SERİ B CİLT 42 SAYI 1-2 1992

İSTANBUL ÜNİVERSİTESİ

ORMAN FAKÜLTESİ

D E R G İ S İ

KOLMOGOROV-SMIRNOV TESTİ İLE TOPLUM YAPISININ SINANMASI

Y. Doç. Dr. Ercan TANRITANIR[1)]

Kısa Özet

Toplum yapısının sınanmasında genellikle Khi-Kare Testi uygulanmaktadır. Halbuki Kolmogorov-Smirnov Testi hemen hemen her durumda Khi-Kare Testi'nden daha etkilidir. Bu amaçla sözkonusu test hakkında bilgi verilerek, bir modülün talep yapısı ve ebatlama makinası verileri bu yöntem ile test edilmiştir.

1. GİRİŞ

Bireyler üzerinde belirli ve durağan koşullarda yapılan ölçümler eşit olacaktır. Fakat doğada sonsuz sayıda küçük ve rastgele nedenler yüzünden tam bir eşitlik gözlenememektedir. Örneğin, hepsinde aynı olması beklenen bir x değerinde;

$$x = \mu \pm \in$$

şeklinde sapmalar görülmektedir. Eğer bu farklar sadece rastlantısal nedenlerden ileri geliyorsa, bu farklılıkların önemsiz olduğu, dolayısıyla yaklaşık olarak eşit olacakları kabul edilmekte veya aralarında rastlantısal olmayan önemli etkilerin bulunduğu ve eşit sayılamayacakları yargısına varılmaktadır.

Farkların önemsiz olduğu veya sıfır kabul edilebileceği durumlar için sıfır varsayımı, farklılaşmanın önemli ve anlamlı olduğu durumlarda ise alternatif varsayım kabul edilmektedir. Bu durum **"Uygunluk Testi (Goodness of Fit Testing)"** ile sınanmakta ve bu amaçla;

- Grafik Çizim
- İstatistiksel Ölçüleri Karşılaştırma
- Khi-Kare Testi
- Kolmogorov-Smirnov Testi

kullanılmaktadır (KALIPSIZ 1981).

---

1) İ. Ü. Orman Fakültesi, Orman Endüstrisi, Makinaları ve İşletme Anabilim Dalı.

Yayın Komisyonuna Sunulduğu Tarih : 11.12.1995

2. DAĞILIMLAR

İstatistik dağılımlar; rastgele nedenlerin etkilediği gerçek olay ve nesnelerden alınan örnekler üzerinde gözlem, ölçüm ve deney yapılarak tümevarım yöntemiyle oluşturulmuşlardır. Ancak bu dağılımlar, olayları etkileyen ve denetlenemeyen etmenler yüzünden gerçeği tam olarak yansıtmamaktadırlar.

Teorik dağılımlar; olasılık oranları belirli topumlardan yararlanılarak tümdengelim yolu ile matematik modelleri kurulmak suretiyle türetilen fonksiyonlardır. Bu fonksiyonlar parametreleri ile tanımlanmaktadırlar.

Teorik dağılımlar, x rastlantı değişkeninin nitel veya nicel oluşuna göre **"Sürekli"** ve **"Süreksiz"** diye ikiye ayrılmaktadırlar.

Süreksiz dağılımların sonuçları olumlu-olumsuz, tek-çift, sağlam-hasta; kusurlu-kusursuz, geçer-kalır gibi çoğu kez birbirine karşıt iki terime indirgenmektedir. Olumluluğun olasılık oranı p, olumsuzluğun oranı ise q'dur. Bu oran Hipergeometrik ve Multinomial dağılımda ise iki durumlu olmayıp tavla zarı ve tombalada olduğu gibi çok durumludur. Süreksiz dağılımlara örnek olarak Binomal, Hipergeometrik, Poisson, Uniform ve Multinomial Dağılımlar gösterilebilir.

Sürekli dağılımlar sadece net değerler almayıp, ara değerler de alabilmektedirler. Örneğin arabanın tamir süresi, ağacın boyu ve çapı gibi. Sürekli dağılımlara Normal, Gamma, Beta ve Üssel Dağılımlar örnek verilebilir (KALIPSIZ 1981).

3. ÜSSEL (EXPONENTIAL) DAĞILIM

Bu dağılım, simülasyon deneylerinde benzer olayların gelişleri arasındaki zaman aralığını açıklamak için kullanılmaktadır. Verilen bir zaman aralığındaki olayın meydana gelme olasılığı az ve diğer olaylardan istatistiksel olarak bağımsız ise, olayların oluşları arasındaki zaman aralıkları üssel dağılım ile açıklanabilir (HALAÇ 1982, KARAYALÇIN 1979, KARAYALÇIN/KÜLÜR 1968).

Üssel dağılımın olasılık yoğunluk fonksiyonu;

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

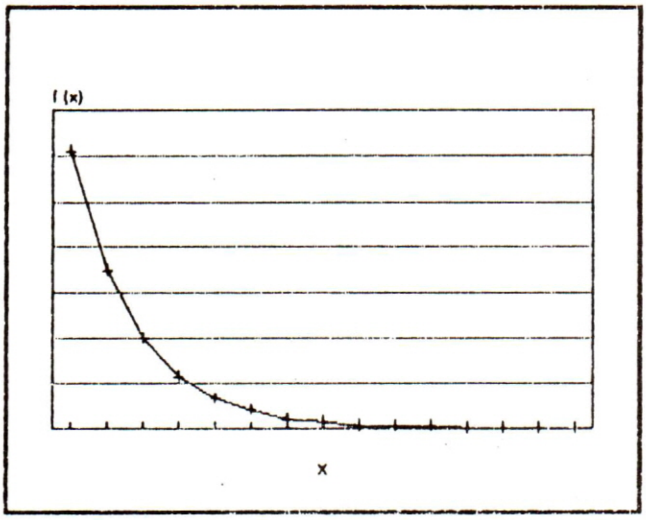
birikimli olasılık fonksiyonu;

$$F(x) = 1 - e^{-\lambda x} \text{ 'dır.}$$

Bu dağılımın olasılık fonksiyonu süreklidir ve parametre olarak λ ile gösterilmektedir. Dağılımın aritmetik ortalaması;

$$\bar{x} \approx \mu = \sigma = 1 / \lambda$$

bağıntısından bulunmaktadır (KALIPSIZ 1981).



Şekil 1: Üssel Dağılım

4. MATERYAL VE YÖNTEM

4.1. Materyal

Bu çalışmanın materyalini bir mobilya fabrikasının ürettiği modüllerden birine ilişkin günlük talep değerleri ve ebatlama makinasından elde edilen hatalı kesim verileri oluşturmaktadır.

4.2. Yöntem

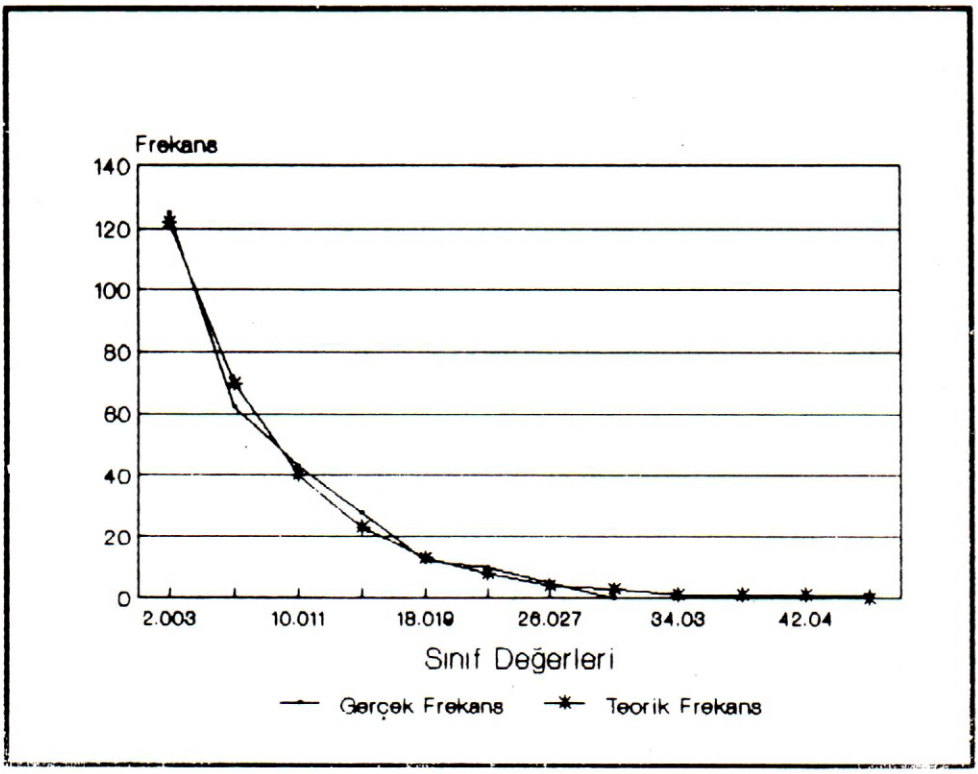
Bu yöntem, herhangi bir frekans dağılımının belirli bir teorik dağılıma uygunluğunu sınamak için geliştirilmiştir. Burada örnek büyüklüğünün yetersiz olduğu durumlarda da ($n < 30$) güvenilir sonuç alınmaktadır. Teorik dağılımdan farklılığın ölçülmesinde Kolmogorov-Smirnov Testi, frekans dağılımının düzensizliğini ortaya çıkarmada ise Khi-Kare Testi önerilmektedir (KALIPSIZ 1981).

Kolmogorov-Smirnov testinde her basamak için gerçek ve teorik birikimli frekanslar oluşturulmakta, her basamak için bu değerlerin sapması alınarak tablo değeri ile kıyaslanmaktadır. Hesaplanan farklar, tablo değerinden küçük ise bu iki dağılımın birbirine uygun olduğu kanısına varılmaktadır (Çizelge 1). Kolmogorov-Smirnov testinde tablo değerleri serbestlik derecesinin 35 olduğu duruma kadar verilmektedir. Bu sayı aşıldığında ise tablo değerleri, farklı güvenirlik düzeyleri için farklı formüller yardımı ile bulunmaktadır (HALAÇ 1982).

Örnek:

Çizelge 1: Kolmogorov-Smirnov Testi'ne Göre Dağılım Testi.

Sınıf Başlangıç Değeri	Sınıf Değeri	Sınıf Bitiş Değeri	Gerçek Frekans	Gerçek Birikimli Olasılık	Teorik Frekans	Teorik Birikimli Olasılık	Fa
0.001	2.003	4.004	125	0.4371	122	0.4624	0.0
4.005	6.007	8.008	62	0.6538	70	0.6710	0.0
8.009	10.011	12.012	43	0.8042	40	0.8113	0.0
12.013	14.015	16.016	28	0.9021	23	0.8918	0.0
16.017	18.019	20.020	12	0.9441	13	0.9379	0.0
20.021	22.023	24.024	10	0.9790	8	0.9644	0.0
24.025	26.027	28.028	5	0.9965	4	0.9796	0.0
28.029	30.031	32.032	0	0.9965	3	0.9883	0.0
32.033	34.035	36.036	0	0.9965	1	0.9933	0.0
36.037	38.039	40.040	0	0.9965	1	0.9961	0.0
40.041	42.043	44.044	0	0.9965	1	0.9978	0.0
44.045	46.047	48.048	1	1.0000	0	0.9987	0.0
			N = 286		N = 286		



Şekil 2: Gerçek ve Teorik Birikimli Frekanslar

$$\mu = \frac{(2.003)\,125 + (6.007)\,62 + (10.011)\,43 + (14.015)\,28 + (18.019)\,12 + (22.023)\,10 + (26.027)\,5 + (30.031)\,0 +}{286}$$

$$\frac{(34.035)\,0 + (38.039)\,0 + (40.041)\,0 + (42.043)\,0 + (46.047)\,1}{} = 7.196 \text{ adet/gün}$$

$$\lambda = \frac{1}{\mu} = \frac{1}{7.196} = = 0.1389 \text{ gün/adet}$$

Teorik Birikimli olasılık : $e^{-\lambda x} = e^{-\lambda\,(\text{Sınıf Bitiş Değeri})}$

$e = 2.71828$

Talebin 4.004'den aşağı olma olasılığı:

$F\ (4.004) = 1 - e^{-0.1389\,(4.004)} = 1 - e^{-0.5562} = 1–0.573 = 0.4264$

Teorik Frekans = 286 (0.4264) = 122 adet

Talebin 8.008'ten aşağı olma olasılığı:

$F(8.008) = 1 - e^{-0.1389\,(8.008)} = 1\text{-}e^{-1.1123} = 1\text{–}0.329 = 0.671$

Teorik Frekans = 286 (0.6710 – 0.4264) = 70 adet

Talebin 12.012'den aşağı olma olasılığı

$F(12.012) = 1\text{–}e^{-0.1389\,(12.012)} = 1\text{–}e^{-1.6685} = 1\text{–}0.1885 = 0.8115$

Teorik Frekans = 286 (0.8115 – 0.6710) = 40 adet

Talebin 48.048'den aşağı olma olasılığı:

$F(48.048) = 1\text{–}e^{-0.1389\,(48.048)} = 1\text{–}e^{-6.6738} = 1\text{–}0.0013 = 0.9987$

Teorik Frekans = 286 (0.9987 – 0.9978) = 0.0009 ≈ 0 adet

Çalışmamızda örnek sayısı 286'dır. Bu durumda tablo değerimiz sabit hale gelmektedir. Çünkü serbestlik derecesi 35'i aştığında α 0.05 için kontrol değeri;

$$D = \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{286}} = 0.080 \text{ olmaktadır.}$$

0.080 > 0.0172 olduğundan sözkonusu dağılım üssel bir dağılımdır.

5. UYGULAMA

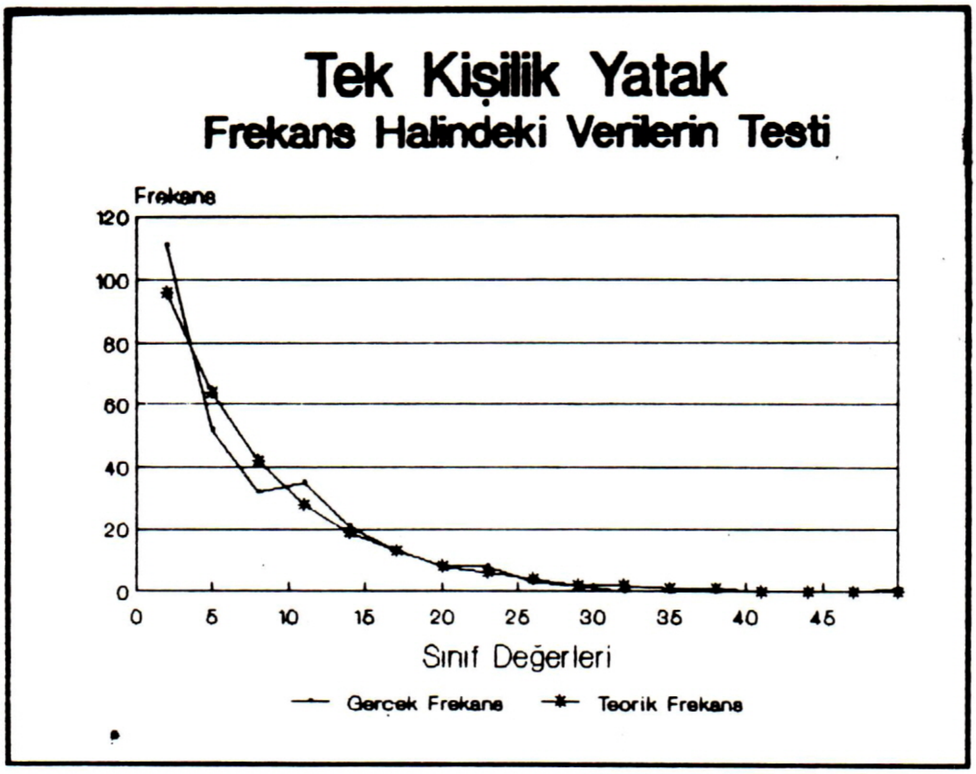
Fabrikadan elde edilen değerlere bilgisayar desteğinde Kolmogorov-Smirnov Testi uygulanmış ve aşağıdaki sonuçlar elde edilmiştir:

Çizelge 2: Frekans Halindeki Verilerin Dağılım Testi.

Sınıf Başlangıç Değeri	Sınıf Değeri	Sınıf Bitiş Değeri	Gerçek Frekans	Gerçek Birikimli Olasılık	Teorik Frekans	Teorik Birikimli Olasılık	
1.000	2.000	3.000	111	0.3881	96	0.3338	0.
4.000	5.000	6.000	52	0.5699	64	0.5562	0.
7.000	8.000	9.000	32	0.6818	42	0.7043	0.
10.000	11.000	12.000	35	0.8042	28	0.8030	0.
13.000	14.000	15.000	21	0.8776	19	0.8688	0.
16.000	17.000	18.000	13	0.9231	13	0.9126	0.
19.000	20.000	21.000	8	0.9510	8	0.9418	0.
22.000	23.000	24.000	8	0.9790	6	0.9612	0.
25.000	26.000	27.000	3	0.9895	4	0.9741	0.
28.000	29.000	30.000	2	0.9965	2	0.9828	0.
31.000	32.000	33.000	0	0.9965	2	0.9885	0.
34.000	35.000	36.000	0	0.9965	1	0.9924	0.
37.000	38.000	39.000	0	0.9965	1	0.9949	0.
40.000	41.000	42.000	0	0.9965	0	0.9966	0.
43.000	44.000	45.000	0	0.9965	0	0.9977	0.
46.000	47.000	48.000	0	0.9965	0	0.9985	0.
49.000	50.000	51.000	1	1.0000	0	0.9990	0.
			N = 286		N = 286		

En Büyük Sapma Değeri : 0.0543 Kontrol Değeri : 0.080

Aritmetik Ortalaması 7.386 ve Standart Sapması 6.932 olan bu dağılım, % 95 güvenilirlikle Üssel Dağılım'dır.



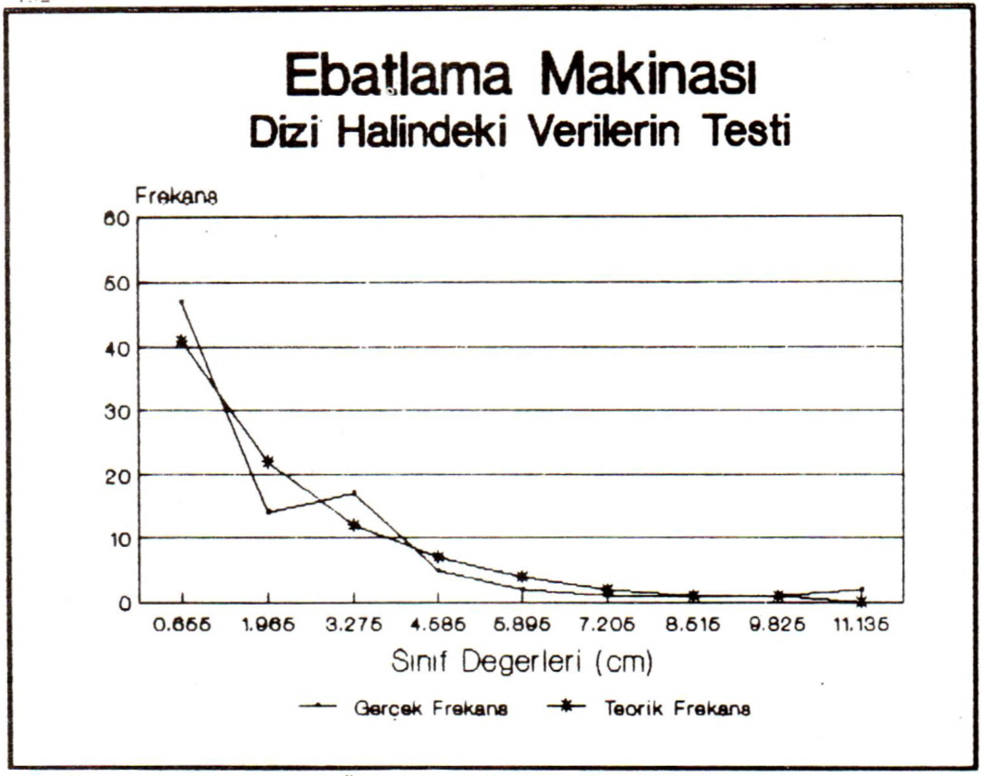
Şekil 3: Üssel Dağılımda Gerçek ve Teorik Frekans.

Çizelge 3: Dizi Halindeki Verilerin Dağılım Testi.

Sınıf Başlangıç Değeri	Sınıf Değeri	Sınıf Bitiş Değeri	Gerçek Frekans	Gerçek Birikimli Olasılık	Teorik Frekans	Teorik Birikimli Olasılık	F
0.000	0.655	1.310	47	0.5222	41	0.4509	0.
1.311	1.965	2.620	14	0.6778	22	0.6984	0.
2.621	3.275	3.930	17	0.8667	12	0.8344	0.
3.931	4.585	5.240	5	0.9222	7	0.9091	0.
5.241	5.895	6.550	2	0.9444	4	0.9501	0.
6.551	7.205	7.860	1	0.9556	2	0.9726	0.
7.861	8.515	9.170	1	0.9667	1	0.9849	0.
9.171	9.825	10.480	1	0.9778	1	0.9917	0.
10.481	11.135	11.790	2	1.0000	0	0.9955	0.
			N = 286		N = 286		

En Büyük Sapma Değeri : 0.0714 Kontrol Değeri : 0.143

Aritmetik Ortalaması 2.187 ve Standart Sapması 2.292 olan bu dağılım, % 95 güvenilirlikle Üssel Dağılım'dır.



Şekil 4: Üssel Dağılımda Gerçek ve Teorik Frekans.

6. SONUÇ VE ÖNERİLER

Yukarıda anlatılan Kolmogorov-Smirnov Testi'nin, Khi-Kare testine göre daha etkin olduğu görülmektedir. Bunun gerekçeleri;

1) Kolmogorov-Smirnov testinde sınıf büyüklüğü ile ilgili sınırlama yoktur.
2) Değerlerin toplamı yerine bireysel sınıf aralıkları ile ilgilenildiğinden Khi-Kare testinde olduğu gibi bilgi kaybı yoktur.
3) Kolmogorov-Smirnov Testi, incelenen değişkenin sürekli dağılım gösterdiği durumda kulanılır (HALAÇ 1982).

Herhangi bir üretim sisteminin uygulanabilirliğinin simülasyon dilleriyle test edilmesinde çok sayıda verinin birlikte kullanılması mümkün olmayabilir. Çünkü, simülasyon amacıyla geliştirilmiş paket programlarda bellek sorunu nedeniyle sınırlı sayıda tanımlama yapma zorunluluğu vardır.

Bu durumda, ürünleri ilişkin bir özelliğin istatistiksel dağılımı bulunup, ana kütleyi temsil etmek için bu dağılımın parametreleri kullanılabilir.

KAYNAKLAR

HALAÇ, O., 1983, İşletmelerde Simülasyon Teknikleri, İ. Ü. İşletme Fakültesi, Yayın No: 130, İstanbul.

KALIPSIZ, A., 1981, İstatistik Yöntemler, İ. Ü. Orman Fakültesi, Yayın No: 294, İstanbul.

KARAYALÇIN, İ., 1979, Yöneylem Araştırması, İTÜ, Yayın No: 1132, İstanbul.

KARAYALÇIN, İ. ve KÜLÜR, C., Matematik İstatistik, Makina Mühendisleri Odası, Yayın No: 33, İstanbul.