

Benzetim

-5-

Girdi Analizi

- Varışlar arası zamanlar veya talep genişlikleri gibi rassal girdileri kullanan bir benzetimi gerçekleştirmek için bu girdilerin olasılık dağılımlarının belirlenmesi gerekir.
- Bir benzetim modeline girdi rassal değişkenlerini temsil eden özel dağılımlar verildiğinde, benzetimin zaman boyunca çalışmasında bu dağılımlardan üretilen rassal değerler kullanılır.
- Bir benzetim, olasılık dağılımları ile sistemin her bir Rassallık kaynağının gösterilmesi gerekir.
- Uygulamada karşımıza çıkan sistemler genellikle bir ya da daha fazla rassallık kaynağına sahiptirler.

SİSTEM	RASSALLIK KAYNAĞI
Üretim	İşlem zamanları , arızalanma aralıkları, tamir süresi
Bilgisayar	İşlerin varışlar arası zaman aralığı , iş tipleri , işlem zamanı
Haberleşme	Mesajların varışlar arası zaman aralığı, mesaj tipleri, mesaj uzunlukları

Girdi Dağılımlarının Belirlenmesinde 4 Adım:

- 1) veri toplama
 - 2) dağılım ailesinin belirlenmesi (üstel, normal , vb.)
 - 3) parametre tahmini
 - 4) uygunluk testi
- Uygunluk testi ile seçilen dağılım kabul edilmez ise , 2. adıma geri dönülür ve farklı bir dağılım seçilerek prosedür tekrarlanır.
 - Toplanan veri bilinen dağılımlardan hiçbirine uymuyor ise , AMPİRİK DAĞILIM tanımlaması yapılır.

1. Veri Toplama

- İncelenen sistem için, bir benzetim modeli geliştirildikten sonraki adım, modelde kullanılacak girdiler için sistemden verilerin toplanmasıdır.
- **Veri toplamada uyulması gereken kurallar:**
 - ✓ Sistem önceden gözlenmeli ve hangi verilerin toplanması gerektiğine , hangi zamanlarda verinin toplanacağına karar verilmelidir. Veri toplamak için gerekli formlar hazırlanmalıdır.
 - ✓ Girdi dağılımını belirlemek için yeterli verinin toplanması gerekir.
 - ✓ Sistemi iyi temsil edecek şekilde veri (homojen veri) toplanmalıdır.
 - ✓ Homojenliği kontrol etmek için kullanılan testlerden biri 2 örnekli t-testidir. Bu test ile dağılımların ortalamalarının eşit olup olmadığı test edilir.
 - ✓ İki değişken arasında bir ilişkinin olup olmadığının belirlenmesi gerekir. Scatter diyagramları kullanarak ilişkinin varlığı gözlemlenebilir. Regresyon analizi de değişkenler arasında ilişkinin belirlenmesinde kullanılmaktadır

2. Dağılım Ailesinin Belirlenmesi

- Dağılım ailesinin seçiminde;

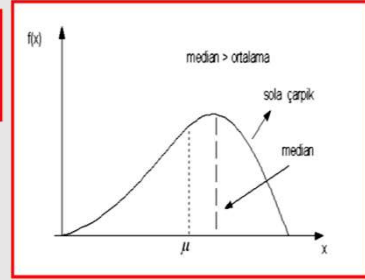
Nokta İstatistikleri

- Bazı özel dağılımlar özel istatistik değerlere sahiptir. Bu istatistikler veriden elde edilir ve teorik dağılımın nokta istatistikleri ile karşılaştırılır.

Ortalama, Medyan ve Varyans

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \rightarrow \text{ortalama}, \quad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \rightarrow \text{varians}$$

$$x_{0,5}(n) = \begin{cases} x_{(n+1)/2}, & n \text{ tek sayı} \\ \frac{x_{(n/2)} + x_{(n/2)+1}}{2}, & n \text{ çift ise} \end{cases} \rightarrow \text{medyan}$$



2. Dağılım Ailesinin Belirlenmesi

Değişim Katsayısı ve Lexis Oranı

- Değişim katsayısı sürekli dağılımın şekli hakkında bilgi sahibi olmayı sağlar.

$$\delta = \frac{\sqrt{s^2(x)}}{\bar{x}} \quad x_1, x_2, \dots, x_n$$

- Üstel dağılım için,

$$\delta = \frac{\sqrt{\frac{1}{\beta^2}}}{\frac{1}{\beta}} = 1$$

- Değişim katsayısı "1" ise , dağılımın üstel olduğunu gösterir.

2. Dağılım Ailesinin Belirlenmesi

- Bazı Sürekli Dağılımlar İçin Değişim Katsayısı

Dağılım	$\hat{\delta}$	$\hat{\delta}$ 'in açıklığı
$U(a,b)$	$\frac{b-a}{\sqrt{3(a+b)}}$	$(-\infty, \infty]$
$Expo(\beta)$	1	{ 1 }
$N(\mu, \sigma^2)$	$\frac{\sigma}{\mu}$	$(-\infty, \infty)$ (0 hariç)
$Beta(\alpha_1, \alpha_2)$	$\left[\left(\frac{\alpha_1}{\alpha_2} \right) (\alpha_1 + \alpha_2 + 1) \right]^{-\frac{1}{2}}$	$(0, -\infty)$
$\tilde{Üçgen}$ (a,b,c)	$\frac{(a^2 + b^2 + c^2 - ab - ac - bc)^{\frac{1}{2}}}{\sqrt{2}(a+b+c)}$	$(-\infty, \infty)$ (0 hariç)

2. Dağılım Ailesinin Belirlenmesi

Lexis Oranı

- kesikli dağılımlar için kullanılır. Değişim katsayısı ile aynı işleve sahiptir.

$$\hat{r} = \frac{s^2}{\bar{x}} \quad \hat{r} = \begin{cases} =1 & \text{Poisson} \\ >1 & \text{Negative Binomial} \\ <1 & \text{Binomial} \end{cases}$$

Çarpıklık ve Basıklık Katsayısı

- Çarpıklık katsayısı , bir dağılımın simetrikliğinin ölçüsüdür.

- ❑ $\alpha_3 = 0$ ise dağılım simetrik; $\alpha_3 > 0$ ise dağılım sağa çarpık ;
- ❑ $\alpha_3 < 0$ ise dağılım sola çarpıktır.
- ❑ normal dağılımda $\alpha_3 = 0$
- ❑ üstel dağılımda $\alpha_3 = 2$

$$\alpha_3 = \frac{E[(x-\mu)^3]}{(\sigma^2)^{\frac{3}{2}}}, \quad \hat{\alpha}_3 = \frac{\sum_{i=1}^n [x_i - \bar{x}_{(n)}]^3}{[s^2(n)]^{\frac{3}{2}}}$$

2. Dağılım Ailesinin Belirlenmesi

Basıklık Katsayısı

- dağılımın yüksekliğinin ölçüsüdür.
- Normal dağılımda $\alpha_4 = 3$
- Uniform Dağılımda $\alpha_4 = 1.8$
- Üstel Dağılımda $\alpha_4 = 9$

$$\alpha_4 = \frac{E[(x - \mu)^4]}{(\sigma^2)^2} \quad , \quad \hat{\alpha}_4 = \frac{\sum_{i=1}^n [x_i - \bar{x}_{(n)}]^4}{[s^2(n)]^2}$$

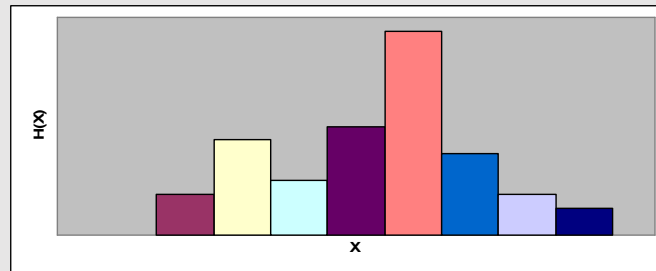
Histogramlar

- Bir histogram , toplanan verinin dağılımı ile ilgili olasılık fonksiyonunun grafiksel tahminidir.
- Bir histogram, veri için uygun bir model olarak araştırılan dağılımlar ile ilgili iyi bir ipucu verir.
- Veriden yararlanılarak elde edilen histogram teorik dağılımın şekli ile karşılaştırılır.

2. Dağılım Ailesinin Belirlenmesi

Histogramlar

- x_1, x_2, \dots, x_n gözlemler olsun. Açıklık eşit uzunlukta k aralığa bölünür.
- Bir aralığın genişliği Δb olsun; $\Delta b : [b_0, b_1], [b_1, b_2], \dots, [b_{k-1}, b_k]$
- $h_j : [b_{j-1}, b_j]$ aralığına düşen x oranını gösterebilir.
- Çizilen histogramlar , teorik dağılımın şekli ile karşılaştırılarak verinin hangi dağılımdan geldiği belirlemeye çalışılır



2. Dağılım Ailesinin Belirlenmesi

Aralık Sayısının Belirlenmesi

Montgomery'e göre aralık sayısı $\rightarrow \sqrt{n}$
 Blank'e göre aralık sayısı:
 $n > 50$ ise $10-20$
 $n \leq 50$ ise $6-10$ arasında belirlenebilir

Aralık Genişliğinin Belirlenmesi

$$\Delta_d = \frac{[\text{max değer}] - [\text{min imum değer}]}{\text{aralık sayısı}}$$

- Bir histogramın çiziminde aralık genişliğinin belirlenmesi önemlidir.
- Aralık genişliğinin çok büyük ya da küçük alınması ile çizilen histogram , verinin hangi dağılımdan geldiğine ilişkin iyi bir bilgi vermez.

3) Parametre Tahmini

- Veri için uygun bir dağılım belirlendikten sonra , bu dağılımın benzetimde kullanımı için parametre değerlerinin belirlenmesi gerekir.
- Elde edilen x_1, x_2, \dots, x_n veri seti, dağılımın parametrelerinin tahmin edilmesinde kullanılır.
- Bir tahminci, verinin numerik bir fonksiyonudur. Bu dağılımın parametresini (ya da parametrelerini) tahmin etmek için kullanılan çeşitli metotlar vardır.

Parametre Tahmini Metotları

- maximum likelihood tahmin edici (MLE)
- en küçük kareler tahmin edici
- moment metodu

3) Parametre Tahmini

Maximum Likelihood Tahmin Edici (MLE)

- Veriye uydurulan dağılımın bir kesikli veya sürekli dağılım olduğunu Kabul edelim. Bu dağılımın bir parametresi ; θ olsun
- $P_{\theta}(x)$: Dağılımın olasılık fonksiyonu
- x_1, x_2, \dots, x_n : gözlemlenen değerler verildiğinde ;
- Likelihood fonksiyonu $L(\theta)$ aşağıdaki gibi tanımlanır.

$$L(\theta) = P_{\theta}(x_1) P_{\theta}(x_2) \dots P_{\theta}(x_n)$$

- $L(\theta)$; "bileşik olasılık fonksiyonu" dur ve gözlemlenen verinin elde edilme olasılığını verir.
- θ ' nın bilinmeyen değerinin MLE; $L(\theta)$ ' u maksimize eden θ değeri olarak tanımlanır ve $\hat{\theta}$ ile gösterilir.

$$\text{MLE} \rightarrow \max L(\theta) \rightarrow \hat{\theta}$$

3) Parametre Tahmini

Örnek 1: Üstel dağılımın parametresini MLE ile tahmin ediniz

3) Parametre Tahmini

Örnek 2: Bernoulli dağılımın parametresini MLE ile tahmin ediniz

3) Parametre Tahmini

Distribution	Parameter(s)	Suggested estimators
Bernoulli	p	$\hat{p} = \hat{X}$
Poisson	λ	$\hat{\lambda} = \hat{X}$
Exponential	λ	$\hat{\lambda} = \frac{1}{\hat{X}}$
Normal	μ, σ^2	$\hat{\mu} = \hat{X}$ $\widehat{\sigma^2} = S^2$

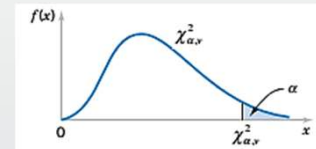
4) Uygunluk Testleri

- Uyum testleri, verilerin seçilen dağılıma ne kadar iyi uyduğunu gösterir.
- Bir uygunluk testi aşağıdaki hipotezi test etmek için kullanılır.
- H_0 : x_i gözlemleri , F^{\wedge} dağılım fonksiyonu ile bağımsız özdeş dağılmış rassal değişkenlerdir.
- Hipotez testi prosedürü, kitleden alınan bir rasgele örneklemedeki bilginin kullanılmasına dayanır.
- Eğer bu bilgi hipotezle tutarlı ise, hipotezin doğru olduğu sonucuna; eğer bu bilgi hipotez ile tutarlı değilse, hipotezin yanlış olduğu kararına varırız.
- Verilerin uyumu,
 - ❑ Ki-kare (χ^2) (Kesikli ve Sürekli dağılımlar)
 - ❑ Kolmogorov Smirnov(Sadece Sürekli dağılımlar)
 - ❑ Anderson Darling(Sadece Sürekli dağılımlar)

testleriyle kontrol edilir.

Ki-Kare (χ^2) Testi

- Ki kare testi şu adımları gerçekleştirerek yapılır:
 1. Örnek verilerin tüm parametrelere verilmiş ve ya tahmin edilmiş olan bir dağılımdan geldiği hipotez kurulur.
 2. Test edilen dağılımın aralığı, m adet alt bölge ve aralığa bölünür.
 3. Her bir aralığa denk gelen teorik frekans değeri belirlenir. Burada her bir guruba düşen frekans sayısının veya örnek adedin 5'ten az olmaması gerekir. Eğer 5'ten az ise bir önceki gurup ile birleştirilir.
 4. Ki-kare değerini hesapla aşağıdaki formül ile hesapla:
 - $\chi_0^2 = \sum_{j=1}^m \frac{(G_i - B_i)^2}{B_i}$
 - G_i : i . sınıf aralığında gözlenen frekans,
 - B_i : i . sınıf aralığında beklenen frekans.
 5. Belirlenen güvenilirlik düzeyi ve serbestlik derecesine göre tablodan ki-kare değeri okunur.



Serbestlik derecesi= $m - (\text{test edilen dağılımın parametre sayısı}) - 1$

6. Hesaplanan ki-kare değeri, tablodan okunan değerden küçük ise hipotez KABUL edilir. Aksi halde RED edilir.

Örnek 1: Belli bir ebattaki metal levha üzerindeki hata sayılarının Poisson dağılımına uyup, uymadığını araştıralım. 60 birimlik rassal örneklem alınmış ve aşağıda verilen hata sayıları gözlenmiştir.

Hata Sayısı	Gözlenen Frekans
0	32
1	15
2	9
3	4

Poisson dağılımının parametresini nasıl tahmin edelim ?

- Hipotezlenen $\lambda=0.75$ hata/levha parametrelili Poisson dağılımından i. sınıf aralığıyla ilgili P_i olasılıklarını aşağıdaki gibi hesaplayabiliriz.

$$f(x) = P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x=0,1,2,\dots$$

$$p_1 = P(X=0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X=1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X=2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041$$

- Beklenen frekansları hesaplamak için örneklem büyüklüğü $n=60$ ve P_i olasılıkları çarpılır.

$$B_i = n \cdot P_i$$

Hata Sayısı	Olasılık	Beklenen Frekans
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (veya daha fazla)	0.041	2.46

Eğer beklenen frekans 5'ten küçükse, önceki sınıfla birleştir:

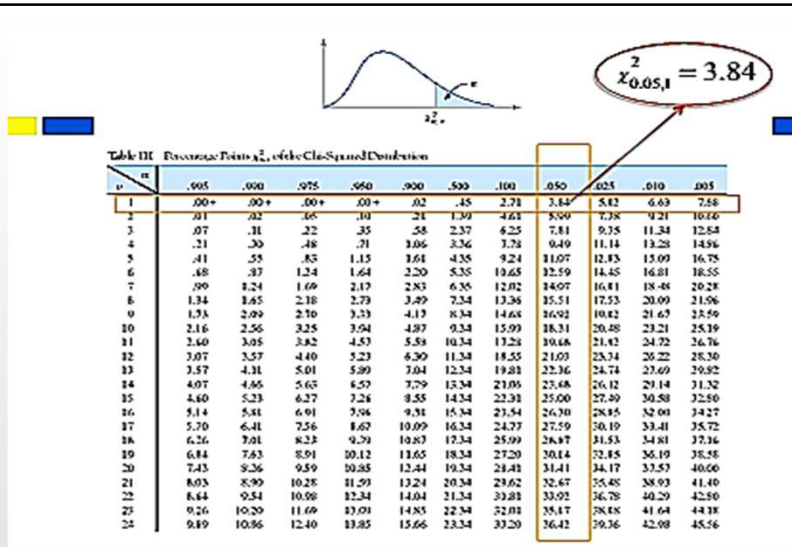
Hata Sayısı	Gözlenen Frekans	Beklenen Frekans
0	32	28.32
1	15	21.24
2, (veya daha fazla)	13	10.44

- H_0 : Levha üzerindeki hata sayısı Poisson dağılımına uyar.
- H_1 : Levha üzerindeki hata sayısı Poisson dağılımına uymaz.

$$\alpha=0.05$$

$$\text{Test İstatistigi: } \chi_0^2 = \sum_{j=1}^m \frac{(G_j - B_j)^2}{B_j}$$

$$\chi_0^2 = \frac{(32-28.32)^2}{28.32} + \frac{(15-21.24)^2}{21.24} + \frac{(13-10.44)^2}{10.44} = 2.94$$



Örnek 2:

Bir eczane müşterilere servis süreleri (dk.) rassal olarak gözlemlenmiş ve aşağıda verilen 100 örneklem verisi oluşturulmuştur.

Servis sürelerinin Üstel dağılıma uyup uymadığını Ki-kare Uyum testiyle kontrol edin.

1,02	4,65	3,24	0,92	6,64
0,98	1,18	2,53	0,97	1,45
6,25	13,04	3,09	10,68	2,29
3,26	0,77	2,01	6,82	2,86
3,1	2,11	6,13	1,45	2,26
4,48	1,83	2,45	0,98	4,86
4,29	4,9	1,23	1,19	3,12
7,93	7,23	7,35	2,03	2,9
1,29	3,52	5,5	1,62	15,19
1,48	4,44	1,63	5,02	3,36
1,8	1,2	0,89	1,15	2,5
1,58	1,09	4,67	4,61	2,72
4,07	1,19	2,05	8,3	11,3
3,84	2,59	0,5	3,88	4,26
6,18	1,39	8,59	0,87	6,6
2,67	2,47	0,94	2,4	3,03
1,75	0,77	1,25	1,44	9,23
1,72	1,29	3,14	1,78	2,05
2,34	1,74	4,63	3,06	1,49
3,23	5,86	1,59	3,65	9,11

$$k = \sqrt{n} = \sqrt{100} = 10$$

(k: sınıf sayısı)

$$S = \frac{X_{\max} - X_{\min}}{k} = \frac{15,19 - 0,5}{10} = 1,469 \approx 1,5$$

(S: sınıf genişliği)

Sınıflar	f_i
2,0'den az	37
2,0 - 3,5	28
3,5 - 5,0	15
5,0 - 6,5	6
6,5 - 8,0	6
8,0 - 9,5	4
9,5 - 11,0	1
11,0 - 12,5	1
12,5 - 14,0	1
14,0'ten çok	1

Üstel dağılımının parametresini nasıl tahmin edelim ?

$$\hat{\lambda} = \frac{1}{\bar{x}} = 0,285$$

H_0 : Müsterilerin servis süreleri Üstel dağılıma uyar.

H_1 : Müsterilerin servis süreleri Üstel dağılıma uymaz.

X (Servis Süresi) – Üstel(λ)

$$P(a < X < b) = \int_a^b f(x) dx = \int_a^b \lambda e^{-\lambda x} dx$$

$$f(x) = \lambda e^{-\lambda x} \quad x > 0$$

$$P(a < X < b) = F(b) - F(a)$$

$$F(x) = 1 - e^{-\lambda x}$$

$$F(b) = 1 - e^{-\lambda b} \quad ; \quad F(a) = 1 - e^{-\lambda a}$$

(Beklenen Frekansların Belirlenmesi)

Sınıf	$F(x)$	$P(a < X < b)$	E_i
<2	0,435	0,435	43,5
2 - 3,5	0,632	0,197	19,7
3,5 - 5	0,76	0,128	12,8
5 - 6,5	0,844	0,084	8,4
6,5 - 8	0,898	0,054	5,4
8 - 9,5	0,934	0,036	3,6
9,5 - 11	0,957	0,023	2,3
11 - 12,5	0,972	0,015	1,5
12,5 - 14	0,982	0,01	1
>14	1	0,018	1,8

$100 * 0,435 = 43,5$

Beklenen frekans 5'ten küçük sınıflar, önceki sınıfla birleşir

Sınıf	G_i	E_i
<2	37	43,5
2 - 3,5	28	19,7
3,5 - 5	15	12,8
5 - 6,5	6	8,4
6,5 - 8	6	5,4
>8	8	10,2

$F(2) = 1 - e^{-10,285 \cdot 2} \approx 0,435$; $F(3,5) = 1 - e^{-10,285 \cdot 3,5} \approx 0,632$

$P(2 < X < 3,5) = F(3,5) - F(2) = 0,632 - 0,435 = 0,197$

H_0 : Servissüreleri $\lambda=0,285$ olay/dk. olan Üstel dağılıma uyar.

H_1 : Servissüreleri $\lambda=0,285$ olay/dk. Olan Üsteldagılıma uymaz.

$\alpha=0.01$

Serbestlik derecesi = n -(test edilen dağılımın parametre sayısı)-1

$$v=6-1-1=4$$

Test İstatistigi:
$$\chi_0^2 = \sum_{j=1}^m \frac{(G_i - B_i)^2}{B_i}$$

$$\chi_0^2 = \frac{(37-43.5)^2}{43.5} + \frac{(28-19.7)^2}{19.7} + \dots + \frac{(8-10.2)^2}{10.2} = 6,073$$

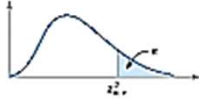


Table III: Percentage Points χ^2_{α} of the Chi-Squared Distribution

α	.995	.990	.975	.950	.900	.800	.700	.600	.500	.400	.300	.200	.100	.050	.025	.010	.005
1	.004	.005	.008	.016	.032	.054	.078	.104	.135	.171	.215	.268	.337	.424	.541	.675	.842
2	.010	.012	.016	.020	.029	.041	.054	.070	.088	.110	.136	.168	.207	.259	.325	.401	.496
3	.016	.018	.024	.029	.039	.052	.067	.085	.104	.128	.158	.194	.238	.292	.359	.438	.539
4	.216	.230	.248	.270	.297	.329	.364	.403	.447	.496	.550	.609	.674	.746	.825	.912	1.026
5	.412	.428	.446	.467	.493	.523	.557	.595	.638	.686	.739	.797	.860	.928	1.002	1.083	1.181
6	.485	.502	.520	.542	.568	.598	.632	.670	.713	.761	.814	.871	.934	1.002	1.077	1.159	1.266
7	.559	.576	.594	.616	.642	.672	.706	.744	.787	.835	.888	.945	1.008	1.077	1.159	1.241	1.348
8	.634	.651	.669	.691	.717	.747	.781	.824	.867	.915	.968	1.025	1.088	1.159	1.241	1.323	1.430
9	.710	.727	.745	.767	.793	.823	.857	.899	.942	.990	1.043	1.096	1.159	1.241	1.323	1.405	1.512
10	.788	.805	.823	.845	.871	.901	.935	.977	1.020	1.068	1.121	1.174	1.237	1.319	1.401	1.483	1.590
11	.867	.884	.902	.924	.950	.980	1.014	1.056	1.100	1.148	1.201	1.254	1.317	1.399	1.481	1.563	1.670
12	.946	.963	.981	1.003	1.029	1.059	1.093	1.135	1.179	1.227	1.280	1.333	1.396	1.478	1.560	1.642	1.749
13	1.026	1.043	1.061	1.083	1.109	1.139	1.173	1.215	1.259	1.307	1.360	1.413	1.476	1.558	1.640	1.722	1.829
14	1.106	1.123	1.141	1.163	1.189	1.219	1.253	1.295	1.339	1.387	1.440	1.493	1.556	1.638	1.720	1.802	1.909
15	1.186	1.203	1.221	1.243	1.269	1.299	1.333	1.375	1.419	1.467	1.520	1.573	1.636	1.718	1.800	1.882	1.989
16	1.266	1.283	1.301	1.323	1.349	1.379	1.413	1.455	1.499	1.547	1.600	1.653	1.716	1.798	1.880	1.962	2.069
17	1.346	1.363	1.381	1.403	1.429	1.459	1.493	1.535	1.579	1.627	1.680	1.733	1.796	1.878	1.960	2.042	2.149
18	1.426	1.443	1.461	1.483	1.509	1.539	1.573	1.615	1.659	1.707	1.760	1.813	1.876	1.958	2.040	2.122	2.229
19	1.506	1.523	1.541	1.563	1.589	1.619	1.653	1.695	1.739	1.787	1.840	1.893	1.956	2.038	2.120	2.202	2.309
20	1.586	1.603	1.621	1.643	1.669	1.699	1.733	1.775	1.819	1.867	1.920	1.973	2.036	2.118	2.200	2.282	2.389
21	1.666	1.683	1.701	1.723	1.749	1.779	1.813	1.855	1.899	1.947	2.000	2.053	2.116	2.198	2.280	2.362	2.469
22	1.746	1.763	1.781	1.803	1.829	1.859	1.893	1.935	1.979	2.027	2.080	2.133	2.196	2.278	2.360	2.442	2.549
23	1.826	1.843	1.861	1.883	1.909	1.939	1.973	2.015	2.059	2.107	2.160	2.213	2.276	2.358	2.440	2.522	2.629
24	1.906	1.923	1.941	1.963	1.989	2.019	2.053	2.095	2.139	2.187	2.240	2.293	2.356	2.438	2.520	2.602	2.709

$\chi_{0.01,4}^2 = 13.28$

$$6.073 < 13.28 \rightarrow \chi_0^2 < \chi_{0.05,1}^2 \rightarrow H_0 \text{ Kabul edilir}$$