

KESTİRİSEL İSTATİSTİK

Arş. Gör. Şener BÜYÜKÖZTÜRK*

Giriş

Bilimsel araştırmalarda yöntem ve özellikle veri çözümü konusu üzerinde sürekli tartışılan önemli sorunlardan biri olagelmıştır. Veri çözümü öz olarak, verileri uygun istatistik tekniklerle işleyerek verilerden anlamlı sonuçlar çıkarma sürecini anlatır.

Veri çözümü süreci doğal olarak araştırma ile toplanan verilerin sınıflandırılması süreci ile başlar. Bunu, verilere uygun çeşitli istatistik yöntemlerin uygulanması izler. Verilerin istatistiksel bakımdan çözümlenmesinde iki temel yaklaşım vardır. Birincisi, evrenin parametrelerine doğrudan ulaşılabilirdiği ya da parametrelerin hesaplanabilirdiği "Betimsel İstatistik", ikincisi araştırma evrenine doğrudan ulaşamadığı durumlarda evren parametrelerinin evrenden yansız olarak çekilen örneklemelerden hesaplanan istatistiklere dayanılarak kestirilebildiği "Kestirisel İstatistik"dir.

Araştırmacılar, genelde, istatistiksel çözümlemede evrene ulaşmanın zorluğundan dolayı kestisel istatistik yaklaşımını izlerler. Bu çalışmada kestirisel istatistik istatistiksel kestirim, null hipotezi, karşı hipotez, istatistiksel manidarlık testi, manidarlık testlerinin yorumu, manidarlık testi türleri, istatistiksel güç çözümü ve istatistiksel çözümleme ünitesi konuları ile sınırlandırılarak tartışılmıştır.

İstatistiksel Kestirim

İstatistiksel kestirim, kestirisel istatistiğin bir konusudur. Esasen, kestirisel istatistik, "evren hakkında ne biliyorum?" sorusunu cevaplamaya çalışır. Bunu cevaplarken de görüldüğü gibi, yani özelden genele doğru bir mantık (tümevarım) yürütür. Araştırmacı, örneklemden hesapladığı bir istatistikten yararlanarak parametreyi kestirmeye çalışabileceği gibi, bulduğu bir istatistiksel farkın ya da ilişkinin evrende gerçekte olup olmadığını da öğrenmek isteyebilir. Birinci yöntem istatistiksel kestirim (tahmin), ikinci yöntem ise hipotez

* A.Ü. Eğitim Bilimleri Fakültesi Eğitim Yönetimi ve Planlaması Bölümü.

testi (manidarlık testi) olarak isimlendirilir (Glass ve Stanley, 1970, s. 240).

Araştırmacı, evren üzerinde çalışmanın mümkün olmadığı durumlarda, parametre hakkında bilgilenmek ve evrene ilişkin genellemeye varmak istediğinde öncelikle evrenden yansız olarak çektiği örneklemden ilgili istatistikleri hesaplar. İstatistikler, gerçekte şans faktörü dışta tutulduğunda parametrelerle tam olarak uyuşmadığından araştırmacı, söz konusu şans faktörü için bir tolerans miktarı belirler. Tolerans, daima bir risk unsurunu ima eder. Dolayısıyla de evrene yönelik genellenenin olasılığa dayalı olması gerekliliği ortaya çıkar. Buna göre genellemede, yani istatistiksel kestirime dayalı kararlarda belli bir yüzde hata olasılığı olacaktır. İstatistiksel kestirim, araştırmacının şanstın kaynaklanan farkları, ilişkiyi gerçeklerden ayırmasını sağlar. Şanstın kaynaklanan farklar, kontrol edilemeyen bazı değişkenlerin etkisiyle oluşur (Mouly, 1963, s. 145).

Bu bağlamda istatistiksel kestirimin amacının örneklemin karakterlerinden evrenin karakterlerini tanımak, kestirmek ya da ilgilenilen değişkenler bakımından örnekleme gözlenen değerler arasındaki ilişkiye dayanarak evrendeki ilişki hakkında kestirim yapılmasını sağlamak olduğu söylenebilir (Williamson ve diğerleri, 1977, s. 437). Johnson (1980, s. 293) *kestirimi*, bir evren parametresinin değeri ile ilgili olarak sorulan bir sorunun cevaplandırılmasında kullanılan işlemler olarak tanımlamaktadır. Kestirim iki türlü olup, bunlar nokta ve aralık kestirimleridir.

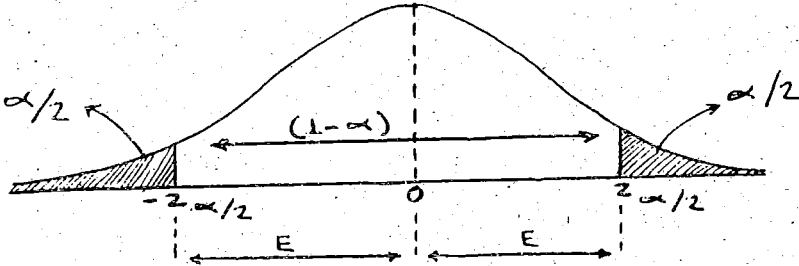
Nokta kestirimi, bir evren parametresine karşılık gelen istatistiktir. Örneğin, \bar{X} , ortalaması μ olan bir evrenden yansız olarak seçilen bir örneklemin ortalaması ise, μ 'nün nokta kestirimi $\mu = \bar{X}$ 'dir. *Aralık kestirimi*, bir evren parametresinin değerinin kestiriminde kullanılan ve örneklemden hesaplanan istatistiğe dayalı olarak bulunan ve iki değerle sınırlandırılan bir aralıktır. Örneğin, \bar{X} , ortalaması μ , varyansı σ^2 olan bir normal dağılıma sahip evrenden yansız olarak seçilen bir örneklem ortalaması ise evren ortalaması μ 'nün aralık kestirimi $(1-\alpha)$ güven düzeyi için $\bar{X} \mp z_{\alpha/2} \times (\sigma / \sqrt{N})$ 'dir (Johnson, 1980, s. 294; Tanis, 1987, s. 258). Başka bir anlatımla aralık kestirimi, istatistiğin değer bölgesine ait bir aralığı oluşturma işlemidir. Burada aralığın parametreyi içerme olasılığı da açıklanır (Özden, 1981, s. 210).

Aralık kestirimiyle saptanan güven aralığı için belirlenen güven düzeyi $(1-\alpha)$, seçilmiş örneklemin sunduğu sınır değerlerin olasılığını verir. Sınır değerleri, kestirilen parametrenin zıt taraflarından uzanır.

Güven düzeyi bazen güven katsayısı olarak da isimlendirilir. Bu kapsamda güven aralığı, belirlenmiş bir güven düzeyi ile ilgili olarak bir aralık kestirimini gösterir (Johnson, 1980, s. 294).

Evren parametrelerinin aralık kestirimlerinde izlenen yöntem benzer olmakla birlikte kullanılan formüller değişmektedir. Sonuçta, araştırmacı, istatistikle aynı olan parametreye ulaşmayı beklemez, O, sadece belli olasılıklarla parametreyi içeren güven aralığının oluşturulmasıyla ilgilenir (Meyer, 1962, s. 311). Nokta kestiriminin yapılmasında ise, kestirimin standart hatasının da birlikte verilmesi zorunluluğu unutulmamalıdır. Çünkü, tek örneklemden bulunacak bir değer, ilgili istatistiğin alabileceği değerlerden yalnızca birini gösterir. Bunun için bu yaklaşım çoğunlukla kullanılmaz (Özden, 1981, s. 209).

Kestirimin maksimum hatasını gösteren E, güven aralığının yarısına eşittir. Sapma miktarı olarak da isimlendirilen bu değer, standart hatanın katsayısı olarak da görülmektedir. $E = z_{\alpha/2} \times (\sigma / \sqrt{N})$ formülasyonu kullanılır. Hata, örneklem büyüklüğü ile ters, varyans ile doğru orantılıdır (Johnson, 1980, s. 279).



Şekil 1. Güven Aralığı ve Maksimum Hata (Johnson, 1980, s. 297)

Örnek: Bir okul yöneticisi okulunda fen bilimleri öğretiminde uygulanan mevcut A1 yöntemi yerine, daha başarılı sonuçlar beklediği A2 yöntemini uygulamayı planlamaktadır. Bu bağlamda yönetici, A2 yönteminin A1 yöntemine olan üstünlüğünü test etmek ister. Bunun için yansızlık kuralına göre seçilen 40 kişilik deney grubuna A2, 40 kişilik kontrol grubuna da A1 yönteminin uygulanmasına karar verir. Uygulamaya geçilmeden önce geliştirilen standart bir ön test her iki gruba verilir ve alınan cevaplar değerlendirildiğinde iki grubun aritmetik ortalama ve standart sapma değerlerinin birbirine çok yakın olduğu görülür. Bu durum, örneklemelerin yansız ve ait oldukları evreni

temsil ettikleri şekilde algılanır. Aynı öğretim programı, aynı koşullarda her iki gruba farklı yöntemlerle uygulanır. Uygulama sonunda daha önce ön test olarak uygulanan aynı test her iki gruba da aynı koşullarda verilir. Son teste verilen cevaplardan bulunan bazı istatistikler şunlardır.

$$\bar{X}_{A2} = 75, \bar{X}_{A1} = 70, S_{A2} = 5, S_{A1} = 7.$$

Örnek dikkate alındığında, A2 öğretim yöntemi örneklemin çekildiği evrenin tümünde uygulansaydı evren parametresi μ 'nin .95 güven aralığı (GA .95) ne olurdu? Başka bir deyişle μ , .95 güvenle hangi aralığa düşerdi? Ya da μ 'nün alt ve üst sınırları ne olurdu?

(1- α) güven düzeyi için $GA_{.95} = \bar{X} \mp z_{\alpha/2} (\sigma / \sqrt{N})$ 'idi.

$$GA_{.95} = 75 \mp 1.96 \times 0.79 = 75 \mp 1.5484$$

Üst sınır: 76.5484, Alt sınır: 73.4516 olur.

$$76.5484 \quad | \text{-----} | \quad 73.4516$$

Yani, evren parametresi μ , % 95 güvenle (yüz kezde doksanbeş kez) 73.4516'dan küçük, 76.5484'den büyük olamazken % 5 hatayla (yüz kezde beş kez) bu aralığın sınırları dışında kalırdı.

Null Hipotezi

Daha önce vurgulandığı üzere kestirisel istatistiğin diğer bir konusu da hipotez (manidarlık) testidir. İleride tartışılacağı gibi hipotez testinde kurulan hipotezlerden istatistiksel olarak test edileni null hipotezidir.

Null hipotezi, iki ya da daha çok örneklem istatistiği arasında gerçek bir ilişki ya da farkın olmadığını ifade eder. Bazen de sadece bir istatistiğin ona ilişkin parametreden farklı olmadığını gösterecek şekilde ifade edilir (Meyer, 1962, s. 309). Başka bir anlatımla H_0 sembolü ile gösterilen null hipotezi farksızlığı ya da ilişkisizliği açıklar. Buna göre, null hipotezi çoğu kez geçerliği test edilen bir evren parametresi hakkında kurulan ve araştırmacının önerilen karşıt hipotez lehinde reddetmeyi umduğu hipotezdir (Tanis, 1987, s. 305).

Daha önce belirtilen örnek dikkate alındığında null hipotezi, "fen bilimleri öğretiminde A2 ve A1 öğretim yöntemlerinin başarıları arasında bir fark yoktur" şeklinde ifade edilebilir.

Karşıt Hipotez

Karşıt hipotez, çoğu kez araştırmacının doğrulanmasını umduğu bir gerçektir. H_a ya da H_1 sembolü ile gösterilir ve null hipotezinin reddi durumunda kabul edilecek olan hipotezdir (Tanis, 1987, s. 305).

Karşıt hipotez, araştırma hipotezi olarak da isimlendirilmektedir ve null hipotezinde kullanılan aynı evren parametresine ilişkin bir öneridir. Karşıt hipotez, örneklemeden elde edilen iki istatistik arasında bir farkın ya da ilgilenilen değişkenler arasında bir ilişkinin olduğunu gösterir. Başka bir anlatımla karşıt hipotez, null hipotezinde verilen değerden farklı bir değere sahip olan evren parametresine ilişkin bir öneridir (Johnson, 1980, s. 274).

Örneğimizde karşıt hipotez, “fen bilimleri öğretiminde A2 öğretim yöntemi A_1 öğretim yönteminden daha başarılı sonuçlar verir” şeklinde ifade edilebilir.

İstatistiksel Manidarlık Testi

Araştırmada, evrenden yansız olarak çekilen örneklemeler arasında ilgilenilen değişkenlere göre bulunan bir ilişkinin ya da istatistiksel bir farkın, örnekleme ve ölçme hatasından kaynaklanıp kaynaklanmadığına ya da belli bir güvenle deneysel koşulların sonucu olup olmadığına karar verilmesi gereklidir. Bu tür kestirimler yapmayı içeren tekniklere *İstatistiksel Manidarlık Testleri* denir. Bu amaçla yapılan işlemlere de *Manidarlık Testi* adı verilir. Testler birbirleriyle ilişkili ya da ilişkisiz örneklemelerden elde edilen istatistikler için yapılır (Peaker, 1988, s. 765). Manidarlık testi, örneklemede gözlenen bir farkın gerçekte, evrende olup olmadığına, yani farkın şans dalgalanmalarından kaynaklanıp kaynaklanmadığına karar vermeyi kapsar (Fox, 1969, s. 57).

Manidarlık testinde gerçekte test edilen karşıt hipotez olmakla birlikte uygulamada null hipotezidir. Null hipotezi karşıt hipoteze karşı test edilir. İstatistiksel manidarlık testi sonuçta, null hipotezin doğru olmasının olasılığını bulmamıza yardım eder. Manidarlık testi, null hipotezinin belli olasılıklara göre reddedilebilir olup olmadığına karar vermede kullanılır (Borg ve Gall, 1983, s. 372).

Örneğimiz düşünülüğünde iki öğretim yönteminin ortalama puanları arasında gözlenen 5 puanlık bir farkın evrende olup olmadığı araştırılabilir. Gerçekte böyle bir fark olmazsa bile, evrenden aynı genişlikte alınan birçok örneklemeler için belli olasılıklarla böyle bir fark

bulunabilir. Araştırmacının ilgi odağı null hipotezinin doğru olduğunda onu reddedebilmenin olasılığının ne olduğudur.

Test işlemi null hipotezinin doğru olduğu sayılışına dayanır. Dolayısıyla, null hipotezin doğruluğunun kanıtlanmasına çalışılır. Bunun için veri toplanır ve kanıt olarak uygun bir test istatistiği ve onun örnekleme dağılımının bilinmesi gerekliliği vardır. (Gamgam, 1989, s. 13-14).

Herhangi bir istatistiksel hipotezin testinde araştırmacının kararı, kesinlikle hipotezin doğru ya da yanlış olduğu şeklinde oluşturulamaz. Araştırmacı daima doğru olmayan bir kararın riskini, olasılığını bilir ve onu dikkate alır. (Glass ve Stonley, 1970, s. 275). Buna göre manidarlık testi işleminde null hipotezinin doğru olduğu sayılış altında hareket edilir ve işlemin sonunda iki olası karardan – “null hipotezi reddedilecek ya da reddedilemeyecek” – biri kabul edilir (Johnson, 1980, s. 275).

Manidarlık testi (hipotez testi), birbirine bağlı çeşitli işlem adımlarından oluşur. Bu adımlar değişik şekillerde sınıflandırılmakla birlikte özde tümünün aynı işlemleri kullandığı görülür. Bu çalışmada Johnson'un (1980, s. 273) beş adımda topladığı sınıflandırması esas alınmıştır. Ancak beş adımın ilk ikisini oluşturan null hipotezi ve karşıt hipotezin belirlenmesi adımları, hipotezlerin birlikte formüle edilmelerinin uygunluğu göz önünde tutularak tek adımda toplanmıştır. Böylece manidarlık testinin işlem adımları dörde indirgenmiştir.

1. Adım : Null Hipotezi ve Karşıt Hipotezin Belirlenmesi

Araştırmacının araştırmasında desteklenmesini beklediği araştırma hipotezine dayanılarak null hipotezi ve karşıt hipotez işlemsel olarak tanımlanmalıdır. Bu tanımlama, araştırmacının kurduğu hipotezin yapısına bağlı olarak bir ya da iki yönlü şekilde ifade ve test edilebilir. Araştırma hipotezi yönlü ise bir yönlü, yönsüz ise iki yönlü manidarlık testi kullanılır (Williamson ve diğerleri, 1977, s. 439).

Başka bir anlatımla araştırmacının sorunun çözümünün yönü hakkında bir beklentisi yok ise, yani onun için bir farkın ya da bir ilişkinin olup olmaması önemliyse *İki Yönlü Test* uygulanır. Buna karşılık araştırma hipotezinde bir yön belirtiliyorsa *Bir Yönlü Test* uygulanır. Görüldüğü gibi iki yönlü testte örneklemeden hesaplanan istatistiklerin gerçekte birbirlerinden farklı olup olmadıkları; bir yönlü testte farkın

kestirilen yönü, yani istatistiklerden birinin diğerinden büyük ya da küçük olup olmadığı dikkate alınır ve test edilir. (Fox, 1969, s. 253- 256).

Araştırma hipotezinin kuruluş yapısı, H_0 ve H_1 'in formüle edilmelerini etkiler. Literatürde hipotezlerin formüle ediliş şeklinde iki yönlü testler için bir ortaklık (uyuşma) sağlanmakla birlikte bir yönlü testlerde iki farklı yaklaşım görülmektedir.

İki yönlü testte örneğimiz dikkate alındığında H_0 ve H_1 şöyle formüle edilebilir.

$H_0: \mu_{A2} = \mu_{A1}$, $H_1: \mu_{A2} \neq \mu_{A2}$ ' dir (iki öğretim yönteminin ortalama puanlarının yalnızca birbirinden farklı olup olmadıkları araştırılırsa).

Glass ve Stanley (1970, s. 289), Johnson (1980, s. 282), Kerlinger (1973, s. 202), Roscoe (1975, s. 176) ve Tanis (1987, s. 321) gibi bazı yazarlar H_0 'ın formülasyonunun bir yönlü testte de değişmeyeceği, yani iki yönlü testteki formülasyonun aynısı ile yazılması gerektiği görüşündedirler. Buna göre, $H_0: \mu_{A2} = \mu_{A1}$ yazılışı bir yönlü test için de geçerli kabul edilir. Karşıt hipotez ise yöne bağlı olarak $H_1: \mu_{A2} > \mu_{A1}$ (örneğimiz için uygun yazılım) ya da $H_1: \mu_{A2} < \mu_{A1}$ şeklinde ifade edilir. H_0 'ın formülasyonunda kimi yazarlara, örneğin Ferguson'a (1976, s. 163) göre, H_1 'in yönü esas alınır. Yani H_0 , H_1 'in yönüne göre ve onun tersi olacak şekilde oluşturulmalıdır. Örneğin, $H_0: \mu_{A2} \leq \mu_{A1}$ $H_1: \mu_{A2} > \mu_{A1}$ ya da $H_0: \mu_{A2} \geq \mu_{A1}$ $H_1: \mu_{A2} < \mu_{A1}$ şeklinde formüle edilmelidir.

Genelde kabul gören anlayışa göre null hipotezi, bir evren değerini sadece bir değerini tanımlar ya da istatistiklerin aynı olduğunu açıklar ve kurulurken de eşitlik işaretinin kesinlikle olması gereklidir. Başka bir anlatımla null hipotezi iki istatistiğin aynı olduğunu, eşitliğini ya da sadece belli bir evren parametresinin değerine eşit olduğunu açıklar ($H_0: \mu_1 = \mu_2$; μ_0 kestirilen ya da olduğu kabul edilen değer, $H_0: \mu = \mu_0$ gibi). Buna karşılık karşıt hipotez, üç değişik şekilde formüle edilebilmektedir. Birincisinde iki istatistiğin farklı ya da bir istatistiğin evrenin belli bir değerine eşit değil ($\mu_2 \neq \mu_1$, $\mu \neq \mu_0$), ikincisi ve üçüncüsünde ise birinin diğerinden ya da belli bir değerden büyük ya da küçük olduğu ($\mu_2 > \mu_1$, $\mu_2 < \mu_1$; $\mu > \mu_0$, $\mu < \mu_0$) şeklindedir (Johnson, 1980, s. 287).

2. Adım: Test Ölçütünün Belirlenmesi

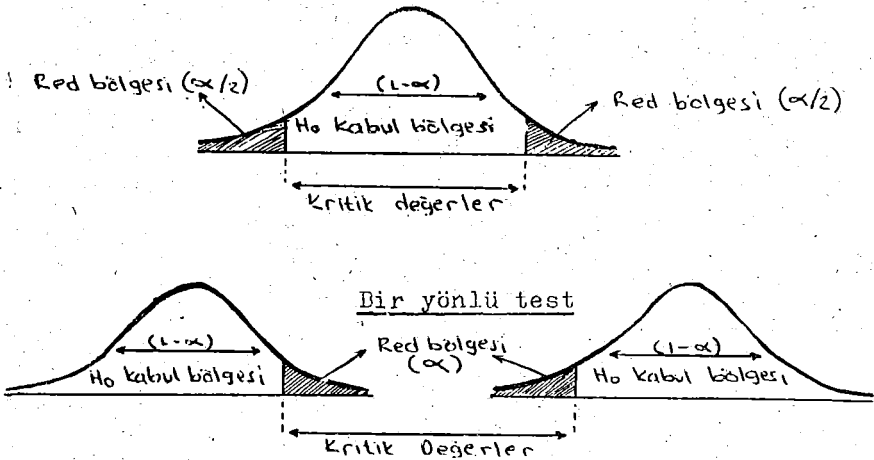
Üç test ölçütü vardır: Test istatistiği, manidarlık düzeyi (α), kritik bölge ve kritik değer.

Test istatistiği, null hipotezin reddi ya da reddedilemeyeceğine ilişkin karar vermede kullanılacak olan yansız bir değişkene ait değerdir (Johnson, 1980, s. 276). Başka bir anlatımla test istatistiği, null hipotezi ya da karşıt hipotez lehinde karar vermek için kullanılan ve gözlemlere dayandırılan bir istatistiktir (Tanis, 1987, s. 306). Test istatistiği, Manidarlık Testi Türleri başlığı altında daha geniş bir şekilde tartışılmıştır.

Manidarlık düzeyi, α sembolü ile gösterilir ve I. tür hatayı yapmanın olasılığını verir (Johnson, 1980, s. 277; Tanis, 1987, s. 306).

Kritik ya da red bölgesi, null hipotezin hangi koşullar altında reddedileceğini tanımlayan ve deney öncesinde seçilen test istatistiğinin değerlerinin ya da alabileceği değerlerin bir kümesidir (Johnson, 1980, s. 277; Tanis, 1987, s. 306). Başka bir anlatımla kritik bölge, null hipotezin yanlışlığına karar verilmesine yol açacak örneklem istatistiğinin bir değerler kümesidir (Glass ve Stanley, 1970, s. 280). *Kritik değer* ise kritik bölge içindeki ilk değerdir ve test istatistiğinin örneklem dağılımı üzerinde verilen α manidarlık düzeyi ve serbestlik derecesine göre beklenen değerini gösterir. Gözlenen değer (test istatistiği), kritik değerden büyük çıkarsa, yani kritik-red bölgesi içine düşerse null hipotezi reddedilir (Johnson, 1980, s. 277).

İki yönlü test



Şekil 2. İki Yönlü ve Bir Yönlü Testlerde Red Bölgesi ve Kritik Değerler (Fox, 1969, 254-255)

Şekil 2'de görüldüğü gibi null hipotezini reddetmenin olasılığı α , iki yönlü testlerde örneklem dağılım eğrisinin, iki kuyruğun, uç kısımlarına eşit olarak bölünmüştür. Bir yönlü testte aynı olasılık testin yönüne göre sadece kuyruğun bir tarafında toplanmıştır. Buna göre aynı koşullarda ve manidarlık düzeyinde null hipotezini reddetmenin olasılığı iki yönlü testlere göre bir yönlü testlerde çok daha fazladır (Fox, 1969, s. 255).

3. Adım: Test İstatistiğinin Değerlerinin Hesaplanması

İkinci adımda kullanılmasına karar verilen test istatistiği (z , t , F , χ^2 gibi), gözlemlere dayanılarak hesap edilir.

4. Adım: Kararın Verilmesi ve Yorumlanması

Null hipotezin testinde değişik koşullara göre üç farklı biçimde karar oluşturulabilir (Gamgam, 1989, s. 15):

Birincisi, kritik değerleri ilgili istatistiğin örnekleme dağılımı üzerinde bulmak. Örneğin, pearson korelasyon katsayısının manidarlık değerleri verilen α manidarlık düzeyi ve serbestlik derecesi $(n-2)$ 'ne göre bulunur ve hesap edilen ya da verilen korelasyon değeri ile karşılaştırılabilir. Hesap edilen ya da verilen korelasyon katsayısı kritik değere (beklenen değere) eşit ya da büyük ise gözlenen ilişkinin manidar olduğu söylenir.

İkincisi, istatistiği bir başka istatistiğe, test istatistiğine dönüştürmek ve kritik değerleri, elde edilen yeni istatistiğin örnekleme dağılımından bularak karşılaştırmak. Hesaplanan test istatistiği kritik değere eşit ya da büyük ise gözlenen ilişki ya da istatistiksel farkın manidar olduğu söylenir.

Üçüncüsü bir p değeri hesaplayarak karar vermektir. p değeri, gözlenen değerlerden hesap edilen bir test istatistiğinin ortaya çıkış olasılığına eşit ya da karşıt hipoteze uygun olarak daha uç değerler alma olasılığıdır. p değeri, null hipotezinin doğru olma olasılığını gösterir ve verilen α manidarlık düzeyinde red bölgesinin toplam olasılık değerine eşit ya da küçük olması durumunda null hipotezinin reddini gerektirir. Aksi halde, yani $p > \alpha$ ise null hipotezi reddedilemez. p değeri küçüldükçe null hipotezini reddetmenin olasılığı artar.

Kısacası, yukarıda tartışılan birinci ve ikinci çözümlemede istatistiğin ya da test istatistiğinin beklenen değerine ya da kısacası tablo

değerine eşit ya da büyük olması durumunda null hipotezi reddedilir. Üçüncü yaklaşımda ise bulunan p değerinin red bölgesinin ortaya çıkma olasılığından eşit ya da küçük olması durumunda null hipotezi reddedilebilir (Williamson ve diğerleri, 1977, s. 440).

Hipotez testinde karar verilirken iki tür hata söz konusudur. Null hipotezi doğru iken reddedilme durumunu gösteren *I. Tür Hatanın* ortaya çıkış olasılığı α manidarlık düzeyi ile gösterilir. Null hipotezi yanlış iken kabul edilme durumunu gösteren *II. Tür Hatayı* yapmanın olasılığı β ile gösterilir. Bu iki hatanın dışında iki doğru karar verme olasılığı söz konusudur. Yani, bir bakıma araştırmacının önünde olası dört farklı karar vardır. Araştırmacı, *I. tür hatayı* azaltmak için α düzeyini küçültülebilir. Ancak bu defa da onunla negatif ilişkili olan β olasılığını, yani *II. tür hatayı* artırmış olur. Her iki hatanın ortaya çıkış olasılıkları daha sonraki konularda yer yer tartışılacağı gibi Şekil 3 üzerinde de gösterilmiştir. Araştırmacının hangi tür hataya önem verdiğine bağlı olarak manidarlık düzeyi (α) artırılabilir ya da küçültülebilir. Örneğin açımlayıcı türden çalışmalarda α daha yüksek düzeyde tutulmaktadır. Bu anlamda araştırmacının dikkati α manidarlık düzeyi üzerindedir (Borg ve Gall, 1983, s. 373; Tanis, 1987, s. 307).

		GERÇEK	
		H_0 doğru	H_0 yanlış
KARAR	H_0 red	I. Tür Hata α Manidarlık Düzeıı	Doğru Karar ($1 - \beta$) Testin Gücü
	H_0 kabul	Doğru Karar ($1 - \alpha$) Güven Düzeıı	II. Tür Hata β

Şekil 3. Hipotez Testinde Olası Sonuçlar (Tanis, 1986, s. 307)

Mouley'e (1963, s. 152) göre iki tür hatanın aynı anda azaltılmasının iki yolu vardır. Birincisi, daha büyük örneklem almak, ikincisi örnekleme değişkenliğini (varyansı) azaltmak.

Özetle manidarlık testi, ilgilenilen değişkenler bakımından örneklemde bulunan bir ilişkinin ya da iki istatistik arasındaki bir farkın

evrende olma olasılığının ne kadar olduğuna ya da gerçekte bir ilişki ya da fark olmadığında bunun olasılığını açıklayabilen (örnekleme hata terimlerini kullanarak) bir karar verme sürecidir (Williamson ve diğerleri, 1977, s. 440).

Yukarıda manidarlık testi ve işlemsel adımları bir bütün olarak tartışıldı. Konuyu daha da somutlaştırmak için çalışmanın başında verilen örnekte yer alan yöneticinin, "fen bilimleri öğretiminde A2 öğretim yönteminin A1 öğretim yönteminden daha başarılı sonuçlar verir" şeklindeki kanısını test etmek gerekirse:

Birinci adım: $H_0: \mu_{A2} = \mu_{A1}$, $H_1: \mu_{A2} > \mu_{A1}$

İkinci adım: Büyük örneklem için ilişkisiz z-test istatistiği (kritik oran) seçilir. Manidarlık düzeyi $\alpha = .05$ alınacak olursa bir yönlü test için kritik değer $z = 1.65$ olur.

Üçüncü adım: z-test istatistiğinin bilinen (gözlenen) değerlerden hesap edilmesi,

$$\sigma \bar{X}_{A2} - \bar{X}_{A1} = \sqrt{\frac{S^2_{A2}}{n_2} + \frac{S^2_{A1}}{n_1}} \text{ ve } z = \frac{(\bar{X}_{A2} - \bar{X}_{A1}) - (\mu_{A2} - \mu_{A1})}{\sqrt{S^2_{A2}/n_2 + S^2_{A1}/n_1}} \text{ dir}$$

(Johnson, 1980, s. 349, 352).

$$SH \bar{X}_{A2} - \bar{X}_{A1} = \sigma \bar{X}_{A2} - \bar{X}_{A1} = \sqrt{\frac{7^2}{40} + \frac{5^2}{40}} = 1.36, z = \frac{(75-70)-0}{1.36} = 3.676.$$

Test işleminde null hipotezi esas alındığından $(\mu_{A2} - \mu_{A1}) = 0$ 'dır.

Dördüncü adım: $z_{hes.} = 3.676 > z_{.05} = 1.65$ olduğundan ortalama puanlar arasında A2 öğretim yöntemi lehine gözlenen 5 puanlık farkın .05 düzeyinde manidar olduğu söylenir.

Manidarlık Testlerinin Yorumu

Manidarlık testlerinde kararların yorumlanmasını birkaç biçimde ifade etmek olasıdır. Örneğin, .01 manidarlık düzeyinde null hipotezi reddedilmiştir, .01 manidarlık düzeyi ile null hipotezi reddedilmiştir, % I manidarlık düzeyinde null hipotezi reddedilmiştir ya da .01'lik I. tür hata yapma olasılığıyla null hipotezi reddedilmiştir (Glass ve Stanley, 1970, s. 283). Anlaşılacağı üzere α manidarlık düzeyi yanlış yorumlamanın olasılığını gösterir (Borg ve Gall, 1983, s. 374).

Manidarlık testlerinde verilen kararın ifade edilmiş biçimi de önemlidir. İfadede kesin bir dilin kullanımı yanlıştır. Çünkü, daima bir hata

payı söz konusudur. Dolayısıyla null hipotezin doğru ya da yanlış olduğu değil, reddedilip edilememesi önemlidir (Glass ve Stanley, 1970, s. 283).

Yorumlamada bir başka önemli hata da null hipotezinin kabulü durumunda olmaktadır. Yorumda, iki istatistiğin tam olarak aynı, özdeş olduğu söylenemez. Böyle bir olasılık var olmakla birlikte çok düşüktür. Dolayısıyla yorumda iki istatistiğin birbirinden manidar bir şekilde farklı olmadığı, birbirine oldukça yakın olduğu, farkın önemsiz olduğu biçiminde bir ifade kullanılmalıdır (Williamson ve diğerleri, 1977, s. 440).

Borg ve Gall'a (1983, s. 374-375) göre yanlış yorumlamanın beş tipine rastlanmaktadır. Bunlar:

1. Manidarlık düzeyinin, yani P ile gösterilen yanlış yorumlama olasılığının (sayısının), gruplararası bulunan bir farkın ya da ilişkinin şans ile oluşmasının olasılığını (sayısını) verdiği şeklinde yorumlanması olmaktadır. Bu durum araştırmacıyı null hipotezi reddediğinde, örneğin $\alpha = .01$ için, farkın ya da ilişkinin yüz kezde bir kez şanstın kaynaklandığı ifadesine götürür ki bu yanlıştır. Çünkü, α manidarlık düzeyi I. tür hatayı, yani H_0 reddediğinde dahi onun reddedilemeyeceği, yani kabul edileceği olasılığı gösterir.

2. Manidarlık düzeyinin araştırma hipotezinin doğru olma olasılığını ya da doğru olmasının ne kadar olası olduğunu verdiği şeklindeki bir ifade olmaktadır. Oysa, α düzeyi araştırma hipotezinin doğrulanması üzerinde dolaylı bir anlamı vardır, o yalnızca H_0 'ın reddedilmesine ilişkin bir karar vermeye yardım eder.

3. Yanlış yorumlama olasılığı P'nin yapılan çalışmanın tekrarlanması durumunda, aynı araştırma sonuçlarının bulunması olasılığını vereceği ifade edilmektedir. Oysa çalışmanın bir tekrarından diğerine farkın miktarında önemli bir değişiklik olabilir. Manidarlık düzeyi gelecekteki çalışmaların kestiriminde kullanılmaz, sadece H_0 'ı red etmede kullanılır.

4. Araştırma sonuçlarının pratik ve kuramsal manidarlığı ile manidarlık düzeyinin karıştırılması durumudur. Oysa bu iki kavram birbirinden farklıdır. Örneğin, $\alpha = .01$ 'de 1000 denekli bir örnekleme $r = .08$ manidardır, ancak bu ilişki düzeyinin pratik ve kuramsal manidarlığı yoktur. Böyle bir ilişki son derece önemsizdir ve özellikle sosyal bilimlerde eğitimde pratik bir manidarlığı yoktur. Ayrıca, böyle bir ilişki miktarının evrende sıfıra daha da yaklaşacağı ve sıfır olacağı unutulmamalıdır.

5. Manidarlık düzeyi sadece belirlenmiş bir evrenden yansız bir biçimde seçilen örneklemelerden elde edilen istatistikler arasındaki fark ya da gözlenen bir ilişkinin evrende hangi olasılıklarla olacağına ilişkin null hipotezin testinde kullanılır. Belirlenmemiş (kesinlikle) ya da yansız olarak seçilmemiş örneklemelerde α kullanılamaz, yorumlanamaz. Ayrıca manidarlık düzeyi α , belirlenmiş tüm evren üzerinde yapılan bir çalışmada da kullanılmaz. Çünkü, böyle bir durumda istatistiksel manidarlık testinin kullanılması anlamsızdır, aradaki fark ya da ilişki zaten evren için var olan fark ya da ilişkidir.

Örnekte verilen manidarlık kararı, fen bilimleri öğretiminde A2 öğretim yönteminin A1 öğretim yönteminden % 95 güvenle daha başarılı sonuçlar vereceği ya da gerçekte evrende .05'lik bir hata ile daha başarılı sonuçlar vereceği ifadesi kullanılabilir. Başka bir deyişle de .05 manidarlık düzeyinde (yüz kezde beş kez hata ile) A2 öğretim yönteminin ortalama puanının A1 öğretim yönteminin ortalama puanından daha yüksek çıkacağı ve farkın da manidar olduğu söylenebilir.

Manidarlık Testi Türleri

Daha önce belirtildiği gibi manidarlık testi, kullanılan test istatistiğine bağlı olarak değişmektedir. Ancak kullanılan yaklaşım tüm test istatistikleri için aynıdır. Manidarlık testi kullanılan test istatistiklerine göre sınıflandırılabilmektedir. Test istatistiği, null hipotezi ve karşıt hipotez arasında karar vermek için kullanılan ve gözlemlere dayalı olan bir istatistiktir (Tanis, 1987, s.315). Gerek parametrik, gerek parametrik olmayan test istatistiklerinden sıkça kullanılanları aşağıda özetlenmeye çalışılmıştır.

Parametrik İstatistikler:

Kritik oran (z) ve t-testi. Her iki test istatistiği 1- İki istatistik (\bar{X} , r ya da değişim katsayısı gibi istatistikler) arasındaki farkın, evren farkını yansıtıp yansıtmadığına karar vermede kullanılır, 2- Bazen de bir istatistiğin, örneklemin yansız olarak çekildiği belirlenmiş evrenin bir değerinden (kestirilen, kabul edilen ya da bilinen) manidar biçimde farklı olup olmadığına karar vermede kullanılır (Borg ve Gall, 1983, s. 377).

z-testine genelde normal dağılım testi de denilmektedir. Her iki test istatistiğinin hesaplanmasında evrenin bilinmeyen standart sapması yerine örneklemin standart sapması kullanılır (Mouly, 1963,

s. 156). z ve t-test istatistikleri aynı kullanım alanlarına sahip olmakla birlikte z-testi yalnızca $n > 30$ olan örneklerde, t-testi ise küçük veya büyük tüm örneklerde kullanılır. Hesaplama kullanılan formüller örneklerin ilişkili olup olmamasına bağlı olarak değişir (Johnson, 1980, s. 314).

Z ve t testinin üç temel sayıltısı vardır. (Borg ve Gall, 1983, s. 548); 1) puanlar aralık ya da oran ölçeğindedir, 2) puanların evrendeki dağılımı normaldir, 3) Araştırmaya dahil evrenlerin varyansları eşittir.

F-testi. Bu teste, çoklu t-testi karşılaştırması da denilir ve varyans analizi yapılmak istendiğinde kullanılır. Yani F-testi, üç ya da daha fazla grubun eş zamanlı karşılaştırılmasında ve basit ya da çoklu deney ve kontrol gruplarının karşılaştırıldığı kovaryans analizinde kullanılır (Mouly, 1963, .156). F-testi örneklem ortalamaları arasında manidar bir fark olup olmadığını test eder. Ancak hangi ortalamalar arasında manidar bir farkın varolduğunu belirlemez. Bunu, F-testi manidar çıkarsa, Fisher'in en küçük manidar fark (LSD), Duncan'ın çoklu genişlik, Scheffe test ve Turkey'nin uygun manidar fark (HSD) testleriyle buluruz (Borg ve Gall, 1983, s. 377).

F-testinin diğer bir kullanım alanı, parametrik testlerin temel sayıltısından biri olan "örneklerin ait oldukları evrenlerin varyanslarının eşit olduğu"nun testinde uygulanmasıdır (Johnson, 1980, s. 356-357). Varyans, kovaryans analizi ve varyansların eşitliğinin test edilmesinde F dağılımı ve F-testi kullanılır. Bunun dışında varyans analizi, iki ya da daha çok faktörlerin arasındaki ortak etkinin istatistiksel bakımdan manidarlığının testi (iki ya da daha çok boyutlu çözümlenme) için de kullanılmaktadır. Kovaryans analizi de, kontrol değişkenleriyle ilgili olarak gruplararası başlangıçta varolan farklılıkları kontrol ederek grupları eşitlemede kullanılır (Borg ve Gall, 1983, s. 543).

F-testi nadir olarak hipotezde öngörülen eğilimin istatistiksel bakımdan manidar olup olmadığına karar vermede de kullanılır ki buna Trend Analizi denir (Borg ve Gall, 1983, s. 377).

Verilen test istatistiklerinin Ho'm doğruluğu şartı altında beklenen kuramsal örneklem dağılımları vardır. Test işlemi, bir bakıma gözlenen değerlerle beklenen değerlerin karşılaştırılmasıdır. Parametrik z ve t testinin sayıltıları F-testi için de geçerlidir.

Parametrik Olmayan İstatistikler:

Kay-Kare X^2 testi (Chi-Square test). X^2 test istatistiği, temel olarak k olası sonuçların ($k = 1, 2, \dots, n$) her biri için sonuçların beklenen değerleri (sayıları) ile gözlenen değerlerin bir karşılaştırmasını yapmak amacıyla kullanılır. Bu kapsamda X^2 testi belli iki ayrı özelliğe göre gözlenen değerlerin manidar bir şekilde farklılaşp farklılaşmadığına ve iki özelliğin birbirleriyle ilişkili (bağımlı) olup olmadığına bakmak için kullanılır. Bağımsızlık testi de denilen X^2 testinde red bölgesi daima tek yönlüdür (Tanis, 1987, s. 437).

Frekanslarla (sayılarla) çahşan X^2 testi temel olarak sınıflama düzeyinde ölçülen değişkenler için kullanılır. Ancak tekniğin sıkça kullanılan diğer bir kullanımını da bir bağımsız değişkenin (sınıflamalı ölçek) farklı düzeylerinin bağımlı değişken (sıralamalı ölçek) üzerinde ölçüldüğü durumlarda kullanılmaktadır.

İşaret testi (The Sign test). İlişkili iki örneklemden elde edilen iki ayrı ölçüme ilişkin her çiftin iki değerinin birbirleriyle karşılaştırılarak sıralamanın mümkün olduğu durumlarda kullanılır. Veri olarak niceliksel ölçüler yerine artı (+) ve eski (—) işaretleri kullanılır. Araştırmacı iki koşulun, örneğin aynı gruba (örnekleme) verilen ön test ile son testin sonuçlarının farklı olup olmadığını görmek istediğinde işaret testini kullanır. Bu test, ilişkili t-testine karşılık olarak kullanılır ve o teste karşı $N = 6$ gibi küçük örneklemlerde .95 gibi yüksek bir etkililik oranına sahiptir (Siegel, 1977, s. 74-79, 103).

Wilcoxon İşaretili Sıralar testi (Wilcoxon Signed Ranks test). İlişkili örneklemler için kullanılır. İşaret testinde sadece çiftler arası farkın yönü hakkındaki bilgi önemli iken bu testte farkların yönü kadar miktarı da dikkate alınır ve incelenir. Wilcoxon işaretili sıralar testinde hem çiftler arası hem de çiftlerin her biri için ölçme düzeyi sıralayıcı ölçekte yapılır. İlişkili t-testine karşılık olup, etkililik oranı .95'dir (Siegel, 1977, s. 83). Bu test, benzeri işaret testi için de düşünüleceği gibi temelde ilişkili iki örnekleme ait ölçümlerin ortalama değerlerinin birbirinden manidar bir şekilde farklı olup olmadığına karar vermede kullanılır (Borg ve Gall, 1983, s. 379).

Ortanca testi (Median test). İlişkisiz iki örneklemin merkezi eğilimler bakımından farklılaşp farklılaşmadığını test etmede kullanılan ortanca testinde ölçümlerin (gözlemlerin) en az sıralayıcı bir ölçekte ölçülmüş olması gerekliliği vardır. Ortanca testi, ilişkisiz iki örneklemin aynı ortancaya sahip evrenlerden gelip gelmediklerine karar vermede kullanılır (Siegel, 1977, s. 122).

Mann-Whitney U-testi (Mann Whitney U test). Bu test, en az sıralama düzeyinde ölçmeyi gerektirir ve ilişkisiz iki örneklemin aynı evrenden gelip gelmediklerini test etmede kullanılır (Siegel, 1977, s. 128). Mann-Whitney U testi, ilişkisiz t-testinin parametrik olmayan karşıtıdır ve etkililik oranı .95'tir (Johnson, 1980, s. 515). Test, temel olarak iki ilişkisiz örneklem ortalama değerinin birbirinden manidar bir şekilde farklı olup olmadığına karar vermede kullanılır (Borg ve Gall, 1983, s. 379).

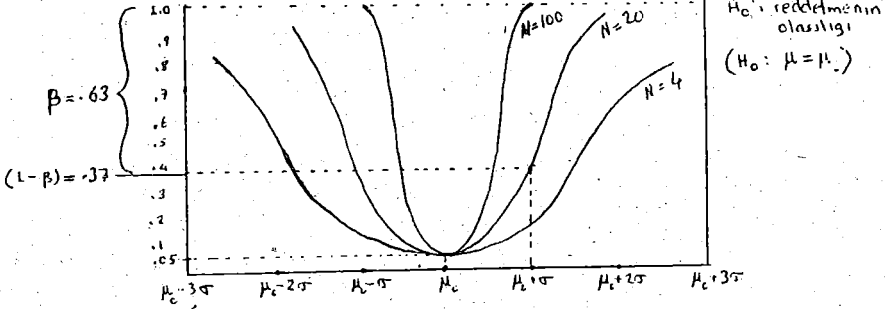
Kruskal-Wallis testi (Kruskal-Wallis test). İlişkisiz iki ya da daha fazla örneklemin farklı evrenlerden gelip gelmediğine karar vermede kullanılır ve sıralar yoluyla tek boyutlu varyans analizi olarak da isimlendirilir (Siegel, 1977, s. 202). Kruskal-Wallis testi, temel olarak ilişkisiz üç ya da daha fazla örneklem sonuçlarının (ortalama değerlerinin) birbirinden manidar bir şekilde farklı olup olmadığını test eder (Borg ve Gall, 1983, s. 379). Bu test, tek boyutlu F-testine karşılık olup, etkililik oranı .95'tir (Gamgam, 1989, s. 21).

Spearman Testi (Spearman test). Test, evrenden yansız olarak çekilen bir örneklemden iki değişkene göre gözlenen iki değer-ölçümler seti arasındaki ilişkinin (her bir ölçüm dizisi için ayrı ayrı bulunacak büyüklük sıralamasından yola çıkılarak hesaplanan spearman sıra farkları korelasyon katsayısının), kuramsal örneklem dağılımında örneklem genişliği (n) ve manidarlık düzeyine (α) göre bulunan tablo değeri ile karşılaştırılarak test edilmesinde kullanılır (Gamgam, 1989, s. 213). Bu test, pearson r'sinin testine karşılıktır ve etkililik oranı .91'dir. (Johnson, 1980, s. 515).

İstatistiksel Güç Çözümlemesi

Araştırmalarda araştırmacının ilgi odağı genelde null hipotezinin doğrulanması değil, aksine null hipotezinin reddinin olabirliğinin maksimize edilmesidir. Bu bağlamda *İstatistiksel Güç Çözümlemesi*, bir null hipotezini reddetmek için yeterli olacak belli bir istatistiksel manidarlık testinin olabirliğini çözümüleme işlemidir. *İstatistiksel Güç* ya da *Testin Gücü* ise, belli bir istatistiksel manidarlık testinin null hipotezi yanlış iken onun raeddine yol açması olasılığıdır. (Borg ve Gall, 1983, s. 379). Başka bir anlatımla bir testin gücü, bir ya da daha çok parametrenin bilinmeyen değerinin bir fonksiyonu olarak kabul edilir. Değeri bilinmeyen parametreler için güç hesap edilemez, karar verilemez (Roscoe, 1975, s. 184).

Verilen bir istatistiksel testin gücü Şekil 4'de görüldüğü üzere kullanılan manidarlık düzeyinde örneklem genişliğinin bir fonksiyonudur ve bu fonksiyon null hipotezinin ne kadar yanlış olduğunun olasılığını verir. (Roscoe, 1975, s. 185).



Şekil 4. Bir İstatistiksel Testin Güç Eğrileri.

(Roscoe, 1975, s. 185)

Bir testin gücünü etkileyen etkenleri beş grupta toplamak olasıdır. Bunlar (Borg ve Gall, 1983, s. 379; Howell, 1987, s. 195; Roscoe, 1975, s. 185):

1. *Örneklem genişliği.* Büyük örneklerde null hipotezini reddetmek için gerekli olan ilişki ve farklar çok daha küçüktür. Örneğin, $r_{xy} = .25$ 'lik bir ilişki $N = 47$ 'de reddedilemezken $N = 62$ 'de red edilir.

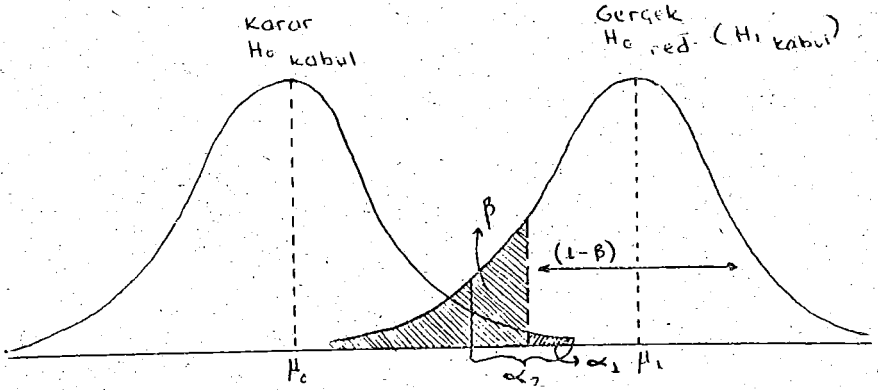
2. *Null hipotezinin reddedileceği manidarlık düzeyi.* Testin gücü manidarlık düzeyinin yükseltilmesi ile artırılabilir. Örneğin, $\alpha = .10$ olan bir test $\alpha = .05$ olan bir testten (diğer koşullar aynı kalmak koşuluyla) red bölgesi daha geniş olacağından da güçlüdür.

3. *Araştırma hipotezinin kuruluş yapısı.* Bir yönlü testler iki yönlü testlerden red bölgeleri daha geniş olduğundan daha güçlüdürler.

4. *İstatistiksel testin özellikleri, sayıtları.* Parametrik testler, parametrik olmayan testlerden daha güçlüdür.

5. *Parametrenin büyüklüğü.* İstatistiğin elde edildiği örneklemin yansız olarak seçildiği evrende, parametrenin büyük olması testin gücünü artırır. Bu çerçevede tartışılan diğer bir kavram da etki genişliğidir (effect size). Etki genişliği, bir örneklem ya da evrendeki bir farkın ya da ilişkinin büyüklüğünün ne olduğuna ilişkindir. Örneğin, r 'nin evrendeki değeri yüksek ise örneklemelerde de benzer büyüklükte r değerleri sunma eğilimi vardır. Bu ise, null hipotezinin reddini kolaylaştırır. Etki genişliği deneysel çalışmalarda hesaplanabilmektedir.

Yukarıda verilen ilk üç etken araştırmacının kontrolindedir.



Şekil 5. I. Tür ve II. Tür Hata ve Bir Testin Gücü.
(Glass ve Stanley, 1970, s. 285; Howell, 1987, s. 195)

Şekil 5'de görüleceği gibi α_1 yerine α_2 , manidarlık düzeyi alındığında I. tür hata artmakta, II. tür hata (β) azalmakta, buna karşın testin gücü ($1-\beta$) artmaktadır.

Özden (1981, s. 261) de, α aynı iken β 'yi düşürmenin dolayısıyla da testin gücünü artırmanın iki yolundan birinin örneklemin genişliğini artırmak, diğerinin de standart sapmayı azaltmak olduğunu ifade etmektedir.

İstatistiksel Çözümleme Ünitesi

İstatistiksel çözümleme ünitesi ya da birimi, bir araştırmada araştırma probleminin somutlaşmış şekli olan hipotezdeki değişkenlere göre saptanan örnekleme ünitesi-birimidir. Yani çözümleme ünitesi,

üzerinde veri toplanan, gözlemde bulunulan birimi ya da birimleri gösterir. Birim, bireylerden oluşacağı gibi gruplardan da oluşabilir. Ancak kimi araştırmalar vardır ki bu birim küçük bir grubun ötesinde bir okul, mahalle, bölge ve hatta ülkeler olabilir. Araştırmacı çözümleme biriminin ne olacağına karar verirken çözümleme ve yorumlamada yapacağı karşılaştırmaları (değişkenlere göre) dikkate alır (Borg ve Gall, 1983, s. 394-395).

Örnek hatırlanacak olursa $\bar{X}_{A_2} = 75$ ve $\bar{X}_{A_1} = 70$ 'idi. Buna göre $H_0: \mu_{A_2} = \mu_{A_1}$ kurulmuş ve bu hipotez $H_1: \mu_{A_2} > \mu_{A_1}$ 'ya karşı test edilmişti. Bu örnekte çözümleme ünitesi bireylerden oluşmuş ve deney grubunda yer alan bireyler ile kontrol grubunda yer alan bireylerin ortalama puanları arasındaki fark test edilmişti. Araştırma deseni, öğretim yönteminin yanı sıra ailenin sosyo-ekonomik durumu da dikkate alınarak değiştirilirse, ortaya 2×2 'lik bir faktöryel desen çıkar ve her bir gözeneğe de bir çözümleme ünitesini oluşturur.

Sonuç olarak bu çalışmada, kestirisel istatistik konusu bir bütün olarak ele alınmış; ancak konunun çok geniş olması dolayısıyla konuya ilişkin bazı temel kavramlar seçilmiş ve incelenmiştir.

KAYNAKÇA

- Borg, Walter R., and Meredith Gall., *Educational Research: An Introduction*. Fourth Edition. New York: Longman Inc, 1983.
- Ferguson, George., *Statistical Analysis in Psychology and Education*. Fourth Edition. USA: Mc Graw-Hill Inc, 1976.
- Fox, David J., *The Research Process in Education*. USA: Holt, Rinehart and Winston Inc, 1969.
- Gamgam, Hamza., *Parametrik Olmayan İstatistiksel Teknikler*. Ankara: Gazi Üni. Fen-Ede. Fak. Yayınları No: 21, 1989.
- Glass, Gene V., and Jullian C. Stanley., *Statistical Methods in Educaion and Psychology*. New Jersey: Prentice Hall Inc, 1970.
- Howell, David G., *Statistical Methods for Psychology*. Second Edition. USA: PWS Publishers, 1987.
- Johnson, Robert R., *Elementary Statistics*. Third Edition. California: Wadsworth Inc, 1980.
- Kerlinger, Fred N., *Foundations of Behavioral Research*. Second Edition. London: Holt, Rinehart and Winston Inc, 1973.
- Meyer, William J., "Inferential Statistics" *Psychology and Human Development Understanding Educational Eesearch*. USA: Mc. Graw-Hill Book Company, 1962.

Mouly, George J., *The Science of Educational Research*. New York: American Book Company, 1963.

Özden, Hayriye., *İstatistik Kuramı ve Uygulamaları*. Ankara: Hacettepe Üni. Fen. Fak. Yayınları No: 13, 1981.

Peaker, G.F., "Statistical Analysis in Educational Research" *Educational Research Methodology and Measurement: An International Handbook*. Edited by John P. Keeves. Great Britain: Books for Professionals Inc, 1988.

Roscoe, John T., *Fundamental Research Statistics for The Behavioral Sciences*. Second Edition. New York: Holt, Rinehart and Winston Inc., 1975.

Siegel, Sindney., *Davranış Bilimleri için Parametrik Olmayan İstatistikler*. Çev. Yurdal Topsever. Ankara: A.Ü. DTCF yayınları No: 274, 1977.

Tanis, Elliot A., *Statistics II. Estimation and Tests of The Hypothes*. New York: Books for Professionals Inc, 1987.