{(38-36,09)^2}{36,09} + \frac{(30-31,9)^2}{31,9} + \frac{(46-52,01)^2}{52,01} + \frac{(52-45,98)^2}{45,98} + \frac{(54-49,9)^2}{49,9} + \frac{(40-44,1)^2}{44,1}$$

$$\chi^2_h = 2,41$$

$$\chi^2_{2;0,025} = 7,378$$



(9) İlişkilerin ölçülmesi

İlişki katsayıları

I. Her iki değişken sınıflama ölçeğinde ise

a) Olumsallık katsayısı

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$ Pearson tarafından önerilmiştir.

b) Cramer V katsayısı

$$V = \sqrt{\frac{\chi^2}{N \cdot \min\left[(I-1),(J-1)\right]}}$$

$0 < V < 1$
$0 < C < 1$

0 ilişki yok
1 tam ilişki

c) ϕ Phi katsayısı

V'nin özel durumu

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Örnek! N = 165 kişi işe kabul durumu olumsallık katsayısını ve V katsayısını hesaplayın

	İtalyanca	İngilizce	Çince
İşe kabul edildi	10	65	18
İşe kabul edilmedi	28	33	11

a) $\chi^2 = 18.29$ $C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{18,29}{18,29 + 165}} = 0,32$

İlişki var zayıf

b) $V = \sqrt{\frac{\chi^2}{N \cdot \min\{(I-1),(J-1)\}}} = \sqrt{\frac{18,29}{165 \cdot \min(3-1, 2-1)}} = 0,33$

(10) Her iki değişken sıralama ölçme düzeyinde ise

a) Gamma katsayısı

$$\gamma = \frac{P-Q}{P+Q} \qquad -1 < \gamma < 1$$

P: Uyumlu çift sayısı konkordans

~~Her satır hanede sütünların birhücrenin sıra indisi diğer hücrenin sıra indisinden küçükse~~

Bir hücrenin bulunduğu sütün ve satır kopatıldığında o hücrenin sağında ve altında kalan hücreler, uyumlu

Q: uyumsuz çift sayısı diskonkordans

Bir hücrenin bulunduğu sütun kopatıldığında solunda ve altında kalan hücreler uyumsuzdur.

Hücre değeri (Uyumlu hücrelerin toplamı) = P

Hücre değeri (Uyumsuz hücrelerin toplamı) = Q

b) Kendall'ın tau-b katsayısı $-1 < tau_b < 1$

$$tau\text{-}b = \frac{P-Q}{\sqrt{(P+Q+T_{sa})(P+Q+T_{sü})}}$$

T_{sa}: Satıra dağılı çiftlerin toplam sayısı

$T_{sü}$: Sütuna dağılı çiftlerin toplam sayısı

~~Hafif şiddette~~

~~Orta şiddette~~

~~Aşırı şiddette~~

c) $tau\text{-}c = \dfrac{2.S\,(P-Q)}{N \times N\,(S-1)}$

S → küçük olan satır yada sütün sayısı

11) Örnek! Bir araştırmada sürücülerin karıştırdıkları kazaların şiddetiyle sürücülerin yaşları arasındaki ilişki ortaya konmaya çalışılıyor. Tabloya göre a) Gama b) Kendall tau b c) Kendall tau c katsayılarını hesaplayınız.

	Genç	Orta	Yaşlı
Hafif şiddet	80	40	35
Orta şiddet	56	10	6
Ağır kaza	18	8	5

$N=258$

P Uyumlu çiftlerin sayısı

1. satır için $80(10+6+8+5)=2320$

$40(6+5)=440$

2. satır için $56(8+5)=728$

$10(5)=50$

$3538=P$

Q Uyumsuz çiftlerin sayısı

1. satır için $35(56+10+18+8)=3220$

$40(56+18)=2960$

2. satır için $6(18+8)=156$

$10(18)=180$

$6516=Q$

$$\gamma = \frac{3538-6516}{3538+6516} = -0{,}30$$

(12)

b) $tau\,b = \dfrac{P - Q}{\sqrt{(P + Q + T_{sa}) \cdot (P + Q + T_{sa})}}$

	Genç	Orta	Yaşlı
Holist	80	40	35
Orta	56	10	6
Ağır	18	8	5

$T_{sa} =$ $80 \times 40 = 3200$, $80 \times 35 = 2800$, $40 \times 35 = 1400$ } 1. satır

$56 \times 10 = 560$, $56 \times 6 = 336$, $10 \times 6 = 60$ } 2. satır

$18 \times 8 = 144$, $18 \times 5 = 90$, $8 \times 5 = 40$ } 3. satır

$T_{sa} =$ $80 \times 56 = 4480$ $80 \times 18 = 1440$ $56 \times 18 = 1008$ 1. satır

$40 \times 10 = 400$ $40 \times 8 = 320$ $10 \times 8 = 80$ 2. satır

$35 \times 6 = 210$ $35 \times 5 = 175$ $6 \times 5 = 30$ 3. satır

$T_{sa} = 8143$ $T_{sa} = 8630$

$tau\text{-}b = \dfrac{3538 - 6516}{\sqrt{(3538 + 6516 + 8630)(3538 + 6516 + 8143)}}$ $\rightarrow -0,16$

$tau\,c = \dfrac{2 \cdot S(P-Q)}{N \cdot N(S-1)} = \dfrac{2 \cdot 3(3538 - 6516)}{(258)(258)(3-1)} = -0,13$