İSTATİSTİK TÜME VARIMDA HİPOTEZ TESTLERİ [1]

Mustafa AYTAÇ *

1- İSTATİSTİK KARAR

Herhangi bir konuda karar kelimesinin kullanılması olası iki veya daha fazla hareket durumu arasından birinin seçimini ifade eder [2]. Karar alma ise, belirlilik ve belirsizlik altında olmak üzere iki şekilde olur. Olası bütün durumlar hakkında kesin bilgiler var ise, herhangi bir hareket yöntemini seçmekle bir hata yapma olasılığı yoktur. Başka bir ifade ile belirlilik altında karara varılır. Fakat hakkında karar alınacak olaylar genellikle kesin olay olma niteliğinden uzaktır. Bu durumlarda belirsizlik altında karar almaktan söz edilir. Belirsizlik, gelecek hakkında eksik bilgi sahibi olmak şeklinde tanımlanabileceğinden belirsizlik altında karar almak, her zaman yanlış bir karar verme tehlikesini de içermektedir.

Genel olarak istatistik yöntemlerden, başka deyişle olasılık kuramından yararlanarak karar alınıyorsa, "istatistik karar"dan söz edilir. Bu geniş anlamda karardır. Çağdaş anlamda istatistik "rastlantısal olaylar karşısında iyi kararlar alınmasını sağlayan yöntemler kümesi" diye bilinir [3].

Nokta veya aralık şeklinde tahmin ile hipotez testleri yardımı ile varılan kararlara,dar anlamda karar alma ya da klasik karar alma adı verilir. Çünkü, bunlar yardımıyla birçok seçenek arasından değil, yalnızca iki seçenek arasından seçim yapmak söz konusudur.

İstatistik karar almada kullanılan yöntemlerden birisi de "Bayes Karar Alma" dır. Bayes karar alma yaklaşımı, klasik karar alma yaklaşımının bir uzantısı olarak, hipotez testleri ve tahmin problemlerinde bu yaklaşımları ile aynı teknikleri kullanmasına rağmen, bu tekniklerin arkasındaki felsefe bakımından her iki yaklaşım arasında önemli ayrıcalıklar vardır [4]. Bayes karar almanın üstünlüklerini

a) Karar alma problemlerinde iktisadi yönleri de ele alması,

b) Subjektif olasılıkları kullanabilmesi,

---

* *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Yardımcı Doçenti*

1 Bu makalede büyük ölçüde basılmamış olan "Tek Örnekli Durumlarda Parametrik olmayan İstatistik Testleri ve Bazı Uygulamalar" adındaki doktora tezinden yararlanılmıştır.

2 Kenan Ural, İstatistik ve Karar Alma, İstanbul, İktisat Fakültesi, Yayın No. 324, 1974, s. 4.

3 Merih İpek, İstatistiğe Giriş II, Ders Notları, İstanbul Üniversitesi İktisat Fakültesi, 1979, s. 160.

4 Mehmet Genceli, Karar Almada Bayes Yaklaşımı ve Bazı Uygulamalar, Basılmamış Doktora Tezi, İst. Üniv. İktisat Fakültesi, 1976, s. 95.

c) Sonuca ulaşmak için her türlü bilgiyi kullanması noktalarında toplayabiliriz. Yığın özelliklerinden herhangi birisinin (dağılımı, ortalaması, varyansı vb.) belirlenmesinde veya belirli bir konuda bir şeyin geçerliliğini olasılık esaslarına göre araştırmak için araç olarak istatistik testlerden yararlanırız. Amaç bu araştırmada ileri sürülen hipotezlerimizin geçerliliğini saptamaktır. Bu hipotezlere istatistik hipotezi adı verilir. Şunu özellikle vurgulamamız gerekir ki, istatistik hipotezi örnekten elde edilen verilerin değerlendirilmesiyle ana kütleye ilişkin olarak yapılır.

İster tam sayım, isterse örnekleme yöntemi ile olsun, toplanan verilerin anlamlılığı hakkında varacağımız karar, ileri sürdürdüğümüz hipotezin saklı tutulmasına, değiştirilmesine veya reddine yol açabilir. Tüm bu işlemleri yaparken objektifliğe ağırlık vermemiz gerekmektedir. Başka bir deyimle subjektif yargılardan kaçınmamız zorunludur. Objektifliğe ağırlık vermemizin nedeni elde ettiğimiz sonuçlara, bilimsel yöntemlerle herkesin ulaşabilmesi ve koşullar aynı kaldığı sürece sonuçların tekrarlanabilir olmasındandır.

Bunun yanında objektif kararlara en iyi şekilde varabilmek için istatistik verilerin elverişlilik, açıklık, ölçülebilirlik ve karşılaştırmalara elverişli olma özelliklerini de içermesi gerekir.

Parametrik ve/veya parametrik olmayan istatistikleri test etmek istediğimiz zaman, çoğunlukla birbirini izleyen birkaç adım vardır. Bu basamakları tek tek ele alıp çözümleyerek bir sonuca varmaya çalışırız.

2. PARAMETRİK VE/VEYA PARAMETRİK OLMAYAN İSTATİSTİK TESTLERDE İZLENECEK ADIMLAR

Herhangi bir istatistik testini kullanırken birbiri arkasından gelen belirli adımlarla sonuca varılır. Sonuçta aynı yere varmalarına rağmen bazı yazarlar bu adımları beş, kimileri de dört veya altı başlık altında yorumlamışlardır. Biz, istatistik testleri altı adımda sonuçlandırmaya çalıştık. Bunları şöyle sıralayabiliriz:

2.1. Sıfır hipotezinin (H_0) belirlenmesi,

2.2. Uygun istatistiksel testin seçimi,

2.3. Örnek sayısı n ile birinci ve ikinci tip hataların (α ve β) belirlenmesi,

2.4. Sıfır hipotezi altında örnekleme dağıtımına dayanılarak ana kütle dağılımının belirlenmesi.

2.5. Reddetme bölgesinin tanımlanması.

2.6. Karar verme.

Bu adımları sırasıyla açıklamaya çalışalım.

2.1. Sıfır Hipotezinin (H_0) Belirlenmesi

Sıfır hipotezi H_0, belli bir amaç ile reddedilmesi veya kabul edilmesi için formüle edilmiştir. Bu test reddedilirse, almaşık hipotez H_1 kabul edilir. Yani H_0 ve H_1 testlerinden birisi reddedilirse diğeri kabul edilir.

H_0 hipotezi geçerli olup olmadığı araştırılan öngörüştür. Diğer bir ifade ile H_0'ı, H_1'e karşı test ederiz.

Test edilecek hipotezlerin kurulması, araştırmanın başarı oranı ile yakından ilgilidir. Hipotezin örnek alınmadan önce kurulması gerekir. Ayrıca neyin red hipotezi olarak konacağı da çok önemlidir. Genellikle çürütmek istediğimiz hipotezi red hipotezi yaparız [5].

Hipotez testleri, tiplerine ve içerdikleri parametre sayılarına göre başlıca iki gruba ayrılabilir:

İçerdikleri parametre sayılarına göre sınıflandırmada, tek parametre ihtiva eden hipotez testleri — ki, uygulamada daha çok bu tip problemlerle karşılaşılır — ile iki veya daha fazla parametre ihtiva eden hipotez testleri diye ikiye bölünebilir [6].

Hipotez testleri, tiplerine göre de dört değişik grupta sınıflandırılabilir:

2.1.1. Basit Bir Hipotezin, Basit bir Almaşığa Karşı Testi

Gerek H_0, gerekse H_1 hipotezleri tek bir değeri test ediyorlarsa, bu tip hipotezlere basit bir hipotezin, basit bir almaşığa karşı hipotez testi denir. Formüle edersek;

$$H_0 : \mu = 10 \qquad H_1 : \mu = 15$$

2.1.2. Basit Bir Hipotezin, Karmaşık Bir Almaşığa Karşı Testi

H_0 basit bir hipotez ise, yani bir tek değeri test ediyorsa; buna karşılık H_1 hipotezi, H_0 hipotezinin test ettiği değer dışında bütün değerleri test ediyorsa — bir diğer anlatımla H_1 karmaşıksa bu sınıflandırmaya sokulabilir. Bu tip testlere tek yanlı veya iki yanlı testler adı da verilmektedir. Yani

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0 \quad (\mu > \mu_0 \text{ veya } \mu < \mu_0 \text{ olabilir.})$$

2.1.3. Karmaşık Bir Hipotezin, Basit Bir Almaşığa Karşı Testi

Uygulamada az rastlanmakla beraber, karmaşık bir hipotezin basit bir almaşığa karşı testi de olasıdır ve şöyle ifade edilir.

$$H_0 : \mu < \mu_0 \qquad H_1 : \mu = \mu_0$$

Açıklıkla görüldüğü gibi H_0'da μ_0 değerinden küçük değerler test edildiği halde, H_1'de bir tek değer μ_0 test edilmektedir.

2.1.4. Karmaşık Bir Hipotezin, Karmaşık Bir Almaşığa Karşı Testi

H_0 ve H_1'in ikisi birden karmaşık ise bu gruba girer. Bir başka anlatımla;

$$H_0 : \mu < \mu_0 \qquad H_1 : \mu \geqslant \mu_0$$

Uygulamada ise daha çok 2.1.1 ve 2.1.2 tipindeki hipotez testlerine rastlanılmaktadır.

2.2. Uygun İstatistiksel Testin Seçilmesi

İstatistik testlerin kullanım alanları, özellikle II. Dünya Savaşından sonra çok büyük gelişmeler göstermiştir. Bunun sonucu olarak da oluşturulan herhangi bir hipotez hakkında karar vermede kullanılabilecek seçenekli istatistiksel testler vardır.

---

5 Uğur Korum, Matematiksel İstatistiğe Giriş, T.O.D.A.İ.E., Ankara, 1977, s. 266.

6 Herman Chernoff ve Lincoln E. Moses, Elemantary Decision Theory, John Wiley, New York, 1959, s. 246.

O zaman akla şöyle bir soru gelebilir. "Bu seçenekli testlerden hangilerini ne gibi kriterler sonucunda seçeceğiz?"

Bu kriterlerden birincisi güç kriteridir. Bunu bir sonraki basamakta ele alıp, geniş şekilde açıklamaya çalışacağız.

İkinci kriter ise "İstatistiksel Modeldir". Yığının yapısını ve örnekleme yöntemini belirledikten sonra istatistiksel modeli kurmamız gerekir. Her istatistiksel test belli bir model ve ölçme altında geçerlidir. Bu her test için değişebilir. Bazen bir istatistik testdeki modelin gerektirdiği koşulları kontrol edebilmemize rağmen, çoğu kez bu koşulların doğruluğunu varsayarız. Bunun sonucu olarak en kapsamlı olan testler en güçlü testlerdir. Varsayımlar daha az veya daha zayıf oldukları zaman, elde ettiğimiz istatistik testin sonucu daha genel olur [7].

Herhangi bir testin kullanılışıyla varılan bütün kararlar şu koşulu da beraber taşırlar: "Eğer kullanılan model doğruysa ve ölçme gereksinimi yerine getirildiyse, o zaman......" [8].

Üçüncü kriter ise, güç etkinliğidir. (Power-Efficiency). Güç etkinliği kavramı, test T_1'i test T_2 kadar güçlendirmek için gerekli olan örnek sayısının artış miktarı ile ilgilidir.

Örneğin T_1-testi modele uygun koşulları yerine getirildiği zaman, kendi tipindeki testler içinde en kuvvetli test olarak biliniyorsa ve aynı araştırma düzeninin T_2 testi, n_2 kadar olayı (birimi) içerdiğinde, n_1 olayı kapsayan test kadar kuvvetli ise, o zaman

$$T_2\text{- Testinin Güç Etkinliği} = \%\,100 \,.\, \frac{n_1}{n_2}\text{ ' dir.}$$

Güç etkinliği kavramının ve formülünün istatistik testlerin kullanılışında çok önemli bir yeri vardır. Bu yöntemle farklı bir test seçmek ve daha büyük örnek almakla güç kaybetmeksizin en güçlü testlerin bazı varsayımlarını yerine getirme zorunluluğunu ortadan kaldırabiliriz.

Güç etkinliği ile ilgili bir diğer kavramda Bağıl Asimtotik Etkinliktir (Asymtotik Relative Efficiency). Kısaca ARE olarak gösterilen ve Pitman Etkinliği veya Asimtotik Etkinlik olarak da bilinen bu kavram dört değişik durumda kullanılabilir. Bunları şöyle sayabiliriz; güven aralıklarında yararlanılan ARE, herhangi bir parametre tahmininin ARE'si, çoklu karşılaştırma yöntemlerinin ARE'si ve testlerin ARE'si.

Çalışmamızda testlerle ilgilendiğimiz için testlerin ARE'si üzerinde kısaca duralım. T_1 ve T_2 testleri α-anlamlılık seviyesinde H_1 almaşığı ile H_0 hipotezlerinin birbirini tutan iki testi olsun. T_2 testine göre, T_1 testinin ARE'si n_2/n_1 örnek sayı-

7 Burada kısaca "tutucu test" kavramından sözedelim. Birinci tip hata işleme olasılığı (α), herhangi bir test istatistiğinin tablosu, tarafından verilen değerlerden daha küçük çıkarsa, bu tip testlere tutucu test adı verilir. Başka bir deyimle H_0 hipotezini redetme olasılığı azalacaktır. Fakat H_0 hipotezi bu testle rededilirse elde edilen sonuç çok güvenilirdir.

8 Sidney Siegel, Nonparametric Statistics for Behavioral Sciences, Mc Graw Hill, New York, 1956, s. 18.

lar oranının limit değerleridir. Yani örnek sayıları n_1 ve n_2 sonsuza giderken ($n_1 \rightarrow \infty$, $n_2 \rightarrow \infty$) ve H_1 hipotezi H_0 hipotezine yaklaşırken ($H_1 \rightarrow H_0$); T_1 testinin, T_2 testine eşit olabilmesi için T_1 testi tarafından talep edilen örnek sayısıdır [9].

Uygun bir istatistiksel testin seçilmesinde dördüncü kriter ise ölçmedir. Oluşturulan herhangi bir hipotez için bize gerekli olan o yığının belirli özellikleridir. Yani verilerin bir kısmı bizim için çok gereklidir.

Verileri kısaca sayılabilir ve ölçülebilir veri türleri diye iki kısma ayırabiliriz.

Sayılabilir veriler, gözlenen birimlerin herbirisinin belli özelliklere sahip olan bir gruba katılması veya katılmamasıdır.

Diğer veri türleri, ölçme altındaki tesadüfi değişkenin belli değerlerine göre sınıflandırılabilir. Eğer tesadüfi değişken yalnız bir boyutlu ise bir boyutlu değişken, iki boyutlu ise iki boyutlu değişken, ikiden fazla boyutlu ise çok boyutlu değişken olarak adlandırılır. Böyle durumlarda elde edilen örnek verileri de aynı şekilde sınıflandırılır. Tek boyutlu verilerde tek örneklem durumu — bu tip örneklemlerde değişken sürekli veya kesikli olabilir — söz konusudur. İki boyutlu verilerde k-örneklem durumu adlarını alırlar. İki ve k, boyutlu durumlarda değişkenler sürekli, kesikli veya hem sürekli hem de kesikli olabilirler.

Verilerle ilgili bilmek zorunda olduğumuz diğer bir kavram da ölçme düzeyidir. Çünkü sonuçların ortaya çıkmaları ölçme düzeyinin etkileri ile yakından ilgilidir. Ölçme, verilere belirli bir özelliğe sahip oluş derecelerini belirtmek için belli kurallara uyarak sembolik değerler verme işlemidir.

Ölçmeyi başlıca dört gruba ayırabiliriz. Bunlar sınıflayıcı, sıralayıcı, aralıklı ve oranlı ölçek çeşitleridir [10]. Bunları kısaca açıklamaya çalışalım.

Sınıflayıcı ölçek, sayılar, işaretler veya simgelerle verileri farklı gruplara ayırmada kullanılır. Bu ölçmede ölçekleme işleminin yapılabilmesi için belirli bir sınıfı eşdeğerlilik ilişkileri altında farklı alt sınıflara bölmek gerekir ve aynı numara veya simge birden fazla veriye ya da sınıfa verilmemelidir. Yani her alt sınıflarda bulunan birimlerin gösterdikleri özellik, ölçeğe vurulduklari zaman eşit olmalıdır. Egemen değer, kontenjans katsayısı ve diğer parametrik olmayan teknikler sınıflayıcı ölçekle ölçülmüş verilere uygulanabilir.

Sıralayıcı ölçek, ölçme sonuçlarını karşılaştırmada kullanılır. Bu ölçek, verileri veya sınıfları numaralamakla kalmayıp, onları miktarlarına göre sıralama olanağı da yaratmaktadır. Bu nedenden dolayı sıralayıcı ölçek, sınıflayıcı ölçekten gerek kullanılabilecek istatistik teknik sayısı gerekse tekniklerin güvenilirlikleri açısından daha üstündür. Sıralayıcı ölçekle ölçülmüş verilere ortalayan, sıra korelasyonu katsayısı ile çok sayıda parametrik olmayan istatistik teknikleri uygulanabilir.

---

9 Daha fazla bilgi için bakınız:
Myles Hollander and Douglas A Wolwe, Nonparametric Statistical Methods, John Wiley, New York, 1973, s. 438-440.
Jean Dickinson Gibbons, Nonparametric Statistical Inference, Mc Graw Hill 1971, s. 18 ve s. 273-294.

10 Bu ölçeklerin dışında çoğu kez psikolojik niteliklerin ölçülmesinde kullanılan ölçekler de vardır. Bunlar; Skalogram analizi, Boyutsal ayırma ölçeği, Q tekniği, Thurstone aralıklı ölçeği ve Likert toplama ölçekleridir.

Sıralayıcı bir ölçeğin bütün özelliklerine sahip bir ölçekte, planladığımız ölçme üzerinde bütün özelliklerin arasındaki uzaklıkların genişliğini bize bildirecek kadar dakikse, aralıklı ölçmeyi başarmış oluruz.Böylece bir birimin diğer bir birimden büyük veya küçük olduğunu değil, bu farklılığın ne kadar olduğunu da söyleyebiliriz. Aralıklı bir ölçek, sıralanmış bir kümede bütün olay çiftlerine bir gerçek sayı veren ortak ve sabit bir ölçme birimi ile karakterize edilir ve bu ölçme birimi (aynı zamanda sıfır noktası) keyfidir. Parametrik istatistik yöntemlerinin çoğunun uygulanabilmesi için gözlem sonuçları hiç olmazsa bir aralıklı ölçekle ölçülebilmelidir. Bunların arasında aritmetik ortalama, standart sapma,korelasyon katsayısı, t ve F testleri sayılabilir.

Oranlı ölçek ölçme düzeyi en yüksek olan ölçektir. Eğer bir ölçek aralıklı ölçeğin bütün özelliklerini taşıyorsa; sıfır noktası ve/veya sabit bir başlangıç noktasına sahipse buna oranlı ölçek adı verilir. Dikkat edilecek bir diğer nokta da oranlı ölçekte, herhangi iki ölçek noktasının birbirine oranının, ölçme biriminden bağımsız olmasıdır. Örneğin, sıfır sabit bir başlangıç noktası olduğu sürece, uzunluk ve ağırlık birer oranlı ölçeklerdir. Geometrik ve harmonik ortalamalar, değişim katsayısı, yüzde değişmeleri ve bütün istatistik testler oransal ölçekle ölçülmüş verilere uygulanabilir.

Yukarda kısaca anlatmaya çalıştığımız ölçekler arasındaki önemli bir bağıntı da şudur: Üst düzeydeki ölçekler alt ölçek cinsindende ifade edilebilir. Bu durumun tersi ise geçerli değildir. Yani oranlı ölçeği, sınıflayıcı, sıralayıcı veya aralıklı, ölçek, cinsine dönüştürebildiğimiz halde; bunlardan birisi yardımıyla oranlı ölçmeye geçme olanağımız yoktur.

2.3. Örnek Sayısı n ile Birinci ve İkinci Tip Hatanın (α ve β) Belirlenmesi

Herhangi bir araştırmada hipotezler belirlendikten ve uygun istatistiksel test seçildikten sonraki aşamada n ile α ve β'nın belirlenmesidir.

α'yı araştırmayı yapan kişinin kendisi belirler ve çoğu kez 0.01,005 ve 0,010 olarak ele alınır. Bunun nedeni ise, yapacağı araştırmada iddiasını belirli ölçüler içinde tutma isteğidir.

Belli bir hipotez testi yapıldığında H_0 hipotezinin kabul edilmesine veya reddine karar verileceği zaman dört durum ortaya çıkar.

a) H_0 hipotezi gerçekte doğrudur ve kabul edilmiştir.

b) H_0 hipotezi gerçekte doğrudur ve reddedilmiştir.

c) H_0 hipotezi gerçekte yanlıştır ve kabul edilmiştir.

d) H_0 hipotezi gerçekte yanlıştır ve reddedilmiştir.

Görüldüğü gibi (a) ve (d) deki kararlarda hiçbir hata payı yoktur. Fakat kalan iki seçenekte hata vardır. (b)'deki hataya, yani H_0 hipotezi doğru olduğu halde reddedilmesine α tipi veya birinci tip hata adı verilir. (c)'deki hataya ise β tipi veya ikinci tip hata adı verilir.

Bu hata tiplerini ve doğru karar vermeleri aşağıdaki tablo 1 yardımı ile daha net olarak görebiliriz.

Tablo 1'den de görüleceği gibi hipotez testlerinin incelenmesinde ortaya çıkan dört durumdan herbirisi koşullu olasılıklardan başka bir şey değildir. Yani,

Tablo: 1
Birinci ve İkinci Tip Hatalar
KARAR

		H_0 Kabul	H_0 Red
Hipotez	H_0 Doğru	Doğru Karar Verme Olasılığı = $1-\alpha$	Birinci Tip Hata Yapma Olasılığı = α
	H_0 Yanlış	İkinci Tip Hata Yapma olasılığı = β	Doğru Karar Verme Olasılığı = $1-\beta$

$P(H_0 \text{ kabul edilme}/H_0 \text{ doğru}) = 1-\alpha$
$P(H_0 \text{ red edilme}/H_0 \text{ doğru}) = P(\text{birinci tip hata}) = \alpha$
$P(H_0 \text{ red edilme}/H_0 \text{ yanlış}) = 1-\beta$
$P(H_0 \text{ kabul edilme}/H_0 \text{ yanlış}) = P(\text{ikinci tip hata}) = \beta$

Açıkca görüleceği gibi H_0 hipotezi doğru olduğu zaman bir istatistiksel testden alınan değerin ortaya çıkışına ilişkin olasılık α'ya eşit veya ondan küçükse H_0 hipotezini rededer, H_1 hipotezini kabul ederiz. Bu olasılık α'dan büyükse H_0 hipotezini kabul ederiz yerine, H_0 hipotezini redetmek için yeterli neden bulunamamıştır demek daha doğrudur.

Örnek sayısı veri iken α ve β tipi hatalardan sadece bir tanesini kontrol altında tutabiliriz. Her iki tip hatanın aynı anda kontrolünü ancak örnek sayısını gerekli olan seviyeye yükseltmekle yapabiliriz [11]. Bu nedenden dolayı önce α ve n belirlenir, sonra β hesaplanır. Önce α'nın saptanmasının nedeni ise, hipotez testlerinde genellikle α-hatası üzerinde durulmasıdır. α'nın küçük olmasının ve önce seçilmesinin nedenlerinden birisi de doğru olan hipotezi redetme durumuna düşmemektedir.

Önemli bir nokta da, α ve β tipi hatalar arasında ters bir orantının olduğudur. Yani bu iki tip hatadan biri azalırken veya artarken, diğeri artar veya azalır.

Örnek sayısı n ile α ve β arasındaki ilişkiyi ise şöyle gösterebiliriz [12]:

$$P\left\{\sqrt{n}\,\frac{(\mu_0-\mu_1)}{\sigma_x} - t_\alpha\right\} = 1-\alpha$$

Formüldeki μ_0 ve μ_1; $H_0: \mu = \mu_0$, $H_1: \mu = \mu_1$ hipotezinde kullanılan değerlerdir. Bu tanımdan yararlanarak n, α ve β arasındaki en iyi kombinasyonu kurmak olasıdır. Örneğin, n sabitse β'nın küçülmesi α'nın büyümesine bağlıdır. α sabitse, β'nın küçülmesi n'in büyümesine bağlıdır.

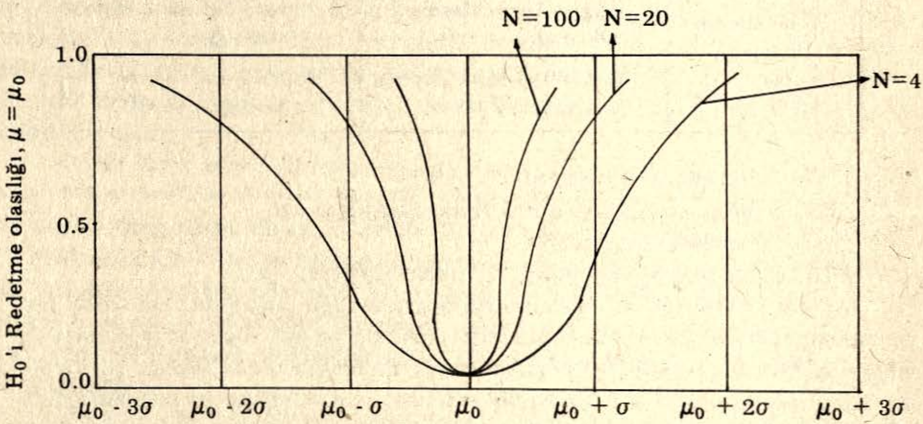
Uygun bir istatistiksel testin seçilmesinde dikkat edilecek kriterlerden birisi de bir testin gücü dediğimiz kavramdır.

Bir testin gücü H_0 hipotezi gerçekten doğru olmadığı zaman onun rededilmesi olarak tanımlanır. Yani,

$$\text{Güç} = G = 1-\beta$$

11 Uğur Korum, A.g.e. s. 70.
12 Kemal Yoğurtçugil, Örnekleme, İstanbul Üniversitesi, İktisat Fakültesi, İstanbul, 1976, s. 65-66.

n ile β arasında ters bir orantı vardır, n arttıkça ikinci tip hatayı işleme olasılığı azalmaktadır. Bunun sonucu olarak da n ile G arasında doğru orantı vardır. Bu ilişki Şekil 1'de de görülmektedir.



Kaynak: Siegel, S., A.g.e., s. 11.

Şekil: 1

Değişen Örnekleme Sayılarına Göre $\alpha = 0.05$ Önem Seviyesinde İki Yanlı Testden Elde Edilen Güç Fonksiyonları

Bir testin kuvveti, seçilen istatistiksel testin yapısına olduğu kadar H_1 almaşık hipotezin özelliklerine de bağlıdır. Eğer H_1 yön belirtiyorsa tek yanlı bir test kullanılır. Herhangi bir noktadaki testin gücü, o parametre noktasındaki güç eğrisinin değeri olarak tanımlanır.

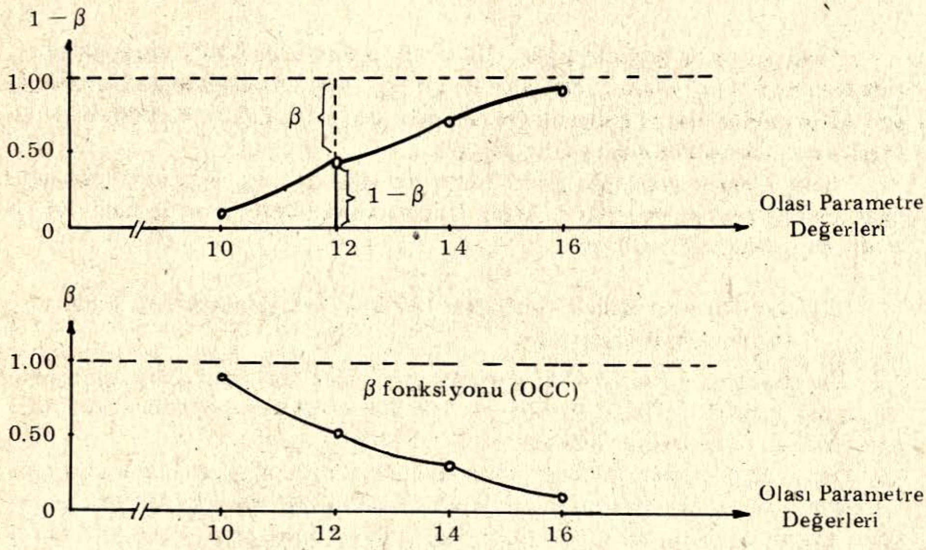
Güç fonksiyonu μ_0'da minimum bir değere sahipse, $H_0: \mu = \mu_0$ hipotez testi tarafsızdır denir ve bu μ_0 değerine göre güç fonksiyonu simetriktir [13].

Bir sonraki adıma geçmeden önce bir testin gücü ile yakından ilgili olan, "Karakteristik hareket doğrusu" (Operating Characteristic Curve = OCÇ) ve "tek düze en güçlü test" (Uniformly Most Powerful Test = UMPT) üzerinde de kısaca duralım.

Karakteristik hareket doğrusu güç eğrisinin tersidir. Diğer bir ifade ile güç eğrisinin grafiği $1 - \beta$ ise, β'nın grafiği de OCC'ü verir. Teorik olan istatistik çalışmalarda genellikle OCC üzerinde fazla durulmaz. Fakat pratik istatistiğin uygulandığı bazı durumlarda, OCC'nün kullanılması amaçlara daha uygun düşebildiği için sıkça kullanılabilir. OCC, Şekil 2'de grafikle gösterilmiştir.

Daha önce değindiğimiz gibi $1 - \beta$'ya güç eğrisi adı verilir. Akla hemen şöyle bir soru gelebilir: Güç eğrisini — yani $1 - \beta$ değerini — yükseltebilen bir karar kriteri var mıdır? Böyle bir şey yapılabilirse, bu en iyi karar verme şekli olabilir. Hemen

13 Brunk, H.D. An Introduction to Mathematical Statistics, Blaisdel Publishing Co., Second Edition, 1965, s. 257.

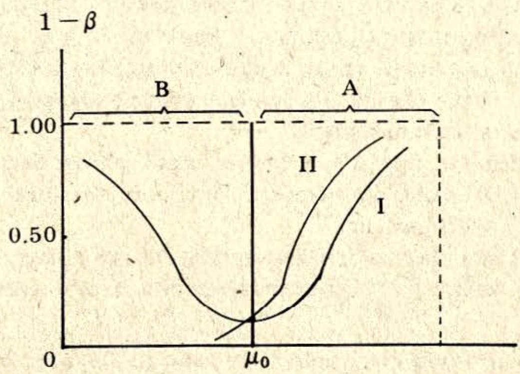


Kaynak: Wonnacat, T.H. and Wonnacat R.J., Intr. Statistics for Business and Econ., John Wiley, Toronto, 2nd Ed., 1977, s. 266.

Şekil: 2

Güç ve OCC Eğrileri

belirtmemiz gerekir ki, bunu yalnızca tek yanlı testler için yapabiliriz. Çift taraflı testler için böyle bir olanağımız yoktur [14]. Bir test diğer testler tarafından oluşturulan güç eğrilerinden daha yüksek bir güç eğrisine sahipse (yani $1-\beta$ değeri diğer testlerin oluşturduğu $1-\beta$ değerinden büyükse), bu test UMPT olarak bilinir.



Kaynak: Taro, Yamane, A.g.e., s. 268.

Şekil: 3

Güç Eğrilerinin Karşılaştırılması

---

14 Taro Yamane, Statistics and Int. Analysis, Harper International Edition, Singapore, 1973, s. 252.

Şekil 3'den de görüldüğü gibi çift taraflı testler için UMPT oluşturulamaz. Hipotezimiz $\mu = \mu_0$ olsun. Şekil 3'de de çizildiği gibi (A) bölgesinde II numaralı eğri ile gösterilen test, I numaralı testden daha güçlüdür. Fakat (B) Bölgesinde de I testi aynı nedenlerden dolayı daha güçlüdür.

Şekil 3'dende görüldüğü gibi I testin güç eğrisi $\mu = \mu_0$ değerinde minimum olup, bu değere göre simetriktir. Ayrıca I numaralı güç eğrisine sahip olan test tarafsız olup, diğer test taraflıdır.

2.4. Sıfır Hipotezi Altında Örnekleme Dağılımına Dayanılarak Ana Kütle Dağılımının Belirlenmesi

Oluşturulan hipotez testini sonuçlandırabilmek için, ana kütle dağılımını belirlemek gerekebilir. Bunu yapabilmek için ise, örnekten elde edilen sonuçlara göre birtakım testlere veya yöntemlere başvurulabilir.

Örnekleme dağılımı kuramsal bir dağılım olup, aynı yığından herbirini tesadüfi çekmek yöntemi ile aynı sayıdaki olanaklı bütün örneklerden hesaplanan bir değerin alabileceği olası bütün değerlerin oluşturduğu dağılımdır. Yığın büyüdükçe bunu yapmanın güç, hatta olanaksız olduğu açıkca görülmektedir. Bu durumlarda temeli varsayımlardan oluşan bir dizi ispatlanmış matematiksel teoremlerin güvenilirliğine dayanırız. Bunlardan en önemlileri "Chebyshev Eşitsizliği" ve "Merkezi Limit Teoremi" dir.

Merkezi limit teoreminin varsayımları daha güçlü ve normal dağılımla ilgili olduğu için pratik bakımdan çok önemlidir [15]. Bu kadar önemli olan ve sürekli yararlandığımız merkezi limit teoremini şöyle tanımlarız:

Bir tesadüfi değişken ortalama (μ) ve varyans (σ^2) değerlerine sahipse ve bu yığından sayıları n olan tesadüfi örnekler alınırsa, bu örneklerin ortalamaları X yeterince büyük örnek sayıları için ortalaması (μ) ve varyansı (σ^2/n) olmak üzere normal bir dağılıma yaklaşırlar. Örnek ortalamalarının ($\overline{X}$) dağılım fonksiyonu n sonsuza giderken standart normal dağılıma yaklaşırlar.

Merkezi limit teoremi ile yukarıda da görüldüğü gibi standart normal dağılıma geçilebilmektedir. Bu ise olasılıklarla konuşma olanağı verdiği için araştırmacılara çok büyük kolaylıklar sağlamaktadır.

Örnek ortalamaları için n'in yeterince büyük olması demek, n'in 30'a eşit veya 30'dan fazla birim alınması demektir. Yani aritmetik ortalamalar için geçerli olan örnek sayısı n = 30 olmalıdır [16].

Yığındaki dağılım normal ise, örnek ortalamalarının sahip olduğu dağılımının da normal olacağı açıktır [17]. Yığındaki dağılımın tamamen normalden farklı oldu-

15 Bu iki teoremin kullanılış alanları için bakınız: Emanuel Parzen Modern Probability Theory and Its Applications, John Wiley, New York, 1960, s. 430-433.

16 Çoğu istatistik kitaplarında örnek sayılarının 25'e eşit veya 25'den fazla birim alınmasının da yeterli olduğu ileri sürülmektedir.

17 Örnek ortalamalarının kuramsal dağılımını üç temel başlık altında toplayabiliriz:

i) Ana kütle dağılımı normal ve ana kütle varyansı belli ise örnek sayısı n'nin tüm değerleri için örnek ortalamalarının dağılımı normaldir.

ğu durumlarda $n \geqslant 30$ koşulu sağlandığı sürece, alınan örnek sayısı amacımıza ulaşmak için yeterli olabilecektir. Örneğin, yığın dağılımı $p = 1/2$ ile bir binom dağılımı olsun. Örnek yığınının büyüklüğü $n \geqslant 30$ ise, merkezi limit teoremini kullanarak örnek ortalamaları dağılımını elde etmenin maksimum hatası 0.07'dir. Ve bu hata bazı düzeltmeler yapılarak daha da düşürülebilir [18].

2.5. Redetme Bölgesinin Tanımlanması

İstatistiksel testi sonuçlandırmak için en önemli işlemlerden birisi de H_0 hipotezi altında redetme bölgesinin belirlenmesidir.

Örnekleme dağılımı, H_0 hipotezi koşulu altında bir test istatistiğinin alabileceği tüm değerleri içerir [19]. Örnek dağılım bölgesi iki temel bölgeden oluşur. Bunlar redetme ve kabul bölgeleridir ve herbirisi örnek dağılım bölgesinin bir alt bölgesidir.

Redetme bölgesi içindeki herhangi bir değerin olasılığı, anlamlılık seviyesi olan α'ya eşit veya ondan küçüktür. Yani gözlemlediğimiz örnekteki değişkenin H_0 hipotezi ve α-anlamlılık seviyesinde rededilmesi için yeterli neden gösterilememesine yol açan değerlerin tümüne kabul bölgesi; rededilmesine yol açan değerlerin tümüne red bölgesi veya kritik bölge adı verilir. Başka bir anlatımla, örnek değerinin kabul aralığına düşme olasılığı $1 - \alpha$, dışına ya da redetme bölgesine düşme olasılığı α olarak tanımlanır. α'ya göre belirlenen kabul alanının uzunluğuna güven aralığı, bunun uç değerlerine de kritik değerler denir.

Redetme bölgesi, H_1 hipotezine bağlı olarak örnekleme dağılımının ya tamamen bir tarafında (tek yanlı testler için) ya da iki tarafında bulunabilir (iki yanlı testler için). Yani tek ve iki yanlı testlerde gösterilen redetme bölgeleri yer bakımından birbirinden farklılıklar gösterirler. Fakat Şekil 4'de de gösterildiği gibi, örnekleme dağılımları normal bir yapıya sahipse; tek ve iki yanlı testlerin alan genişlikleri birbirine eşittir.

Şekil 4'den de görüldüğü gibi redetme bölgesinin genişliği anlamlılık seviyesi ile gösterilmektedir. Ayrıca test yönü daha önceden tahmin edildiği taktirde, tek yanlı testler iki yanlı testlere göre daha güçlüdürler.

---

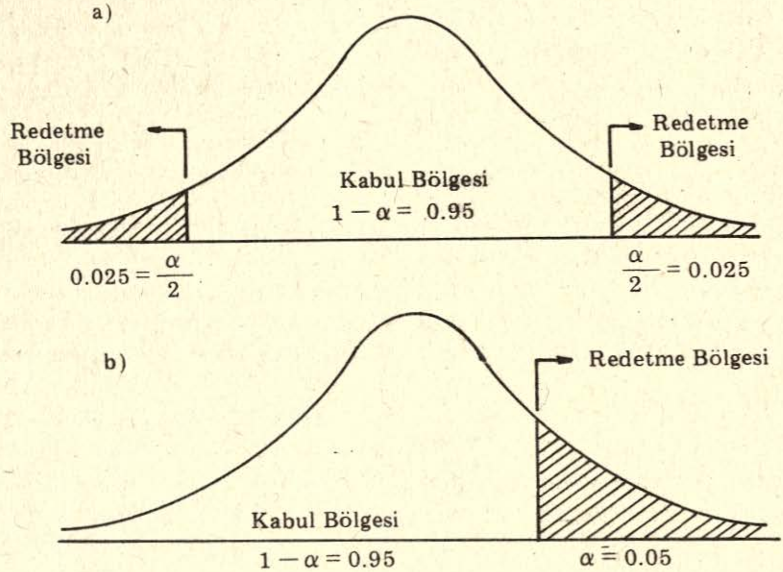
ii) Ana kütle dağılımı normal fakat ana kütle varyansı belli değilse, $n < 30$ değerleri için $n - 1$ serbestlik dereceli student-t dağılımı verir. $n > 30$ ise dağılım normale yaklaşır ve bu yaklaşım n; büyüdükçe artar.

iii) Ana kütle dağılımı normal değilse, az asimetrik ve basıklığı 3'den az fark etmek koşuluyla dağılımın n'in büyüyen değerleri için normale yaklaşımı kabul edilir.

Bunların yanında ana kütle normal değilse, varyans bilinmiyorsa ve n de yeteri kadar büyük değilse, ana kütle hakkındaki bilgiler ancak parametrik olmayan istatistik yöntemlerin kullanılması ile sağlanır.

18 Korum, A.g.e. s. 187.

19 Test istatistiği, herhangi bir hipotez testinin redetme bölgesini belirleyen bir istatistiktir.



Şekil: 4
α = 0.05'e Göre İki Yanlı (a) ve Tek Yanlı (b) Hipotez Testlerinde Redetme ve Kabul Bölgeleri

2.6. Karar Verme

İstatistiksel testi sonuçlandırmanın en son aşaması daha önceki bilgilerin ışığında sonuca ulaşmaktır.

Yapılan bir hipotezin doğruluğu veya yanlışlığı ilgili ana kütle tamamıyla incelenmeden bilinemez. Bu da çoğunlukla olası değildir. Örnekten elde edilecek değerlere dayanarak hipotezin reddine veya kabul edilmesine karar verilir. Bir istatistik hipotezinin kabul edilmesi, reddi için yeterli neden olmadığını ve aksine red edilmesi de kabulu için yeterli neden bulunmadığını gösterir.

Karar verme aşamasında sonuç olarak aynı yargıya götüren birbirinden farklı iki yöntemle bunu sağlayabiliriz.

Bunlardan birincisi olasılık olarak karara varmaktır. İstatistiksel testin gözlenen bir değerine ilişkin hata payı, α-önem seviyesine eşit veya ondan büyükse, H_0 hipotezin rededilmesi kararına varırız. Gözlenen böyle bir değer "önemli fark" diye adlandırılır. Böyle bir durumda H_0 hipotezi gerçekten doğru ise, hata yapma olasılığımız α dır.

İkinci yöntem ise, örneklerden gidilerek hesaplanmış bir test istatistiği değeri t_{hes} ile ($n \geqslant 30$ ise Z-testi; $n < 30$ ise t-testi ile bulunur). α-anlamlılık seviyesinde tablodan bulunmuş (uygun olduğu örnekleme dağılımı ile ilgili tablodan) bir tablo değerinin ($t_{tab,\,\alpha}$) karşılaştırılması ile karar alma yoludur.

İki yanlı bir test için kabul bölgesi ($-t_{tab,\alpha}$.........., $t_{tab,\alpha}$) tek yanlı bir test için H_1 hipotezine bağlı olarak ya ($-\infty$ ...., $-t_{tab,\,\alpha}$) veya ($t_{tab,\alpha}$......, $+\infty$)

şeklindedir. Yığından seçilen örnek değeri t_{hes}. (H_1'e bağlı olarak) bu aralıklar arasına düşüyorsa H_0 hipotezini kabul; bu aralıklar arasına düşmüyorsa, α- önem seviyesinde H_0 hipotezini reddederiz.

Örneğin $H_0: \mu = \mu_0$ hipotezini α- anlamlılık seviyesinde aşağıdaki gibi de inceleyebiliriz [20].

$H_1: \mu \neq \mu_0$ olarak belirlenmişse, $t_{hes} < t_{tab,\alpha}$ veya $t_{hes} > t_{tab,\alpha}$ yani, $|t_{hes}| > |t_{tab,\alpha}|$ ise H_0 hipotezini reddederiz.

$H_1: \mu > \mu_0$ ve $t_{hes} > t_{tab,\alpha}$ ise H_0 hipotezini reddederiz.

$H_1: \mu < \mu_0$ ve $t_{hes} < t_{tab,\alpha}$ ise H_0 hipotezini reddederiz.

Buraya kadar açıklamaya çalıştığımız istatistik tümevarım da hipotez testleri adımlarını bir problemde uygulamaya çalışalım.

1981 yılının Ekim ayına kadar B. Ü. İktisadi ve Sosyal Bilimler Fakültesinden mezun olan 333 öğrencinin bölümlere göre dağılımı Tablo 2'de verilmiştir.

Tablo: 2

Mezun Olan Öğrencilerin Bölümlere Göre Dağılımı

Bölümler		
İktisat	40	12.01
İşletme	209	62.76
Sos. Siy. ve Koop.	33	9.91
Siyasal Bil.	51	15.32
TOPLAM	333	100.00

1981 Ekim sınavı döneminde mezun olan öğrencilerden ilk 50'sinin bölümlere göre dağılımı ise Tablo 3'de gösterilmiştir.

Tablo: 3

Ekim 1981 de Mezun Olan İlk 50 Öğrencinin Bölümlere Göre Dağılımı

Bölümler	Mezun olan öğrenci sayısı
İktisat	4
İşletme	30
Sos.Siy. ve Koop.	6
Siyasal Bilimler	10
TOPLAM	50

20 Zeki Avralıoğlu, İstatistik, Ankara, 1976, s. 177-179.

2.1 Hipotezimiz;

H_0: 1981 Ekim ayına kadar mezun olan öğrencilerle Ekim ayı sınavlarında mezun olan öğrencilerin bölümlere göre oranlarında bir fark yoktur.

H_1: Bir fark vardır.

2.2. İki veri grubu arasındaki uyumluluğu ölçeceğimiz veya belirleyeceğimiz için Ki-Kare uygunluk testinden yararlanabiliriz. Bu dağılım için tablo değerleri herhangi bir istatistik kitabı sonunda serbestlik derecesi otuzdan küçük olmak koşulu ile (s.d = 30) bulunabilir.

2.3. Örnek sayımız 50 tanedir. Kurmuş olduğumuz H_0 hipotezini $\alpha = 0.05$ anlamlılık seviyesinde test etmeye çalışalım.

2.4. Daha önceki tablolardan elde ettiğimiz değerleri aynı örnek büyüklüğünden Tablo 4'de birleştirerek Ki-Kare dağılımının test istatistiğini hesaplamaya çalışalım.

Tablo: 4

Bölümler	1981 Ekim Ayına Kadar Mez.Ol.= G_i	1981 Ekim Döneminde Mez. Ol.= B_i
İktisat	6	4
İşletme	31	30
Sos. Siy. ve Koop.	5	6
Siyasal Bil.	8	10
TOPLAM	50	50

Test istatistiğimiz Ki-Kare değerini şu şekilde buluruz:

$$X^2 = \sum_{i=1}^{4} \frac{(G_i - B_i)^2}{B_i} = \frac{(6-4)^2}{4} + \frac{(31-30)^2}{30} + \frac{(5-6)^2}{6} + \frac{(8-10)^2}{10} = \frac{96}{60}$$
$$= 1.600$$

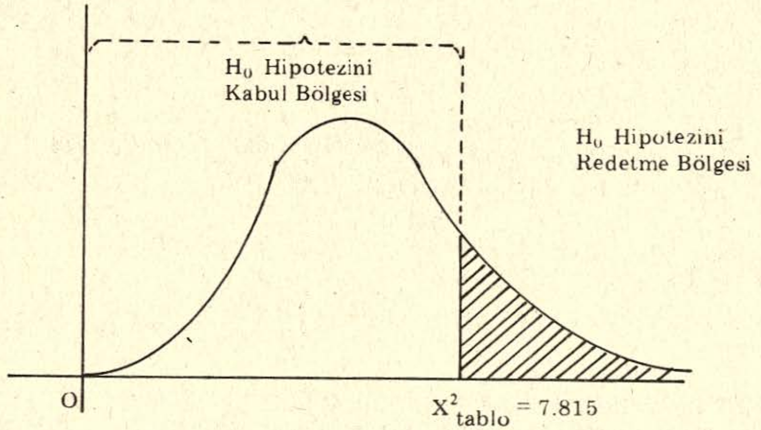
2.5. Redetme bölgemizi ise şöyle oluşturabiliriz.

s.d = k = 4 − 1 = 3. Serbestlik derecesi (s.d) birden büyük olduğu için Yates'in süreklilik düzeltmesini, test istatistiği değerini bulurken dikkate almadık.

$\alpha = 0.05$ ve s.d = 3'e göre Ki-Kare tablosundan bulduğumuz tablo kritik değeri 7.815'dir. Bu verilere göre redetme ve kabul bölgelerimizi Şekil 5 yardımı ile daha kolayca görebilir ve bir karara varabiliriz.

2.6. Karar Verme:

Hesapladığımız Ki-Kare değeri kabul bölgesine düştüğü için − bir diğer ifade ile $X^2_{hes} < X^2_{tablo}$ olduğunda − H_0 hipotezini kabul ederiz. Yani, bu dönemlerde mezun olan öğrencilerin bölümlere göre oranlarında bir fark olmadığını redetmek için yeterli nedenimiz yoktur.



Şekil: 5

FAYDALANILAN KAYNAKLAR

Avralıoğlu, Z.: İstatistik, Ankara, 1976.

Brunk, H.D.: An Introduction to Mathematical Statistics, Blaisdell Publishing Co., Second Edition, 1965.

Chernoff, Herman and Moses, Lincoln, E.: Elemantary Decision Theory Wiley, New York, 1959.

Genceli, M.: Karar Almada Bayes Yaklaşımı ve Bazı Uygulamalar, Basılmamış Doktora Tezi, İstanbul Üniversitesi İktisat Fakültesi; 1976.

Gibbons, J.D.: Nonparametric Statistical Inference, Mc Graw Hill, 1971.

Hollander, Myles and Wolfe, Douglas, A.: Nonparametric Statistical Methods, John Wiley, New York, 1973.

İpek, M.: İstatistiğe Giriş II, Ders Notları, İst. Üniversitesi İktisat Fakültesi, 1979.

Korum, U.: Matematiksel İstatistiğe Giriş, T.O.D.A.I.E., Ankara, 1977.

Parzen, E.: Modern Probability Theory and Its Applications, John Wiley, New York 1960.

Siegel, S.: Nonparametric Statistics for Behavioral Sciences, Mc Graw Hill, New York, 1956.

Ural, K.: İstatistik ve Karar Alma, İstanbul Üniversitesi İktisat Fakültesi, İstanbul, 1973.

Yamane, T.: Statistics and Introductory Analysis, Harper International Edition, Singapore, 1973.

Yoğurtçugil, M. Kemal: Örnekleme. İstanbul Üniversitesi İktisat Fakültesi, İstanbul, 1978.