

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Burak KESKİN

İSTATİSTİKSEL GÜÇ ANALİZİ: SOSYAL BİLİMLER ALANINDA
BİR UYGULAMA

İşletme Ana Bilim Dalı

Yüksek Lisans Tezi

Antalya, 2012

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Burak KESKİN

İSTATİSTİKSEL GÜÇ ANALİZİ: SOSYAL BİLİMLER ALANINDA
BİR UYGULAMA

Danışman

Prof. Dr. Ayşe ANAFARTA




İşletme Ana Bilim Dalı

Yüksek Lisans Tezi

Antalya, 2012

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Burak KESKİN'in bu çalışması jürimiz tarafından İşletme Ana Bilim Dalı Yüksek Lisans Programı tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Mualla AKSU 
Üye (Danışmanı) : Prof. Dr. Ayşe ANAFARTA 
Üye : Doç. Dr. Can Deniz KÖKSAL 

Tez Başlığı: İSTATİSTİKSEL GÜÇ ANALİZİ: SOSYAL BİLİMLER
ALANINDA BİR UYGULAMA

Onay : Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 15/06/2012

Mezuniyet Tarihi : 27.06/2012

İÇİNDEKİLER

ŞEKİLLER LİSTESİ	iv
TABLolar LİSTESİ	v
ÖZET	vii
ABSTRACT	ix
ÖNSÖZ	xi
GİRİŞ.....	1

BİRİNCİ BÖLÜM

İSTATİSTİKSEL GÜÇ ANALİZİ

1.1	İstatistiklerin Kullanılması.....	3
1.2	Hipotez Test Etme Süreci	4
1.3	Hipotez Test Etme Sürecinde Yapılan Hatalar	6
1.4	İstatistiksel Güç Kavramı.....	11
1.5	Neyman-Pearson Teorisi ve Fisher Teorisi Arasındaki Tartışmalı Konular	12
1.6	İstatistiksel Güç Analizi.....	17
1.7	İstatistiksel Gücü Doğrudan Etkileyen Faktörler.....	19
1.7.1	Anlamlılık Seviyesi(α).....	20
1.7.1.1	Alfa ve Beta Arasındaki İlişki	22
1.7.2	Etki Büyüklüğü.....	22
1.7.2.1	Etki Büyüklüğü İndeksleri.....	28
1.7.2.1.1	"d" Ailesi: Gruplar Arasındaki Farklılıkların Ölçülmesi.....	28
1.7.2.1.2	"r" Ailesi: Değişkenler Arasındaki İlişkilerin Ölçülmesi.....	31
1.7.2.2	Etkinin Yorumlanması.....	36
1.7.2.3	Etki Büyüklüğü İçin Güven Aralığı (Confidence Interval) Tahmini.....	38
1.7.3	Örneklem Sayısı	40
1.7.3.1	Belirlenebilecek En Küçük Etki.....	48

1.8	İstatistiksel Gücü Dolaylı Olarak Etkileyen Faktörler.....	50
1.8.1	Alternatif Hipotezin Tek Yönlü veya Çift Yönlü Olması	50
1.8.2	Kullanılan İstatistiksel Testin Parametrik veya Non-Parametrik Olması.....	50
1.8.3	Örnekleme Hatası	50
1.8.4	Ölçüm Hatası	50
1.8.5	Deney ve Kontrol Gruplarının Örneklem Sayılarının Eşit Olması.....	51
1.8.6	Varyans Analizinde Kullanılan Grup Sayısı	51
1.8.7	Çoklu Regresyon Analizinde Kullanılan Bağımsız Değişken Sayısı.....	52
1.9	İstatistiksel Güç Hesaplanmasına İlişkin Örnekler	53
1.10	İstatistiksel Güç Analizi Çeşitleri	56
1.10.1	Teorik Güç Analizi (Priori Power Analysis-Prospective Power Analysis).....	57
1.10.2	Deneysel Güç Analizi (Post-Hoc Power Analysis-Retrospective Power Analysis).....	59
1.10.3	Karşılaştırmalı Güç Analizi (Compromise Power Analysis)	60
1.11	Yetersiz ve Aşırı İstatistiksel Gücün Neden Olduğu Sorunlar	61
1.12	Güç Analizi Değerlerinin Yorumuna İlişkin Literatürdeki Tartışmalar	68
1.13	Literatürdeki Çalışmaların Güç Açısından İncelenmesi.....	68
1.14	İstatistiksel Gücü Artırma Yolları.....	77
1.15	Araştırmacıları İstatistiksel Gücü İhmal Etmeye İten Nedenler.....	79
1.16	İstatistiksel Güç Hesaplamaya Yönelik Bilgisayar Programları.....	81
1.16.1	G-POWER: Genel Bir Güç Analizi Programı.....	81
1.16.2	PASS ve STATISTICA	86

İKİNCİ BÖLÜM
İSTATİSTİKSEL GÜÇ BELİRLEMeye İLİŞKİN
SOSYAL BİLİMLER ALANINDA BİR UYGULAMA

2.1 Araştırmanın Amacı, Kapsamı ve Önemi	87
2.2 Araştırmanın Yöntemi	87
2.2.1 Araştırmanın Örneklemi	87
2.2.2 Veri Toplama Araçları	89
2.2.3 Veri Analiz Teknikleri	89
2.2.4 Araştırmanın Bulguları ve Yorumları	90
2.2.4.1 İstatistiksel Güç Analizi	92
2.3 İstatistiksel Güç ile Anlamlılık Testlerinden Elde Edilen Sonuçlar Arasındaki İlişki	103
SONUÇ	105
KAYNAKÇA	108
EK-1	114
ÖZGEÇMİŞ	116

ŞEKİLLER LİSTESİ

Şekil 1.1 H_0 ve H_1 Hipotez Dağılımlarında I. Tip Hata - II. Tip Hata ve İstatistiksel Güç Bölgelerinin Gösterilmesi.....	10
Şekil 1.2 H_0 ve H_1 Hipotezlerinin Tepe Noktaları Arasındaki Farkın İstatistiksel Güç Üzerindeki Etkisi	11
Şekil 1.3 Alfa Seviyesinin İstatistiksel Güç Üzerindeki Etkisi.....	21
Şekil 1.4 Etki Büyüklüğünün İstatistiksel Güç Üzerine Etkisi.....	26
Şekil 1.5 İstatistiksel Güç – Örneklem Sayısı Arasındaki İlişkinin Çift Yönlü Farklı Alfa Seviyelerinde Gösterilmesi.....	41
Şekil 1.6 Alternatif Hipotezi Tek Yönlü veya Çift Yönlü Olmasının İstatistiksel Güç Üzerindeki Etkisi	50
Şekil 1.7 Deney ve Kontrol Gruplarının Örneklem Sayılarının Eşit Olup Olmamasının İstatistiksel Güç Üzerindeki Etkisi	51
Şekil 1.8 Varyans Analizindeki Grup Sayısının İstatistiksel Güç Üzerindeki Etkisi.....	52
Şekil 1.9 Çoklu Regresyon Analizinde Kullanılan Bağımsız Değişken Sayısının İstatistiksel Güç Üzerine Etkisi	53
Şekil 1.10 G-Power 3.1.3 Programından Güç Analizi Öncesi Bir Kesit.....	83
Şekil 1.11 G-Power 3.1.3 Programından Güç Analizi Sonrası Bir Kesit.....	84
Şekil 1.12 G-Power 3.1.3 Programından Belirlenebilecek Etkiye Dair Bir Kesit.....	85

TABLOLAR LİSTESİ

Tablo 1.1 Hipotez Test Sürecindeki Hatalar	7
Tablo 1.2 Cohen'in Etki Büyüklüğü Sınıflandırmaları	28
Tablo 1.3 J. Rossi Tarafından Sağlık Psikolojisi Alanı İçin Revize Edilen Cohen'in Etki Büyüklüğü Değerleri.....	36
Tablo 1.4 Pearson Korelasyon Katsayısı İçin Minimum Örneklem Sayıları ($\alpha_1 = 0.05$ ve tek yönlü)	44
Tablo 1.5 Pearson Korelasyon Katsayısı İçin Minimum Örneklem Sayıları ($\alpha_2 = 0.05$ ve çift yönlü).....	45
Tablo 1.6 İki Grup Arasında Anlamlı Bir Farklılık Belirlemek İçin Gerekli Örneklem Sayıları ($\alpha_1 = 0.05$ ve tek yönlü)	46
Tablo 1.7 İki Grup Arasında Anlamlı Bir Farklılık Belirlemek İçin Gerekli Örneklem Sayıları ($\alpha = 0.05$ ve çift yönlü).....	47
Tablo 1.8 Verilen Bir Örneklem Sayısı İçin Belirlenebilecek En Küçük Etki Seviyeleri.....	49
Tablo 1.9 Alfa ve İstatistiksel Güç İçin Bazı Faydalı Eşdeğer z Skorları.....	54
Tablo 1.10 Simülasyon Sonuçları.....	56
Tablo 1.11 Farklı Alanlarda Yapılmış Çalışmaların İstatistiksel Güçlerinin İncelenmesi	73
Tablo 2.1 Araştırmaya Dahil Edilen Tezlerin Yüksek Lisans ve Doktora Dağılımı.....	89
Tablo 2.2 İstatistiksel Anlamlılık Testlerinin Dağılımı.....	90
Tablo 2.3 Araştırmaya Dâhil Edilen Tezlerde Kullanılan Gerçek Örneklem Sayıları ve Optimum Örneklem Sayıları	91
Tablo 2.4 Gerçek ve Optimum Örneklem Sayılarının Test Başına Dağılımları	92
Tablo 2.5 Araştırmaya Dâhil Edilen Tezlerin Küçük, Orta ve Büyük Seviyedeki Etki Büyüklüklerine Göre Hesaplanan Ortalama İstatistiksel Güçleri	92

Tablo 2.6 Tezlerde Uygulanan İstatistiksel Anlamlılık Testlerinin Güç Ortalamalarının Frekans ve Kümülatif Yüzdeleri.....	93
Tablo 2.7 Her Bir İstatistiksel Anlamlılık Testi İçin Ortalama İstatistiksel Güç Seviyesi.....	94
Tablo 2.8 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin ANOVA Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı	95
Tablo 2.9 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Bağımsız Gruplar t-Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı.....	96
Tablo 2.10 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Çoklu Regresyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı	97
Tablo 2.11 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Basit Regresyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı.....	98
Tablo 2.12 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Kruskal Wallis Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı.....	99
Tablo 2.13 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Mann-Whitney U Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı.....	100
Tablo 2.14 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Pearson Korelasyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı.....	101
Tablo 2.15 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Ki Kare Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı	102
Tablo 2.16 İstatistiksel Anlamlılık Testlerinin Sonuç Dağılımları.....	103

ÖZET

Güç analizi, güvenilirliği, geçerliliği yüksek bir araştırma planlamayı ve araştırma sonucunda alınacak kararların geçerliğini, güvenilirliğini ve duyarlılığını garanti altına almayı sağlayan bir yöntemdir. İstatistiksel güç, araştırmada bir amacın denetlenmesi için kullanılan istatistiksel testler sonucunda varılan kararın ne kadar güvenilir, geçerli olduğunu olasılık olarak tahmin eden bir yaklaşımdır. Güç analizi, diğer istatistiksel yaklaşımlardan önemli ölçüde farklıdır. Birçok istatistiksel analiz mevcut veri ile analize başlamakta ve sonuçların yorumuna odaklanmaktadır. Hâlbuki güç analizi veri toplama öncesinde anlamlıdır ve hipotez ifade sürecinde önemlidir.

Bu bağlamda tezin amacı;

- i. Sosyal bilim çalışmalarında güç analizinin ne olduğu konusunda farkındalık oluşturarak araştırmalara katkı sağlamak,
- ii. Bir sosyal bilim çalışmasında güç analizinin nasıl kullanılacağı konusunda araştırmacılara rehberlik etmektir.

Bu çalışmada, 2005-2012 yılları arasında “İnsan Kaynakları” alanında başarıyla tamamlanmış olan Yüksek Öğretim Kurulu (YÖK) Ulusal Tez Merkezi’ne kayıtlı yüksek lisans ve doktora tezleri, istatistiksel güç seviyesi belirlemek amacıyla incelenmiştir. Çünkü ilgili yazında böyle bir güç analizi araştırmasına rastlanmamıştır ve bu çalışmanın yazındaki mevcut istatistiksel güç bilgisine önemli ölçüde katkı yapması beklenmektedir.

Tezin yazın taraması bölümünde istatistiksel güç analizi konusunda ayrıntılı bilgi verilirken, uygulama bölümünde bir örnek üzerinde istatistiksel güç analizi sürecine yer verilmiştir. Cohen’in güç tablolarından ve G-Power 3.1.3 istatistiksel güç analizi programından yararlanılarak gerçekleştirilen istatistiksel güç analizi sonucunda araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinin istatistiksel güçleri sırasıyla küçük, orta ve büyük seviyedeki etki büyüklükleri için 0.23, 0.72 ve 0.92 olarak bulunmuştur. Bu güç değerleri, büyük seviyedeki etki büyüklüğünü belirlemek dışında küçük ve orta seviyedeki etki büyüklüklerini belirleyebilmek için Cohen’in tavsiye ettiği %80 seviyesinin altındadır. Buna karşılık, uygulamada elde edilen güç değerleri, yazında mevcut olan istatistiksel güç analizi çalışmalarında elde edilen değerlerden nispeten daha yüksek çıkmıştır. Ayrıca uygulama kısmında, tezlerde uygulanan anlamlılık testlerinden elde edilen sonuçlar

incelenmiş ve toplamda uygulanan 2970 adet anlamlılık testinden 1685'i (%56,73) anlamsız olarak, 1285'i (%43,27) de anlamlı olarak bulunmuştur. Uygulanan anlamlılık testlerinin sonuçlarının genellikle anlamsız olması, bunun nedeninin düşük istatistiksel güç olabileceği ihtimalini artırmaktadır. Buradan hareketle, anlamsız olarak sonuçlanan anlamlılık testlerinin, yeterli istatistiksel güç sağlandıktan sonra tekrar edilmeleri faydalı olacaktır.

ABSTRACT

The power analysis is a method that enables to plan a highly valid and reliable research and to guarantee the validity, reliability and sensibility of the results of a research. Statistical power, in addition, is an approach which is used to evaluate to what degree the decisions obtained as a result of statistical tests used to test an aim are valid and reliable in terms of probability values. The power analysis is different from the other statistical methods to a great extent. Various statistical analyses begin the analysis with the existing data and focuses on the comments of the consequences. Nevertheless, power analysis is meaningful before data collection and significant in the process of setting the hypothesis.

In this context, the major purposes of this thesis are;

- i. To contribute researches in social sciences by creating awareness about what the power analysis means.
- ii. To guide researchers about how the power analyses can be employed in a social science study.

In this study, the master and PhD dissertations which are successfully completed between 2005-2012 years in “Human Resources” area and registered to the Council of Higher Education (YÖK) National Thesis Center are investigated to detect statistical power levels. This study is expected to contribute to the existing statistical power knowledge significantly, since there is no exist other study similar to this research.

In the literature part of thesis detailed information has given about statistical power and also in the application part statistical power analysis process on the an example has given. Cohen’s power tables and G-Power 3.1.3 statistical power analysis program, the statistical powers of the masters and PhD dissertations included in the study are found as 0.23, 0.72 and 0.92 for small, medium and large effect sizes respectively. These power values are less than 80% level which is suggested by Cohen to be able to determine small and medium levels effect sizes except large effects. On the other hand, the power values attained as a result of this application are relatively greater than those of the existing literature. Moreover, in the application part, the outcomes of the significance tests employed in the dissertations are examined and 1685 of the tests (56.73%) are found to be insignificant and 1285 (43.27%) are found to be significant out of 2970 total significance tests. The reason behind the failure of the majority of the significance tests may have been associated with low statistical power. Thus,

the significance tests which result in insignificant results will be better if they are repeated after adequate statistical power is assured.

ÖNSÖZ

İstatistiksel güç analizi ile ilgili tez konusunu seçmemde büyük katkısı olan, tez çalışmalarım boyunca bana güvenen, destek veren, beni yönlendiren ve hoşgörüsünü bir an olsun esirgemeyen değerli tez danışmanım Prof. Dr. Ayşe ANAFARTA'ya teşekkürlerimi bir borç bilirim.

Çalışmam boyunca bana yardımcı olan, ulaşamadığım kaynaklara ulaşmamı sağlayan ve istatistiksel güç analizi ile ilgili her türlü soruma içtenlikle cevap veren Dr. Joseph S. ROSSI'ye, Dr. Scott MAXWELL'e, Dr. Daniel O'KEFEE'ye, Dr. Mark HALLAHAN'a ve Araş. Gör. Fatih ŞAHAN'a çalışmama olan katkılarından ötürü müteşekkirim.

Tüm okul hayatım boyunca bana maddi ve manevi her türlü desteği sağlayan, en umutsuz olduğum anlarda bile beni yüreklendiren, eğitim hayatım boyunca beni hep daha iyiye ulaşmak için teşvik eden, her zaman yanımda olduklarını bildiğim çok değerli aileme en içten saygı, sevgi ve teşekkürlerimi sunarım.

GİRİŞ

Araştırmacılar, yıllardır araştırma yaptıkları alanlar ile ilgili verileri analiz etmek için istatistik tekniklerini kullanmaktadır. İstatistikler, araştırmacılara verilerin analizi sürecinde objektif ve doğru sonuçlar elde etmelerini sağlamakta ve onların doğru bir şekilde bu sonuçları yorumlamalarına yardımcı olmaktadır. Araştırmacılar, istatistiksel teknikleri örneklem verilerini analiz etmek ve yapılan işlemlerin anakütle üzerindeki etkilerini test etmek amacıyla kullanmaktadır. Bu etki, araştırılan anakütleden elde edilen iki değişken arasındaki ilişki veya iki grup arasındaki farklılık olabilmektedir.

Bir araştırmanın kalitesini birçok faktör etkilemektedir. Araştırmanın tasarımı, veri toplama yöntemleri, veri analiz teknikleri ve araştırmacının bakış açısı gibi faktörler araştırmanın doğruluğunu etkilemektedir. Bir istatistiksel anlamlılık testi kurulduğunda ise araştırmanın doğruluğu açısından birkaç ilave faktör daha devreye girmektedir. Bu faktörler anlamlılık seviyesi, örneklem sayısı, etki büyüklüğü ve istatistiksel güçtür (Deng, 2000, s. 2). Anlamlılık testinde, gerçekte yanlış olan bir sıfır hipotezinin reddedilme olasılığı *istatistiksel güç* olarak tanımlanmaktadır. Belirli bir anlamlılık seviyesi, örneklem sayısı ve etki büyüklüğü ile çalışmanın istatistiksel gücüne Jacop Cohen'in 1988 yılında yayımladığı "*Statistical Power Analysis for the Behavioral Sciences*" isimli kitabında yer verdiği güç tabloları, formüller veya geliştirilen bilgisayar yazılımları kullanılarak karar verilebilmektedir. Bu süreç ise "*İstatistiksel Güç Analizi*" olarak adlandırılmaktadır (Hallahan ve Rosenthal, 1995, s. 492).

İstatistiksel güç, araştırılan anakütlede, araştırmacının bulmak istediği veya bulmayı umduğu gerçek etkiyi ortaya çıkarabilme ihtimalidir. Bulmak istenilen gerçek etkiden kasıt, şansa bağlı olarak elde edilmeyen gerçek bir değerdir. Güç analizi, diğer istatistiksel yöntemlerden farklı bir öneme sahiptir. Birçok istatistiksel analiz, verilerin elde edilmesiyle başlarken güç analizi farklıdır. Araştırmanın planlama aşamasında gerçekleştirilecek güç analizi, araştırmacının elinde veri olmasını gerektirmez. Kısaca güç analizi, hipotez ifade sürecinin bir parçasıdır. Güç analizinin diğer istatistiksel yöntemlerden bir diğer farkı da elde edilen sonuçların yorumlanma aşamasındadır. Çoğu istatistiksel yöntem, sonuçların veya bilgisayar çıktılarının yorumlanması için epey vakit ayrılmasını gerektirirken, güç analizi sonuçlarının yorumlanması için çok az zaman ve dikkat gerektirmektedir (Aberson, 2010, s. ix). Araştırmacı aradığı gerçek etkiyi bulduğunda, bu artık onun araştırılan sıfır hipotezini (H_0) reddedebileceği anlamına gelmektedir. Eğer bir araştırmacı çalışmasındaki istatistiksel gücü artırırsa, gerçek etkiyi bulma ve yanlış bir H_0 hipotezini reddetme şansını da artıracaktır.

Aksine arařtırmacı, arařtırmasında yeterli gücü saęlayamazsa bu onun anakütle üzerindeki doęru etkiyi bulmasını ve yanlış bir H_0 hipotezini reddetme şansını oldukça azaltacaktır.

Yeterli gücün saęlanamaması ile ilgili olarak ortaya çıkan problemleri ilk kez 1962 yılında Jacop Cohen, *Journal of Abnormal and Social Psychology* isimli dergide yayımladığı ve güç analizi konusunda iz bırakan bir çalışma olan “*The Statistical Power of Abnormal-Social Psychological Research: A Review*” isimli makalesinde tanımlamıştır. Cohen, istatistiksel güç ile ilgili gerekli noktalara o tarihte değinmiştir. Ancak çoęu arařtırmacı tarafından yeterli güç saęlama sorununa doğrudan deęil de arařtırmada kullanılacak örneklem sayısı endişesi nedeniyle dolaylı olarak yaklaşılmıştır.

Cohen’in 1962 yılındaki çalışmasından sonra ilerleyen yıllarda bazı istatistikçiler yaptıkları çalışmalarda güç analizi konusundaki endişelerini dile getirmeye başlamıştır. Shavelson bu konuda řu şekilde görüş belirtmiştir; “Arařtırmacılar, çalışmalarının güç ile ilgili olan kısmını dikkate almalılar ve çalışmalarını mümkün olduęunca fazla istatistiksel güç üzerine inşa etmelidirler.” Dięer birçok istatistikçi gibi Shavelson da arařtırmanın tasarım aşamasında güç adımının dikkate alınarak, zaman ve maliyet açısından tasarruflu davranmanın çok mantıklı bir davranış olacaęını belirtmiştir (Tener, 2000, s. 6).

Sosyal bilimler alanındaki çalışmalarda istatistiksel güç sürecini uygulamalı olarak anlatmayı amaçlayan tezin birinci bölümünde, istatistiksel güç analizini saęlıklı bir şekilde gerçekleřtirebilmek için bilinmesi gereken temel istatistiksel kavramlara yer verilmiştir. Bu bölümde genel olarak hipotez test etme sürecinden ve hipotez test etme sürecinde yapılan hatalardan bahsedilmiştir. Ayrıca, istatistiksel güç kavramı detaylı bir şekilde açıklanmıştır. Keza istatistiksel güç, istatistiksel güç analizi, istatistiksel gücü etkileyen faktörler, istatistiksel güç analizi çeřitleri ve daha önce gerçekleştirilmiş istatistiksel güç analizi çalışmaları hakkında geniş bir bilgiye yer verilmiştir.

İkinci bölümde ise istatistiksel güç analizi konusunun daha iyi anlaşılması amacıyla uygulama çalışması yapılmıştır. Uygulamada, Yükseköęretim Kurulu’nun Ulusal Tez Merkezi’ne kayıtlı 2005-2011 yılları arasında “İnsan Kaynakları” alanında tamamlanmış olan yüksek lisans ve doktora tezleri incelenmiş ve istatistiksel güçleri hesaplanmıştır.

Sonuç bölümünde ise arařtırma bulguları, deęerlendirme ve istatistiksel güç analizi konusunda arařtırmacılara bazı önerilere yer verilmiştir.

BİRİNCİ BÖLÜM

İSTATİSTİKSEL GÜÇ ANALİZİ

Bu bölüm, çalışmayı desteklemek için gerekli yazın bilgisini içermektedir. İstatistiksel güç ve istatistiksel güç analizi kavramının daha iyi anlaşılması amacı ile bu bölümde önce hipotez test süreci ve hipotez test sürecinde yapılan hatalardan bahsedilmiştir. Daha sonra istatistiksel güç, istatistiksel güç analizi, istatistiksel gücü etkileyen faktörler, istatistiksel güç analizi çeşitleri ve daha önce gerçekleştirilmiş istatistiksel güç analizi çalışmaları hakkında geniş bir bilgiye yer verilmiştir.

1.1 İstatistiklerin Kullanılması

İstatistikler, herhangi bir araştırma için toplanan verileri yorumlamanın bir yoludur ve araştırmacılar istatistikleri genellikle 3 yolla kullanmaktadırlar. Bunlardan birincisi, üzerinde çalışılan araştırma örnekleminin ve verilerin karakteristiklerini açıklamaktır. Bu karakteristikleri açıklamanın en yaygın yolu ise mod, medyan ve standart sapma gibi eğilim ölçüleridir (Tener, 2000, s. 15).

İstatistiklerin bir başka kullanım yolu ise veri setleri arasındaki ilişkileri kurmak ve bunları test etmektir. Buradaki ilişkiden kasıt, değişkenler arasındaki korelasyondur ve bunun da en yaygın tipi “Pearson product moment r korelasyonu” yani basit korelasyondur. Pearson r korelasyonunun korelasyon katsayısı olarak adlandırılan sayısal değeri, iki değişken arasındaki ilişkinin ölçülmesidir. Herhangi iki değişken arasında pozitif veya negatif yönde korelasyon olabilmektedir. Bu yüzden korelasyon katsayısının değeri +1 ile -1 arasında değişir. “+1”, iki değişken arasında mükemmel pozitif korelasyon olduğu, “-1”, iki değişken arasında mükemmel negatif korelasyon olduğu ve “0” ise iki değişken arasında herhangi bir ilişkinin olmadığı anlamına gelmektedir (Ellis, 2010, s. 11).

İstatistiklerin kullanıldığı üçüncü yol ise veri grupları arasındaki farklılıkları ölçmektir. Bu süreç, araştırma için toplanan veri gruplarından bir değer hesaplanmasını ve yapılan bu hesaplamalardan sonra elde edilecek bulguların, eğer araştırılan veri grupları arasında anlamlı farklılıklar varsa karşılaştırılmasını gerektirmektedir. Yaygın olarak bu ölçümü yapmak için kullanılan testler; t testi (iki grup arasındaki farklılıkların ölçülmesi) ve varyans analizidir (ANOVA, iki veya daha fazla sayıdaki bağımsız değişken arasındaki farklılıkların ölçülmesi)(Sipahi vd., 2008, s. 118,124)

Geliştirilen istatistiksel yöntemler araştırmacılara çok sayıda verinin analiz edilmesi ve yorumlanması imkânı sağlamaktadır. Örneğin, 200 kişilik lise öğrencisinin oluşturduğu bir grup üzerinde alkol ve sigara kullanma oranını belirlemek amacıyla bir araştırma yapılacaktır. Toplanan veriler analiz edildikten sonra ortalamalar, standart sapma vs. gibi istatistiksel bilgiler araştırmacıya sonuçları yorumlama imkânı sağlayacaktır. Anakütleyi temsilen seçilen örneklemden elde edilen istatistiklerin amacı, sonuçları doğru yorumlayarak anakütle üzerine genelleştirme yapabilmektir. Çünkü bu konuda araştırma yapan birisi mevcut tüm lise birinci sınıf öğrencileri ile alkol ve sigara kullanma konusunda görüşme yapamayacağı için, belirli bir gruptan elde edeceği verileri tüm lise birinci sınıf öğrencileri üzerine genelleştirecektir (Coblick, 1998, s. 66).

Genellikle istatistiksel tekniklerin kullanılması konusunda yanlışlıklar vardır. Betimleyici ve anlam çıkarıcı istatistikler sanki birer istatistiksel teknikmiş gibi tanımlanmaktadır. Betimsel istatistikler (descriptive statistics), belirli bir gruba ait gözlemlere ilişkin sayısal verilerin özetlenmesi, düzenlenmesi ve sınıflanmasıyla ilgilenen istatistiklerdir. Anlam çıkarıcı istatistikler (inferential statistics) ise tanımlanmış bir evrenin yansız bir örnekleme ait istatistiklere dayalı olarak evren hakkında genellemeler ve tahminler yapmayı amaçlayan istatistiklerdir (Büyüköztürk, 2010, s. 11). Eğer örneklem uygun seçildiyse ve seçilen bu örneklem anakütleyi temsil ediyorsa o zaman elde edilen bulgular anakütle üzerine genelleştirilebilir. Anlam çıkarıcı istatistik örneklem üzerine yoğunlaşmaktadır, kullanılan istatistiksel teknik ile bir ilgisi yoktur.

1.2 Hipotez Test Etme Süreci

Herhangi bir araştırma için veri toplama sürecinden önce hipotezler veya fikirler oluşturulmalıdır. Araştırmacılar hipotez testinde, düşündüğü yargı hakkında karar verebilmek için istatistiksel tekniklerle topladığı veri setinden hareketle iddialarını test etmektedirler (Sipahi, vd., 2008, s. 11).

Hipotezlerin, anakütleden seçilen örneklem yardımıyla istatistiksel olarak incelenmesine “*hipotez testi*” denir. Araştırmalarda test edilmesi gereken asıl hipotez sıfır hipotezidir (H_0). Araştırmacılar, H_0 hipotezi ile araştırılan anakütlede herhangi bir etkinin var olmadığını varsayarlar. İstatistikler, gruplar arası farklılıkları ölçmek için kullanıldığında araştırmacı, eğer grupların ortalamaları arasında anlamlı bir farklılık varsa bunu test eder. İstatistikler, değişkenler arasındaki ilişkileri test etmek için kullanılırsa araştırmacı, araştırılan değişkenler arasında ilişki olup olmadığını göstermeye çalışır (Gürkan, 2007, s. 124).

Araştırmacılar, araştırma yaptıkları herhangi bir alanda araştırmak istedikleri olayları hipoteze dönüştürürler ve bağımlı veya bağımsız değişkenler arasındaki ilişkileri, farklılıkları eğer gerçekte var ise belirlemeye çalışırlar. Klasik istatistiksel testlerde araştırmacılar araştırdıkları olayın anakütle üzerinde var olmadığını veya değişkenler arasında herhangi bir ilişkinin olmadığını varsayarak sıfır hipotezini kurarlar ve bunu reddetmeye çalışırlar (Cashen ve Geiger, 2004, s. 154).

Araştırmacılar tarafından, sıfır hipotezinin yanında ayrıca bir de alternatif hipotez (H_1) belirlenmekte ve bu alternatif hipotez ile araştırma altındaki anakütle üzerinde bir etkinin var olduğu kabul edilmektedir. Bir araştırmacı, grup ortalamalarının eşit olmadığını veya bir grubun ortalamasının diğer gruptan daha büyük olduğunu söyleyebilir. Ayrıca araştırmacı ilişkiler üzerinde çalışıyorsa “araştırılan iki değişken arasında ilişki vardır” veya “ilişki pozitifdir, ilişki negatiftir” şeklinde hipotezler oluşturabilir. Eğer kurulan hipotezler ile araştırmacının iddiası “büyüktür, küçüktür, büyük eşittir, küçük eşittir” şeklinde yön gösteriyor ise bu tarz hipotez testine “*tek yönlü hipotez testi*” denir. Tersine araştırmacının hipotezleri yön göstermeyip sadece eşitlik içeriyorsa bu tarz hipotezlere de “*çift yönlü hipotez testi*” denir (Sipahi vd., 2008, s. 12).

Araştırmacıların alternatif hipotezlerini seçerken mutlaka ve mutlaka önceden yapılmış çalışmaları incelemeleri gerekmektedir. Eğer daha önce yapılan çalışmaların herhangi birinde çalışılmak istenen konu ile ilgili bir hipotez test edilmişse, daha detaylı bir araştırma yapmaya olanak sağlayacak bir alternatif hipotez seçilmelidir. Her durumda H_0 hipotezi reddedilmek için kurulmaktadır. Cohen (1990), H_0 hipotezine neden sıfır hipotezi denildiğini şu şekilde açıklamıştır: “Bu hipotez sıfır hipotezi olarak adlandırılır. Çünkü uygulanan bütün stratejiler bu hipotezi boşa düşürmek, hükümsüz kılmak, anlamsız kılmak, etkisiz kılmak içindir. Eğer sıfır hipotezinin reddedilmesi başarılırsa, dolaylı olarak alternatif hipotez kabul edilmiş olacaktır” (Cohen, 1990, s. 8).

Anlamlılık seviyesi ise veri toplamadan önce belirlenmesi gereken bir diğer önemli konudur. Alfa (α) ile gösterilen ve olasılığı “ p ” olan anlamlılık seviyesi, araştırmacı için sıfır hipotezinin reddedilip reddedilmeyeceği konusunda bir kriterdir. Belirlenen anlamlılık seviyesi, daha sonra alınacak kararları da etkileyeceği için dikkatle seçilmelidir. Eğer örneklemden elde edilen değer, anlamlılık seviyesinden küçük veya eşit ise sıfır hipotezi reddedilir. Bu değer anlamlılık seviyesinden büyük ise sıfır hipotezi reddedilemez.

Özetle, arařtırmacı önce test edilecek bir sıfır hipotezini daha sonra iliřki veya farklılıkları test etmek için bir anlamlılık seviyesi belirler. Arařtırmacı daha sonra elde ettiđi verilere belirli istatistiksel teknikler uygular ve uygulanan bu tekniklerin sonucunda belirli deđerler elde eder. Daha sonra elde edilen bu deđerler anlamlılık seviyesi ile karřılařtırılır. Ayrıca bilgisayarlar aracılıđıyla son yıllarda kullanılmakta olan istatistik yazılımları ile belirli istatistik deđerlerinin olasılıkları herhangi bir tabloya ihtiyaç duyulmadan kolayca hesaplanabilmektedir. Eđer hesaplanan bu olasılık deđeri anlamlılık seviyesine eřit veya ondan daha az ise ($p \leq \alpha$) ise H_0 reddedilmektedir (Tener, 2000, s. 20).

Hipotez testi ile ilgili verilen eđitimlerde çođu zaman H_0 hipotezinin reddedildiđi anda arařtırmanın “*herhangi bir etki bulunamamıřtır*” řeklinde sonuçlanacađı söylenmektedir. Bu řekilde yapılan bir yorum birçok kusur barındırmaktadır. “ p ” deđerı, genellikle yanlış olarak sıfır hipotezinin reddedilmesi veya kabul edilmesine karar vermek için kullanılır. Bu noktada eđer test sonuçları istatistiksel olarak anlamlı ise “*etki vardır*”, test sonuçları istatistiksel olarak anlamlı deđil ise “*etki yoktur*” řeklinde bir sonuca varılmaktadır. “ p ” deđerinin kullanılarak etki büyüklüđu hakkında yorum yapmak yanlıřtır. Eđer yeterli istatistiksel güç var ise bir etki ortaya çıkarılabilmek mümkündür. İstatistiksel olarak anlamlı bir sonuç, gerçek bir etkinin var olacađı anlamına gelmez. Bazı durumlarda II. Tip hata yapma olasılıđı kaçınılmazdır. Hangi sonucun gerçekten gerçek etkiye sahip olup olmayacađına karar verebilmek için yapılan testler farklı zamanlarda tekrar edilmelidir. Bu nedenle sadece “ p ” deđerine bakarak arařtırmayı sonuçlandırmak oldukça yanlış bir harekettir (Ellis, 2010, s. 49).

Güçlü bir anlamlılık testi, iki grup arasındaki çok küçük farklılıkları da belirlemektedir. Böylece sıfır hipotezini reddetmek de kolaylařacaktır. Bu nedenle güçlü bir anlamlılık testi, arařtırmacılara sıfır hipotezinin reddedilmesi konusunda iyi bir yol gösterici olacaktır (Deng, 2000, s.3).

1.3 Hipotez Test Etme Sürecinde Yapılan Hatalar

Arařtırmacıların, dođru yöntemler uygulamalarına rađmen hipotez test sürecinde hata yapma olasılıkları her zaman vardır. Ayrıca benzer problem sıfır hipotezinin reddedilmesinde veya reddedilememesinde meydana gelebilecek bir yanlışlık ile de oluşabilmektedir. Bazen arařtırmacı, gerçekte dođru olmasına ve reddedilmemesi gerekmesine rađmen sıfır hipotezini reddeder. Bu hata “*I. Tip Hata (false positive)*” olarak bilinmektedir. Bazen ise sıfır hipotezi, gerçekte yanlış olmasına ve reddedilmesi gerekmesine rađmen eldeki mevcut bulgulara dayanılarak reddedilemez. Bu tür bir hata da “*II. Tip Hata (false negative)*” olarak

adlandırılmaktadır (Ottenbacher, 1982, s. 15; Mone, Mueller ve Mauland, 1996, s. 103; Mazen, Graf ve Kellogg, Hemmasi, 1987, s. 369; Balkin ve Sheperis, 2011, s. 268; Rotenberry ve Patterns, 1985, s. 164).

Lieber, başka bir bakış açısıyla iki tip hatadan bahsetmiştir. Gerçekte sıfır hipotezi ya doğrudur ya da yanlıştır ve araştırmacı sıfır hipotezini ya kabul eder ya da reddeder. Bu kararların sonucunda kombinasyon olarak olası dört durum vardır. Bunlardan ikisi doğru diğer ikisi de yanlış kararlar olacaktır. Yani sıfır hipotezi yanlış olduğunda reddedilmesi ve sıfır hipotezi doğru olduğunda reddedilmemesi doğru kararlardır. Tablo 1.1, olası durumları ve doğru kararları özetlemektedir. Lieber ayrıca II. Tip hatayı kontrol etmenin önem olarak I. Tip hataya eşit hatta I. Tip hatadan daha önemli olduğunu belirtmiştir. Araştırmacılar, elbette ne I. Tip hatayı ne de II. Tip hatayı yapmak istemektedir. Fakat araştırma sonucunda elde edilen “ p ” değeri, basit bir şekilde araştırmada yapılan I. Tip hata oranını göstermektedir. Araştırmacılar çalışmalarını “ $p < 0.05$ olduğundan, çalışmada %5’ten daha az oranda I. Tip hata yapıldı ve bu nedenle anlamlı sonuçlar elde edildi” şeklinde sonuçlandırmaktadırlar (Lieber, 1990, s. 305).

Ağıdaki tabloda hipotez testi sürecinde yapılabilecek olası hatalar gösterilmiştir.

Tablo 1.1 Hipotez Test Sürecindeki Hatalar

<i>Kararlar</i>	<i>H₀ Doğru</i>	<i>H₀ Yanlış</i>	<i>Kontrol Aracı</i>
H ₀ Red, H ₁ Kabul	I. Tip Hata	Doğru	Anlamlılık Seviyesi
H ₀ Kabul, H ₁ Red	Doğru	II. Tip Hata	İstatistiksel Güç

Kaynak: Abraham ve Russel, 2008, s. 284.

Mantıken, aynı hipotez testinde sadece tek tip hata meydana gelmektedir. Her iki tip hata eşzamanlı olarak aynı anda meydana gelmez. Doğası gereği II. Tip hata sıfır hipotezinin reddedildiği bir durumda, I. Tip hata ise sıfır hipotezinin reddedilmediği bir durumda meydana gelmez (Clark, 1996, s. 10).

Bir anlamlılık testinde sıfır hipotezi, alternatif hipotezi geçerli kılmak için reddedilecek hedef olarak kurulmaktadır. Sıfır hipotezi reddedilirken, araştırmacılar I. Tip veya II. Tip hata yapma olasılıklarının farkındadırlar. Bu farkındalığı göstermek için araştırmacılar

çalışmalarını gerçekleştirmeden önce kabul edilebilir bir seviyede I. Tip hata yapma olasılıklarını anlamlılık seviyesi (α) olarak belirlerler. Fakat I. Tip hata seviyesi çalışma öncesinde belirlenebilirken aynı şey II. Tip hata için geçerli değildir. Çünkü II. Tip hata yapma olasılığına (β) sıfır hipotezi ve alternatif hipotezin değerleri belirlenene kadar karar verilemez (Deng, 2000, s. 14).

Araştırmacı, bu iki tip hata arasındaki ters ilişkiyi göz önüne alarak araştırmanın tasarım aşamasında dikkatli kararlar vermelidir ve her iki tip hatayı da göz önünde bulundurmalıdır. Araştırmacı, bu kararları vermeden önce hangi tip hatanın daha ciddi sonuçlar doğuracağına karar vermelidir. I. Tip hata, araştırmacıyı anakütle üzerinde var olmayan bir etkiye güvenmesi gibi bir çıkmaz yola sürükleyebilir. Özellikle keşfedici (exploratory) çalışmalarda II. Tip hata, gelecekte kullanılacak bazı faktörlerin veya bütünüyle teorinin elimine olmasına neden olabileceği için çok maliyetli sonuçlar doğurabilmektedir (Mazen vd., 1987, 369).

Tener'in (2000) aktardığına göre, Keppel ve Shavelson, araştırmacıların I. Tip ve II. Tip hatalarının meydana gelmesi konusunda sağduyulu olmalarından, olası hatalara karşı araştırmanın tasarım aşamasında dikkatli olmalarından ve buna göre ciddi önlemler almaları gerektiğinden bahsetmişlerdir. Keppel ve Shavelson herhangi bir araştırma üzerinde araştırmacının her iki tip hatayı da kontrol etmeyi sağlayacak yeterli bir tasarım yapmaya ihtiyacının olduğunu ve tasarım aşamasında dengeli bir yaklaşım (balanced approach) gerekli olmasına rağmen, bunu söylemesinin yapmasından daha kolay olduğunu belirtmişlerdir (s. 24).

Geleneksel olarak, davranış bilimleri ve sosyal bilimler alanındaki araştırmacılar I. Tip hata yapma ihtimaline, II. Tip hata yapma ihtimalinden daha fazla önem vermektedirler. Gidişat bu şekilde olmasına rağmen araştırmacılar her iki hataya da önem vermek zorundadırlar (Ottenbacher, 1982, s. 16).

Araştırmacılar, genellikle alfa ve beta arasındaki zıt ilişkinin farkındadır. Birisi artarsa diğeri azalmaktadır. Genellikle araştırmalarda bütün dikkat I. Tip hataya verilmekte ve bu durum kişiyi sadece I. Tip hataya karşı korumakta, II. Tip hataya karşı bir koruma sağlamamaktadır. En iyi araştırma tasarımı, her iki tip hatanın da dikkate alınarak yapılan tasarımıdır. I. Tip hata H_0 doğru, II. Tip hata H_0 yanlış iken yapılır. H_0 aynı anda hem doğru hem yanlış olmayacağı için her iki hata da aynı anda ortaya çıkmaz. Eğer araştırma altında bir etki gerçekten var ise I. Tip hata yapılma olasılığı yoktur. Bir etki elde edildiğinde sadece II. Tip hata yapılma olasılığı vardır ve bu sadece araştırma düşük istatistiksel güç ile

gerçekleştiriliyorsa mümkündür (Ottenbacher, 1982, s. 15; Ellis, 2010, s. 52; Balkin ve Sheperis, 2011, s. 269).

Davranış bilimleri arařtırmalarında, yanlış bir sıfır hipotezinin kabul edilmesi (II. Tip hata) genellikle doğru bir sıfır hipotezinin reddedilmesinden (I. Tip hata) daha ciddi sonuçlar doğurmaktadır. Örneğın, davranış bilimleri arařtırmalarında yanlış bir sıfır hipotezi reddedilmezse, gerçekte uygulanan mevcut yöntem veya uygulamalardan daha iyi olan yöntemler kabul edilmeyecek ve böylece bu durum, yeni yöntem ve uygulamaların geliştirilmesi, tasarlanması aşamasında harcanan kaynakların boşa gitmesine neden olacaktır. Eğer doğru bir sıfır hipotezi reddedilirse, mevcut yöntem ve uygulamalardan daha iyi olmayan yöntem ve uygulamalar kullanılacak, bu durumda yeni yöntem geliřtirmek için harcanan kaynaklar boşa gitmemiş olacaktır (Deng, 2000, s. 2).

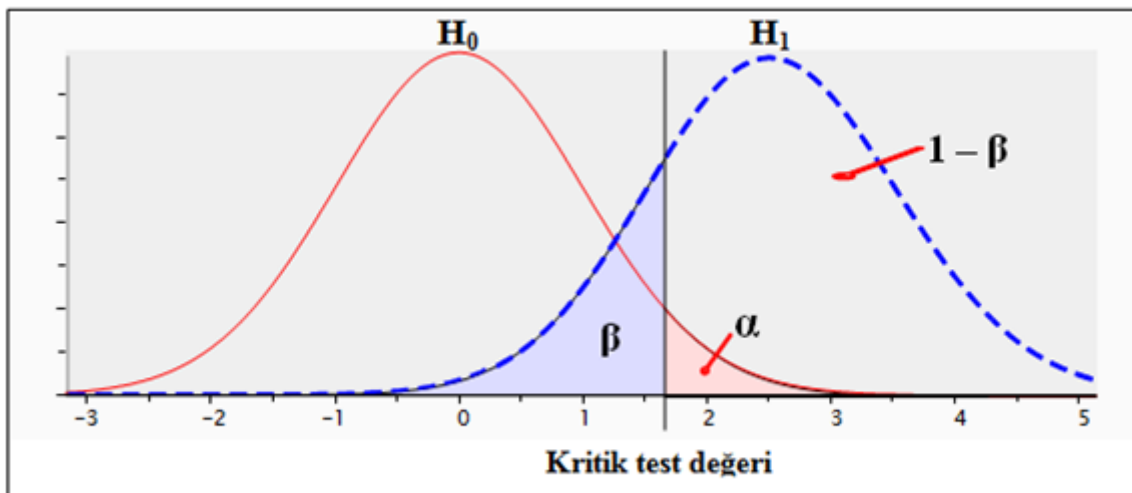
Tener'in (2000) aktardığına göre, Huck, Bornier ve Bounds, Keppel, Shavelson, Thomas ve Nelson, hata kontrolü konusunda řu şekilde açıklama yapmışlardır: "Arařtırmacılar aynı zamanda anlamlılık seviyesi için de kullanılan alfayı kullanarak I. Tip hatayı kontrol altında tutabilirler. Alfa, sadece anlamlılık seviyesi kararında kullanılmaz. Aynı zamanda doğru olan bir sıfır hipotezinin reddedilme ihtimalini de temsil eder. Alfa, 0.10'dan 0.05'e doğru veya 0.01'e doğru azaldığında, arařtırmacı I. Tip hata yapma olasılığını da azaltmış olur ve aynı zamanda sıfır hipotezi doğru olduğunda doğru karar verme olasılığını da artırmış olur" (s. 25).

Arařtırmacılar, beta seviyesini de II. Tip hatanın kontrol altında tutulması konusunda kullanabilirler. Beta'nın sınırlanması, kısıtlanması II. Tip hata yapma olasılığını azaltacak, aynı zamanda sıfır hipotezinin yanlış olduğu durumda doğru karar verme olasılığını artıracaktır.

Cashen ve Geiger, yazdıkları bir makalede yazında I. Tip hatanın II. Tip hatadan daha ciddi olarak kabul gördüğünü belirtmişlerdir. Bunu göze alan arařtırmacılar, I. Tip hataya daha katı kurullarla yaklaşmakta ve kendilerini bu şekilde savunmaktadırlar. Ancak bazı yazarlar hata riskinin II. Tip hata payını da göz önüne alıp her iki hataya da dağıtılması gerektiğini savunmuşlardır. Bu konuda (β/α) oranının makul bir düzeyde ayarlanması gerekmektedir (Cashen ve Geiger, 2004, s. 154).

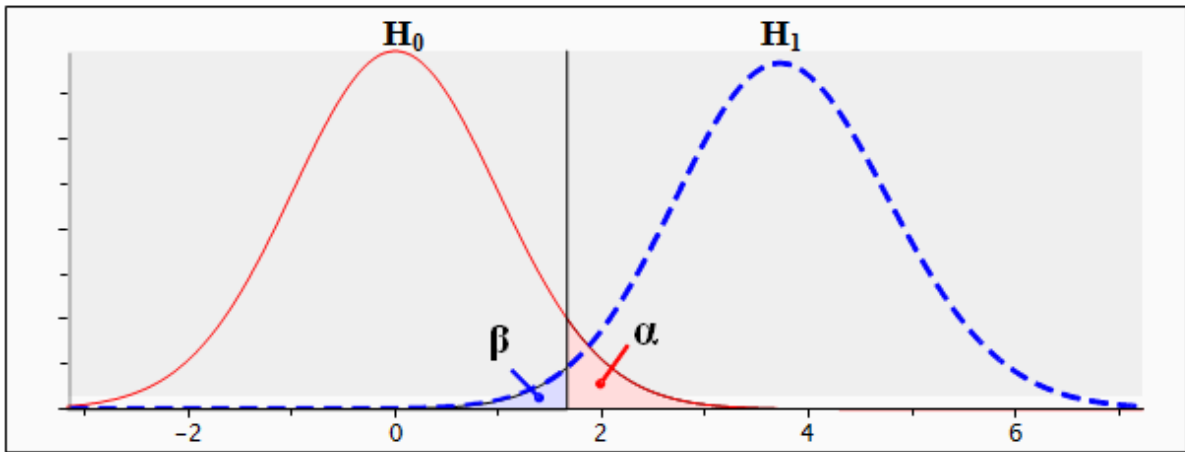
Hangi tip hatanın daha ciddi olduğunun arařtırmanın çeřitine ve arařtırmacının insiyatifine bağılıdır. Örneğın, akarsu içindeki toksin miktarını test etmek için planlanan bir çalışmada II. Tip hata yapmanın maliyeti (normalde akarsu içinde toksin varken arařtırmayı toksinin olmadığı şeklinde sonuçlandırmak), I. Tip hata yapmanın (akarsu içinde normalde toksin yok

iken arařtırmayı toksinin olduđu yönünde sonuçlandırmak) maliyetinden çok daha fazladır. Bu şekilde yapılacak bir II. Tip hata ekolojik çevrenin zarar görmesine ve hatta suda yaşayan birçok canlının ölümüne neden olacaktır. Bir başka arařtırmada ise yeni uygulanacak bir makineli tarım sisteminin önceden uygulanan fiziksel güce dayanan sistemden daha iyi olup olmadığı arařtırılmak istensin. Bu arařtırmada yapılacak olan I. Tip hatanın maliyeti (gerçekte daha iyi olmamasına rağmen yeni teknolojinin daha iyi olduđu sonucuna varmak) daha fazladır. Yeni teknolojinin uygulanması, bir sistemin tamamen deđişmesi ve insanların iş kayıpları yaşaması anlamına geleceđi için büyük bir maliyet demektir. II. Tip hata durumunda ise (normalde daha iyi olmamasına rağmen eski sistemin daha iyi olduđu sonucuna varmak) çok fazla maliyet olmayacaktır. Bu nedenle hangi tip hatanın daha ciddi olduđuna karar vermek, büyük ölçüde arařtırmanın tipine bađlıdır (Stephano, 2003, s. 707).



Şekil 1.1 H_0 ve H_1 Hipotez Dađılımlarında I. Tip Hata - II. Tip Hata ve İstatistiksel Güç Bölgelerinin Gösterilmesi

Şekil 1.1'de $\alpha = 0.05$ ve $\beta = 0.20$ olduđu varsayılırsa güç = $1 - \beta = 0.80$ olacaktır. Şekil 1.2'de ise H_0 ile H_1 hipotezlerinin tepe noktaları arasındaki fark artmış (Şekil 1.1'de H_1 ortalaması yaklaşık 2.5 iken, Şekil 2.2'de H_1 ortalaması 4 olmuştur), daha büyük bir etki belirlenmiş, II. Tip hata oranı düşmüş ve tüm bunların sonucu olarak istatistiksel güç artmıştır.



Şekil 1.2 H_0 ve H_1 Hipotezlerinin Tepe Noktaları Arasındaki Farkın İstatistiksel Güce Etkisi

1.4 İstatistiksel Güç Kavramı

“Güç analizine rastladığımda, sanki ölmüş ve cennete gitmiş gibiydim.” (Jacop Cohen)

Cohen, istatistiksel gücü; “yanlış bir sıfır hipotezinin reddedilme olasılığı ($1-\beta$)” olarak tanımlamıştır. Dahası bir istatistiksel testin gücü, istatistiksel olarak anlamlı sonuçlar sağlayıp sağlamayacağı olasılığıdır (Cohen, 1988, s. 1; O’Keefe, 2007, s. 291; Balkin ve Sheperis, 2011, s. 268). Joseph Rossi ise olması gerektiğini düşündüğü güç tanımını “bir testin, anakütle üzerinde gerçek bir etki olduğu varsayımı altında anlamlı bir sonuç sağlayıp sağlamayacağı olasılığıdır” şeklinde yapmıştır (Rossi, 2012, s. 21). Jones ve Sommerlund (2007) ise istatistiksel gücü, bir araştırmanın sağlığının ölçülmesi olarak tanımlamışlardır (s. 225).

İstatistiksel güç kavramının geliştirilmesi Jerzy Neyman ve Agon S. Pearson’a atfedilebilir. Onlar bu kavramı ilk olarak 1928 yılında tanıtmışlardır. Jerzy Neyman ve Agon S. Pearson, 1920’li yılların sonundan 1930’lu yılların ortasına kadar geçen sürede kendi istatistiksel teorileri içinde güç, beta ve II. Tip hata kavramlarından bahsetmişlerdir. Fakat bu kavramlar, tüm zamanların bilinen en iyi istatistikçisi ve varyans analizi dâhil olmak üzere birçok istatistiksel tekniğin mucidi olan R.A. Fisher tarafından yaşamı boyunca katı bir şekilde reddedilmiştir. Fisher, kendi teorisinde, sadece sıfır hipotezinin kullanılması gerektiğini savunurken, Neyman-Pearson ise ortaya attıkları teorilerinde hem sıfır hipotezinin hem de alternatif bir hipotezin olması gerektiğini savunmuşlardır.

Her iki teorinin de ortaya çıkmasından sonra güç konusunun ihmal edilmeye devam edilmesinin nedenlerinden biri de her iki teorinin ileriki yıllarda “*Hibrit Teori*” adı altında birleştirilmiş olmasıdır. Araştırmacılar tarafından önce sadece Fisher’ın teorisi, 1946 yılından sonra ise Fisher’ın teorisi ile birlikte Neyman-Pearson teorisi de kullanılmaya başlanmıştır. Neyman-Pearson’ın teorisinin de kullanılmaya başlanmasıyla araştırmacıların ve öğrencilerin kafası iyice karışmıştır. Çünkü Fisher’ın teorisinde güç hesaplama diye bir kavram yoktu ve bu teoride güç hesaplanmazdı. Ayrıca Neyman-Pearson’ın teorisi araştırmacılara tamamen anlatılmamış ve her iki teorinin çatıştığı noktalar aydınlatılmamıştı (DiLullo, 1987, s. 38)

Neyman-Pearson teorisi ile Fisher’ın teorisi çok sayıda noktada çatışmasına rağmen, sorunun çözülemeyen en önemli noktası hipotez test etme konusundadır. Aşağıda her iki teorinin birbirinden ayrıldığı konular açıklanmıştır.

1.5 Neyman-Pearson Teorisi ve Fisher Teorisi Arasındaki Tartışmalı Konular

- 1) **Hipotez Testi:** Neyman-Pearson teorisi H_0 ve H_1 hipotezlerini kullanmaktadır. H_0 hipotezi deney ve kontrol grupları arasında herhangi bir farkın olmadığını ifade ederken, deney ve kontrol grupları arasındaki beklenen fark H_1 hipotezi ile ifade edilmektedir. Fisher’ın teorisinde ise sadece H_0 hipotezi kullanılmaktadır. Bu teoride de H_0 hipotezi deney ve kontrol grupları arasında farkın olmadığını göstermektedir. Fakat gruplar arasında herhangi bir fark elde edilirse bu durum sadece “gruplar arasında bir fark vardır” şeklinde ifade edilmekte, ancak bu farkın büyüklüğü ile ilgili bir ifadeye yer verilmemektedir.
- 2) **Hipotez Testinin Yorumlanması:** Neyman-Pearson teorisinde H_0 ’ı reddetmek için bir karar testi yapılmaktadır. Bu karar özel bir gerekçeye dayanarak verilmektedir. Fisher’ın teorisinde ise H_0 hipotezini reddetmek için bir anlamlılık testi yapılmaktadır.
- 3) **Anlamlı Olmayan Sonuçların Yorumlanması:** Neyman-Pearson teorisinde araştırmacı anlamlı olmayan bir sonuç ile karşılaştığında özel bir gerekçeye dayanarak H_0 hipotezini kabul etmektedir. Fisher’ın teorisinde ise araştırmacı derhal H_0 hipotezini kabul etmekte ve çalışmasını “deney ve kontrol grupları arasında herhangi bir fark yoktur” diyerek sonlandırmaktadır.
- 4) **Anlamlı Sonuçların Yorumlanması:** Neyman-Pearson teorisinde araştırmacı bu durumda özel bir nedene dayanarak H_0 hipotezini reddeder ve kontrol grubu üzerinde bir etkinin var olduğunu belirtir. Fakat H_0 hipotezinin ne derece geçersiz olduğu veya

başka bir deyişle kontrol grubu üzerinde var olduğu düşünülen etkinin boyutuna karar verme işlemi yapılan bir dizi istatistiksel testten sonra gerçekleştirilir. Fisher'ın teorisinde ise araştırmacı çalışmasını H_0 geçersizdir, deney ve kontrol grupları arasında bir fark vardır diyerek sonlandırır. Fakat aradaki bu belirlenen farkın büyüklüğü konusunda herhangi bir açıklama yapılmaz.

- 5) **Etki:** Neyman-Pearson teorisinde H_0 ve H_1 hipotezleri arasında fark etki olarak açıklanmaktadır. Deney ve kontrol grupları arasındaki farkın nedeni bu teoriye göre bir “etki” dir. Bu nedenle etkiyi ifade etmek için “etki büyüklüğü” terimi kullanılmaktadır. Fisher'ın teorisinde ise etki diye bir kavramdan bahsedilmemektedir.
- 6) **Alfa:** Alfa, Neyman-Pearson teorisinde çalışma gerçekleştirilmeden önce belirlenmekte ve tüm istatistiksel testler boyunca sabit kalmaktadır. Fisher ise alfa seviyesinin 0.05 olması gerektiğini, çalışmadan önce ifade edilmesine gerek olmadığını ve çalışma süresince sabit olmayabileceğini ifade etmiştir.
- 7) **I. Tip Hata:** Neyman-Pearson teorisinde I. Tip hatanın zararlı olduğu ifade edilmektedir. Bu hata, yapılan istatistiksel testler sonucunda H_0 hipotezinin kabul edilmesi gerekirken yanlış bir şekilde reddedilmesi anlamına gelmektedir. Fisher ise bu hatanın yapılabilecek zararlı bir işlem olduğunu ve istatistiksel testlerin sonuçlarının yorumlanması açısından dikkate alınır, araştırmacının bu hatayı yapma olasılığını göze aldığını belirtmektedir.
- 8) **II. Tip Hata:** Neyman-Pearson teorisinde II. Tip hatanın da zararlı olduğu ifade edilmektedir. Bu hata, yapılan istatistiksel testler sonucunda yanlış olan H_0 hipotezinin reddedilmesi gerekirken yanlış bir şekilde reddedilememesi anlamına gelmektedir. Fisher'ın teorisinde ise bu tip bir hataya zararlı olarak yer verilmemiştir. Çünkü bu teoride alternatif hipotez olmadığından II. Tip hata hesaplanmamaktadır.
- 9) **Güç:** Neyman-Pearson teorisinde güç, yanlış olan bir H_0 hipotezinin reddedilme olasılığını ifade etmektedir. Bu teoride güç seviyesine etki büyüklüğü, örneklem sayısı ve alfa seviyesi kullanılarak karar verilmektedir. Fisher'ın teorisinde ise güç diye bir kavram yoktur ve hesaplanmamaktadır.

Cohen (1965), anlamlı olmayan sonuçların yorumlanması noktasında Neyman-Pearson teorisinin daha avantajlı olduğunu ifade etmiştir. Çünkü Cohen'e göre, sıfır hipotezinin kabul edilmesi, araştırma yapılan örneklem üzerinde herhangi bir etkinin kesinlikle var olmadığı

anlamına gelmemektedir. Bazen etki gerçekten yoktur. Bazen ise vardır ama çok küçüktür. Bu etkinin boyutu Neyman-Pearson teorisinde açıklanmaktadır. Ancak Fisher'in teorisinde etki büyüklüğü konusuna yer verilmediği için kabul edilen bir H_0 hipotezi sadece etki yoktur anlamına gelmektedir. Bu konuda Cohen şu örneği vermiştir: “ β 'yı α ile aynı seviyede belirleyerek bir çalışma gerçekleştirelim (bu demektir ki çalışma %95 gibi yüksek bir güç ile gerçekleştirilecektir). Belirlemek istediğimiz etkinin değerini ise D olarak düşünelim. Eğer çalışma sonucunda elde edilecek etki büyüklüğü değeri bu D değerinden düşük olursa H_0 hipotezi reddedilemeyecek ve gözlemlenen etki ihmal edilebilecektir. Bu noktada H_0 hipotezinin reddedilmesi etkinin olmadığı anlamına değil, belirlenmek istenen seviyede olmadığı anlamına gelmektedir. Eğer gözlemlenecek etki D seviyesine eşit olursa bu %95 olasılık ile H_0 hipotezini reddedebileceğimiz anlamına gelmektedir”(DiLullo, 1997, s.34-37).

1962 yılında Jacop Cohen, güç konusundaki çığır açıcı çalışmasını yayımlamış ve istatistiksel güç konusunun yeniden doğmasına öncülük etmiştir. Bir dergide istatistiksel güç ile ilgili “*The Statistical Power of Abnormal-Social Psychology Research: A Review*” isimli istatistiksel gücü anlatan makalede Cohen şunları amaçlamıştır:

- i) Güç konusuna araştırmacıların, araştırma yaptıran müşterilerin, araştırmanın planlama ve tasarım aşamasında değerlendirme yapan kişilerin (sponsorlar, dergi editörleri vs.) dikkatini çekmek,
- ii) Araştırmalarını güç analizi üzerine inşa etmek isteyen araştırmacılara rehberlik yapmak ve işlerini kolaylaştırmak amacıyla güç tabloları oluşturmak,
- iii) Güç analizi üzerine inşa edilecek olan çalışmalarını değerlendirebilmek için belirli seviyelerde güçler belirlemek.

Cohen'in bu çalışması sadece psikoloji alanındaki araştırmacılara değil, birçok alandaki araştırmacılara kendi güç analizlerini yapmaları konusunda esin kaynağı olmuştur. Ayrıca şunu da belirtmek gerekir ki, Cohen'in makalesinde verdiği güç ile ilgili değerler bir standart olarak kabul edilmiş ve çoğu araştırmacı, Cohen'in değerlerine bağlı kalarak araştırmalarını farklı alanlarda gerçekleştirmişlerdir (Abraham ve Russel, 2008, s. 283).

Bir araştırmanın tasarım aşamasındaki güç analizinin amacı, anakütlede var olduğu düşünülen etkinin, mümkün olan en yüksek olasılık ile belirlenebilmesini garantilemektir. Araştırma yapılan anakütle üzerinde bir etki varsa, yapılacak olan istatistiksel testler

neticesinde hesaplanacak istatistiksel güç, elde edilecek bulguların istatistiksel olarak anlamlı sonuçlar verip vermeyeceği olasılığıdır (Tener, 2000, s. 26).

İstatistiksel güç, anakütlede gerçekte var olduğu düşünülen özgün bir etkinin doğru olarak belirlenebilme becerisi olarak da tanımlanabilir. Güç, teknik olarak yanlış olan bir H_0 hipotezinin reddedilmesindeki başarısızlık olasılığının tümleyenidir yani $1-\beta$ 'dir. Bu nedenle güç, II. Tip hata yapma olasılığı ile ters orantılıdır. Güç, sayısal olarak 0 ile 1 arasında değişen değerler almaktadır. "1" değeri mükemmel güç anlamına, "0" değeri ise gücün olmadığı anlamına gelmektedir. Cohen, araştırmacılara sağlıklı veriler elde edebilmeleri için 0.80 seviyesindeki güç ile çalışmalarını tavsiye etmiştir. Cohen' e göre bu seviyedeki bir güç, araştırmacıya H_0 hipotezinin kabul edilip edilmemesi konusunda doğru karar verebilme yeterliliği sağlayacaktır. Eğer araştırmanın gücü %80 ise, o zaman araştırmacının H_0 hipotezini reddetmek için %80 oranında şansı vardır anlamına gelmektedir (DiLullo, 1997, s. 3).

Teorik olarak istatistiksel testin gücü, $1-\beta$ şeklindeki matematiksel tanımın ötesindedir. Güç, araştırma altındaki anakütlede bulmak istenilen ilişkilerin veya farklılıkların derecesi ya da sağlanacak sonuçların daha sonra yapılması olası çalışmalarda kullanılabilmesi ihtimalidir. Bu yüzden güç analizi, araştırmacıya yanlış olan bir sıfır hipotezinin reddedilmesine olanak sağlayan gerçek farklılıkların ya da ilişkilerin belirlenmesi yeterliliğini içermektedir. Cohen'e göre güç analizinin anakütle üzerinde bir etki belirlenmesine izin veren istatistiksel testlerin yanında bir faydası da yeterli gücün araştırmacıya H_0 hipotezini kanıtlama fırsatı vermesidir.

İstatistiksel güç, II. Tip hata ile ters ilişkiye sahiptir. Gücün artması, II. Tip hatanın yapılma olasılığını azaltacağı için yeterli güç ile çalışılarak II. Tip hata kontrol altına alınabilir. Alfa'nın 0.05 olarak seçilmesinden başka, yeterli gücü sağlamak için nelerin gerekli olduğu konusunda bir standart yoktur. Güç seviyesi olarak önerilen seviyeler kesinlikle %50 den aşağıda olmamalı ve olabildiğince %95'e yakın olmalıdır. %70 ile %80 arasındaki güç seviyesi tatmin edici olarak nitelendirilmektedir. Bu konuda Cohen, olması gereken güç seviyesinin %80 olduğunu belirtmiştir (Olejnik, 1984, s. 41).

Güç konusunda sık sık üzerinde durulan konu, araştırılan olayın anakütle üzerinde tam anlamıyla var olabilmesi için gücün en az %80 olması gerektiğidir. Ayrıca istatistiksel güç, orijinal çalışmanın sonuçlarının tekrarlanması da içermektedir. Örneğin, bir araştırmacı çalışmasında istatistiksel gücü %50 olarak bulduysa, araştırmacı yanlış bir sıfır hipotezini reddetmek için %50 şansa sahiptir. Eğer çalışma 100 kez tekrar edilseydi, araştırmacı 50

tanesinde yanlış olan sıfır hipotezini doğru bir şekilde reddetme şansına sahip olacaktı (Tener, 2000, s. 26).

Steidl, Hayes ve Schaubert (1997), güç analizinin uygun bir şekilde nasıl kullanılacağını, istatistiksel gücü kullanarak araştırmanın nasıl daha verimli olarak tasarlanacağını veya verimliliğinin nasıl artırılacağı konusunda açıklama yapmışlardır. Ayrıca istatistiksel güç analizinin kullanılmasıyla araştırmadan elde edilen sonuçların netliğinin artırılacağını vurgulamışlardır. Steidl, Hayes ve Schaubert (1997) 'e göre güç analizi, araştırmanın planlama aşamasında anlamlı bir etki belirleyebilmek için gerekli örneklem sayısının tahmin edilmesi konusunda çok önemlidir. Onlara göre;

- i) İstatistiksel güç analizi, araştırmanın planlama aşamasında bir rutin olarak kullanılmalıdır.
- ii) Tüm güç analizleri için anlamlı etki büyüklükleri kullanılmalıdır.
- iii) Post-hoc güç analizleri için güven aralığı kullanılmalıdır.
- iv) Eğer post-hoc güç analizlerinde güç tahminleri kayıt altına alınıyorsa, o zaman alfa seviyesi, etki büyüklüğü ve hesaplamada kullanılan örneklem sayısı da mutlaka kayıt edilmelidir.

Bu öneriler güç analizinin uygulamalı araştırmalardaki rolünün netliği açısından önemlidir.

Bir istatistiksel testin gücünü tahmin etmenin önemi özellikle uygulamalı istatistik, eğitim-öğretim, psikoloji ve tıp alanlarında artmıştır. Yeni istatistiksel modüller ve yazılım paketleri güç ve örneklem sayısı hesaplama konusunda araştırmacılara oldukça yardımcı olduğu kadar bu araçlara ulaşmak günümüzde oldukça kolaydır. Araştırmacılar, akademisyenler, lisansüstü öğrenciler araştırmalarının özellikle tasarım aşamalarında bu yazılımları kullanarak ön çalışma ile gerekli örneklem sayısına kolayca karar verebilmektedirler. Ne yazık ki, yazında yapılmış olan birçok araştırma, güç ve örneklem sayısına karar verme konusunda bu kadar ilerlemiş düzeyde yazılımlar olmasına rağmen hala araştırmacıların güç konusuna yeterli ilgiyi vermediklerini göstermektedir. Bu nedenle anakütle ortalamalarına göre oldukça anlamlı farklılıklar mevcut olmasına rağmen, örneklem ortalamalarında bu farklılıklar ortaya çıkarılamamaktadır (Rogers ve Hopkins, 1988, s. 647).

İstatistiksel güç, etki büyüklüğü, alfa ve örneklem sayısı çalışmanın tasarım aşamasında birlikte düşünülmelidir. Tener'in (2000) aktardığına göre, Thomas (1997), bu faktörlerden

üçünün bilinmesi durumunda veya üçünün tahmin edilmesi durumunda diğer dördüncü faktörün hesaplanabileceğini belirtmiştir. Araştırmacı istatistiksel gücü, etki büyüklüğünü, alfa ve örneklem sayısını ve bunlar arasındaki ilişkiyi uygun bir araştırma tasarlayabilmek için kullanmalıdır. Bu parametreler araştırma yapılmadan önce dikkate alınmalıdır (s. 30).

Araştırmacıların, araştırmalarının tasarım aşamalarında güç konusunu dikkate almalarını gerektirecek üç önemli neden vardır;

- i) Güç, farklılıkların (eğer gerçekte varsa) ortaya çıkarılma olasılıklarını temsil etmektedir. Eğer araştırma düşük güç ile gerçekleştirilirse, gerçekte anlamlı olan farklılıklar düşük güç nedeniyle anlamsız olarak bulunabilir.
- ii) Bir araştırmacı çalışmasını %50 gibi bir düşük seviyede güç üzerine inşa ederse, bu onun araştırma yaptığı grup üzerinde %50 olasılıkla doğru bir şekilde anlamlı bir ilişki veya farklılık bulacağı anlamına gelmektedir.
- iii) Bir araştırma gereğinden çok fazla sayıda örneklem üzerine kurulmuşsa, bu büyük ölçüde kaynak ve zaman israfı anlamına gelmektedir. Araştırma öncesinde yapılacak bir güç analizi ile gerekli optimum örneklem sayısına karar verilerek zaman ve kaynak israfından kaçınılabılır (Coblick, 1998, s. 82).

1.6 İstatistiksel Güç Analizi

Güç analizi, istatistiksel anlamlılık testleri ile kıyaslandığında yeni yeni dikkat çekmeye başlayan bir konudur. İlk sistematik güç analizi Jacop Cohen tarafından 1962 yılında 70 adet makale incelenerek gerçekleştirilmiştir. Bu çalışma ile Cohen, araştırmacılara güç analizi ile ilgili tekniklerden bahsederek bu konudaki farkındalığı artırmaya çalışmıştır (Clark, 1996, s. 19).

Güç analizi, dört değer belirlenmesini veya bilinmesini gerektirmektedir. Bu değerler örneklem sayısı, alfa, etki büyüklüğü ve güçtür (Mazen vd., 1987, s. 369). İstatistiksel güç, alfa, örneklem sayısı ve etki büyüklüğü ile Cohen dört olası güç analizi tipi açıklamıştır (burada bir parametre, diğer üçünün bir fonksiyonu olarak belirlenmektedir):

- 1) Güç; alfa, etki büyüklüğü ve örneklem sayısının bir fonksiyonudur. Bu analiz tipi günümüz çalışmalarında en yaygın olarak kullanılan güç analizidir.
- 2) Örneklem Sayısı; alfa, etki büyüklüğü ve gücün bir fonksiyonudur. Bu analiz tipi, çalışma öncesinde gerekli olan örneklem sayısının belirlenmesinde kullanılmaktadır.

3) Etki Büyüklüğü; alfa, güç ve örneklem sayısının bir fonksiyonudur. Bu analiz tipi yazındaki çalışmaların karşılaştırılması için kullanılmaktadır.

4) Alfa; güç, örneklem sayısı ve etki büyüklüğünün bir fonksiyonudur. Bu analiz tipi, araştırma gelenekleri yüzünden çok yaygın olmayan bir güç analizi çeşitidir. Çünkü araştırmacılar geleneksel olarak alfayı 0.10, 0.05 ve 0.01 değerlerinden biri olarak seçerler. Bu değerler neredeyse bir standart haline gelmiştir. Araştırmacılar genellikle bilinmeyen değerlerdeki alfa seviyelerini kullanmaktansa II. Tip hata yapmayı göze almaktadırlar.

Cohen, birinci ve ikinci analizlerin en yaygın kullanılan tipler olduğunu belirtmiştir. Analiz 1, tamamlanan bir araştırmanın gücünü belirlemek için (Post-Hoc Power Analysis) alfa, örneklem sayısı ve etki büyüklüğü kullanılarak uygulanır. Araştırmacı yaptığı analizlerde örneklem sayısı, anlamlılık seviyesi olan alfayı ve etki büyüklüğünü kullanarak gerekli olan gücü sağlayabilir. Eğer bu kıstaslar kullanılarak yeterli güce ulaşılamadıysa, araştırmacı bunlardan herhangi birini artırarak gerekli olan güce ulaşabilir. Analiz 2 ise çalışma öncesinde gerekli örneklem sayısını belirlemek için (Priori Power Analysis) uygulanır. Araştırma inşa edilmeden önce belirli bir seviyede etki büyüklüğü beklentisinde olan araştırmacı, belirli bir anlamlılık seviyesi ve gücü kullanarak gerekli olan örneklem sayısına karar verebilir. Cohen, Analiz 3'ün diğerlerine oranla genellikle daha az kullanıldığını fakat belirli alanlarda; örneğin yazındaki daha önce yapılmış olan çalışmaların sonuçlarının karşılaştırılmasında yoğun olarak kullanıldığını belirtmiştir. Cohen, ayrıca Analiz 4'ün çok az kullanıldığını, çünkü araştırmacının alfa seviyesini anlamlılık kriterinden başka daha geniş bir alanda kullanmak istemeyebileceğini belirtmiştir (Cohen, 1988, s. 14-15).

Çok sayıda makale çeşitli istatistiksel testler için gücün nasıl artırılacağını açıklarken, daha hızlı ve daha basit yöntemler olan bir tablo aracılığı ile ve bazı bilgisayar programları yardımıyla güç tahmini gerçekleştirilmektedir. Çok sayıda tablo var olmasına rağmen Cohen'in oluşturmuş olduğu güç tabloları (Cohen, 1988), güç analizi alanında standart bir kaynak olarak kullanılmaktadır.

Güç analizi araştırmanın hem tasarım aşamasında hem de yorum aşamasında araştırmacıya yarar sağlamaktadır. Tasarım aşamasında güç analizi, araştırmacıya çalışmanın anlamlı sonuçlar üretip üretmeyeceğine dair bir ön bilgi sağlamaktadır. Bu aşamada düşük güç elde edilmesi halinde ya parametreler üzerinde değişikliğe gidilerek çalışma yeniden yapılmakta ve güç artırılmaya çalışılmakta ya da çalışma tamamen terk edilmektedir. Bu sayede aşırı

düşük güç olması halinde güç analizi araştırmacıya daha etkili ve daha verimli bir çalışma tasarlayarak zaman, para ve emek konusunda yardımcı olmaktadır (Clark, 1996, s. 22).

Güç analizinin öneminin anlaşılabilmesi için aşağıdaki örnek son derece uygundur:

Bir gökbilimci, uzak galaksiler üzerinde çalışmak için teleskop almayı planlamıştır. Bu tasarımdaki en kritik faktör teleskopun büyütme gücünün ne olacağıdır. Yeterli büyütme gücü olmayan bir teleskop ile çalışmak galaksinin anlaşılabilir bir bulanıklık ile görünmesine neden olacaktır. Sağlıklı bir gözlem yapabilmek için ne kadar büyütme gücüne sahip bir teleskopa ihtiyacı olduğunu belirlemek yerine bu gökbilimci, elindeki mevcut kaynağın yettiği kadar büyütme gücüne sahip bir teleskop almak istemiştir. Bu noktada gökbilimci ne kadar büyütme gücüne sahip bir teleskop ile çalışması gerektiğini bilmemekte fakat tek bildiği elindeki mevcut bütçedir. Bu nedenle gökbilimci sahip olduğu bütçenin tamamını kullanarak imkân dâhilinde en iyi teleskopu sipariş edecektir.

Sosyal bilimler alanındaki bu gökbilimciye benzer tavır sergileyen araştırmacı, sahip olduğu bütçeyi temel alarak örneklem büyüklüğüne karar veren kişidir. Bu araştırmacıya “örneklem sayın ne kadar olmalı?” tarzında bir soru sorulduğunda “bütçem dâhilinde elde edebileceğim en fazlası” şeklinde bir cevap gelecektir.

Kaynak kısıtları araştırma hayatının bir gerçeğidir. Eğer kısıtlı imkânlara sahip bir araştırma yapılmak isteniyorsa esas olan araştırma altında ortaya çıkarılmak istenen etkiyi belirlemek için gerekli gücün ne olacağı sorusuna cevap bulmaktır. Hipotezleri sağlıklı bir şekilde test etmek için ne kadar örnekleme ihtiyaç var? Araştırma altında ortaya çıkarılmak istenin etkinin anakütle üzerinde var olduğu biliniyorsa, bu durumda etkiyi ortaya çıkarabilmek için araştırma yeterli şansa sahip mi? Bu şans nasıl artırılabilir? Eğer örneklem sayısı sınırlı ise (ki bu çoğu durumda karşılaşılabilecek bir durumdur) istatistiksel testleri gerçekleştirebilmek için yeterli güç var mı? İşte istatistiksel güç analizi bu tür sorulara cevap sağlamaktadır (Ellis, 2010, s. 47).

1.7 İstatistiksel Gücü Doğrudan Etkileyen Faktörler

Bir araştırmacının istatistiksel gücünü doğrudan etkileyen faktörler;

- i) Anlamlılık seviyesi
- ii) Etki büyüklüğü
- iii) Örneklem sayısıdır.

Diğer parametreler sabit olarak düşünülürken bu üç parametrenin artırılması istatistiksel gücü artıracaktır. Yani bahsi geçen her bir parametre ile güç arasında doğru orantılı bir ilişki vardır (Mazen vd., 1987, s. 370; O’Keefe, 2007, s. 292). İstatistiksel gücü doğrudan etkileyen faktörlerin yanında dolaylı olarak etkileyen,

- i) Hipotezin tek yönlü mü yoksa iki yönlü mü olduğu,
- ii) Kullanılan istatistiksel test,
- iii) Örneklem hatasının düşük olması,
- iv) Ölçüm hatası,
- v) Varyans analizinde kullanılan grup sayısı,
- vi) Çoklu regresyon analizinde kullanılan bağımsız değişken sayısı (predictors); daha fazla grup ve daha fazla bağımsız değişken sayısı daha az güç demektir,
- vii) Örneklem sayılarının deney ve kontrol grubu olarak eşit sayıda bölünmesi gibi faktörler de mevcuttur (Clark, 1996, s. 17).

1.7.1 Anlamlılık Seviyesi (α)

Güçü etkileyen parametrelerden birincisi anlamlılık seviyesidir. Anlamlılık seviyesi, araştırmacı tarafından veriler toplanmadan önce kurulmaktadır. Alfa (α), doğru bir H_0 hipotezinin reddedilmesi ihtimalidir yani I. Tip hatadır. Başka bir deyişle daha küçük bir α , daha az olasılık ile H_0 hipotezinin hatalı bir şekilde reddedileceği anlamına gelir. Buna karşılık, diğer parametreler sabit iken daha küçük bir anlamlılık seviyesi, daha az güç demektir. Başka bir deyişle, I. Tip Hata yapma olasılığı konusunda daha rahat olan bir araştırmacı daha yüksek seviyede bir güç elde eder.

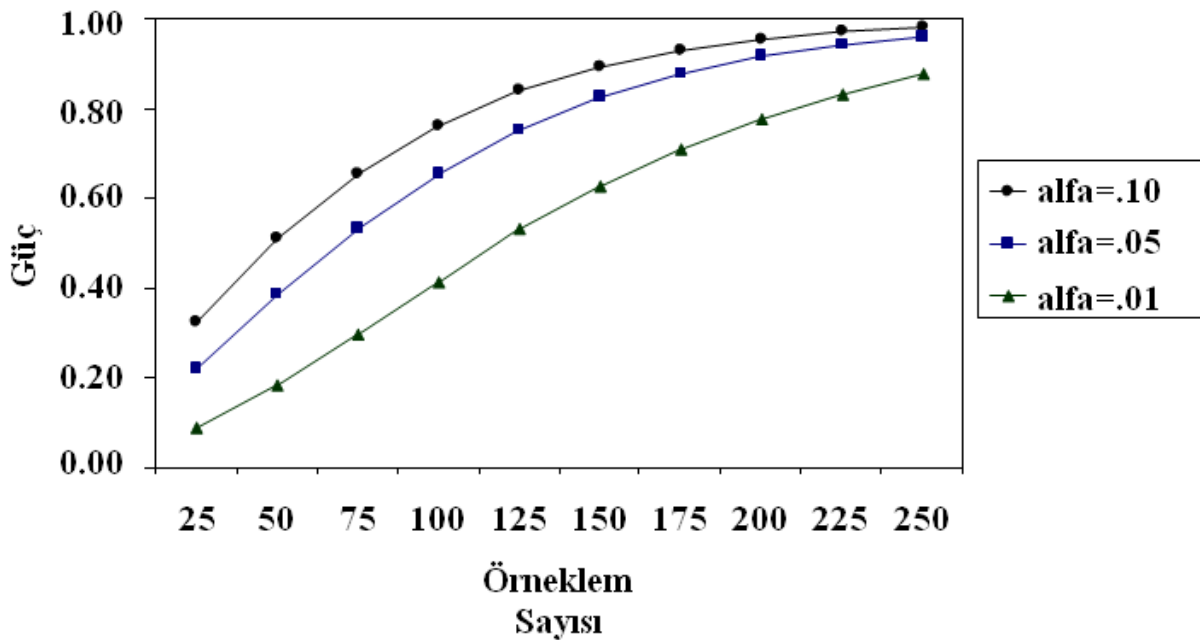
Unutulmamalıdır ki I. Tip hatayı alanında çok iyi olan bir kişi de yapabilmektedir. İşte bu nedendir ki tüm zamanların en iyi istatistikçilerinden birisi olarak kabul edilen Sir Ronald Aymler Fisher, uzun zaman önce araştırma sonucunda elde edilen bulguların bir “*istatistiksel olarak anlamlı*” etiketi elde edebilmesine karar verebilmek için bazı standartlara ihtiyaç olduğunu belirtmiştir. İşte bu standart “*alfa*”dır. Fisher’dan sonra alfanın standart değeri 0.05 olarak kabul görmüştür (Ellis, 2010, s. 48).

İlk bakışta burada gücü artırmak için sadece anlamlılık seviyesinin ve örneklem sayısının artırılması gerektiği anlamı çıkarılabilir. Fakat alfa, genellikle ya 0.01 olarak ya da 0.05 olarak

kurulur ve genellikle 0.05 seviyesinden daha yukarı seviyelerde kurulmaz. Eğer $\alpha = 0.05$ olarak kurulursa bu %5 olasılıkla veya 100 kez tekrarlanan bir deney içerisinde 5 kez doğru bir H_0 hipotezinin reddedilmesi olasılığının olduğu anlamına gelmektedir. Başka bir deyişle eğer H_0 hipotezi gerçekte doğru ise bu araştırma gerçekleştirilen anakütle içerisinde sadece %5'lik bir kısmın bu anakütlenin karakteristiklerini yansıtıyor anlamına gelmektedir ve H_0 hipotezinin yanlışlıkla reddedilmesi olasılığı %5'tir. Eğer $\alpha = 0.01$ ise bu yapılan bir araştırmada 100 kez tekrarlanan bir deney içerisinde 1 kez doğru bir H_0 hipotezinin reddedilmesi olasılığının bulunduğunu göstermektedir.

Bazen hangi seviyede anlamlılık seviyesi seçileceği konusu, yapılacak olan araştırmanın çeşidine göre de değişebilmektedir. Örneğin, 1.000.000 TL tutarındaki bir fon ile ulusal bir eğitim programının desteklenme kararı 0.05 anlamlılık seviyesi yerine 0.01 anlamlılık seviyesi seçilmelidir. Çünkü bu durumda devletin kaynaklarının israf edilmesi olasılığının çok düşük seviyelerde olması gerekmektedir (DiLullo,1997,s. 5).

Ayrıca yapılan hipotez testinde alfa seviyesinin istatistiksel güç üzerinde bir etkisi vardır. Bu etki Şekil 1.3'te gösterilmiştir.



Şekil 1.3 Alfa Seviyesinin İstatistiksel Güç Üzerindeki Etkisi

1.7.1.1 Alfa ve Beta Arasındaki İlişki

I. Tip Hata yapma olasılığı olan α ile II. Tip Hata yapma olasılığı olan β arasında zıt yönlü bir ilişki vardır. Bu nedenle her ikisi de aynı anda aşırı küçük olacak şekilde belirlenemez. Ayrıca her ikisinin temelindeki koşullar farklı olduğundan toplamı 1 olacak şekilde ayarlanamazlar. α , H_0 hipotezinin doğru olduğu varsayımına dayanırken, β ise H_0 hipotezinin yanlış olduğu varsayımına dayanmaktadır. Hipotezin doğru veya yanlış olduğu araştırma öncesinden asla kesin olarak bilinemez. Bu yüzden araştırmacının her zaman olası bir hata yapma riski vardır. Her iki tip hatayı da aynı anda aynı test içinde yapmak mümkün değildir. α , etki büyüklüğünün sıfıra eşit olduğu durumlarda bile geçerliyken, β ancak etki büyüklüğünün sıfırdan farklı olduğu durumlarda geçerlidir (McKean, 1990, s. 16). Eğer bir araştırmacı α seviyesinin 0.01 ve gücün 0.50 olduğu bir araştırma gerçekleştiriyorsa, bu durumda araştırmacının anakütle üzerinde gerçekte var olmayan yanlış bir etkiyi bulma durumu, gerçekte anakütle üzerinde olan bir etkiyi bulamama durumundan 50 kat daha ciddidir.

1.7.2 Etki Büyüklüğü

Gücü etkileyen öncelikli parametrelerden ikincisi etki büyüklüğüdür. Etki büyüklüğü bilinen en yaygın tanımıyla araştırılan konunun, araştırılan anakütle üzerinde ne derece var olduğudur (Cohen, 1988, s. 10). Etki büyüklüğü için Sawyer ve Ball ise “o, iki veya daha fazla değişken arasındaki bir ölçümü, bir gücü veya bir ilişkiyi açıklar” demiştir (Sawyer ve Ball, 1981, s. 275). Aberson (2010) ise etki büyüklüğü için “anlamlılık testleri sadece etkinin olup olmadığı sorunun cevabını verir. Etki büyüklüğü ise bu etkinin ne derecede olduğunu anlatır” demiştir (Aberson, 2010, s. 2).

En basit tanımı ile etki büyüklüğü, H_0 ile H_1 hipotezlerinin tepe değerleri arasındaki farktır. Gücü tahmin etmek veya belirli bir güç seviyesini yakalama amacıyla örneklem sayısını belirlemek için öncelikle yapılması gereken işlerden birisi anakütle üzerindeki etki büyüklüğünü tahmin etmektir. Bunu tahmin etmek, bir araştırmacının planlama aşamasında yapması gereken en önemli ve en zor işlerden biridir. Belirlenen bir etki büyüklüğü ile güç analizine başlamak için büyük bir engel ortadan kaldırılmış olur (Chase ve Chase, 1976, s. 234; Mazen vd., 1987, s. 370; Cohen, 1992, s. 156; Thomas ve Juanes, 1996, s. 856; Borkowski vd., 2001, s. 65; Lewis, 2006, s. 607; Rose ve Bowen, 2009, s. 44; Rossi, 2012, s. 17).

İstatistiksel anlamlılık testleri, arařtırmacılara elde ettikleri sonuçların gerek mi yoksa tesadüfen mi elde edildiđine dair bilgiler sađlamaktadır. Etki büyüklüğü ise yapılan arařtırmanın anakütle üzerinde ne derece etkili olduđunu arařtırmacıya açıklamaktadır. Etki büyüklüğü, özellikle herhangi iki alıřmanın sonuçlarının karşılařtırılmasında önemlidir (Thalheimer ve Cook, 2002, s. 2).

Arařtırmacıların öncelikli amacı gerek dünyada var olan iliřkileri veya iliřkilerin gücünü veya farklılıkları tahmin etmek, bunların etkilerini ortaya ıkarmaktır. Bahsi geen etki bir tedavi, bir deneme, bir karar, bir strateji, bir yenilik, bir icat, bir operasyon dolayısıyla her řey olabilir. Bu arařtırmanın türüne bađlıdır. Küçük etkileri ortaya ıkarmak daha zordur ve daha fazla örneklem sayısı gerektirmektedir (Rotenberry ve Patterns, 1985, s. 164).

Etki büyüklüğüne hayatın her alanında rastlanılabilir. Örneđin bir kilo kontrol tedavi yöntemi kiřilere bir ay içinde 10 kilo zayıflama garantisi verdiđinde bu, uygulanacak tedavi yönteminin etki büyüklüğüdür. İlaveten, “servis sađlayıcınızı deđiřtirin ve %20 tasarruf sađlayın, botoks ile 10 yıl gençleřin” gibi ifadeler etki büyüklüğüne bir sonucudur (Ellis, 2010, s. 6)

İstatistiksel anlamlılık, arařtırma sonucunda elde edilen bulguların gerek hayatta uygulanıp uygulanmayacađına karar verme konusunda etkilidir. Bu noktada sorulacak soru řudur: “Elde edilen sonuçlar etki hakkında ne söylemektedir?”

2000-2009 yılları arasında yayımlanan kitaplarda yapılan arařtırmalara göre etki büyüklüğünden ve onun nasıl yorumlanacađından bahsedilme oranı Ellis’e (2010) göre sadece %13 tür. Ancak sosyal ve davranıř bilimlerinde etki büyüklüğünün rapor edilmesi konusunda özellikle dergi editörleri tarafından gittike artan bir baskı vardır. Buna rađmen etki büyüklüğünün dođru tahmin edilmesi ve raporlanması konusuna hala yeterince önem verilmemektedir. 2008 yılında Osborne, yaptıđı bir arařtırmada eđitim psikolojisi alanında sadece %17 oranında etki büyüklüğünden bahsedildiđi, Armstrong ve Henson (2004) ise inceledikleri makalelerin sadece %5’inde etki büyüklüklerinin rapor edildiđi sonucuna ulařmışlardır (Ellis, 2010, s. xiii).

“Arařtırmam, aradıđım etkiyi bulabilmem için yeterli güce sahip mi?” sorusu arařtırmanın planlama ařamasında sorulması gereken bir sorudur. İstatistiksel güç, eđer anakütle üzerinde gerekten bir etki varsa bunun ortaya ıkarılma olasılıđını açıklamaktadır. Yapılan incelemeler özellikle sosyal bilimler alanında güç konusuna gereken önemin verilmediđini

ortaya koymaktadır. Bunun negatif sonucu olarak araştırılan etkilerin ortaya çıkarılma şansları oldukça düşüktür (Ellis, 2010, s. xv).

Etki büyüklüğü anakütle üzerinde elde edilen bir değerdir ve sadece örneklem üzerinden tahmin edilebilir (Abraham ve Russell, 2008, s. 285). Eğer, yapılan bir araştırmada anakütle üzerinde bir etki varsa, o zaman bu etki büyüklüğü H_0 hipotezi ile H_1 hipotezi arasındaki farkları açıklamaya yardımcı olur. Etki büyüklüğünün anlamı, araştırmanın içeriğine bağlıdır. Yapılan bir araştırmada gruplar arasında farklılık yok ise veya örneklemden elde edilen istatistikler ile anakütle parametreleri arasında anlamlı bir ilişki yok ise o zaman etki büyüklüğünün değeri sıfırdır. Eğer bazı istatistiksel anlamlılıklar elde edildiyse, o zaman anakütle üzerinde bir etki vardır ve etki büyüklüğünün değeri sıfırdan farklıdır. Bir verimlilik araştırmasında, etki büyüklüğü kavramsal olarak anakütle anlamları arasındaki farklılıklar olarak ifade edilebilir. Eğer birinci grup, ikinci gruptan daha fazla puan aldıysa, araştırma ile ilgilenen kişilerin aradaki bu farkın anlamlı mı yoksa tesadüfen mi olduğunu bilmeye ihtiyaçları vardır (Chase ve Chase, 1976, s. 234; McKean, 1990, s. 25).

Etki büyüklüğünün rapor edilmesi ve yorumlanması için üç önemli sebep vardır:

- i) Araştırma bulgularının pratik önem açısından yorumunu kolaylaştırması,
- ii) Etki büyüklüğü de dikkate alınarak çalışmada kullanılması gereken optimum örneklem sayısına ulaşılmasını sağlaması,
- iii) Gerçekleştirilen farklı çalışmaların sonuçlarının karşılaştırılmasında kullanılabilir olmasıdır (Ellis, 2010, s. 24).

Fern ve Monroe (1996), metodolojik ve teorik olarak etki büyüklüğünün yorumlanmasına etki eden faktörleri incelemişlerdir. Yapılan incelemeler neticesinde araştırmacıların dolaylı veya doğrudan, yapılan araştırmalardan elde edilen etkiye müdahale ettiklerini belirlemişlerdir. İstatistiksel anlamlılık ve etki boyutunun yanında araştırmanın içeriği, araştırma alanındaki teorik gelişmelerin seviyesi vs. etki büyüklüğünü etkilemektedir. Küçük örneklem sayısı yüzünden II. Tip hatadan kaçınmak için araştırmacılar, güç analizinde kullanılacak veya kullanılan etki büyüklüğüne etki eden faktörleri dikkate almalıdırlar (Fern ve Monroe, 1996, s. 89). McKean'in (1990) aktardığına göre, Lipsey (1990) "araştırmanın tasarım aşamasında etki büyüklüğü hesaplanamaz, sadece tahmin edilmiş olmalıdır ve bu tahmin güce karar verme aşamasındaki en meşakkatli faktördür" demiştir (s. 26).

Bir etki, iki grup üzerinde uygulanan bir tedavinin sonucu olabilir veya iki ilgili değişken arasındaki ilişkinin derecesi olabilir. Etki büyüklüğünün tahmin edilmesi elde edilen sonuçların doğru yorumlanması açısından gereklidir. APA tarafından yayımlanan *Publication Manual, 5th Edition*, etki büyüklüğünün rapor edilmesi konusundaki başarısızlığın editörler tarafından en çok gözlemlenen yedi kusurdan biri olduğunu belirtmiştir. Çalışma bulgularının öneminin daha rahat anlaşılabilmesi için hemen hemen her zaman etki büyüklüğünü gösteren bazı indeksler verilmelidir (APA, 2001, s. 25). American Educational Research Association (AERA) da araştırmacılara istatistiksel sonuçlar ile birlikte etki büyüklüğünü de içeren sayısal bilgilerin verilmesi gerektiğini tavsiye etmiştir (AERA, 2006, s. 37).

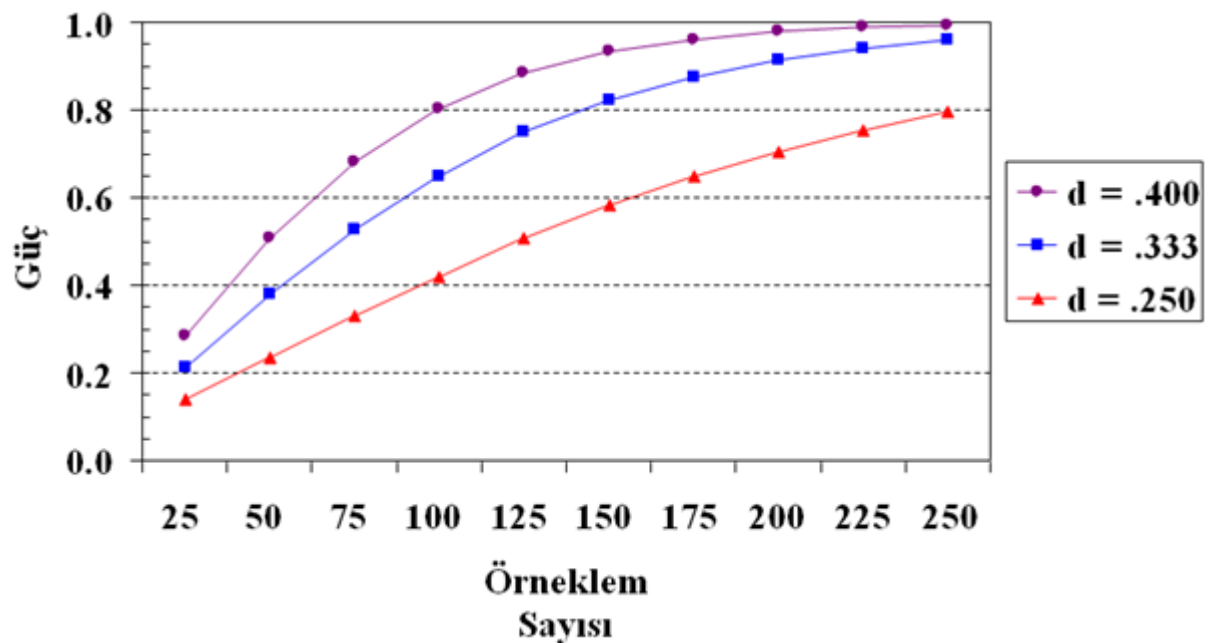
Etki büyüklüğünü belirlemenin en etkili yolu tüm anakütle üyelerini incelemektir. Fakat bu durum genellikle mümkün olmayacağı için etki büyüklüğünün tahmin edilmesi gerekmektedir. Yapılacak olan tahminin netliği kullanılan yöntemlere ve seçilen örneklemin anakütleyi iyi bir şekilde temsil etmesine bağlıdır. Tam bu noktada iki soru sorulması gerekmektedir. Birincisi, ne kadar büyüklükte bir etkinin var olduğu, ikincisi ise tahmin edilen etkinin ne kadar doğru olduğu sorusudur. Sokaktaki insan için birinci sorunun cevabı daha önemli iken, araştırmacı ve öğrenciler için ikinci sorunun cevabı daha önemlidir. Bu noktada bir örnek vermek daha açıklayıcı olacaktır.

Bir araştırmacı, Alzheimer hastası olan 12 hastayı eşit iki gruba ayırarak, deney grubuna uygulayacağı bir tedavinin hastaların IQ seviyelerine olan etkisini incelemek istemektedir. Yapılan uygulamalar sonucunda deney grubundaki hastaların IQ seviyelerinde, başlangıç seviyelerine oranla ortalama 13 puanlık bir artış meydana gelmiştir. Bu artış, tedavinin işe yaradığının bir göstergesidir. Fakat uygulanan istatistiksel testlerin sonucu ($t = 1.61$ ve $p = 0.14$, $\alpha = 0.05$) istatistiksel olarak anlamsız çıkmıştır. İstatistiksel olarak anlamlı olmayan bu sonuç tedavinin işe yaramadığını değil, şansa bağlı olarak nitelendirilemeyecek kadar büyük bir etki göstermediği anlamına gelmektedir. Daha büyük bir örneklem ile istatistiksel olarak daha anlamlı sonuçlar elde edilebilir. Bu nedenle araştırmanın tekrar edilmesi gerekmektedir. Buradaki sonuç, pratik olarak anlamlı fakat istatistiksel olarak anlamsızdır. Pratik anlamlılık sonucuna etki büyüklüğüne bakılarak karar verilirken, istatistiksel olarak anlamlılığa ise tahminin netliğine bakılarak karar verilmektedir.

Herhangi bir sonucun istatistiksel olarak anlamlılığı hem etki büyüklüğünden hem de kullanılan örneklem sayısından etkilenmektedir. Nispeten daha küçük bir örneklem, gerçekleştirilen bir çalışmada daha az olasılıkla elde edilecek sonuçların istatistiksel olarak

anamlı olacağı anlamına gelmektedir. Bu durum, istatistiksel anlamlılık testinden pratik anlamlılığa dair bir sonuç çıkarılamayacağını göstermektedir (Ellis, 2010,s. 5).

Sawyer ve Ball (1981) ve Kosciulek ve Szymonski (1983), etki büyüklüğünün gruplar arasındaki farklılıkların derecesi veya değişkenler arasındaki ilişkinin gücü olduğunu belirtmişlerdir. West (1985), etki büyüklüğünü doğru olan bir alternatif hipotezin anakütle üzerindeki boyutu olarak tanımlamıştır. Katzer ve Sordt (1973) etki büyüklüğünü, araştırmacıya araştırma sonucunda bulduğu sonuçların ne kadar önemli olduğunu anlatan bir kavram olarak tanımlamışlardır. Cohen (1992) ve Shavelson (1996) etki büyüklüğünü, sıfır hipotezi ile alternatif hipotez arasındaki farklılığın bir göstergesi olarak belirtmişlerdir. Bu gösterge ya da belirti bir sayısal değer ile açıklanabilmektedir. Eğer sıfır hipotezi doğru ise, o zaman etki büyüklüğünün değeri “0”dır. Yani araştırılan anakütle üzerinde herhangi bir etki yoktur. Aksine, eğer sıfır hipotezi yanlışsa, o zaman etki büyüklüğünün değeri sıfırdan farklıdır ve anakütle üzerinde bir etki vardır ve bu etki sıfır hipotezinin ne derece yanlış olduğunu gösteren belirli bir değerdir. Etki büyüklüğünün değeri ne kadar büyük ise sıfır hipotezi de o derece yanlıştır ve araştırılmak istenen olay, olgu vs. araştırma yapılan anakütle üzerinde o derecede vardır. Ayrıca aşağıdaki Şekil 1.4’te gösterildiği gibi, etki büyüklüğünün değeri arttıkça, anlamlılık seviyesi ve örneklem sayısı sabit tutulduğunda, istatistiksel güç artmaktadır. İlâveten, daha büyük bir etki daha küçük bir örneklem sayısını gerektirmektedir (DiLullo, 1997, s. 6).



Şekil 1.4 Etki Büyüklüğünün İstatistiksel Güç Üzerine Etkisi

Cohen(1962), etki büyüklüğünün ne kadar olması gerektiğine karar vermek veya ne kadar büyüklükteki bir etkinin gerçekte anakütle üzerinde var olduğuna karar vermek araştırmacı için en zor konulardan biri olduğunu belirtmiştir. Ancak Cohen, bu konunun araştırmacılar için karar vermesi en zor konulardan biri olmasına rağmen aynı zamanda en az dikkate alınan konulardan da birisi olduğunu vurgulamıştır (Cohen, 1962, s. 146; Cohen, 1988, s. 10).

Etki büyüklüğü, araştırma sonuçlarının yorumlanması açısından çok önemlidir. Tener'in (2000) aktardığına göre, Ahrens (1971), "araştırmacı, anlamlılığı test etmek için sınırlanmamalı, aksine ilişkilerin büyüklüğünü de ölçmelidir" demiştir (s. 28).

Cohen, etki büyüklüğünün tahmini konusunda araştırmacılara bazı tavsiyelerde bulunmuştur. Birincisi, bir araştırmacı, çalışması hakkındaki bilgisini mümkün olduğunca çalışma öncesinde kullanmalıdır. Buna karşılık, en azından sosyal bilimler alanındaki araştırmalarda aşağıdaki nedenlerle yetersiz bilgiye sahip olunması nedeniyle etki büyüklüğünün doğru bir şekilde tahmin edilmesi konusu çok zordur. Bu nedenler:

1. Sosyal bilimler alanındaki teorilerin nadiren etki büyüklüğünün tahmini konusunda ipuçları vermesi,
2. Güç analizi için gerekli olan bilginin genellikle yürütülen araştırmada yetersiz olması,
3. Elde edilen bulguların çalışma yeterli bilgiyi içerse bile manipülasyonlar, ölçümler, değişkenler vs. gibi nedenlerle araştırmadan araştırmaya çok farklılık göstermesidir.

Bu eksikliklerden kaynaklanan zorluğu giderebilmek için Sawyer ve Ball (1981), araştırmacılara doğru bir etki büyüklüğü tahmini yapabilmek için çalışma öncesinde anakütleden elde edilecek bir altküme üzerinde bir ön çalışma yapılarak etki büyüklüğünün tahmin edilmesini önermişlerdir (s. 277).

İkinci olarak, eğer ön çalışma kısmında araştırmacı etki büyüklüğünün tahmini konusunda yeterli bir bilgiye sahip olamıyorsa, bu durumda Cohen'in etki büyüklüğü hakkında yapmış olduğu geleneksel bir sınıflandırma olan küçük, orta ve büyük seviyelerden birisini seçebilir. Fakat etki büyüklüğü, kullanılan istatistiksel teste göre değişiklik göstermektedir. Tablo 1.2'de, Cohen'in sunmuş olduğu küçük, orta ve büyük seviyelerdeki etki büyüklükleri çeşitli istatistiksel testler için gösterilmektedir (Sacco, 1982, s. 755; Mazen vd., 1987, s. 370; DiLullo, 1997, s. 7-8; Borkowski vd., 2001, s. 63; Abraham ve Russell, 2008, s. 285).

Tablo 1.2 Cohen'in Etki Büyüklüğü Sınıflandırmaları

Test	İlgili Etki Büyüklüğü	Etki Büyüklüğü Sınıfları		
		Küçük	Orta	Büyük
Bağımsız Ort. Karşılaştırılması	$d, \Delta, \text{Hedges' } g$.20	.50	.80
Korelasyonların Karşılaştırılması	q	.10	.30	.50
Oranlar Arasında Fark	Cohen's g	.05	.15	.25
Korelasyon	r	.10	.30	.50
	r^2	.01	.09	.25
Çapraz Tablolar	w, ϕ, V, C	.10	.30	.50
ANOVA	f	.10	.25	.40
	η^2	.01	.06	.14
	R^2	.02	.13	.26
Çoklu Regresyon	R^2	.02	.13	.26
	f^2	.02	.15	.35

Not: Etki büyüklüğü sınıflandırmaları ve aralarındaki ilişkiler ile ilgili detaylı bilgiler Cohen'in (1988) *Statistical Power Analysis for the Behavioral Sciences* isimli kitabında şu sayfalarda yer almaktadır: Cohen'd (s. 40), q (s.115), Cohen's g (s. 147-149), r ve r^2 (s. 79,80), Cohen's w (s. 224-227), f ve η^2 (s. 285-287), R ve R^2 (s. 413-414).

1.7.2.1 Etki Büyüklüğü İndeksleri

Etki büyüklüğünün sembol olarak çok sayıda gösterimi vardır. Ancak genel olarak yaygın bir şekilde kullanılan 2 çeşit etki büyüklüğü indeksi vardır. Bunlar “*d family*” ve “*r family*” olmak üzere yani d ve r aileleri olarak gösterilmektedir. “*d family*”, iki değişken arasındaki farkın büyüklüğünü gösterirken, “*r family*” iki değişken arasındaki ilişkinin kuvvetini göstermektedir (Nakagawa, 2004, s. 1044).

1.7.2.1.1 “d” Ailesi: Gruplar Arasındaki Farklılıkların Ölçülmesi

Gruplar arasındaki farklılığın ölçülmesinde d ailesi, değişkenler arasındaki ilişkilerin ölçülmesinde r ailesi kullanılmaktadır. Ortalamalar arasındaki farkın, standart sapmaya bölünmesiyle d indeksi yaklaşık olarak tahmin edilmektedir. En yaygın etki büyüklüğü indeksi “d” dir. “d”, t-testi gibi ortalamalar arasındaki farkın açıklanmasına dayanan çalışmalarda etki büyüklüğünü ifade etmek için kullanılmaktadır.

İki grup arasındaki farkı hesaplamak için bir grubun ortalaması (M_1) diğer grubun ortalamasından (M_2) çıkarılır ve elde edilen sonuç, grupların örneklem olarak seçildiği anakütlenin standart sapmasına bölünür. Burada hesaplanması en zor olan değer anakütlenin

standart sapmasıdır. Bu indeksi geliştiren Cohen, bu problemin nasıl çözüleceğine ilişkin bir çözüm önermemiştir. Fakat daha sonra bu problemin çözümüne ilişkin 3 çözüm ortaya atılmıştır. Bunlar *Cohen's d*, *Glass's delta* ve *Hedges's g* dir.

$$\text{Cohen's } d = \frac{M_1 - M_2}{\sigma_{pooled}}$$

$$\text{Glass's } \Delta = \frac{M_1 - M_2}{\sigma_{kontrol}}$$

$$\text{Hedges's } g = \frac{M_1 - M_2}{\sigma_{pooled}}$$

Yukarıda gösterilen denklemlerden de anlaşılacağı gibi aralarındaki tek farklılık standart sapmanın hesaplanmasından kaynaklanmaktadır. Bu formüllerden hangisinin kullanılması gerektiğine karar verme işi her bir grubun standart sapmalarının incelenmesini gerektirir. *Cohen's d* indeksinde yer alan anakütlenin standart sapması, eğer standart sapmalar farklı ise aşağıdaki denklem ile hesaplanmaktadır:

$$\sigma_{pooled} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \quad (\text{Cohen's } d \text{ için})$$

Eğer her iki grubun standart sapmaları birbirinden farklı ise Glass, denklemde kontrol grubunun standart sapmasının kullanılması gerektiğini, çünkü üzerinde herhangi bir işlem yapılmayan kontrol grubunun standart sapmasının anakütlenin standart sapmasına daha yakın olacağını belirtmiştir (Ellis, 2010, s. 10).

Bir başka yaklaşım ise grupların örneklem sayıları ile sahip oldukları standart sapmaların ağırlıklandırılarak anakütlenin standart sapmasına ulaşmaktır. Bu yöntem ise Hedges tarafından önerilmiştir.

$$\sigma_{pooled} = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{(n_1+n_2-2)}} \quad (\text{Hedges's } g \text{ için})$$

Ayrıca Hedges, kendi denkleminde elde edilecek olan bir değer için nasıl Cohen'in d indeksine dönüştürüleceğine ilişkin aşağıdaki denklemi vermiştir:

$$g \approx d \left(1 - \frac{3}{4(n_1+n_2)-9} \right)$$

Gruplar arasında daha fazla farklılık demek daha fazla etki büyüklüğü demektir. Gerçekleştirilen iki çalışmada aynı seviyede (örneğin $d=0.50$) etki büyüklükleri elde edildiyse, bu iki çalışma arasında etki büyüklüğü açısından kıyaslama yapılabilir ve aynı etkiye sahiptirler denilebilir.

Aşağıdaki formüller kullanılarak etki büyüklüğü, iki grup arasındaki farkın t -test değeri kullanılarak hesaplanabilir (Becker, 2000, s. 2).

$$d = \frac{2t}{\sqrt{df}}$$

df: degrees of freedom (serbestlik derecesi), t: test değeridir.

Bu formül, grupların örneklem sayılarının eşit olduğu durumda kullanılır. İki grubun örneklem sayıları farklı olduğu durumda ise etki büyüklüğünü hesaplamak için aşağıdaki formül kullanılır:

$$d = \frac{t(n_1+n_2)}{\sqrt{df(n_1n_2)}}$$

n_1 : birinci grubun örneklem sayısı, n_2 : ikinci grubun örneklem sayısı, t : test değeri, df : serbestlik derecesidir.

1.7.2.1.2 “r” Ailesi: Değişkenler Arasındaki İlişkinin Ölçülmesi

İkinci etki büyüklüğü ailesi, iki veya daha çok değişken arasındaki ilişki ölçmeye yarayan çok sayıda ölçüm çeşidine sahiptir. Bunların çoğu ise korelasyon katsayısının çeşitli türleridir. Korelasyon katsayısı (r), X ve Y gibi iki değişken arasındaki ilişkinin gücünü ve yönünü nitelemektedir. Korelasyon -1 ile $+1$ arasında değişmekte ve “ -1 ”, mükemmel zıt ilişki anlamına gelirken “ $+1$ ” ise mükemmel pozitif ilişki anlamına gelmektedir. “ 0 ” ise ilişki yok anlamındadır. Korelasyon katsayısı, çoğu kişinin onu etki büyüklüğü indeksi olduğunu bilmemesine rağmen iyi bir etki büyüklüğü ölçөгüdür (Ellis, 2010, s. 11).

Paylaşılan varyans oranı (PVO) veya başka bir deyişle açıklanan varyans oranı da bazı disiplinlerde iki değişken arasındaki ilişkiyi açıklamak için kullanılmaktadır. Örneğin, korelasyon katsayısı r 'nin PVO'su r^2 dir. Eğer X ve Y , $r = -0.60$ değerinde bir korelasyona sahip iseler, o zaman $PVO = 0.36$ olacaktır. Bu toplam varyansın %36'sı X ve Y değişkenleri tarafından paylaşılıyor anlamına gelmektedir. Ya da Y değişkeni, X değişkeninde meydana gelecek bir değişimi %36 oranında açıklıyor denilebilir. PVO, 0 ile 1 arasında değişir.”0”, paylaşılan varyans olmadığı için, “1” ise varyansın tamamen paylaşıldığının göstergesidir (Ellis, 2010, s.12).

Bir başka etki büyüklüğü indeksi ise *Cohen's f* ve f^2 dir. Bu indeksler ANOVA ve Çoklu Regresyon Analizleri için kullanılmaktadır. *Cohen's f*, *Cohen's d* indeksinin bir değişik versiyonudur. *Cohen's d*, iki değişken arasındaki fark için kullanılırken, *Cohen's f* ise 3 veya daha fazla grup arasındaki ayrışmayı ortaya çıkarmak için kullanılmaktadır. *Cohen's d* ile *Cohen's f* arasında indeks değerleri olarak,

$$d = 2f$$

şeklinde bir eşitlik vardır. Bir F testinden elde edilen F istatistiğinin, *Cohen's d* dönüşümü aşağıdaki denklem ile gerçekleştirilmektedir (Thalheimer ve Cook, 2002, s. 8):

$$d = \sqrt{F \left(\frac{n_1 + n_2}{n_1 n_2} \right)} \sqrt{\frac{n_1 + n_2}{n_1 + n_2 - 2}}$$

F : Test istatistiği,

n_1, n_2 : Grupların örneklem sayılarıdır.

Etki büyüklüklerini, indekslerini belirtmeden rapor etmek anlamsızdır. Örneğin bir etki büyüklüğü 0.5 olarak rapor edilmiş olsun. Elde edilen bu 0.5 değeri, etki büyüklüğü indeksinin d ailesine mi yoksa r ailesine mi ait olması durumuna göre oldukça farklı anlamlara gelmektedir. Çünkü d ailesindeki bir değer, r ailesindeki bir değer yaklaşık olarak 2 katıdır. Genellikle kullanılan indeks ölçülen etkinin türüne göre değişiklik göstermektedir. Eğer iki değişken arasındaki bir ilişki ölçülüyorsa r indeksi kullanılmalıdır. Eğer gruplar karşılaştırılıyorsa d indeksi kullanılmalıdır. Her iki indeksin yorumları farklıdır fakat aşağıdaki denklemler ile birbirlerine dönüştürülebilirler:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

t -test değerinin r etki büyüklüğü indeksine dönüştürülmesi ise aşağıdaki formül kullanılarak yapılabilmektedir:

$$r = \frac{\sqrt{t^2}}{\sqrt{t^2 + (df)}}$$

Bir etki büyüklüğünün anlamı, araştırmanın içeriğine, nerede yapıldığına, ne zaman yapıldığına veya kimler ile yapıldığına bağlıdır. Ancak bazı durumlarda etki büyüklüğü için bir kaynağa başvurmak gerekmektedir. Bunun da en basit yolu Cohen tarafından hazırlanmış olan etki büyüklüğü tablolarıdır. Cohen, sosyal ve davranış bilimleri alanlarında araştırmacılara rehber olması için küçük, orta ve büyük olmak üzere 3 farklı seviyede etki

büyüklüğü sınıflandırması yapmıştır. Cohen, bu değerlerin çalışmanın alanıyla ilgili olarak değişebileceğini ve bunların sadece rehber amaçlı kullanılması gerektiği konusunda araştırmacıları uyarmıştır. Ancak Cohen'in uyarılarına karşın araştırmacılar tarafından bu değerler bir gelenek olarak kabul edilmiş ve kullanılmıştır (Abraham ve Russel, 2008, s. 285).

Cohen, küçük seviyedeki etki büyüklüğünün orta seviyedeki etkiden daha küçük fakat önemsiz sayılacak derecede küçük olmadığını ve yaklaşık değerinin $d = 0.20$ veya Pearson product korelasyon katsayısının $r = 0.10$ veya $r^2 = 0.01$ seviyesinde olduğunu belirtmiştir. Burada sadece toplam varyansın %1'i bağımlı ve bağımsız değişken arasındaki ilişki tarafından açıklanmaktadır. Bu oran çok önemsiz olarak görülebilir. Fakat 15 ile 16 yaşındaki kızların boy uzunlukları ortalamaları arasındaki fark, küçük seviyedeki etki büyüklüğüne örnektir.

Lieber, ortaya çıkarılmak istenen etki küçük bir etki ise bunu belirleyebilmek için araştırmacıların çok sayıda örnekleme ihtiyaçları olduğunu, eğer ortaya çıkarılmak istenen etki büyük ise bu büyük etkiyi belirleyebilmek için nispeten daha az örneklem sayısı gerektiğini belirtmiştir (Lieber, 1990, s. 306; Olejnik, 1984, s. 42).

Küçük etkiler, eğer sonradan büyük etki doğurma olasılıkları varsa oldukça önemlidirler (small sparks start big fires). 2008 yılı mart ayında Hong Kong'ta bir hastanede kaptığı virüs nedeniyle 7 yaşındaki bir çocuğun ölümü üzerine ülkedeki sağlık bakanlığı yetkilisinin endişe edecek bir durum olmadığını belirtmesine rağmen hükümet, tedbir amaçlı tüm okulları 2 haftalığına kapatmıştır. Yapılan otopsi sonucunda, zaten 2003 yılında SARS virüsünün çıkış yeri olarak kötü bir ün yapmış olan bu ülkede kuş gribi riskinin ciddi boyutlara ulaştığı haber bültenlerinde yer almıştır. İlk bakışta önemsiz gibi görünen bir ölüm nedeniyle alınan önlem, ülkedeki çoğu çocuğun hayatını kurtarmıştır. Aynı şekilde 2 Temmuz 1997'de Tayland hükümetinin para birimi Baht'ı devalüe etmesi Asya finansal krizini tetiklemiştir. 14 Eylül 2008'de finansal hizmet şirketi olan Lehman Brothers'ın iflasını ilan etmesi ile başlayan süreçte dünya büyük bir finansal krizin içine girmiştir. 2008 yılı Amerikan başkanlık seçimlerinde Barack Obama, kampanya döneminde sadece araba lastiklerinin uygun oranda şişirilmesi ile kişilerin yılda %3 oranında yakıttan tasarruf edebileceğini, bunu herkesin uygulaması halinde yıllık ithal edilen petrol miktarında binlerce varil azalma meydana geleceğini belirtmiştir. İşte bu gibi nedenler ile küçük sayılacak etkiler bazen çok büyük öneme sahip olabilirler (Ellis, 2010, s. 36).

Cohen, orta seviyedeki etki büyüklüğünün yaklaşık olarak değerini $d = 0.50$ olarak, $r = 0.3$ veya $r^2 = 0.09$ olarak tanımlamıştır. Ayrıca Cohen, orta seviyedeki etki için “benim niyetim orta seviyedeki etkiyi dikkatli bir gözlemci tarafından çıplak gözle fark edilebilecek bir etkiyi temsil etmesidir” demiştir.

Cohen ayrıca orta seviyedeki etki büyüklüğü için, çeşitli alanlardaki gözlemlenen etkinin yaklaşık olarak ortalamasıdır demiştir (Deng, 2000, s. 8-9). Orta seviyedeki etki büyüklüğü, iki grup arasında gözle görülebilecek kadar farkı ifade etmektedir. 14 ve 18 yaşındaki kızların boy uzunlukları arasındaki fark orta seviyedeki etkiye örnektir.

Büyük seviyedeki etki büyüklüğü için ise Cohen, küçük seviyedeki etki ile orta seviyedeki etki arasındaki mesafe kadar orta seviyedeki etkiden daha büyük bir etkidir demiştir (Deng, 2000, s. 9). Büyük seviyedeki etki değeri ise Cohen tarafından $d = 0.80$ olarak, $r = 0.50$ veya $r^2 = 0.25$ olarak tanımlamıştır. Doktora öğrencileri ile lise öğrencileri arasındaki IQ ortalamaları farkı büyük seviyedeki etkiye örnektir (McKean, 1990, s. 18-20). Ancak Rosenthal (1996), Cohen’in sınıflandırmaları ile yetinmemiş bir de çok büyük etki tanımını yapmıştır. Bu sınıflandırma için Rosenthal, Cohen’in küçük, orta ve büyük etki sınıflandırmasına ilaveten çok büyük etkinin $d = 1.30$ olduğunu belirtmiştir.

Cohen’in etki büyüklüğü sınıflandırmalarının iki önemli noktası vardır. Birincisi, kavraması kolaydır. Elde edilen değerler ile Cohen’in belirttiği değerler kolayca kıyaslanıp yorum yapılabilir. İkincisi, Cohen’in bu değerlerin alandan alana farklılık gösterebileceğini belirtmesine rağmen çoğu araştırmacı tarafından bu değerler mantıklı bulunmuştur. Fakat avantajları olmasına rağmen bazı noktalarda bu sınıflandırmalar üzerinde tartışmalar da mevcuttur. Cohen’in yapmış olduğu bu sınıflandırmalara meta-analizinin geliştiricilerinden birisi olan Gene Glass, bunlar olsa olsa t-shirt ölçüleri olur demiştir. Thompson (2008) ise sınıflandırmanın esnek olmamasını eleştirmiştir (Ellis, 2010, s. 42).

Olejnik, beklenen etki büyüklüğünün Cohen’in belirtmiş olduğu değerler seviyesinde belirlenebileceğini ve bu değerlerin araştırmacılar tarafından bilinen en iyi ve en geniş çapta kabul görmüş bir rehber olduğunu belirtmiştir. İlaveten Olejnik, etki büyüklüğünün araştırmanın planlama aşamasında olması gerekenden düşük belirlenmesinin araştırmaya vereceği zararların, olması gerekenden büyük belirlenmesinden daha az olacağını düşünmektedir. Aşırı yapılan etki büyüklüğü tahmininin istatistiksel gücü azaltacağını belirtmiştir (Olejnik, 1984, s. 42).

Bazı deęişimler olmasına rağmen Cohen tarafından yapılmış olan etki büyüklüğü seviyeleri yıllarca adeta bir standart tanım gibi arařtırmacılar tarafından rehber olarak kullanılmıştır. Cohen'in bu yöntemi kullanılarak anlamlılık seviyesi ve örneklem sayısı gibi veriler varsa istatistiksel gücü belirlemek mümkündür (Deng, 2000, s. 9).

Güç arařtırmalarında kritik bir durum, Cohen tarafından önerilmiş olan küçük, orta ve büyük seviyelerdeki etki büyüklüklerinin kullanılmasıdır. Eğer arařtırmalarda Cohen tarafından önerilen etki seviyelerinden gerçekte daha büyük etkiler olsaydı, güç seviyeleri daha büyük olurdu. Ancak yapılan arařtırmalarda anakütle üzerinde gözlemlenen etki büyüklüklerinin genellikle Cohen'in tavsiye ettiği seviyelere yakın deęerler olduęu görülmektedir. Haase vd. (1982), *Journal of Counseling Psychology* dergisinde 10 yıllık bir süre içinde yayımlanmış olan çalışmalar üzerinde gerçekleřtirdikleri incelemelerde medyan etki büyüklüğünü $d = 0.60$ olarak bulmuşlardır. Bu deęer Cohen'in orta seviyeli etki büyüklüğü sınıflandırmasına çok yakın bir deęerdir. Lipsey ve Wilson, 1993 yılında yaptıkları arařtırmalarda 302 adet meta analizinin sonucunda etki büyüklüğünü ortalama olarak $d = 0.50$ seviyesinde bulmuşlardır. Richard vd., 2003 yılında sosyal psikoloji alanında 25.000 bireysel çalışmayı ve toplamda 8.000.000 örneklemini içeren 100 yıllık bir süreçte yayımlanmış olan çalışmalarını incelemişler ve ortalama etki büyüklüğünü $d = 0.43$ olarak bulmuşlardır. Görüldüğü gibi bu deęerler Cohen'in deęerlerine çok yakındır.

Tüm bu deęerler Cohen'in yaptığı tahminlerin ne kadar güvenilir olduęunu göstermektedir. Ancak Cohen, bu deęerlerin bir alandan dięerine deęişiklik gösterebileceğini ve dolayısı ile tüm alanlar için bir standart deęer olarak algılanmamaları gerektiğini belirtmiştir. Örneğin, psikoloji alanında büyük bir etki yakalamak oldukça zor bir durumdur. Arařtırmacılar ancak laboratuvar ortamındaki bir deneyde büyük etkiler ortaya çıkarabilmektedir. Rutledge ve Loh, 2004 yılında saęlık psikolojisi alanında 14 çalışmayı incelemiş ve ortalama etkinin $d = 0.161$ olduęu sonucuna ulaşmışlardır. Benzer şekilde Yin vd. 2011 yılında yaptığı arařtırmalarda ortalama etkiyi $d = 0.096$ olarak bulmuştur. Görüldüğü gibi Cohen'in yapmış olduęu sınıflandırmaların bir standart olarak kullanılması yerine belirli alanlara göre revize edilmesi gerekmektedir. Tablo 1.3'te saęlık psikolojisi alanında Rossi tarafından 2003 yılında revize edilmiş etki büyüklüğü deęerleri verilmiştir:

Tablo 1.3 J. Rossi Tarafından Sağlık Psikolojisi Alanı İçin Revize Edilen Cohen'in Etki Büyüklüğü Değerleri

Etki Büyüklüğü	Cohen (1988)		Rossi (2003)	
	<i>d</i>	η^2	<i>d</i>	η^2
Küçük	0.20	.010	0.15	.005
Orta	0.50	.059	0.20	.010
Büyük	0.80	.138	0.25	.015

Araştırmacılar için etki büyüklüğünün anlamını bilmeden onu rapor etmek pek alışılmadık bir durum değildir. Örneğin bir araştırmacı çalışmada değişkenler arasındaki ilişkiyi göstermek için korelasyon matrisini oluşturduğunda test istatistiklerini rapor ederken aynı zamanda etki büyüklüğü de rapor etmiş olurlar (R^2 gibi) ancak onu yorumlamazlar (Ellis, 2010, s. 16).

1.7.2.2 Etkinin Yorumlanması

Sosyal bilimler alanındaki araştırmacılar için anlamlılık testi sonuçlarına göre yorum yapmak nadir görülen bir durum değildir. Kimi araştırmacılar yüksek seviyede anlamlı çıkan sonuçlarını, orta seviyede anlamlı olan sonuçlara göre daha önemli olarak nitelendirebilirler. Kimi araştırmacılar ise anlamlı olmayan bir sonuca ulaştıkları anda çalışmalarını herhangi bir etkinin olmadığını belirterek sonuçlandırabilirler. Her iki durumda da yanlış yorum yapılma ihtimali oldukça fazladır. Etki büyüklüğünün yorumu, anlamlılık testinden ayrı olmalıdır. Her ikisi de birbiri ile ilişkili olmasına rağmen, istatistiksel anlamlılık örneklem sayısından etkilenmektedir. Daha büyük örneklem ile daha yüksek ihtimalle anlamlı sonuçlar elde edilebilir. Fakat $p = 0.01$ değeri, $p = 0.05$ değerinden daha önemli olmak zorunda değildir. Fakat nispeten daha büyük etkilerin, küçük etkilerden daha önemli olduğu söylenebilir. Örneğin Cohen's $d = 1$ değeri, $d = 0.2$ değerinden daha önemlidir.

Bir çalışma sonucunun pratik olarak değerlendirilebilmesi için etki büyüklüğünün değerini bilmek yetmez. Aynı zamanda anlam çıkarabilmek için onu yorumlamak gerekir. Peki, anlam çıkarabilmek için ne kadar etki elde etmek gerekir? Etkiler, belirli referans alınan ölçeklere göre yorumlanmadıkça bir anlam ifade etmezler. Örneğin, bir öğrencinin herhangi bir sınavdan 140 puan aldığına kulak misafiri olan bir kişi, bu puanın IQ test puanı mı yoksa

GMAT (Graduate Management Admission Test) puanı mı olduğunu bilmeden bunu yanlış yorumlaması kaçınılmazdır. Eğer bu puan IQ puanı ise oldukça iyi bir seviyedir, GMAT puanı ise bu o kişinin yurt dışında iyi bir üniversitede yüksek lisans yapması için yeterli bir puan değildir. Bu konu ile ilgili bir başka örnek daha verilebilir. Örneğin, bir doktorun yapılan incelemeler sonucunda hastanın vücut-kitle indeksi ve hareketsiz yaşam tarzı puanının, ortalama kalp krizi riski sınırından 2 puan daha aşağıda olduğunu söylemesi muhtemelen hastayı daha sağlıklı bir yaşam tarzına teşvik etmez. Çünkü hastanın bu iki puanın anlamını bilmeye ihtiyacı vardır. Ciddi bir risk var mı? Hastanın endişelenmesi gerekir mi? Hasta, ölçөгünü bilmediğı bir puan ile karşılaştığı için bu durumdan bir sonuç çıkaramaz. Ancak doktor hastaya: “Sizin vücut-kitle indeksi ve hareketsiz yaşam tarzı puanınızda olan insanlar, normal insanlarla ile karşılaştırıldığında, onlara göre sizin 10 yıl içinde 4 kat daha fazla kalp krizi geçirme riskiniz var” şeklinde bir açıklama yaparsa, kişinin tüm dikkatini konu üzerine toplamış ve kesinlikle kişiyi hayat tarzı ile ilgili radikal kararlar almaya zorlamış olur. İşte etki büyüklüğünün rapor edilmesinin yanında iyi bilinen bir ölçөгө göre yorumunun yapılması bu derece bir öneme sahiptir (Ellis, 2010, s. 32).

Etki büyüklüğü ile p değeri arasındaki yorum farkının öneminin daha iyi anlaşılmasını sağlayacak bir örnek aşağıda verilmiştir:

	<i>N</i>	Mean	<i>SD</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Çalışma 1						
A Grubu	15	25	9	1.52	>.05	0.56
B Grubu	15	20	9			
Çalışma 2						
A Grubu	30	25	9	2.15	<.05	0.56
B Grubu	30	20	9			

A ve B grubuna dâhil edilen kişilerin, bilim kurgu filmleri alanındaki bilgileri ölçülmeye çalışılmış ve yapılan t-testi sonuçları yukarıda verilmiştir (alfa = 0.05). Tablodan görüldüğü gibi ortalama puanlar ve standart sapmalar her iki çalışmada da aynıdır. Fakat birinci çalışma $p>0.05$ olduğu için istatistiksel olarak anlamsızdır. Bu sonuç, birinci çalışmayı gerçekleştiren araştırmacıyı her iki grup arasında bilim kurgu alanında sahip olunan bilgi açısından dikkate değer bir farklılık olmadığı sonucuna itecektir. Buna karşılık ikinci çalışmanın sonucu $p<0.05$

çıkacağı için istatistiksel olarak anlamlıdır ve iki grup arasındaki 5 puanlık fark dikkate almaya değerdir. Bu durumda sorulması gereken soru şudur: Nasıl oluyor da her iki çalışmada aynı etki büyüklüğü ($d = 0.56$) elde edilmişken bu şekilde bir farklılık meydana geliyor?

Bu sorunun cevabı anlamlılık testinin sonuçlarının yanlış yorumlanmasıdır. 1. çalışmayı gerçekleştiren araştırmacı etki büyüklüğünü dikkate almamış, sadece p değerine odaklanmıştır. Anlamlı olmayan bir sonucun yanlış yorumlanması nedeniyle anlamlı bir etkinin olmadığı sonucuna ulaşılmıştır. Herhangi bir etki olmayabilir veya etki belirleyebilmek için çalışma, o etkiyi ortaya çıkaracak olan yeterli güç ile tasarlanmamış olabilir. Bu durumda yapılan 2. çalışma, 1. çalışmanın etki belirleyebilmek için yeterli güçten yoksun olduğunu ortaya koymaktadır. Bunun da muhtemel nedeni örneklem sayısının yetersiz olmasıdır. G-Power 3.1.3 programı ile her iki çalışmanın güç seviyeleri hesaplanmış ve 1. çalışmanın istatistiksel gücünün %44, 2. çalışmanın istatistiksel gücünün ise %69 olduğu görülmüştür (Ellis, 2010, s. 34). Sosyal bilimler alanında araştırmacılar kendilerini kesinlikle p değeri ile sınırlamamalı, daha da öteye bakmalıdırlar.

1.7.2.3 Etki Büyüklüğü için Güven Aralığı (Confidence Interval) Tahmini

Bir araştırmacı, deney ve kontrol grupları olmak üzere iki grubu, aralarındaki farklılıkları belirlemek amacıyla karşılaştırmak istesin. Her bir gruba örneklemelerin rassal olarak atandığı, normal dağılım altında varyansların eşit olduğu varsayımları dikkate alınsın. Araştırmacı çalışmasını çift yönlü $\alpha = 0.05$ ile bağımsız gruplar t -testini kullanarak gerçekleştirecek olsun ve %80 güç seviyesini yakalamak istesin. Geriye, belirlenmesi gereken tek değer etki büyüklüğü kalacaktır. Etki büyüklüğü değeri, Cohen'in etki büyüklüğü sınıflandırmalarından yararlanılarak $d = 0.50$ olarak seçilsin. Bu durumda çalışma için gerekli olacak örneklem sayısı $n_1 = 64$ ve $n_2 = 64$ olmak üzere toplamda 128 olacaktır. Çalışma gerçekleştirildikten sonra elde edilen gruplar arası farklılıkların ortalama değeri tam da Cohen'in sınıflandırmasından seçildiği gibi $d = 0.50$ olarak bulunmuş ve istatistiksel olarak anlamlı bir sonuç elde edilmiş olsun. Bu durum tüm araştırmacıların isteyeceği mutlu bir sonudur. Ancak elde edilen etki büyüklüğü değeri örnekleme ait tahmini bir değerdir ve örneklemden örnekleme değişebilir. Ancak birçok kaynak (*Publication Manual of the American Psychological Association*, *American Educational Research Association* vs. gibi) çalışma sonuçlarını içeren raporlarda etki büyüklüğü tahminini içeren güven aralığının da verilmesi gerektiğini belirtmektedir. Wilkinson (1999), güven aralıklarının rapor edilmesinin çalışma sonuçlarının doğru yorumlanması açısından çok etkili bir yöntem olduğunu ve bu yüzden güven aralıklarının da çalışma sonuçları ile birlikte verilmesinin şiddetle tavsiye edildiğini

belirtmiştir. Yapılan tüm güven aralığı tavsiyelerini dikkate alarak örnekte verilen araştırmacının güven aralığını belirlemesi gerekmektedir. Goodman ve Berlin (1994), aşağıdaki basit bir yaklaşım ile güven aralıklarının nasıl belirleneceğini göstermişlerdir:

$$\underline{\text{Etki Büyüklüğü İçin \%95 Güven Aralığı (\%80 Güç)}} \quad (1)$$

$$\text{Etki Büyüklüğü} \pm 0.7 \times \text{Etki Büyüklüğü}$$

$$\underline{\text{Etki Büyüklüğü İçin \%95 Güven Aralığı (\%90 Güç)}} \quad (2)$$

$$\text{Etki Büyüklüğü} \pm 0.8 \times \text{Etki Büyüklüğü}$$

Araştırmacı, çalışmasını %80 güç ile gerçekleştirdiğinden etki büyüklüğü için güven aralığı tahminini yaparken yaklaşım (1)' i kullanmalıdır. %80 güç ve $d = 0.50$ güven aralığı $0.1 < d < 0.85$ şeklinde olacaktır. Güven aralığı dikkate alınarak yapılması gereken yorum, etki büyüklüğünün yaklaşık olarak $0.85 - 0.15 = 0.70$ 'lik bir genişlikte yer aldığı ve tahmin edilen etki büyüklüğü $d = 0.50$ 'nin çok kesin bilgiler vermeyeceğidir. Gerçek etki büyüklüğü 0.16 gibi çok küçük bir etki de olabilir, 0.84 gibi çok büyük bir etki de olabilir. Araştırmacının %80 gibi yeterli bir güç seviyesi ile çalışmasına rağmen bu kadar geniş bir etki büyüklüğü aralığı elde etmesinin nedeni, araştırmaya dâhil edilen 128 kişilik örneklem sayısının yeterli gücü sağlamasına rağmen gerçek etkiyi yüksek bir kesinlik ile ortaya çıkarabilmesinin yeterli olmamasıdır.

Bu örnek şunu göstermiştir ki örneklem sayısı, istenen güç seviyesinin elde edilmesi açısından yeterli olsa bile doğru parametre tahmini açısından yeterli olmayabilir. Eğer bir çalışma sadece büyük etkileri ortaya çıkaracak kadar istatistiksel güce sahip ise güven aralığı bu durumda daha geniş olacaktır.

Bu örnekten çıkarılması gereken bir başka sonuç ise örneklem sayısı planlama aşamasının bazen yanlış bir sıfır hipotezinin reddedilme olasılığını artıracak şekilde, bazen de mümkün olduğunca dar bir güven aralığı elde edilmesini sağlayacak şekilde gerçekleştirilebileceğidir (Maxwell vd., 2008, s. 541-542).

Cohen, arařtırmacıların neden güven aralıęı sonuçlarına alıřmalarının rapor veya sonuç bölümünde yer vermedikleri konusunda “bunun ana nedeninin, güven aralıklarının utan verici olacak řekilde geniř olmasından řüphe ediyorum” demiřtir (Cohen, 1994, s. 1002).

1.7.3 Örnekleme Sayısı

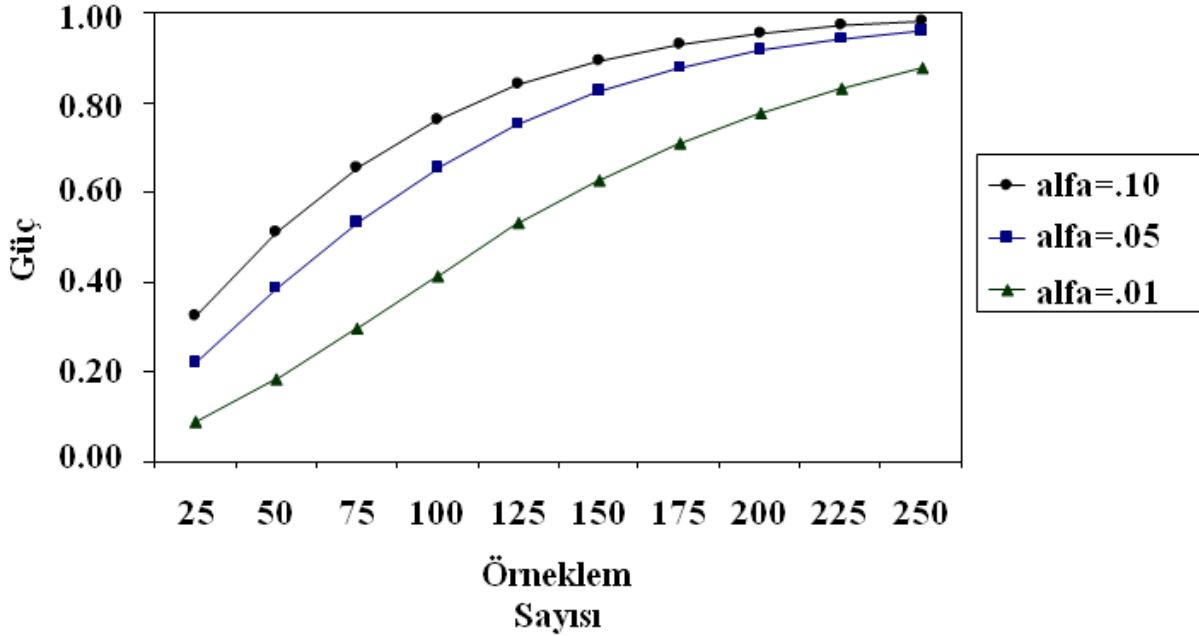
Örnekleme sayısı, gücü doğrudan etkileyen bir dięer parametredir. Örnekleme sayısı, alıřma için kullanılan denek sayısıdır. Örnekleme sayısı, bir istatistiksel testin gücünü etkileyen öncelikli parametredir. DiLullo'nun (1997) aktardığına göre, Keppel (1991), arařtırmacıların maliyet ve ulařılabilirlik, arařtırmanın tasarım ařamasındaki kısıtlar aısından ve yeterli istatistiksel güç aısından uygun bir örnekleme sayısını belirlemeleri gerektiğini belirtmiřtir (s. 8-9).

Arařtırmanın tasarım ařamasındaki önemli konulardan birisi de belirlenen güç seviyesini elde edebilmek için gerekli örnekleme sayısının ne olması gerektiğinin bilinmesidir (MacCallum vd., 1996, s. 142). Arařtırılan konu ile ilgili gözlem sayısı arttıka örneklemden elde edilen verilerin güvenilirlięi de artacaktır. Artan bu güvenilirlięin bir sonucu olarak arařtırmacı yüksek olasılıkla yanlıř olan sıfır hipotezini doğru bir řekilde reddedebilir. Bu durumda örnekleme sayısının artması gücü olumlu yönde etkilemektedir. İdeal olarak bir arařtırmacı alfa, etki büyüklüğü ve arzuladığı güç seviyesini kullanarak alıřması için gerekli optimum örnekleme sayısına ulařabilir (Cashen ve Geiger, 2004, s. 155). Arařtırmalarda anlamlı sonuçlar elde etme olasılıęı büyük örnekleme sayıları ile birlikte artarken, bu olasılık daha küçük örnekleme sayıları ile birlikte azalır. ünkü daha büyük örnekleme, hedef anakütle parametreleri hakkında daha net tahminler yapılmasını saęlar (Balkin ve Sheperis, 2011, s. 269).

Örnekleme sayısı artarsa standart hata azalacağı için güç artmaktadır. Küçük sayıdaki örnekleme ile gerekleřtirilen arařtırmalarda, etki büyüklüğü gerekte büyük olmasına raęmen arařtırmacı sıfır hipotezini reddetme konusunda başarısız olabilir ve mevcut olan önemli farklılıkları belirleyemeyebilir. Aksine, örnekleme sayısı olması gerekenden çok fazla olan alıřmalarda gerekte anlamsız olan etkiler örnekleme sayısının çok fazla olması nedeniyle yanlıř bir řekilde anlamlı olarak bulunabilir. O nedenle mümkün olduęu müddete optimum örnekleme sayısı ile alıřılmalıdır (Sacco, 1982, s. 755; Mazen vd., 1987, s. 370; Coblick, 1998, s. 65; Jones ve Sommerlund, 2007, s. 227; Balkin ve Sheperis, 2011, s. 269).

Anlamlılık seviyesi kurulduğunda ve etki büyüklüğüne bir teori veya yapılan bir ön alıřma ile karar verildiğinde, örnekleme sayısı istatistiksel alıřmanın uygulama ařamasından önce

belirli bir güç elde etmek için kullanılır. Diğer parametreler sabit iken örneklem sayısı ile güç arasında doğru orantılı bir ilişki vardır. Yani örneklem sayısı artarsa güç artmaktadır. Bu durum Şekil 1.5'te gösterilmiştir (DiLullo, 1997, s. 9).



(Etki Büyüklüğü (d) = 0,33)

Şekil 1.5 İstatistiksel Güç – Örneklem Sayısı Arasındaki İlişkinin Çift Yönlü Farklı Alfa Seviyelerinde Gösterilmesi

Aslında burada sorulması gereken soru hipotezin doğru bir şekilde test edilmesi için ne kadar örnekleme ihtiyacı olduğudur. Bu sorunun cevabı, güç analizini kullanmadan, genellikle mantıklı olmayan esaslara dayanmaktadır (Ellis, 2010, s. 61).

Örneklem sayısını ayarlamak, araştırmacılar tarafından gücü ayarlamaktan daha kolay bir iş olarak görülmektedir. Örneğin, genellikle örneklem sayısı geçmişte yapılmış olan çalışmalara, araştırma geleneklerine, araştırmacının mevcut mali gücüne, deneyimlere, araştırma konusu üzerinde derin bilgisi olan uzmanlar ile yapılan müzakerelere ve verilerin mevcudiyetine göre seçilmektedir. Örneklem sayısının seçiminde araştırmacılar böyle basit yöntemleri kullanarak genellikle gereğinden çok daha fazla örneklem ile çalışırlar ki bu da daha fazla zaman ve maliyet külfeti demektir. Ya da olması gerekenden daha az bir örneklem sayısı ile çalışırlar ki bu da düşük güce neden olmaktadır. Bu gibi durumlardan kaçınmak için araştırmacılar, arzu ettikleri gücü, anlamlılık seviyesini ve etki büyüklüğünü kullanarak araştırmanın tasarım aşamasında örneklem sayısını belirlemelidirler (McKean, 1990, s. 22; DiLullo, 1997, s. 10).

Sosyal bilimler alanındaki arařtırmacılar, arařtırmaları için gerekli olan örneklem sayısına karar verme konusunda net bir bilgiye sahip deęillerdir. Genellikle örneklem sayısı konusunda “*ulařabildięin kadar örnekleme kullan*” yöntemi kullanılmaktadır. Ancak bu yöntem çoęu zaman iře yaramamaktadır. Bazen ulařılan örneklem sayısı sıfır hipotezini reddetmek için yeterli olmayabilir. Bazen de normalde ulařılan örneklem sayısından çok daha az sayıdaki örneklem ile arařtırma tamamlanabilmekte ve harcanan emek ve para aısından savurganlık durumun söz konusu olmaktadır (Olejnik,1984, s. 40).

Yazında nitel arařtırma yöntemlerinin planlama ařamasında istatistiksel güç analizinin önemi konusunda çok fazla tartıřma vardır. Olejnik (1984), arařtırmaların planlama ařamasının önemli bir kısmının uygun örneklem sayısına karar vermek olduęunu belirtmiřtir. Arařtırmacılar genellikle kullandıkları örneklem sayısını neden seçtiklerini belirtmezler. Örneklem sayısına karar vermek için dört önemli parametre vardır;

- i) Anlamlılık (alfa) Seviyesi
- ii) İstatistiksel Güç
- iii) Etki Büyüklüęü
- iv) Kullanılan İstatistiksel Test

Olejnik’e göre (1984), alfa, güç ve kullanılacak istatistiksel testin neden seçildięi sorusunun cevabı açıktır. Fakat etki büyüklüęünü seçme konusu çok meřakkatlidir. Bu nedenle Olejnik, arařtırmacılara arařtırma sonuçlarında elde ettikleri etki büyüklüęü ve örneklem sayısını açıklayarak, sonraki arařtırmalarda alıřacak olan arařtırmacılara ışık tutulması gerektięini belirtmiřtir (Olejnik, 1984, s. 41-42).

Ayrıca örneklem sayısının artırılması, gücü artırmak için tek yol deęildir. Etki büyüklüęü ve anlamlılık seviyesini artırmanın yanında rastgele ölçüm hatasını azaltma, yüksek güvenilirlikle arařtırma ve ölçüm araçları tasarlama, uygulamalar yaparak hatasız iřlemler geliřtirme ve anakütlenin homojenlięini artırmak gibi yöntemler kullanılarak da arařtırmanın gücü artırılabilir (DiLullo, 1997, s. 11).

Örneklemden elde edilen deęerlerin güvenilirlięi Cohen tarafından, yaklaşık olarak bulunmak istenen anakütle deęerlerine olan yakınlık olarak tanımlanmıřtır (Cohen, 1977, s. 6). Gerçek anakütle deęerleri, ölçüm birimi vs. gibi dięer faktörlerin de etkilemesine raęmen, sonuçların güvenilirlięini etkileyen en önemli faktör örneklem sayısıdır.

Güç konusu, örneklem sayısına karar verme aşamasında dolaylı olarak dikkate alınmaktadır. Araştırmacılar genellikle örneklem sayısına karar verme aşamasında etkin bir yöntem olan güç analizi yerine geçmişteki tecrübelerle, verilerin ulaşılabilirliğine veya ulaşabildiği kadar örneklem metoduyla çok da verimli olmayan yöntemler kullanmaktadır. Bu şekilde güç analizi üzerine inşa edilmeyen araştırmalar maliyet ve zaman açısından etkin değildir. Bir güç analizi, araştırmacıya yeterli sayıda örneklem ile çalışmayı ve anakütle üzerindeki eğer varsa gerçek etkiyi belirleme şansı vermektedir (Coblick, 1998, s. 83).

Örneklem sayısı kararları genellikle geleneklere, verilerin mevcudiyetine, tecrübe ve müzakerelere ve nadiren de olsa veri toplamadan önce yapılan II. Tip hata bazlı analizlere bağlı olarak alınmaktadır (Cohen, 1962, s. 145). Örneklem sayısına genellikle keyfi ve gelişmiş güzel yöntemlerle karar verilmektedir. McKean'ın (1990) aktardığına göre, Brewer ve Sindelar (1988), kimi durumlarda yazarların örneklem sayısına, karar verilmesi gereken bir değişkenden ziyade sadece örneklem sayısının “fazla” olması gerektiği şeklinde yaklaşımlarının olduğunu ifade etmişlerdir. Buradaki “fazla” kelimesindeki kasıt nedir? Örneğin, medikal araştırmalarda $n=30$ yeterli sayıda örneklem olabilirken, sosyal bilimlerde bu sayı örneklem için oldukça düşüktür. Sosyal bilimler alanında örneklem sayısının $n=30$ ve anlamlılık seviyesinin 0.05 olduğu durumda araştırmacı, yeterli güce sadece büyük seviyedeki etki büyüklüğü ile ulaşabilmektedir. Eğer etki büyüklüğünün seviyesi küçük olursa o zaman elde edilecek güç çok daha düşük olacaktır. Eğer etki büyüklüğü 0.50 olursa güç 0.61, etki büyüklüğü büyük olursa güç yeterli seviye olan 0.80 değerinde olacaktır. Bu nedenle güç analizi, anlamlılık testi için gerekli olan örneklem sayısının belirlenmesinde gerçekçi bir yöntem olarak öne çıkmaktadır (s. 23).

Bir teorik güç analizi, örneklem sayısının ne olacağı sorusuna kolaylıkla cevap vermektedir. Örneğin, $r = 0.40$ etki büyüklüğüne sahip, çift yönlü $\alpha = 0.05$ ve %80 güç ile bağımsız gruplar t-testi için sahip olunması gereken minimum örneklem sayısı yaklaşık olarak $N = 46$ olacaktır. Eğer $r = 0.20$ seviyesinde bir etki belirlenecek ise gerekli olan minimum örneklem sayısı $N = 193$ olacaktır. Görüldüğü gibi daha küçük bir etkiyi belirlemek daha fazla örneklem sayısı gerektirmektedir. Bunun nedeni küçük etkilerin ortaya çıkarılmasının zor olmasından kaynaklanmaktadır. Ancak gerekli örneklem sayısına karar verirken en zahmetli iş doğru bir etki büyüklüğü tahmin etmektir. Yukarıdaki örneklerde de görüldüğü üzere etki büyüklüğünün gerekli olan örneklem sayısı üzerinde çok büyük etkisi vardır. Eğer beklenen yani tahmin edilen etki, var olan gerçek etkiden büyük ise işte o zaman kullanılan örneklem sayısı ile yeterli güç seviyesi sağlanamayacağından daha küçük bir etkiyi ortaya

çıkarma konusunda başarısız olunabilir. En sağlıklı etki büyüklüğü tahmini için yapılması gereken daha önce yapılmış benzer çalışmaları incelemektir (Ellis, 2010, s. 62).

Tablo 1.4 Korelasyon Katsayısı İçin Minimum Örneklem Sayıları ($\alpha_1 = 0.05$ ve tek yönlü)

$r(\alpha_1)$	Güç									
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
0.05	1,083	1,254	1,441	1,648	1,881	2,150	2,471	2,873	3,422	4,324
0.10	271	314	360	412	470	537	616	716	853	1,077
0.15	121	140	160	183	208	238	273	317	377	476
0.20	68	79	90	103	117	133	153	177	211	266
0.25	44	50	58	66	75	85	97	113	134	168
0.30	31	35	40	45	51	59	67	77	92	115
0.35	23	26	29	33	38	43	49	56	67	83
0.40	18	20	23	25	29	32	37	42	50	63
0.45	14	16	18	20	22	25	29	33	39	49
0.50	12	13	14	16	18	20	23	26	31	38
0.55	10	11	12	13	15	16	19	21	25	31
0.60	8	9	10	11	12	14	15	17	20	25
0.65	7	8	9	9	10	11	13	14	17	21
0.70	6	7	7	8	9	10	11	12	14	17
0.75	6	6	6	7	7	8	9	10	12	14
0.80	5	5	6	6	6	7	8	8	10	11
0.85	4	5	5	5	6	6	6	7	8	9
0.90	4	4	4	5	5	5	5	6	6	8
0.95	4	4	4	4	4	4	4	5	5	6

Tablo 1.5 Pearson Korelasyon Katsayısı İçin Minimum Örneklem Sayıları ($\alpha_2 = 0.05$ ve çift yönlü)

$r(\alpha_2)$	Güç									
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
0.05	1,536	1,740	1,959	2,199	2,467	2,774	3,137	3,588	4,198	5,192
0.10	384	435	489	549	616	692	782	894	1,046	1,293
0.15	171	193	217	243	273	306	346	396	462	571
0.20	96	108	122	136	153	171	193	221	258	319
0.25	62	69	78	87	97	109	123	140	164	202
0.30	43	48	54	60	67	75	84	96	112	138
0.35	31	35	39	44	49	54	61	70	81	100
0.40	24	27	30	33	37	41	46	53	61	75
0.45	19	21	23	26	29	32	36	41	47	58
0.50	15	17	19	21	23	26	29	32	37	46
0.55	13	14	15	17	19	21	23	26	30	37
0.60	11	12	13	14	15	17	19	21	24	30
0.65	9	10	11	12	13	14	16	18	20	24
0.70	8	9	9	10	11	12	13	15	17	20
0.75	7	7	8	9	9	10	11	12	14	16
0.80	6	6	7	7	8	8	9	10	11	13
0.85	5	6	6	6	7	7	8	8	9	11
0.90	5	5	5	5	6	6	6	7	8	9
0.95	4	4	4	5	5	5	5	5	6	7

Tablo 1.6 İki Grup Arasında Anlamlı Bir Farklılık Belirlemek İçin Gerekli Örneklem Sayıları ($\alpha_1 = 0.05$ ve tek yönlü)

$d(\alpha_1)$	Güç									
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
0.10	1,084	1,256	1,443	1,650	1,884	2,154	2,475	2,878	3,427	4,331
0.20	272	315	362	414	472	540	620	721	858	1,084
0.30	122	141	162	185	211	241	277	321	382	483
0.40	70	80	92	105	120	136	156	182	216	272
0.50	45	52	60	68	77	88	101	117	139	175
0.60	32	37	42	48	54	62	71	82	97	122
0.70	24	28	31	36	40	46	52	61	72	90
0.80	19	22	24	28	31	36	41	47	55	70
0.90	15	17	20	22	25	29	32	37	44	55
1.00	13	15	16	18	21	23	27	31	36	45
1.10	11	12	14	16	18	20	22	26	30	38
1.20	10	11	12	14	15	17	19	22	26	32
1.30	9	10	11	12	13	15	17	19	22	28
1.40	8	9	10	11	12	13	15	17	19	24
1.50	7	8	9	10	11	12	13	15	17	21
1.60	7	7	8	9	10	11	12	13	15	19
1.70	6	7	7	8	9	10	11	12	14	17
1.80	6	6	7	7	8	9	10	11	13	15
1.90	6	6	6	7	8	8	9	10	12	14

Tablo 1.7 İki Grup Arasında Anlamlı Bir Farklılık Belirlemek İçin Gerekli Örneklem Sayıları ($\alpha = 0.05$ ve çift yönlü)

$d(\alpha_1)$	Güç									
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
0.10	1,539	1,742	1,962	2,203	2,471	2,779	3,142	3,594	4,205	5,200
0.20	387	437	492	552	620	696	787	900	1,053	1,302
0.30	173	196	220	247	277	311	351	401	469	580
0.40	98	111	125	140	157	176	199	227	265	327
0.50	64	72	81	90	101	113	128	146	171	210
0.60	45	51	57	64	71	80	90	102	119	147
0.70	34	38	42	47	53	59	67	76	88	109
0.80	27	30	33	37	41	46	52	59	68	84
0.90	22	24	27	30	33	37	41	47	54	67
1.00	18	20	22	25	27	30	34	38	45	54
1.10	15	17	19	21	23	26	29	32	37	45
1.20	13	15	16	18	20	22	24	28	32	39
1.30	12	13	14	16	17	19	21	24	27	33
1.40	11	12	13	14	15	17	19	21	24	29
1.50	10	11	11	13	14	15	17	19	21	26
1.60	9	10	10	11	12	14	15	17	19	23
1.70	8	9	10	10	11	12	14	15	17	21
1.80	8	8	9	10	10	11	12	14	16	19
1.90	7	8	8	9	10	11	11	13	14	17

Yukarıdaki tablolarda, belirlenmek istenen d ve r ailelerine ait olan farklı etki büyüklükleri ve güç seviyelerine göre minimum örneklem sayıları verilmiştir. Gerekli olan minimum örneklem sayısını belirlemek için istenen güç seviyesi ve etki büyüklüğünün kesiştiği noktaya bakmak yeterlidir. Örneğin, iki grup arasındaki farkın etki büyüklüğü olarak $d = 0.50$ olduğu varsayılırsa ve %80'lik güç seviyesinde bu etkiyi ortaya çıkarmak için gerekli olacak minimum toplam örneklem sayısı 128 dir. Burada iki grup karşılaştırıldığı için bulunan örneklem sayısı toplam örneklem sayısıdır. Yani her bir grup eşit örneklem sayısına sahip ise her biri 64 örneklemden oluşmalıdır. Eğer gerçek etkiler gözden kaçırılmak istenmiyorsa güç seviyesi %90'a çıkarılarak her bir grupta 86 örneklem olmak üzere toplamda 172 örneklem gerekecektir. Buradaki tüm değerler G-Power 3.1.3 programı tarafından hesaplanmıştır (Ellis, 2010, s. 62).

1.7.3.1 Belirlenebilecek En Küçük Etki

Örnekleme sayısının, bir araştırmada belirlenebilecek en küçük etki üzerinde önemli bir role sahip olduğu açıktır. Birçok durumda anlamlı istatistiksel sonuçlara ulaşma veya ulaşamama örnekleme sayısına bağlıdır. Bir araştırmaya başlamadan önce araştırmacıların özellikle bazı faktörler üzerinde önemle durmaları gerekmektedir. Bunlardan birisi tasarlanan çalışmanın belirleyebileceği en küçük etki seviyesinin ne olduğu sorusunun cevabıdır. Güç analizinin başka bir kullanım alanı ise mevcut örnekleme ile ortaya çıkarılabilecek minimum etkiyi belirlemektir (Rose ve Bowen, 2009, s. 44). Aşağıda verilen Tablo 1.4'te standart alfa ve güç seviyeleri sırasıyla 0.05 ve %80 için belirlenebilecek en küçük etki büyüklükleri gösterilmektedir. Bu hesaplamalar G-Power 3.1.3 programı ile bağımsız gruplar t-testi için yapılmıştır. Örneğin, bir kişi örnekleme olarak $n = 100$ kişiye ulaşabildiyse standart iki yönlü alfa ve güç seviyeleri için belirleyebileceği en küçük etki $r = 0.276$ veya $d = 0.566$ olacaktır. Örnekleme sayısı iki katına çıkarılırsa belirlenebilecek en küçük etki seviyeleri $r = 0.197$ ve $d = 0.398$ olacaktır (Ellis, 2010, s. 63).

Tablo 1.8 Verilen Bir Örneklem Sayısı İçin Belirlenebilecek En Küçük Etki Seviyeleri

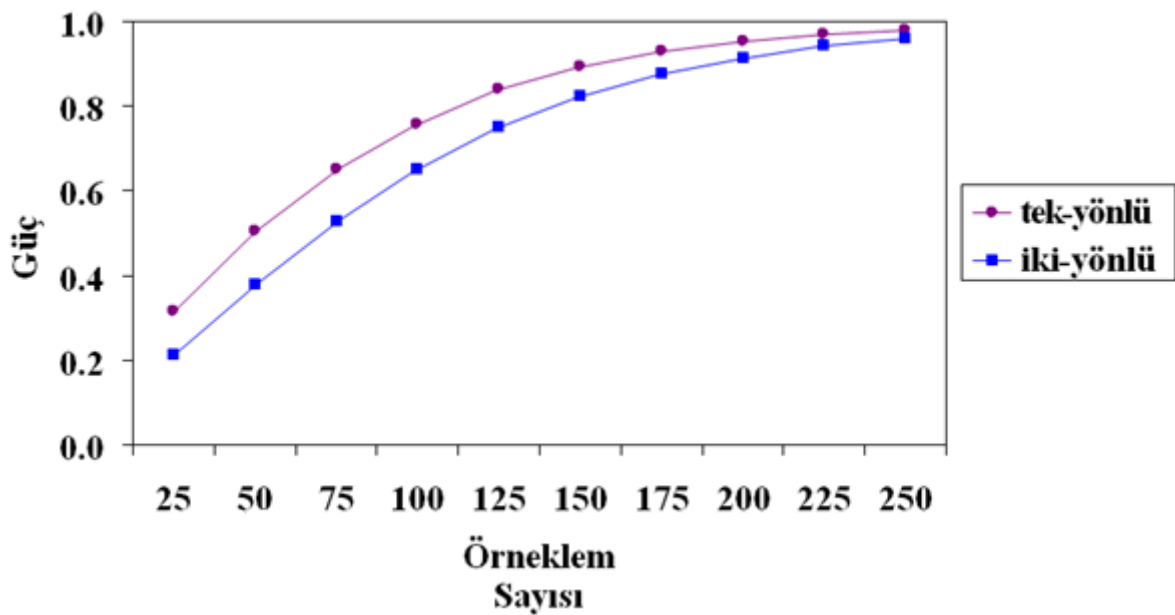
Örneklem Sayısı	<i>r</i>		<i>d</i>	
	Tek yönlü α	İki Yönlü α	Tek Yönlü α	İki Yönlü α
10	.705	.761	1.725	2.024
20	.526	.579	1.156	1.325
30	.437	.485	.931	1.060
40	.382	.426	.801	.909
50	.344	.384	.713	.809
60	.315	.352	.650	.736
70	.292	.327	.600	.679
80	.274	.307	.561	.634
90	.259	.290	.528	.597
100	.246	.276	.501	.566
110	.235	.263	.477	.539
120	.225	.252	.457	.516
130	.216	.243	.438	.495
140	.208	.234	.422	.477
150	.201	.226	.408	.460
160	.195	.219	.395	.446
170	.189	.213	.383	.432
180	.184	.207	.372	.420
190	.179	.202	.362	.409
200	.175	.197	.353	.398
210	.171	.192	.344	.388
220	.167	.187	.336	.379
230	.163	.183	.329	.371
240	.160	.180	.322	.363
250	.157	.176	.315	.356
260	.154	.173	.309	.349
270	.151	.169	.303	.342
280	.148	.166	.298	.336
290	.145	.164	.293	.330
300	.143	.161	.288	.325

Not: Güç = %80, alfa: 0.05'tir.

1.8 İstatistiksel Gücü Dolaylı Olarak Etkileyen Faktörler

1.8.1 Alternatif Hipotezin Tek yönlü veya Çift Yönlü Olması

Alternatif hipotezin tek yönlü veya iki yönlü olmasının istatistiksel güç üzerinde etkisi vardır. Tek yönlü (directional, one-tailed) hipotez testleri, çift yönlü (non-directional, two-tailed) hipotez testlerinden daha güçlüdür. Hipotezin yönlülüğünün istatistiksel güç üzerinde olan etkisi Şekil 1.6'da gösterilmiştir.



Şekil 1.6 Alternatif Hipotezi Tek Yönlü veya Çift Yönlü Olmasının İstatistiksel Güç Üzerindeki Etkisi

1.8.2 Kullanılan İstatistiksel Testin Parametrik veya Non-Parametrik Olması

Parametrik testler, parametrik olmayan testlere göre daha güçlüdür.

1.8.3 Örnekleme Hatası

Büyük örneklemeler daha az örnekleme hatası içerir ve istatistiksel güç daha fazladır.

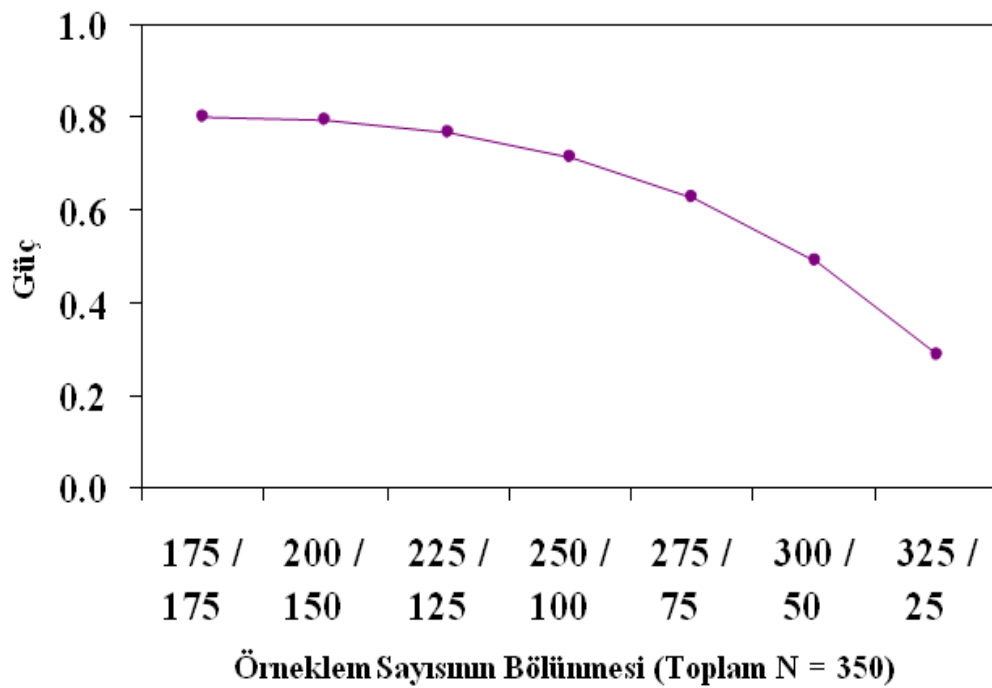
1.8.4 Ölçüm Hatası

İstatistiksel gücü dolaylı olarak etkileyen faktörlerden biri de ölçüm hatasıdır. Güvenilir olmayan ölçümler, veriler üzerinde dalgalanmalar oluşturarak belirlenmek istenen etkinin ortaya çıkarılmasını zorlaştırır. Araştırmacının, ölçüm hatasından kaynaklanan bir güç azalmasını (tahmin edilen etki ortaya doğru bir şekilde çıkarılmak isteniyorsa) diğer faktörleri

kullanarak telafi etmesi gerekir. Ölçüm hatasının tahmin edilen etkiyi ortaya çıkarmak için gerekli olan istatistiksel güç seviyesi üzerinde büyük bir etkisi vardır. Örneğin, gerçek bir $r = 0.20$ etkisini belirlemek için $\alpha = 0.05$ ve güç = %80 seviyeleri için gerekli olan örneklem sayısı 193'tür. Fakat bu $r = 0.20$ seviyesinin, güvenilir olmayan ölçümler nedeniyle $r = 0.14$ gibi bir seviyeye düştüğü varsayılırsa, bu durumda gerekli olan örneklem sayısı $\alpha = 0.05$ ve güç = %80 seviyeleri için 398 olacaktır (Ellis, 2010, s. 66).

1.8.5 Deney ve Kontrol Gruplarının Örneklem Sayılarının Eşit Olması

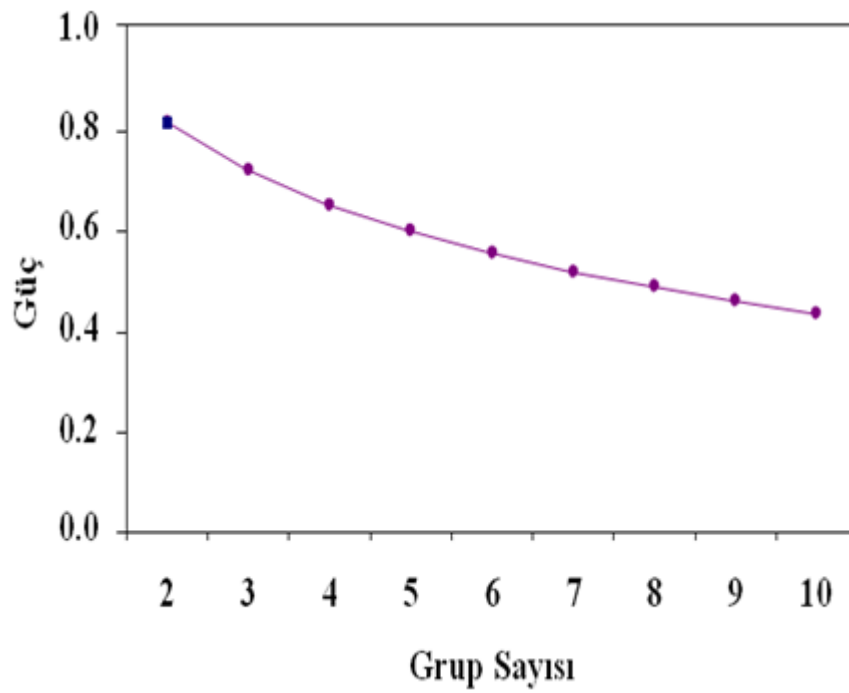
Deney ve kontrol grupları karşılaştırılırken, her bir grubun örneklem sayıları mümkün olduğunca eşit veya birbirine yakın seviyelerde seçilmelidir. Eşit olarak bölünmüş örneklem sayıları ile gerçekleştirilen çalışmalarda istatistiksel güç daha fazladır. Bu durum ilgili Şekil 1.7'de gösterilmiştir.



Şekil 1.7 Deney ve Kontrol Gruplarının Örneklem Sayılarının Eşit Olup Olmamasının İstatistiksel Güç Üzerindeki Etkisi

1.8.6 Varyans Analizinde Kullanılan Grup Sayısı

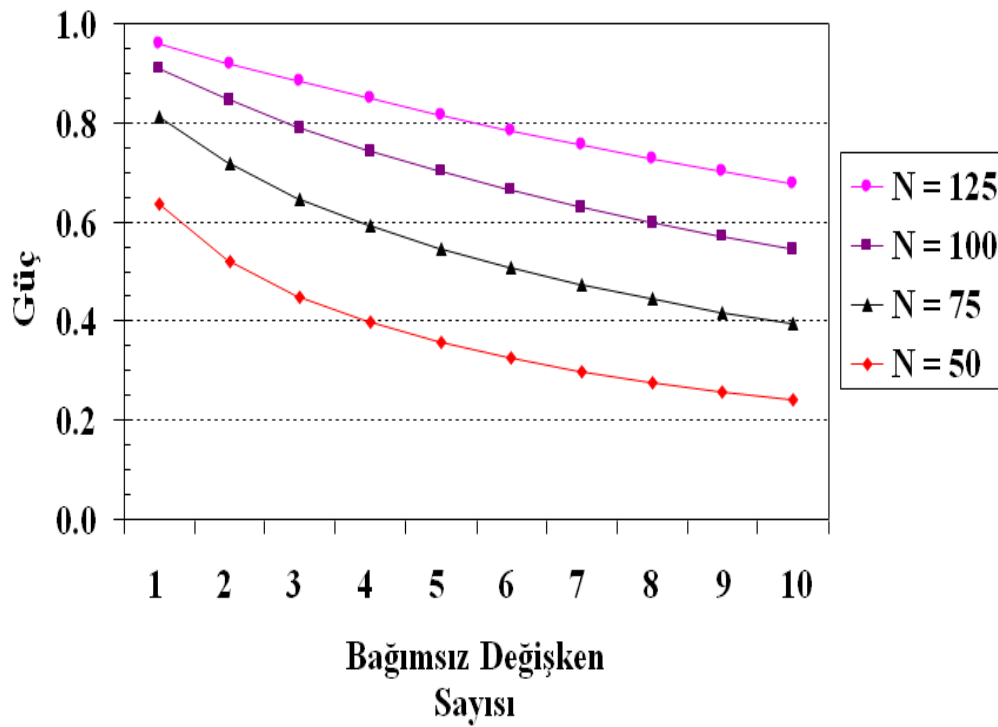
Varyans analizi gerçekleştirilirken kullanılan grup sayısı ile istatistiksel güç arasında ters orantılı bir ilişki vardır. Şekil 1.8'den, grup sayısı arttıkça sabit bir örneklem için güç seviyesinin azaldığı görülmektedir.



Şekil 1.8 Varyans Analizinde Kullanılan Grup Sayısının İstatistiksel Güç Üzerindeki Etkisi

1.8.7 Çoklu Regresyon Analizinde Kullanılan Bağımsız Değişken Sayısı

Çoklu regresyon analizinde kullanılan bağımsız değişken sayısı (predictor) ile istatistiksel güç arasında ters orantılı bir ilişki vardır. Bağımsız değişken sayısı arttıkça hemen her örneklem grubunda güç seviyesi azalmaktadır. Bu durum Şekil 1.9'da gösterilmektedir.



Şekil 1.9 Çoklu Regresyon Analizinde Kullanılan Bağımsız Değişken Sayısının İstatistiksel Güç Üzerine Etkisi

1.9 İstatistiksel Gücün Hesaplanmasına İlişkin Örnekler

Burada basit bir örnek ile güç üzerinde etkisi olan 3 temel faktör kullanılarak istatistiksel güç hesaplaması yapılacaktır. Örnek olarak ise bir tedavinin, sigara içme oranını azaltma yönündeki etkisi araştırılacaktır. Önceki çalışmalardan günlük içilen sigara sayısının 15 adet standart sapma ile 30 adet olduğu varsayalım. Kullanılacak bir tedavi yönteminin ise günlük içilen sigara sayısını 5 adet düşüreceği varsayılmaktadır. *Cohen's d* etki büyüklüğü formülü kullanılarak beklenen etki büyüklüğü $d = (30 - 25) / 15 = 0.333$ olarak bulunacaktır. Araştırmacı, %80 güç ve iki yönlü alfa = 0.05 anlamlılık seviyesinde G-Power 3.1.3 yazılımını kullanarak (for a simple t test of the between the means of two independent groups) gerekli örneklem sayısını her bir grup için 143 olmak üzere toplamda 286 olarak belirleyecektir. Yapılan bu işlemin adı *priori güç analizi (priori power analysis)* veya *ileriye yönelik güç analizi (prospective power analysis)*'dir. İlâveten, gerekli olan örneklem sayısı Guenter (1975,1981) tarafından verilen aşağıdaki basit denklem ile de hesaplanabilmektedir:

$$n = 2 \left(\frac{z_{\alpha} + z_{1-\beta}}{d} \right)^2 + 0.25z_{\alpha}^2$$

n, grup başına örneklem sayısını; z_{α} , alfa için eşdeğer z skorunu; $z_{1-\beta}$, güç için eşdeğer z skorunu göstermektedir. Tablo 1.9’da bazı faydalı z skorları, alfa ve güç için verilmiştir.

Tablo 1.9 Alfa ve İstatistiksel Güç İçin Bazı Faydalı Eşdeğer z Skorları

Alfa	Tek-yönlü z_{α}	Çift-yönlü z_{α}
.01	2.327	2.575
.05	1.645	1.960
.10	1.282	1.645
Güç	$z_{1-\beta}$	
.95	1.645	
.90	1.282	
.80	0.842	
.75	0.675	
.70	0.524	
.60	0.253	
.50	0.000	

Şimdi, denklemde geçen z skorlarının değerleri yukarıdaki Tablo 1.9’dan bulunup yerine koyulursa ($z_{\alpha} = 1.96$, $z_{1-\beta} = 0.842$) her bir grup için gerekli olan örneklem sayısı $n = 142,6$ olarak bulunacaktır. Görüldüğü gibi Guenter’in verdiği çok da fazla bilinmeyen denklem ile G-Power 3.1.3 programının verdiği örneklem sayısı aynen elde edilmiştir. Guenter ayrıca tek örneklem t testi için de aşağıdaki denklemi vermiştir:

$$z_{1-\beta} = d\sqrt{n} - z_{\alpha}$$

Bu denklemde elde edilecek negatif z skorları istatistiksel gücün %50'den aşağıda olduğunu göstermektedir. Şimdi aynı koşullar için araştırmacının mevcut kaynağının toplamda sadece 200 adet örnekleme yettiği varsayımı altında $z_{1-\beta} = 0,395$ ve buna karşılık gelen güç değeri 0.654 olacaktır. Eğer aynı koşullar için araştırmacı tek yönlü hipotez sınaması yaparsa elde edeceği $z_{1-\beta} = 0.710$ ve buna karşılık gelen güç değeri 0.761 olacaktır. Aynı işlemler G-Power 3.1.3 yazılımı ile gerçekleştirilecek olursa elde edilen değerler çok az bir farklılık ile sırasıyla tek yönlü ve çift yönlü hipotez testleri için 0.649 ve 0.759 olacaktır (Rossi, 2012, s. 31-33).

İstatistiksel güç, yanlış bir H_0 hipotezinin doğru bir şekilde reddedilme olasılığı olarak tanımlanmıştır. İstatistiksel gücün, bir çalışma sonucunda elde edilecek anlamlı veya anlamlı olmayan sonuçlara olan etkisini görmek adına Abraham ve Russell (2008), yaklaşık 1000 adet tekrardan oluşan simülasyon ile bir çalışma gerçekleştirmişlerdir. Çalışmada, her bir denemede çekilen deney ve kontrol gruplarının örneklem sayıları eşit ve 40 olarak alınmıştır. Ayrıca örneklemelerin çekildiği anakütle üzerindeki etki büyüklüğü ortalaması $d = 0.50$ 'dir. Bu şartlar altında güç seviyesi G-Power 3.1.3 programı kullanılarak 0.60 olarak hesaplanmıştır. 1000 adet denemeden oluşan simülasyon gerçekleştirilmiş ve aralarından rassal olarak 10 deneme seçilmiş ve incelenmiştir. Yapılan incelemeler sonucunda elde edilen sonuçlar Tablo 1.10'da verilmiştir:

Tablo 1.10 Simülasyon Sonuçları

Örneklem	<i>M</i> Kontrol Grubu	<i>M</i> Araştırma Grubu	<i>M</i> Fark	<i>t</i>	<i>P</i>
754	0.28	0.18	-0.10	-0.43	0.672
023	0.23	0.34	0.11	0.54	0.589
167	-0.01	0.21	0.22	0.95	0.346
838	-0.05	0.42	0.47	1.91	0.060
660	0.19	0.69	0.50	2.34	0.022
947	-0.05	0.51	0.56	2.76	0.007
231	-0.13	0.44	0.57	2.52	0.014
108	-0.15	0.44	0.59	2.61	0.011
009	-0.19	0.49	0.69	3.57	0.001
358	-0.25	0.75	1.00	4.52	<0.001
Ortalama	-0.10	0.55	0.65		

Örneklem olarak belirtilen ilk sütun 1000 adet deneme içinden seçilen 10 adet örneklemin simülasyon içindeki numaralarını ifade etmektedir. ikinci, üçüncü ve dördüncü sütunlardaki M_{Kontrol} , $M_{\text{Araştırma}}$ ve M_{Fark} ise sırasıyla kontrol grubun ortalamasını, araştırma grubunun ortalamasını ve her iki grup arasındaki farklılıkları ifade etmektedir. Tabloda koyu harfler ile yazılan sonuçlar $p < 0.05$ olduğu için anlamlı olan sonuçlardır. Çalışma öncesinde hesaplanan güç değeri 0.60 olduğu için seçilen bu 10 denemeden 6'sının anlamlı sonuçlar verip 4'ünün anlamlı olmayan sonuçlar vermesi yani %60 oranında anlamlı sonuç elde edilmesi kesinlikle tesadüf değildir. Bu durum çalışmanın tasarım aşamasında istatistiksel gücün dikkate alınmasını gerektiren önemli bir örnektir (Abraham ve Russel, 2008, s. 286-287).

1.10 İstatistiksel Güç Analizi Çeşitleri

İstatistiksel güç analizleri genellikle teorik (*prospective power analysis*) ve deneysel (*retrospective power analysis*) olmak üzere iki çeşittir. Bazı kaynaklarda karşılaştırmalı (compromise) güç analizinden de bahsedilmektedir. Prospective güç analizinde istatistiksel güç, ilgili hipotezleri değerlendirmek için yapılacak anlamlılık testlerinden önce hesaplanırken retrospective güç analizinde ise anlamlılık testlerinden sonra hesaplanmaktadır (Cafri vd., 2009, s. 38).

1.10.1 Teorik Güç Analizi (Priori Power Analysis - Prospective Power Analysis)

Birçok kişi tarafından teorik güç analizi ideal güç analizi tipi olarak kabul edilmektedir. Bir teorik güç analizinde araştırmacı önce etki büyüklüğünü, alfa seviyesini ve arzu ettiği güç seviyesini belirler. Belirlenen bu kavramlar ışığında gerekli optimum örneklem sayısına ulaşmak mümkündür. Standart uygulamalar ile etki büyüklüğünün ve alfa seviyesinin seçimi genellikle geleneksel kalıplara dayanmaktadır. Uzunca bir süredir alfa seviyesi ya 0.05 ya da 0.01 olarak seçilmektedir. Etki büyüklüğü ise genellikle Cohen tarafından 1962 yılında tanımlanan küçük, orta ve büyük seviyelerdeki etkilerden birisi olarak seçilmektedir. Tüm bunlara karşın beta seviyesi ile ilgili herhangi bir standartlaşan kalıp yoktur. Fakat Cohen, II. Tip hata oranı olan beta seviyesi ile alakalı 0.20 değerini önermiştir. Buna karşılık bazı yazarlar ise alfa ve beta değerlerini eşit olarak almayı tercih etmektedirler (Erdfelder vd., 1996, s. 2).

Güç analizleri genellikle çalışmanın tasarım aşamasında işe yaramaktadır. Teorik güç analizi, temel olan etki büyüklüğü, örneklem sayısı, güç ve alfa seviyesi gibi parametrelerden örneklem sayısına karar vermek için kullanılır. Eğer beklenen etki büyüklüğü gerçek etkiden büyük ise veya yapılan ölçümler güvenilir değil ise araştırmacının yeterli güce ulaşmak için hesapladığı örneklemden daha fazlasına ihtiyacı olacaktır. Teorik güç analizi aynı zamanda planlanan bir çalışma için II. Tip hata yapma olasılığının belirlenmesi için kullanılmaktadır. Bu işlem sayesinde yeterli güç sağlamayan çalışmalar, araştırmacının başında terk edilebilmekte ve yeni bir tasarım yapılabilir (Lewis, 2006, s. 607).

Teorik güç analizleri özellikle tekrar çalışmaların planlama aşamalarında faydalıdır. Daha önce yapılmış benzer çalışmalardan elde edilecek ortalama etki büyüklüğü ve gerekli örneklem sayısı ile araştırmacı çalışmasına başlar. Örneğin, bir araştırmacı X ve Y arasındaki ilişkiyi araştırmak istesin. Yazında yer alan iki çalışma daha önceden bu ilişkiyi araştırmış ve sırasıyla $r_1 = 0.20$ ve $r_2 = 0.24$ seviyelerinde anlamlı olmayan korelasyonlar bulmuş olsunlar. Araştırmacı, anlamlı olmayan sonuçların yetersiz istatistiksel güçten kaynaklanmış olma ihtimalinden şüphelensin. Sırasıyla X ve Y araştırmalarının örneklem sayıları $n_1 = 78$ ve $n_2 = 63$ olarak alınmış olsun. Hipotez testlerini yeniden yapmadan önce araştırmacı güç tablolarına danışmış ve %80 seviyesindeki istatistiksel gücü elde etmek ve $r_{ortalama} = 0.22$ seviyesindeki etkiyi belirlemek için iki yönlü alfa = 0.05 seviyesinde gerekli olan minimum örneklem sayısının 159 olduğuna karar vermiştir. Araştırmacı yapmış olduğu bu teorik güç analizi ile gerekli olan minimum örneklem sayısını belirlemiş ve bu daha önceden yapılmış çalışmalarda yeterli istatistiksel güç olmadığı için anlamlı sonuçlar edilmediği şüphesinin doğruluğunu

desteklemiştir. Bu noktada arařtırmacı, yazına katkıda bulunmak için iki noktada pozisyonunu almıřtır. Birincisi, daha önce yapılmıř alıřmalarda ortaya ıkarılamayan iliřkiyi eęer gerekten varsa ortaya ıkarabilmek için alıřmasını yeterli istatistiksel g ile tasarlayacaktır. İkincisi, eęer arařtırmacı bekledięi etkiye yakın bir etki elde ederse, önceki alıřmalarda anlamsız olduęu dřnlp, dřk g nedeniyle II. Tip hatanın yapılmıř olma olasılıęının yksek olduęu, anlamlı bir řekilde sonulandırılmayan alıřmaları yeniden yorumlayabilecektir. Belki de bu arařtırmacı, elde edeceęi olumlu sonular ile dięerlerinin gz ardı etmiř olduęu iliřkinin ortaya ıkmıř olmasını saęlayacaktır (Ellis, 2010, s. 57-58).

Teorik g analizleri, etki byklęünün herhangi bir seviyesi için dřk alfa ve beta olasılıklarının kullanılması aısından idealdir. Fakat, genellikle teorik g analizleri ile belirlenmiř olan optimum rneklem sayısı, sosyal bilimlerde zaman ve maliyet kısıtları, veri toplama anında rneklemenin homojen olmaması vs. gibi nedenlerle uygulanandan daha fazladır. Yani alıřmalarda kullanılan rneklem sayısı, bazı sebepler nedeniyle optimum rneklem sayısından daha azdır. Bu nedenlerle teorik g analizinden elde edilen optimum rneklem sayısı genellikle kullanılmamakta, daha az rneklem ile alıřmalar gerekleřtirilmektedir. Benzer bir durumda bir sosyal bilim arařtırmasında, arařtırmacının kısıtların elverdięi lde kullandıęı maksimum rneklem sayısı “N” olsun. Bu rneklem sayısı muhtemelen yapılmıř bir teorik g analizinden elde edilen optimum rneklem sayısından daha dřk olacaktır. Burada arařtırmacı için alıřmasında makul bir g seviyesini yakalamanın tek yolu alfa seviyesini artırmak olacaktır. Ancak 1962 yılında Cohen tarafından verilen g tabloları sadece geleneksel olarak kullanılan alfa seviyelerine gre oluřturulduęundan, istenilen g seviyesi için alfa seviyesinin ne olacaęı sorusu üzerinde dřnlmesi gereken bir konudur. İřte tam bu noktada son yıllarda olduka ileri seviyelerde geliřtirilen g analizi yazılım programları devreye girerek arařtırmacılara standart olmayan seviyelerde alfa kullanma imkânı vermektedirler. rneklem sayısının sabit olduęu durumlarda teorik g analizi yapmak pek faydalı bir iř olmayacaktır. nk analiz sonucunda optimum rneklem sayısı konusunda ne deęer elde edilirse edilsin alıřılacak rneklem sayısı bellidir. Burada deęiřik alfa seviyeleri ile yeterli gce ulařmak mmkndr (Erdfelder vd., 1996, s. 2).

1.10.2 Deneysel Güç Analizi (Post-Hoc Power Analysis-Retrospective Power Analysis)

Güç analizleri genellikle arařtırmaların tasarım ařamalarında faydalıdır. Fakat bazen özellikle istatistiksel olarak anlamlı sonuçlar elde edilemeyen arařtırmalarda, çalıřma tamamlandıktan sonra bir güç analizi gerekleřtirilmesi gerekmektedir. Arařtırma öncesinde yapılacak bir etki büyüklüğü tahminine dayanan teorik güç analizinden farklı olarak deneysel güç analizi, arařtırma sonucunda gözlemlenen etkiye ve gözlemlenen istatistiksel güce dayanmaktadır. Bu tür bir güç analizini yapmanın avantajı, elde edilecek anlamlı olmayan sonuçların dikkate değer olmaları ihtimalinin bulunmasıdır. Çünkü anlamlı olmayan sonuçlar yetersiz güçten veya yetersiz örneklem sayısından kaynaklanabilmektedir. Var olan küçük bir etkiyi belirleyememe nedeniyle anlamsız olarak sonuçlandırılan bir çalıřma, yeterli örneklem sayısı kullanılarak anlamlı olarak bulunabilir (Balkin ve Sheperis, s. 271).

Bir çalıřmada, istatistiksel olarak anlamlı olmayan sonuçlar elde edildiğinde arařtırmacı çalıřmasını yeterli bir istatistiksel güç seviyesinde kurup kurmadığından řüphe etmeli ve “aradıđım etkiyi bulmak için çalıřmam yeterli istatistiksel güce sahip mi? Acaba ne kadar örneklem kullanmalıydım?” gibi soruları kendisine sormalıdır. Bazen bu tür sorular arařtırmacılara dergi editörleri tarafından da sorulmaktadır. Hoenig ve Heisey (2001)’ e göre 19 adet dergi arařtırmacılardan post-hoc güç analizini yapmalarını istemektedir (s. 2). Bunun gerekçesi olarak ise anlamlı olmayan sonuçların II. Tip hata teşkil edecek şekilde düşük istatistiksel güç ile gerekleřtirilmiş olabileceğı gösterilmektedir. Yapılan anlamlılık testi sıfır hipotezini reddetmeye izin vermese de gözden kaçırılan bir gerek etki var olabilir.

Anlamlı olmayan sonuçlar bir arařtırmacı için sorundur ve teorik güç analizi uygulanması her zaman sonuçların beklenildiğı gibi geleceğı garantisini vermez. Teorik güç analizi doğru bir etki büyüklüğü tahminine dayanmaktadır. Eğer var olan gerek etki, arařtırmanın tasarım ařamasında tahmin edilenden düşük ise sahip olunan mevcut güç seviyesinden daha fazlasına ihtiyaç olacaktır. Çalıřma sonucunda elde edilecek etkiye göre deneysel güç analizi yapmak, tahmin edilen güç ile elde edilen güç arasında bir kıyaslama yapma olanağı sağlar. Eğer arařtırmanın tasarım ařamasında tahmin edilen güç düşük çıkarsa, arařtırmacı çalıřmasını řu şekilde sonuçlandırabilir: “Elde edilen sonuçlar anlamlı değıl fakat bu, testlerin yeterince güce sahip olmamasından kaynaklanabilir. Dolayısıyla arařtırmanın tekrarlanmaya ihtiyacı vardır.” Eğer elde edilen gücün yeterli seviyede olduğı kanısına varılırsa “sonuçlar kesinlikle negatiftir” denilebilir (Ellis, 2010, s. 58-59).

Güç analizi, bazen anlamlı olmayan sonuçlar elde edildiğinde yapılmış olan testlerin gücünü ölçmek için kullanılır. Bu şekilde gerçekleştirilen güç analizlerinin amacı yanlış olan bir sıfır hipotezinin yanlışlıkla da olsa doğru olarak açıklanmasını önlemektir. Böylece çalışma sonunda elde edilen sonuçlar yüksek güç sayesinde “etki yoktur” şeklinde etiketlenebilir. Buna karşılık, eğer çalışma sonuçlarında istatistiksel olarak anlamlı sonuçlar elde edildiyse tekrardan güç analizi uygulamanın bir anlamı yoktur. Çünkü bu durumda zaten doğru, anlamlı, gerçek bir etki bulunmuşken, sonuçların anlamlı bir sonuç verme olasılığını tekrardan hesaplamak mantıksızdır. Deneysel güç analizleri sadece istatistiksel olarak anlamlı sonuçlar üretmeyen çalışmalara mahsustur. Ancak yine de bazı araştırmacılar, istatistiksel olarak anlamlı sonuçlara örneklem rastlantısı olarak ulaşılmış olabileceği ihtimaline vurgu yaparak, istatistiksel olarak anlamlı sonuçlar elde edilse bile özellikle küçük sayıdaki örneklem ile gerçekleştirilen çalışmalarda deneysel güç analizinin uygulanması gerektiğini belirtmişlerdir (Ertürk, 2005, s. 743).

Deneysel güç analizinde araştırmacı testin gücünü hesaplamak için alfa, etki büyüklüğü ve örneklem sayısını belirlemektedir. Daha sonra belirlenen bu değerler ışığında testin gücüne ulaşmak mümkündür. Ancak alfa seviyesinin belirlenme durumu biraz karışıktır. Araştırmacı için kullandığı örneklem sayısı açısından hangi alfa seviyesinin daha uygun olduğunu kestirmek adına bir ipucu yoktur. Bu nedenle araştırmacılar olası güç problemleri ile karşılaşmamak için genellikle standart alfa seviyelerini kullanmaktadırlar.

1.10.3 Karşılaştırmalı - Uzlaşmacı Güç Analizi (Compromise Power Analysis)

Karşılaştırmalı güç analizi, standart seviyeler dışında α seviyesi belirlemek için ideal bir yöntemdir. Burada araştırmacı bulmak istediği etki büyüklüğünü, maksimum olasılıkla elde edeceği örneklem sayısını ve β/α oranını belirlemektedir. Belirlenen bu parametreler ışığında geliştirilen güç analizi yazılımları sayesinde standart seviyeler dışında ihtiyaç olan farklı seviyelerde α ve β değerleri hesaplamak mümkündür (Erdfelder vd., 1996, s. 2).

Bazı kaynaklarda güç analizinin üçüncü bir işlevinden bahsedilmektedir. Özellikle klinik araştırmalarda bazı nedenlerden dolayı geleneksel standart α ve β düzeylerini sağlamak bakımından ele alınan etki büyüklüğünü sağlayacak örneklem sayısı çok küçük olabilir. Ayrıca toplanan verilere uyacak modeli belirlerken örneklem sayısı çok büyük ise önemsenmeyecek farklar bile sıfır hipotezinin reddine yol açabilir. Bu gibi durumlarda $q = \beta/\alpha$ oranına uyan optimum bir test istatistiği belirlemek arzu edilmektedir. Bu optimum kritik test istatistiği, ele alınan sabit bir örnek hacminde “düşük α riski, yüksek güç oranı”

isteği arasındaki uzlaşmayı sağlayacak rasyonel uzlaşma değeri olarak ele alınmaktadır. Bu tür problemleri çözümlenmek amacıyla yararlanılan güç analizine *uzlaşmacı güç analizi* adı verilmektedir. Bu yaklaşım G-Power isimli bilgisayar programı tarafından yapılabilmektedir. Ancak bu yaklaşımla ulaşılan çözümler geleneksel istatistiksel çıkarımla sonuçları ile uyumluluk göstermemektedir. Bu nedenle yaygın kullanımı olan bir yöntem değildir (Özdamar, 2003, s.110).

1.11 Yetersiz ve Aşırı İstatistiksel Gücün Neden Olduğu Sorunlar

McKean'in (1990) aktardığına göre, Norton ve Luftig (1980), istatistiksel güç konusunda yazında yaptıkları incelemeler sonucunda, araştırmacıların güç seviyelerini çok fazla şişirdikleri sonucuna varmışlardır. 1971 yılında yaptıkları bir araştırmada hesapladıkları gücün, gerçek gücün yaklaşık olarak iki katı olduğu sonucuna ulaşmışlardır. Araştırmacılar yeterli güce ulaşamadıkları için elde ettikleri gücü şişirerek olması gereken makul seviyeye çekmişlerdir. Bu şekilde gerçekleştirilen bir çalışmadan elde edilecek sonuçların önemliliği ise tartışılacak bir konudur (s. 12).

Güç hesaplama aşaması, istatistiksel çalışmanın tasarım ve uygulama aşamasında diğer kararları etkileyen bir konu olması nedeniyle öncelik ve önem verilmesi gereken bir aşamadır. Yeterli güçten yoksun olan bir çalışmada H_0 hipotezinin reddedilip reddedilmeyeceğine karar verilmesi aşamasında hata yapılma ihtimali oldukça yüksektir. Bir araştırmanın güçten yoksun olması II. Tip hata olasılığının yüksek olması anlamına gelmektedir. Başka bir deyişle, araştırmacı çalışması için yeterli gücü sağlayamadığı durumlarda yanlış olan bir H_0 hipotezini reddetme konusunda oldukça zorluk çekecektir. Böyle durumlarda çok değerli sonuçlar sağlayacak olan bazı araştırmalar yetersiz güç nedeniyle boşa gitmektedir. Burada sadece araştırmanın boşa gitmesi söz konusu değil, aynı zamanda harcanan emek, çaba, zaman ve mali olanakların tümünün boşa gitmesi söz konusudur. Tüm bunlara sebebiyet vermemek için araştırmacı önceden bir pilot çalışma yapmalıdır. Fakat pilot çalışmalar bazen verilecek kararlarda araştırmacıyı yanıltabilir. Çünkü pilot çalışma yaparken düşük sayıda örneklem ile çalışıldığından, olası bir düşük istatistiksel güç nedeniyle buradan elde edilecek sonuçlar araştırmacıyı yanıltabilir. Bu nedenle pilot çalışmalar denenerak yapılmalı ve yapılan tüm pilot çalışmalar sonucunda uygun kararlar verilmelidir (DiLullo,1997, s. 13).

Fagley (1985), istatistiksel güç analizinin kullanımını anlamlı olmayan sonuçların yorumlanması açısından açıklamıştır. Fagley, anlamlı olmayan sonuçların doğru bir şekilde yorumlanabilmesi için yapılan testin yeterli bir güce sahip olması gerektiğini belirtmiştir.

Kişiler, anlamlı olmayan sonuçları değerlendirirken testin gücünü dikkate almalıdırlar. Yapılan bir araştırmanın sonuçları sunulurken yorumlamayı kolaylaştırmak açısından testlerin beklenen güçleri de ilave edilmelidir. Yeterli bir seviyede güç ile çalışma yapılmadıysa o zaman anlamlı olmayan sonuçların doğru bir şekilde yorumlanması çok zordur (Fagley, 1985, s. 391; O’Keefe, 2007, s. 296).

Bir çok araştırmada anlamlılık seviyesi, sıfır hipotezinin reddedilmesi için tek kriter olarak düşünülmektedir. Buna rağmen birçok araştırmada güç analizleri dikkate alınmadan, sıfır hipotezi reddedilmediğinde doğru olarak sonuçlandırılmaktadır. Ancak yeterli güç ile çalışılmadıysa sıfır hipotezinin reddedilemeye nedeni ilişkinin veya farklılığın olmaması değil gücün düşük olması olabilmektedir. Bu nedenle güç konusu, araştırma sonuçlandırılmadan önce kesinlikle göz ardı edilmemelidir (Coblick, 1998, s. 93).

Yeterli istatistiksel güç ile gerçekleştirilen çalışmalar araştırmacılara başarılı olmak için yeterli şansı sağlamaktadır. Düşük güç ile yapılan çalışmalarda şu örnek olaydaki gibi bir sorun vardır:

Ton balığı avlamak için kullanılan ağ ile yem için kullanılan küçük balıkları avlamaya çalışan bir kişi, yanlış tercih yaptığı için muhtemelen hiç balık tutamayacaktır. Fakat bu kişi başarısız olduktan sonra “*bu gölde hiç balık yok*” şeklinde işini sonlandırırsa hata yapmış olur. Çünkü ton balığı yakalamak için kullanılan ağın içinden gölde varsa eğer küçük balıklar kayıp gidecektir. İşte yetersiz güç ile gerçekleştirilen çalışmalar da buna benzemektedir. Eğer bir çalışmadan olumsuz bir sonuç alındıysa bunun nedeni gerçekten bir ilişki veya farklılık olmaması değil yetersiz güç olabilir (Clark, 1996, s. 4).

Düşük gücün getirdiği tehlikeler ile ilgili farkındalık son zamanlarda az da olsa artma eğilimindedir. American Psychological Association (APA)’da bir komisyon, çalışmalar için veri toplamadan önce çalışmanın gücünün değerlendirilmesi gerektiğini tavsiye etmiştir. Çalışmalara maddi olarak fon sağlayan kuruluşlar ve üniversiteler yapılan bir çalışmayı değerlendirmeye almak için çalışmanın güç analizi uygulama sonuçlarını da araştırma teklifleri ile birlikte istemektedirler. II. Tip hata yapma olasılığını dikkate alan birçok dergi editörü artık çalışmaların güç analizini içermesini istemektedirler. Bu şekilde bir önem arz etmesine rağmen güç analizi konusuna verilen önem hala çok azdır. Ekonomiden eğitim bilimleri alanına kadar birçok alanda güç analizinden bahsedilme oranı ne yazık ki sadece %0 - %4 arasında değişmektedir (Ellis, 2010, s. xv-xvi).

Carter (1997), gelişmeler olmasına rağmen araştırmacıların istatistiksel testlerin gücü konusunu çok fazla dikkate almadıklarını ve bununla birlikte araştırma hipotezlerinin reddi konusunda büyük bir risk aldıklarını belirtmiştir (s. 71).

Yazında belirli bir konu üzerindeki bazı araştırmalarda, değişkenler arasında anlamlı veya ters bir ilişkinin varlığından bahsedilirken, bazılarının da ise herhangi bir ilişkinin olmadığı gibi sonuçlara ulaşılmaktadır. İstatistiksel güç analizinin kullanılma nedenlerinden biri mevcut araştırmalardaki tutarsızlıkları değerlendirmektir. Yapılacak bir post-hoc güç analizi ile derlenen bu çalışmalar üzerinde değişkenler arasındaki bu tutarsızlıkların cevabı açıklanabilir. Acaba bu çalışmalardaki tutarsızlıkların nedeni gerçekten istatistiksel güç eksikliğinden mi kaynaklanmaktadır? Eğer öyleyse, durum bu çalışmalara bir priori güç analizi uygulanmasını gerektirmektedir. Bu sayede gerekli etki büyüklüğüne ve yeterli örneklem sayısına ulaşılarak sağlıklı sonuçlar elde edilebilir. Ayrıca yetersiz güç kontrolünün yanında araştırmalardaki istatistiksel tasarım ve yöntemler de değerlendirilmelidir.

Yetersiz gücün neden olduğu bir başka sorun ise anlamlı olmayan sonuçların doğru yorumlanmasını engellemesidir. Chase ve Tucker (1976), priori güç analizi sonucunda elde edilen güç oranı %50 den daha az ise daha ayrıntılı bir araştırma yapmak gerektiğini belirtmişlerdir. Eğer bir araştırma sonucunda anlamlı olmayan sonuçlara ulaşıldıysa, araştırmacı elde ettiği sonuçların geçerliliğini, düşük güç nedeniyle meydana gelip gelmediğini post-hoc istatistiksel güç analizi uygulayarak test etmelidir (Onwuegbuzie ve Leech, 2004, s. 219).

Düşük güçlü ve aşırı yüksek güçlü çalışmalar verimsizdir. Düşük güçlü çalışmalarda yeterli istatistiksel güç sağlanamadığı için sıfır hipotezini reddetmek için sarf edilen kaynaklar israf edilmiş olur. Bu şekilde anlamlı olmayan sonuçlar ortaya çıktığında çalışma yanlış bir şekilde “etki yoktur” şeklinde sonuçlandırılacaktır. Ayrıca düşük güçlü çalışmalar o alanda gelecekte yapılması planlanan araştırmaları da olumsuz yönde etkilemektedir. Aynı şekilde aşırı güçlü çalışmalar da israfa ve yanlış yönlendirmelere neden olabilmektedir. Örneğin, 1000 gözlemden daha fazla örnekleme sahip olan çalışmalar çok küçük etkileri bile ($r < 0.10$ ve $d < 0.20$ gibi) tespit etme şansına sahiptir. Binlerce veri kullanılarak gerçekleştirilen çalışmalar bu olasılığı artıracak ve normalde anlamsız olan bir etkinin istatistiksel olarak anlamlı olmasına yol açacaktır. Elbette etki büyüklüğünü yorumlama konusunda bilgi sahibi olan araştırmacılar doğrudan “ p ” değerinden elde edilecek bir “istatistiksel anlamlılık” tuzağına düşmeyeceklerdir. Aşırı güçlü çalışmaların savurganlığı altındaki gerçek, çalışmalar için toplanması gereken veri miktarının ne olacağı konusunda bir bilgi sahibi olunmamasıdır.

Çoğu zaman kullanılan mantık olan “daha fazla örneklem daha iyidir”, kesinlikle yanlış olan bir mantıktır (Ellis, 2010, s. 52).

Yetersiz güç, birçok yayımlanmamış olan çalışmaların yayımlanmama nedenlerinden biri olabilmektedir. Çoğu dergi, anlamlı sonuçlar vermeyen araştırmaları yayımlamak için isteksizdir ve bu nedenle çoğu araştırmacı da anlamlı sonuçlar vermeyen çalışmalarını yayımlamak için dergilere sunmamaktadır (DiLullo, 1997, s.15).

Davlin, Holcomb ve Guadagnoli isimli araştırmacılar elektromiyografik nöroterapi konusunda gerçekleştirdikleri bir araştırmada rastgele seçilen 36 denek kullanılmış ve denekler 3 gruba ayrılmıştır. Oluşturulan bu gruplara araştırma altındaki incelemek istenen yöntemler ayrı ayrı uygulanmış ve uygulanan yöntemlerin etkilerinin net olarak görülebilmesi için 5 gün boyunca beklenmiştir. 5 günün sonunda Davlin, Holcomb ve Guadagnoli, uyguladıkları testlerden önce ve sonraki durumları karşılaştırdıklarında önemli bulgular elde etmişlerdir. Fakat tüm bunların yanında araştırmaya tabi tutulan 3 grup arasında anlamlı olmayan, önemsiz sonuçlar da elde etmişlerdir. Araştırmacılar bu durumu “gruplar arasında önemsiz sonuçlara ulaşılmasının nedeni örneklem sayısının olması gerekenden çok daha az olmasıydı. Daha olumlu sonuçlar elde edebilmek için mevcut örneklem sayısının 2 katı kadar sayıya ihtiyaçları vardı” şeklinde açıklamışlardır.

Aslında Tener’e göre (2000) asıl sorun 3 deney grubu arasındaki doğru farklılıklar elde edebilmek için uygulanan çalışmanın istatistiksel güçten yoksun olmasıydı. Örneklem sayısındaki yetersizlik bir teorik güç analizi ile belirlenebilir ve yapılacak değişiklik ile araştırmanın daha tasarım aşamasında yeterli güç garanti altına alınarak, araştırmanın gerekliliklerine uygun bir örneklem sayısı belirlenebilirdi (Tener, 2000, s. 6).

Düşük gücün getirdiği negatif sonuçların yanında aşırı yüksek gücün de getirdiği olumsuz sonuçlar vardır. Aşırı güç, gereğinden çok fazla örneklem sayısı demektir ve bu da farazi bir etkinin belirlenmesi olasılığını artırır. Anlamlı sonuçların doğru yorumlanmasını garantilemek için etki büyüklüğü her zaman istatistiksel çalışmalarda kaydedilmelidir. İlaveten, eğer bir araştırmada güç 0.95 seviyesi gibi çok fazla bir seviyedeyseniz (bu durumda $\beta = 0.05$ ve $\alpha = 0.05$ olduğundan) II. Tip hata yapılma olasılığı, I. Tip hata yapılma olasılığı kadar ciddidir. Fakat genellikle kabul edilen görüşe göre I. Tip hata, II. Tip hatadan daha ciddidir. Cohen, II. Tip hata yapma olasılığının, I. Tip hata yapma olasılığına oranının yani β/α oranının 4/1 olması gerektiğini belirtmiştir. Eğer güç %80 ise ve $\alpha = 0.05$ seviyesinde bir anlamlılık seviyesi ile

çalışılıyorsa $\beta/\alpha = 0.20/0.05 = 4$ olacaktır ki bu Cohen'in tavsiyesine göre ideal bir orandır (DiLullo, 1997, s.16).

Cohen'in belirtmiş olduğu $\beta/\alpha = 4$ oranı oldukça ikna edicidir. Birçok araştırmacı şuan Cohen'in tavsiye ettiği %80 güç seviyesinin ve Fisher'in bir standart haline gelen 0.05 anlamlılık seviyesini "5-80 (five-eighty)" olarak adlandırmaktadır. Bir gelenek haline gelmiş bu oran, araştırmacıları oldukça zahmetli bir işten kurtarmıştır. Fakat 4/1 oranını standart olarak her araştırmada yakalamaya çalışmak pek anlamlı değildir. Örneğin, daha önceden yapılmış olan bir çalışmanın tekrarlanması planlanmaktadır. Eğer önceden yapılmış olan bir çalışmada gerçek bir etki tespit edildiyse, yapılacak olan çalışmada gerçekte var olmayan bir etkinin varmış gibi bulunması gibi bir durum olmayacaktır. Yani I. Tip hatanın yapılma olasılığı yoktur. Bu nedenle bu durumda 4/1 gibi bir beta/alfa oranını yakalamaya çalışmak mantıklı değildir. Ayrıca bu durumda alfayı standart bir seviye olan 0.05 seviyesinde tutmak da mantıklı olmayacaktır. Yapılması gereken araştırmacının kendisini II. Tip hataya karşı korumaya almasıdır.

Aşırı güç ile çalışmak anlamsız olan etkilerin ortaya çıkarılmasına neden olacaktır. Fakat az da olsa yüksek güç ile çalışmak gerekmektedir. Çünkü tahmin edilen etki büyüklüğünün, belirlenenden düşük olması durumunda güç seviyesi yeterli kalmayabilir. Bu nedenle biraz yüksek güç ile çalışmakta fayda vardır (%90 gibi) (Rossi, 2012, s. 25).

Medikal çalışmalar, β/α oranını değerlendirmek için oldukça uygundur. Birçok tıbbi test neredeyse hiç II. Tip hata yapılmayacak şekilde tasarlanır. Bu durum ise I. Tip hata yapma olasılığını artırır. Tıp alanındaki bu testleri tasarlayan kişiler, dolaylı olarak sağlıklı bir kişiye yapılan testler sonucunda "hastalık olduğuna dair bulgular tespit edildi, daha fazla test yapılması gerekir" demenin, gerçekte hasta olan kişiye "herşey çok iyi durumda, herhangi bir sorunuz yok" demekten daha iyi olduğunu söylemektedirler (Ellis, 2010, s. 55-56).

İstatistiksel güç konusundaki farkındalığın artması, araştırma sonuçlarının yorumlanmasına çok ciddi katkı sağlayabilir. Örneğin, Amerika'da özel eğitim alanında yapılan araştırmalarda, bu alanın akademik başarılarına katkısının az olduğu sonucuna ulaşılması nedeniyle özel eğitim alanında uygulanan sistem, yöntem vs. inceleme altına alınmıştır. Bunun üzerine eğitimden sorumlu ilgili birimler, bu alandaki eksikliklerin giderilmesi için buraya düzenli eğitim girişimi adı altında yüksek miktarlarda fon aktarmıştır. Ancak daha sonra yapılan ve istatistiksel güç konusunu da içeren araştırmalarda asıl sorunun sistemde veya uygulanan yöntemde olmadığı, sorunun mevcut personelin verimsizliğinden

kaynaklandığı belirlenmiştir. İstatistiksel gücü içermeyen çalışmalar sonucunda elde edilen yanlış saptamalar, hem maliyet hem de zaman açısından büyük kayıplara neden olmuştur. Burada çalışmanın istatistiksel gücünün, mevcut personelin verimsiz olduğu sonucu elde edilmeden önce araştırılması gerekirdi.

Daha iyi güç tahmin etme ihtiyacı son zamanlardaki araştırma teknolojilerindeki gelişmelerle birlikte artmıştır. Özellikle geliştirilen bilgisayar programları sayesinde araştırmacılar çok değişkenli analizler, meta analizi gibi daha karmaşık analizleri başarıyla yapmaya başlamışlardır. Çok değişkenli analizlerdeki yanlış yorumlamalar, istatistiksel güç konusunun daha iyi kullanılması ve anlaşılmasıyla kolayca halledilmiştir.

Araştırmacılar bazen aralarında anlamlı bir fark olup olmadığını görmek adına grupları karşılaştırmakta ve eğer gruplar arasında bir farklılık varsa bunun istatistiksel olarak anlamlı olup olmadığını incelemektedirler. Gözlemlenen bir farklılığın istatistiksel olarak anlamlılığını bir testin istatistiksel gücü etkilemektedir. İstatistiksel güç arttığında, çok küçük etkiler bile istatistiksel olarak anlamlı etiketini alabilmektedir. Bu durum ise normalde birbirine çok yakın olan grupların bile istatistiksel olarak farklı oldukları sonuçlarına yol açabilir. Bu durumda aşırı güç ile çalışmanın getirdiği olumsuz durumu daha iyi görebilmek adına aşağıdaki SPSS çıktısını incelemekte fayda vardır:

t	df	Sig. (2-tailed)	Mean difference
-2.296	999998	.022	.00

Son sütundaki veri, iki grup arasında bir fark olmadığını fakat bu farklılığın istatistiksel olarak anlamlı olduğunu üçüncü sütunda ($p < 0,05$) göstermektedir. Hemen hemen iki özdeş grup arasındaki fark nasıl oluyor da istatistiksel olarak anlamlı çıkıyor? Burada son sütunda gösterilen fark aslında 0,00 değil 0,0046'dır. Fakat SPSS bu sayı çok küçük olduğu için bunu 0.00'a yuvarlamıştır. Birçok kişi bu 0,0046 gibi bir sayıyı anlamlı bir fark olarak kabul etmez. Ancak çalışmada kullanılan yaklaşık 1.000.000 veri, aradaki bu mikroskobik farklılığın anlamlı olarak sonuçlanmasına neden olmuştur. İşte bu örnek, aşırı güç ile çalışmanın ne kadar tehlikeli olduğunu açıkça göstermektedir. “p” değeri konusunda çok hassas olan ve etki

büyüklüğü konusunda pek fazla bilgi sahibi olmayan bir arařtırmacı, normalde anlamlı olmayan bir farklılıđı bu alıřmada anlamlı bir farklılıkmiř gibi gösterecektir.

“İdeal güç seviyesi ne olmalı?” sorusu, cevabı kolay olan bir soru deđildir. Çünkü bu durum risk ve getiri açısından bir mübadele gerektirmektedir. Eđer %50 güç seviyesi ile bir alıřma gerçekleştirilirse bu, yanlış bir sıfır hipotezini reddetmek için arařtırmacının sanki yazı tura atıyormuř gibi %50 řansa sahip olduđu anlamına gelmektedir. Birok emek harcanarak gerçekleştirilen bir alıřmanın başarılı olma durumunun yazı tura atma gibi bir řansa bađlanması açıkası sorgulanması gereken bir durumdur. Eđer güç seviyesi %90 gibi bir seviyede kurulursa, o zaman etki belirlenme řansı olduka artar ve II. Tip hata yapma olasılıđı %10’a düşer. Ancak bu seviyedeki bir istatistiksel güç olduka maliyetlidir. $r = 0.12$ gibi bir etkiyi belirlemek için çift yönlü $\alpha = 0.05$ ve $\beta = 0.10$ ise bu durumda t-testi için gerekli örneklem sayısı $n = 732$ ’dir (Ellis, 2010, s. 53).

Özet olarak, arařtırmanın tasarım ařamasında güç analizini dikkate almak, arařtırmacıya gerekli örneklem sayısının belirlenmesi konusunda yardımcı olmakla kalmaz, aynı zamanda;

1. H_0 hipotezi ile ilgili olası yanlış kararları engeller.
2. Anlamlı olmayan sonuçların dođru yorumlanmasına olanak sađlar.
3. Arařtırmanın yeniden yapılması ihtiyacını azaltarak zaman, aba ve maliyet açısından arařtırmacıya tasarruf sađlar.
4. Kritik öneme sahip arařtırmaların yayımlanmasına olanak sađlar.
5. Büyük örneklem sayısı nedeniyle belirlenebilecek farazi etki olasılıđını azaltır.
6. II. Tip hatadan daha ciddi olan I. Tip hata yapma olasılıđını arařtırmacıların kontrol etmelerine olanak sađlar.
7. Arařtırmacılara H_0 hipotezlerini kanıtlama řansını sađlar.

Buradan hareketle řu açıka söylenebilir ki teorik güç analizi, arařtırmanın tasarım ařamasında kesinlikle dikkate alınmalıdır (DiLullo, 1997, s. 17).

1.12 Güç Analizi Değerlerinin Yorumuna İlişkin Literatürdeki Tartışmalar

Cohen'in 1962 yılındaki çığır açıcı çalışmasından sonra, istatistiksel güç analizinin sonuçlarının yorumlanması konusunda özellikle psikoloji ve eğitim alanındaki dergilerde tartışmalar başlamıştır. Bu tartışmaların çoğu anlamlılık testi yapıldıktan sonra elde edilen sonuçların güç ile ilgili yorumları konusunda olmuştur. McKean'in (1990) aktardığına göre, Cohen (1973), güç ile ilgili yapılan bu tartışma üzerine kısa ve net bir şekilde istatistiksel güç değerinin ne olduğunun anlamlılık testi ile bir alakasının olmadığını ve anlamlılık testinin bundan etkilenmediğini belirtmiştir. Araştırma için gerekli olan veriler toplandıktan ve analiz edildikten sonra güç analizi arka planda kalmaktadır. Eğer anlamlı olmayan sonuçlar elde edilirse o zaman güç konusu gündeme gelir ve tartışılmaya başlar. Bu durumda yorum araştırmanın gücüne bağlıdır. Eğer güç yüksek değil ise araştırma, sonuçlar kesin olsa bile, "bir farklılık yoktur" şeklinde sonuçlandırılmaz. Bu durumda güç yüksek olmadığı müddetçe, sonuçlar fonksiyonel olarak geçersizdir. Herhangi bir farklılık vardır veya oluşan farklılık göz ardı edilebilir şeklinde çıkarımlar yapabilmek için gücün yüksek olması gerekmektedir (s. 29).

1.13 Literatürdeki Çalışmaların Güç Açısından İncelenmesi

İstatistiksel güç analizleri bireysel çalışmalarda uygulanabileceği gibi bir dergide herhangi bir alanda yayımlanmış olan çalışmalar için de uygulanabilmektedir. Bu tür çalışmalarda araştırmacılar genellikle herhangi bir alanın ortalama istatistiksel gücünü belirleme konusunda çalışmakta ve II. Tip hata yapılma oranlarını değerlendirmektedirler. Böyle çalışmalarda araştırmacı, istatistiksel gücü belirlemek için gözlemlenen etki büyüklüğü yerine hipotez kurulmadan önce tahmin edilen etki büyüklüğünü, örneklem sayısını ve alfa seviyesini kullanmaktadırlar. Bu kıstasları kullanarak araştırmacı aslında "eğer bu değerler kullanılsaydı çalışmanın gücü ne olurdu?" sorusunun cevabını bulmaya çalışmaktadır. Bu şekilde gerçekleştirilen çalışmalar, herhangi bir alanda yapılmış olan araştırmaların ortalama istatistiksel gücünü ve yanlış bir sıfır hipotezini reddetme şanslarının ne olduğunu ortaya koymaktadırlar.

Yayımlanmış çalışmalar için yapılan birçok post-hoc güç analizi çalışmasında genellikle Cohen'in 1962 yılında güç analizi konusunda yapılmış ilk çalışmasındaki yöntem izlenmektedir. Bu yöntem aşağıda bahsedildiği gibi dört ana başlık altında toplanabilir:

- 1) Öncelikle güç seviyesi incelenmek istenen alandaki çalışmalar belirlenir. Bu çalışmalar bir dergide yayımlanmış olan belirli dönemler arasındaki makaleler olabildiği gibi, farklı dergilerde farklı zamanlarda aynı alanda yayımlanmış olan makaleler de olabilmektedir. İstatistiksel güç analizi değerlendirmeleri, örneklem sayısı ve etki büyüklüğü tahmini kaydedilen herhangi bir çalışma için yapılabilir. Eğer etki büyüklüğü belirtilmemiş ise Cohen'in geleneksel küçük, orta ve büyük etki büyüklüğü sınıflandırmalarından birisi kullanılabilir.
- 2) Her bir çalışmada kullanılan örneklem sayıları ve kullanılan istatistiksel testler kayıt edilir. Aksi belirtilmedikçe kullanılan testlerin çift yönlü (non-directional, two-tailed) olduğu kabul edilir. Eğer bir çalışmada çok sayıda istatistiksel test kullanılmış ise ana hipotez üzerinde önemli etkisi olan testler seçilir.
- 3) Kaydedilen her bir n , α ve etki büyüklüğü kullanılarak her bir istatistiksel testin gücü hesaplanır. Örneğin, bağımsız ortalamalar t-testi için toplamda $n = 40$ örneklem sayısı ile gerçekleştirilen bir testin $d = 0.20$, $d = 0.50$ ve $d = 0.80$ etki büyüklükleri için sırasıyla istatistiksel güç değerleri 0.09, 0.34 ve 0.69 olacaktır.
- 4) Tüm çalışma boyunca istatistiksel güç değerleri hesaplandıktan sonra etki büyüklüğü seviyeleri için ortalama güçler hesaplanır. Aşırı güçlü ve aşırı düşük güçlü çalışmalar ortalama değeri etkileyebileceği için medyan gücü hesaplamak da mantıklı bir seçenektir (Ellis, 2010, s. 73-74).

1962 yılında ilk güç analizi çalışmasını Jacop Cohen, *Journal of Abnormal and Social Psychology* isimli dergide yayımlanmış olan 70 adet makale üzerinde gerçekleştirmiştir (Chase ve Chase, 1976, s. 234). Cohen, gerçekleştirdiği bu çalışmada kendi oluşturduğu küçük, orta ve büyük seviyelerdeki etki büyüklüklerini kullanarak makaleler üzerinde bir *deneysel güç analizi* hesaplaması yapmış ve bu çalışmanın neticesinde, yapılan istatistiksel testlerin gücü; küçük seviyedeki etki büyüklüğü için 0.18, orta seviyedeki etki büyüklüğü için 0.48 ve büyük seviyedeki etki büyüklüğü için 0.83 çıkmıştır. İncelenen 70 makalenin hiçbirinde küçük seviyedeki etki büyüklüğü sağlayabilmek için tespit edilen güç yeterli değildir. Cohen bu çalışma sonunda araştırmacıların büyük seviyede etki büyüklüğü sağlayamadıkça H_0 hipotezini reddetmek için çok zayıf şansları olduğunu tespit etmiştir. Cohen'in yapmış olduğu bu çalışma, *Journal of Abnormal and Social Psychology* dergisi yazarlarının istatistiksel gücü, çalışmalarında ihmal ettiklerini göstermiştir (McKean, 1990, s.30; Borkowski vd., 2001, s. 69).

Cohen'in 1962 yılında yapmış olduğu çalışmadan sonraki yıllarda yönetim, yönetim ve uygulamalı psikoloji, psikoloji, yönetim bilgi sistemleri, eğitim-öğretim, pazarlama ve iletişim gibi çok farklı disiplinlerde yapılan araştırmalarda Cohen'in araştırmacılara tavsiye ettiği istatistiksel güç analizi konusuna nicel araştırmalarda neredeyse hiç değinilmediği gözlemlenmiştir. Gerçekleştirilen araştırmaların hemen hemen hepsinde elde edilen ortak sonuç güç konusunun neredeyse hiç dikkate alınmamasıdır. Bu durumla alakalı, güç konusunda ilk araştırma yapan kişi olan Jacop Cohen: "Araştırmacıların güç konusunda ihmalkâr davranmalarını anlamak mümkün değil. Bu gidişatın editörler ve yazarlar tarafından bu şekilde kabul edilmiş olması neredeyse bir sırdan daha ötedir" demiştir (Cashen ve Geiger, 2004, s. 151).

Sedlmeier ve Gigerenzer (1989), Cohen'in çalışmasını yaptığı dergiye 24 yıl aradan sonra aynı çalışmayı uygulayarak istatistiksel güç analizi konusunda bir gelişme olup olmadığını "*Do Studies of Statistical Power Have an Effect on the Power of Studies?*" isimli makalelerinde incelemişlerdir. Ancak 1962 yılında Cohen'in çalışmasını yaptığı *Journal of Abnormal and Social Psychology* dergisi, Sedlmeier ve Gigerenzer'in tekrardan incelemek istedikleri dönemde *Journal of Abnormal Psychology* ve *Journal of Personality and Social Psychology* isimli iki dergiye bölünmüştür. Sedlmeier ve Gigerenzer, örneklemi için *Journal of Abnormal Psychology* dergisini seçmişlerdir. Bu çalışmada, Cohen'in değerlerine mümkün olduğunca en yakın değerleri seçmek amacıyla Cohen'in çalışmasında kullandığı kriterlere yakın kriterler belirlenmiştir. Elde edilen sonuçlar istatistiksel güç analizi konusuna olan ilginin, aradan 24 yıl geçmiş olmasına rağmen hiçbir gelişme göstermediği şeklinde olmuştur. Sedlmeier ve Gigerenzer (1989), kötümser bir tabirle güç konusuna neredeyse hiç ilgi olmadığından bahsetmişlerdir. Bahsi geçen bu kötü durum, yapılan çalışmaların yaklaşık olarak %5'inde güçten bahsedildiğini ve hiçbir çalışmanın tasarım aşamasında güç tahmini yapılmadığı gerçeği ile desteklenmiştir (Sedlmeier ve Gigerenzer, 1989; Cohen, 1992, s. 155).

Cohen, 1962 yılında yaptığı çalışmada küçük, orta ve büyük seviyelerdeki etki büyüklükleri için istatistiksel güç değerlerini sırasıyla 0.18, 0.48 ve 0.83 olarak bulmuştur. 24 yıl sonra Sedlmeier ve Gigerenzer, yaptıkları çalışmada ise küçük, orta ve büyük seviyelerdeki etki büyüklükleri için istatistiksel güç değerlerini sırasıyla 0.21, 0.50 ve 0.84 olarak elde etmişlerdir. İlâveten, araştırmaya dâhil edilen çalışmaların sadece ikisinde istatistiksel güç analizinden bahsedilmiş, ancak çalışmaların hiçbirinde uygulanan istatistiksel testlerin gücü rapor edilmemiştir. Dahası, sadece 4 makalede anlamlılık seviyesinden bahsedilmiş fakat hiçbir çalışmada neden belirli bir anlamlılık seviyesi değeri

kullanıldığından veya örneklem sayısının hangi kriterlere göre seçildiğinden bahsedilmemiştir. Ayrıca bu çalışmalarda etki büyüklüğü tahminine dair hiçbir ipucu da elde edilmemiştir (Sedlmeier ve Gigerenzer,1989).

Nickelson (2000), araştırmacıların güç ile olan alakasız olan durumlarını, güç kavramının tam olarak anlaşılmasına ve bu nedenle de çalışmalarda kullanılmamasına bağlamıştır. Mone vd. (1996)'de araştırmacıların güç konusuna olan yaklaşımları konusunda benzer eleştiriler yapmışlardır. Geçmişte yapılan çalışmalardaki istatistiksel güç etkisinin minimal seviyede olduğu, güç kavramının daha fazla dikkate alınması gerektiği belirtilmiştir. Ayrıca gerçekleştirdikleri araştırmada makale yazarlarından 2/3'ünün güç analizini yapmadıklarını tespit etmişlerdir (Mone vd., 1996, s. 111).

Daha sonraki yıllarda çok sayıda istatistiksel güç analizi çeşitli alanlarda yapılmasına rağmen ortalama güç özellikle küçük ve orta seviyedeki etki büyüklükleri için ortalama güç düşük kalmıştır. Küçük seviyedeki etki büyüklükleri 0.10 ile 0.37 arasında değişmektedir ve küçük seviyedeki etki büyüklüğü için de ortalama güç 0.24 olarak hesaplanmıştır. Bu seviye, Cohen'in tavsiye etmiş olduğu %80 güç seviyesinin çok altındadır. Bu seviyedeki bir güç ile II. Tip hata yapılma olasılığı kuvvetle muhtemeldir. II. Tip hata oranı 0.76 ve I. Tip Hata Oranı 0.05 olduğundan β/α oranı 15/1 şeklinde olmaktadır ki bu Cohen'in araştırmacılara tavsiye ettiği $\beta/\alpha = 4/1$ oranından çok çok daha yüksektir (DiLullo, 1997, s. 45).

Orta seviyedeki etki büyüklüğü için ortalama istatistiksel güç 0.63 seviyesindedir. Bu oran, küçük seviyedeki etki büyüklüğü için sağlanmış olan güç seviyesi ile kıyaslanacak olursa (ki bu değer 0.24 idi) güç seviyesinde keskin bir şekilde artış sağlanmıştır. Fakat yakalanmış olan bu keskin artış bile hala sağlıklı sonuçlar elde etmek için yeterli değildir. Bu %63'lük güç oranı, araştırmacıların β/α oranı bakımından 8/1 oranında II. Tip hata yapma olasılıklarının, I Tip hata yapma olasılıklarından daha fazla olduğunu göstermektedir.

Araştırmacılar, çalışmalarını gerçekleştirdikleri anakütle üzerinde büyük etki büyüklüğü belirlemek isteselerdi, yeterli gücü elde edeceklerdi. Büyük etki büyüklüğü için ortalama güç 0.86 seviyesinde elde edilmiştir. Bu nedenle araştırmacılar β/α oranı bakımından 3/1 oranında II. Tip hata yapma olasılığına sahiptirler. Bu oran, Cohen'in tavsiye etmiş olduğu 4/1 şeklindeki orandan daha iyidir. Ayrıca çalışmaların üçünde güç 0.95 civarında olan aşırı seviyelerde sağlanmıştır. Eğer sağlanan güç, %95 ten fazla ise o zaman I. Tip hata ciddi bir hata olarak dikkate alınmaz ve bu durum araştırmacının I. Tip hata yapma olasılığını, II. Tip hata yapma olasılığından daha fazla dikkate aldığını göstermektedir. Örneğin, bir araştırmada

büyük etki büyüklüğü belirlemek istenirse ve bu durumda ortalama gücün 0.98 seviyesinde olduğu varsayılırsa, bu güç değeri araştırmacının β/α oranı bakımından 2/5 oranını kabul ettiğini gösterir. Başka bir deyişle, bu durum araştırmacının H_0 hipotezi ile ilgili, yanlış olan bir H_0 hipotezinin reddedilemediği (II. Tip hata) her 2 yanlış karar için doğru olan H_0 hipotezini reddetme (I. Tip hata) şeklindeki 5 yanlış karar vermeyi göze aldığını göstermektedir (DiLullo, 1997, s. 45-46).

DiLullo'nun (1997) aktardığına göre, bir çalışmanın, içinde istatistiksel gücü barındırması bakımından Katzer ve Sordt da benzer sonuçlar elde etmişlerdir. Araştırmalarını gerçekleştirdikleri 31 adet makalenin sadece ikisinde sonuçların yorumlama aşamasında veya örneklem sayısına karar verilmesi aşamasında güç analizinden bahsedilmiş ve sadece bir çalışmada bu durum ayrıntılı olarak anlatılmıştır. Bu konuda başka bir araştırmayı daha gerçekleştiren Wolly ve Dawson ise yaptıkları incelemeler sonucunda elde ettikleri veriler bakımından diğer araştırmadaki sonuçlar ile aynı kaderi paylaşmışlardır. Çünkü yaptıkları araştırmada 192 adet makale incelemişler ve bunların hiçbirinde güç ile ilgili bir kısma rastlamamışlardır. Bazı makalelerde ise etki büyüklüğü, örneklem sayısı gibi temel istatistik kavramlarına hiç değinilmemiştir (s. 47).

Yapılan incelemelerde güç ile ilgili elde edilen olumsuz veriler, araştırmacıların sıfır hipotezlerini reddetme konusundaki şanslarını da azaltmaktadır. Başka bir ifade ile yetersiz gücün bir sonucu olarak değişkenler arasında gerçekte anlamlı olan farklılıklar veya ilişkiler tespit edilememektedir (Cashien ve Geiger, 2004, s. 152).

DiLullo'nun (1997) aktardığına göre, araştırmaların, diğer araştırmacılar ve okuyucular tarafından uygun bir şekilde yorumlanabilmeleri için ilgili istatistiksel veriler kesinlikle kayıt edilip, açıklanmalıdır. Katzer ve Sordt'a göre birçok bilimsel çalışmanın tam anlamıyla anlaşılabilmesi, araştırmada kullanılan yöntemlerin açıkça anlatılmasına bağlıdır. Kullanılan yöntemlerin net bir şekilde ifadesi, yorumların bağımsızlığı açısından da gereklidir (s. 47).

Peterman, bir ekoloji dergisinde okuduğu makaleler için hayal kırıklığına uğradığını, bunun nedeni olarak makalelerde anlatılanları değil, hiçbirinin istatistiksel güçten bahsetmemiş olmasını göstermiştir. Peterman, ekolojistlerin istatistiksel güç konusuna olması gerekenden çok daha az önem verdiklerini ve bu durumun yapılacak önemli çalışmalar için büyük tehlike arz ettiğini belirtmiştir (Peterman, 1990, s. 2024).

Aşağıda verilen Tablo 1.11, çeşitli araştırmacılar tarafından farklı alanlarda yayımlanmış eserler üzerinde gerçekleştirilen güç incelemelerini özetlemektedir. Gerçekleştirilen bütün

araştırmalarda Cohen'in geleneksel küçük, orta ve büyük seviyelerdeki etki büyüklükleri esas alınmıştır. Küçük ve orta seviyelerdeki etki büyüklüğü için bütün durumlarda yetersiz güç sağlandığı, gerçekleştirilen beş araştırmada ise büyük seviyedeki etki büyüklüğünde dahi yeterli istatistiksel güç sağlanmadığı ortaya çıkmıştır.

Tablo 1.11'den elde edilen veriler, Cohen'in 1962 yılında gerçekleştirmiş olduğu araştırmadaki veriler ile kıyaslanacak olursa aradan geçen yirmi yıla rağmen güç analizi konusunun araştırmacılar tarafından hala ilgi görmediği kolaylıkla söylenebilir. Günümüzde ise Cohen'in 1962 yılında uyguladığı yöntem dikkate alınarak tamamlanmış tezlerin güç açısından incelenmesi işlemi yapılmaktadır (McKean, 1990, s.31).

Tablo 1.11 Farklı Alanlarda Yapılmış Çalışmaların İstatistiksel Güç Ortalamaları

Çalışma	İncelenen Dergiler	İnceleme Yılı	Makale Sayısı	Çalışma Başına Ort. Test Sayısı	Ortalama Güç		
					Küçük	Orta	Büyük
Cohen (1962)	<i>Journal of Abnormal and Social Psychology</i>	1960	70	69.0	.18	.48	.83
Brewer (1972)	<i>American Educational Research Journal</i>	1969–1971	47	7.9	.14	.58	.79
Brewer & Owen (1973)	<i>Journal of Educational Measurement</i>	1969–1971	13	20.5	.21	.72	.96
Katzer & Sodt (1973)	<i>Journal of Communication</i>	1971–1972	31	53.9	.23	.56	.79
Chase & Tucker (1975)	9 communication journals	1973	46	28.2	.18	.52	.79
Kroll & Chase (1975)	2 communication journals	1973–1974	62	16.7	.16	.44	.73
Chase & Baran (1976)	2 communication journals	1974	48	14.6	.34	.76	.91
Chase & Chase (1976)	<i>Journal of Applied Psychology</i>	1974	121	27.9	.25	.67	.86
Christensen & Christensen (1977)	<i>Research Quarterly</i>	1975	43	–	.18	.39	.62
Sawyer & Ball (1981)	<i>Journal of Marketing Research</i>	1979	23	–	.41	.89	.98
Daly & Hexamer (1983)	<i>Research in the Teaching of English</i>	1978–1980	57	21.6	.22	.63	.86
Orme & Combs-Orme (1986)	<i>Social Work Research and Abstracts</i>	1977–1984	79	39.4	.31	.76	.92
Mazen et al. (1987b)	<i>Academy of Management Journal, Strategic Management Journal</i>	1982–1984	44	83.3	.23	.59	.83
Baroudi & Orlikowski (1989)	4 MIS journals	1980–1985	57	2.6	.19	.60	.83
Sedmeier & Gigerenzer (1989)	<i>Journal of Abnormal Psychology</i>	1984	54	–	.21	.50	.84
Rossi (1990)	3 psychology journals	1982	221	27.9	.17	.57	.83
Lindsay (1993)	3 management accounting journals	1970–1987	43	43.5	.16	.59	.83
Kosciulek & Szymanski (1993)	4 rehabilitation counseling journals	1990–1991	32	–	.15	.63	.90
Mone et al. (1996)	7 leading management journals	1992–1994	210	126.1	.27	.74	.92
Clark-Carter (1997)	<i>British Journal of Psychology</i>	1993–1994	54	23.0	.20	.60	.82
Borkowski et al. (2001)	3 behavioral accounting journals	1993–1997	96	18.6	.23	.71	.93
Maddock & Rossi (2001)	3 healthy psychology journals	1997	187	44.2	.36	.77	.92
Bezeau & Graves (2001)	2 neuropsychology journals	1998–1999	66	29.5	.50	.77	.96
Brock (2003)	2 international business & 2 management journals	1990–1997	374	3.0	.29	.77	.93
McSwain (2004)	2 MIS journals	1996–2000	45	3.9	.22	.74	.92

Kaynak: Ellis, 2010, s. 76

Deng'in (2000) aktardığına göre, Tablo 1.11'de bahsi geçen araştırmalardan bazıları şu şekilde gerçekleştirilmiştir;

Katzer ve Sodt (1973), *Journal of Communication* isimli dergide 1971 ve 1972 yılları arasında yayımlanan 31 istatistiksel makaleyi incelemiş ve yapılan inceleme sonucunda sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için ortalama güç 0.23, 0.56 ve 0.79 olarak bulunmuştur.

Chase ve Tucker (1975), *The American Forensic Association Journal*, *Central States Speech Journal*, *Journal of Communication*, *The Quarterly Journal of Speech*, *Southern Speech Communication Journal*, *Speech Monographs*, *The Speech Teacher*, *Today's Speech* ve *Western Speech* isimli dokuz dergide 1973 yılında yayımlanan tüm makaleleri incelemiştir. Hesaplanan ortalama güç sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için 0.18, 0.52 ve 0.79 seviyelerindedir. Elde edilen bu sonuçlar ile daha sonra Cohen'in 1962 yılında psikoloji alanında ve Brewer'ın 1972 yılında eğitim alanında gerçekleştirdiği çalışmalar ile üç farklı disiplin arasında bir karşılaştırma yapılmış ve büyük oranda benzer seviyelerde güç oranlarına rastlanmıştır.

Chase ve Kroll (1975), *Speech Psychology* ve *Audiology Research* isimli dergilerdeki makaleleri incelemiş ve sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için ortalama güç 0.16, 0.44 ve 0.73 şeklinde bulunmuştur.

Tablo 1.11'den de görüldüğü üzere birçok çalışmadaki güç seviyesi oldukça düşük kalmıştır. Araştırmalara dâhil edilen çalışmalardan hiçbiri küçük bir etkiyi belirleme konusunda yeterli değildir. Orta seviyedeki etkiler için durum biraz daha iyi fakat yine de istenilen %80 seviyesinden oldukça uzaktır. Sadece bir dergide (*Journal of Marketing Research*) orta seviyedeki etki için yeterli güç gözlemlenmiştir.

Chase ve Baran (1976), *Journalism Quarterly* ve *Journal of Broadcasting* isimli kitle iletişimi alanındaki dergilerde 1974 yılında yayımlanmış makaleleri incelemiş ve sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için ortalama güç 0.34, 0.76 ve 0.91 çıkmıştır. Elde edilen bu sonuçlar ile daha sonra psikoloji, eğitim, iletişim ve kitle iletişimi gibi disiplinler arasında bir kıyaslama yapılmış ve en yüksek güç seviyelerine kitle iletişimi alanında rastlanmıştır.

Chase ve Chase (1976), *Journal of Applied Psychology* isimli dergide 1974 yılında yayımlanan tüm istatistiksel makaleleri incelemiş ve sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için ortalama güç 0.25, 0.67 ve 0.86 şeklinde bulunmuştur (Deng, 2000, s. 33-36).

Rossi, 1990 yılında o güne kadar yaklaşık 1.500 makale ve 40.000 civarında istatistiksel test üzerinde yapılmış olan 25 güç analizi araştırmasını incelemiş ve ortalama güç seviyelerini çıkarmıştır. Yapılan çalışma sonucunda tüm bu makalelerin ortalama gücü sırasıyla küçük, orta ve büyük seviyelerdeki etkiler için 0.26, 0.64 ve 0.85 çıkmıştır (Rossi, 1990, s. 647).

Tablo 1.11’de yer alan çalışmalarda kullanılan istatistiksel test sayılarının çok fazla olduğu görülmektedir. Bu da araştırmacıların çalışmaları ile ilgili yayımlanabilecek anlamlı sonuçlar çıkarabilmek için adeta balık avına çıktıklarının göstergesidir (Ellis, 2010,s. 77-78).

Çalışmaların çoğunda bulunmak istenen etki büyük olmadıkça düşük güç nedeni ile başarısız olunacaktır. Fakat ortalama güç seviyesi bu kadar düşük olmasına rağmen nasıl oluyor da dergilerde yayımlanan çalışmalarda bir etki belirlenebiliyor? Bu durumun muhtemel iki açıklaması vardır:

- i) Ya bu çalışmalarda belirlenen etkiler çok kolay bir şekilde ortaya çıkarılabilecek etkilerdir (whopper effect),
- ii) Ya da araştırmacılar ortada var olmayan bir etkiyi belirlemektedirler.

Sosyal bilimler alanında yapılan meta analiz sonuçları çalışmaların bazen orta dereceli etki, çoğu zaman ise küçük dereceli etki içerdiğini göstermektedir. Zaten varlığı aşikâr olan büyük seviyedeki etkinin belirlenmesi ise çok da bir anlam ifade etmeyecektir.

Bazı durumlarda aynı veri seti üzerinde çok fazla istatistiksel testin gerçekleştirilmesi sonucunda, elde edilen anlamlı bir etki olmamasına rağmen tesadüfen de olsa bazı sıfır hipotezleri anlamlı sonuçlar verebilir. Bu durum yazında “*curse of multiplicity (çok çeşitliliğin zararı)*” olarak geçmektedir. Örneğin aynı veri seti üzerinde toplamda 14 adet istatistiksel test, her biri $\alpha = 0.05$ anlamlılık düzeyinde gerçekleştirilmiş olsun. Her bir test %5’lik bir hata payı içereceği için tesadüfen de olsa en azından bir adet yanlış olan bir sıfır hipotezi reddedilebilir. Bu durumda istatistiksel güç çok düşük bile olsa bazı anlamlı sonuçlar tesadüfen elde edilebilir. Kullanılan test sayısı ve anlamlılık düzeyine göre en azından bir adet doğru bir sıfır hipotezinin yanlış bulunma olasılığı için aşağıdaki formül verilmiştir:

$$P = 1 - (1 - \alpha)^N$$

N: Çalışmada kullanılan istatistiksel test sayısı, α : Anlamlılık seviyesidir.

Örneğin 14 adet istatistiksel test ile gerçekleştirilen bir çalışmada, $\alpha = 0.05$ için, sıfır hipotezlerinin hepsi doğru bile olsa, tesadüfen en azından bir adet doğru bir sıfır hipotezinin reddedilmesi olasılığı;

$$P = 1 - (1 - 0.05)^{14}$$

P = 0.51 olarak bulunur.

Başka bir çalışmada 35 adet istatistiksel test kullanılsaydı, $\alpha = 0.05$ için bu değer

$$P = 1 - (1 - 0.05)^{35}$$

P = 0.83 olarak bulunurdu.

Görüldüğü gibi aynı veri seti üzerinde kullanılan istatistiksel test sayısı arttıkça, doğru bir sıfır hipotezinin reddedilme olasılığı da artmaktadır.

Son yıllarda yapılan araştırmalarda ise ortalama güç seviyesinin biraz arttığı gözlemlenmiştir. Maddock ve Rossi, 2001 yılında yaptıkları araştırmalarda sağlık psikolojisi alanındaki ortalama güç seviyesinin orta seviyedeki etki büyüklüğü için 0.77 olduğunu ve incelemeye aldıkları çalışmalardan sadece %17'sinin güç değerinin %50'nin altında olduğunu belirlemişlerdir. Bu durum, metodolojik gelişmelerin araştırmacıların alışkanlıklarını etkilemesinin biraz zaman aldığı bir göstergesi olarak düşünülebilir. Ancak burada istisna bir durum da söz konusu olabilir. Çünkü sağlık alanında yapılan araştırmalar devlet tarafından çok büyük fonlar ile desteklenmekte ve yeterli istatistiksel güç ile çalışılması, bunun rapor edilmesi fon sağlanması için bir şart olabilmektedir. Bu nedenle sağlık alanındaki araştırmalarda istatistiksel güce biraz da zorunlu olarak daha fazla önem verilmektedir. Yani araştırmacılar genellikle kullanmak zorunda olduklarında güç analizini başarılı bir şekilde uyguluyorken, güç analizi uygulayıp uygulamama konusu kendi kararlarına bırakıldığında güç analizine başvurmamaktadırlar (Maxwell, 2004, s. 148; Aberson, 2010, s. 9-10). Aynı şekilde 2005 yılında Aguinis vd. tarafından uygulamalı psikoloji alanında yapılan bir araştırmada da güç konusuna verilen önemin arttığı gözlemlenmiştir (Abraham ve Russell, 2008, s. 284). Artan bu önem ile birlikte çoğu dergi editörü, yapılan araştırmaları yayımlamak için yazarlardan güç analizi raporlarını da istemektedir. Dahası araştırmalara fon sağlayan şirketler dahi çoğu zaman araştırma öncesinde yapılmış güç analizi sonuçlarını araştırma teklifleri ile birlikte istemektedirler (Abraham ve Russel, 2008, s. 286)

Cohen'in önermiş olduğu 4:1 beta/alfa oranı, elde edilen sonuçlar ile kişileri kandırmanın riski, var olan bir etkiyi ortaya çıkaramamaktan 4 kat daha ciddi olduğunu göstermektedir.

Tablo 1.11.'de verilen arařtırmalar içinde bu oran ortalama olarak 7:1 seviyesindedir. Güç seviyesinde meydana gelebilecek bir düşüş, bu oranı daha da artıracaktır.

Örneğin, A ve B gibi iki dergi sadece istatistiksel olarak anlamlı sonuçları yayımlasın. A dergisi, yazarlardan 5:80 (five-eighty) kuralına göre (yani alfa = 0.05 ve güç = 0.80) çalışmalarını tasarlamalarını istesin. Bu durumda her 16 çalışmada sıfır hipotezi doğru bir şekilde reddedilecekken 1 kez yanlış olarak reddedilecektir.

B dergisi ise arařtırmacıların sadece $p < 0.05$ kriterine önem vermelerini istesin ve güç konusunda herhangi bir beklentisi olmasın. B dergisinde yayımlanan çalışmalara bir post hoc güç analizi uygulansın ve elde edilen sonuç %40 olsun. Bu durumda 5:40 oranı elde edilecek ve her 8 doğru karar için 1 kez I. Tip hata meydana gelecektir. Düşük gücün bir sonucu olarak B dergisinde, A dergisinde yapılan I. Tip hata oranının iki katı kadar hata yapılacaktır (Ellis, 2010, s. 80).

1.14 İstatistiksel Gücü Artırma Yolları

Bir arařtırma tasarlanırken arařtırmacı, çalışmasının istatistiksel gücünün artmasını sağlayacak yollara aşına olmalıdır. Arařtırmacı diğer parametreleri yani α , etki büyüklüğü ve örneklem sayısını, çalışmasında alternatif hipotezi desteklemek için gereken yeterli istatistiksel gücü sağlamak amacıyla manipüle edebilir (Tener, 2000, s. 31).

Aşırı güç ile çalışmak zararlı olmasına rağmen güç seviyesinin belirli bir seviyeye kadar artması faydalıdır. Arařtırmacılar için gücü artırmak birkaç yol ile mümkündür. Genellikle başvuru yöntem ise büyük etki aramak ve örneklem sayısını artırmaktır. Örneğin, bir şirket yaptığı reklamların şirkete olan olumlu katkısını arařtırıyorsa, reklam-satış gelirleri arasındaki ilişkiye odaklanmamalıdır. Çünkü satış gelirleri iç ve dış olmak üzere birçok faktöre bağlıdır. Bunun yerine daha belirli bir konu olarak reklam-markanın bilinirliğı ilişki arařtırılması daha büyük bir etki elde edilmesini sağlayacaktır. Çünkü markanın bilinirliğini doğrudan etkileyen faktörlerin başında reklam gelmektedir. Bazı durumlarda ise daha büyük ölçüm sayısı elde etme ile de ilgili olabilir. Örneğin, eğitim ile ilgili bir arařtırma yapılıyorsa bir sınıf yerine iki veya daha fazla sınıfın arařtırmaya dâhil edilmesi daha büyük bir etkinin ortaya çıkarılmasına katkı sağlayabilir.

Arařtırmacıların ortaya çıkarmayı tahmin ettikleri etkileri artırma konusunda herhangi bir kontrollerinin olmadığı durumlarda ise gücü artırmak için başvuru bir diğer yöntem örneklem sayısını artırmaktır. Bazı durumlarda örneklem sayısı iki katına çıkarıldığında, elde

edilecek güç seviyesi iki katından fazla artar. Örneğin, çift yönlü bir testte $\alpha = 0.01$, etki büyüklüğü (r) = 0.30, $n = 50$ için güç = 0.33 iken örneklem sayısı iki katına çıkarıldığında güç = 0.68 olmaktadır. Bazı durumlarda ise örneklem sayısındaki artış güç seviyesini pek fazla etkilemeyebilir. Örneğin, çift yönlü bir testte $\alpha = 0.05$, etki büyüklüğü (r) = 0.10, $n = 50$ için güç = 0.11 iken, örneklem sayısı iki katına çıkarıldığında güç = 0.17 olacaktır. Buradan görülüyor ki etki büyüklüğünün güç üzerindeki etkisi daha fazladır (Ellis, 2010, s. 85).

Cohen (1962, 1988 ve 1990), araştırmanın gücünü artırmak için kullanılan yöntemlerden birisinin örneklem sayısını artırmak olduğunu söylemiştir. Eğer α ve etki büyüklüğü sabit tutulursa, araştırmacı örneklem sayısını artırarak çalışmasının istatistiksel gücünü artırabilmektedir. Fakat bu yöntemi kullanmak için zaman, maliyet ve deneklerin ulaşılabilirliği gibi birçok caydırıcı faktör vardır. Bu nedenle araştırmacılar genellikle çalışmalarının istatistiksel gücünü artırmak için α 'yı veya etki büyüklüğünü kullanmalıdırlar.

Alfa seviyesi çalışmanın ihtiyacına göre ayarlanabilir. Tener'in (2000) aktardığına göre, Fraks ve Huck (1986), etki büyüklüğü ve örneklem sayısının sabit tutularak, alfa seviyesinin artırılmasıyla çalışmanın istatistiksel gücünün artırmanın mümkün olduğunu belirtmişlerdir. Eğer örneklem sayısı sınırlıysa veya Tip II hata yapma olasılığının, Tip I hata yapma olasılığından daha maliyetli olacağı durumlarda araştırmacı alfa seviyesini biraz daha yüksek seviyelerde belirleyebilir (0.10, 0.25 gibi). Keppel (1991), bu yöntemin araştırmacının anlamlılık seviyesini artırmadan yeterli istatistiksel gücü sağlayamadığı durumlarda makul bir seçenek olduğunu belirtmiştir (s. 32).

Tener'in (2000) aktardığına göre, Thomasve Nelson (1996), özellikle araştırma altındaki gruplar arasındaki farklılıklar belirlenmeye çalışılıyorsa, mevcut etkiyi artırmak için araştırmacıların başvurabileceği birkaç yöntemin olduğunu belirtmişlerdir. Bunlardan birisi araştırma altındaki deneklere daha güçlü işlemlerin uygulanması iken, diğer bir yöntem ise araştırma altındaki deneklere daha uzun süreli işlemlerin uygulanmasıdır. Yani, 8 haftalık bir işlem yerine 12 haftalık bir işlemin uygulanması, iki grup arasında ölçülebilecek farkları daha net bir şekilde ortaya çıkaracaktır. Bu şekilde elde edilecek daha büyük bir etki, araştırma sonucunda elde edilecek gücün artmasına katkı sağlayacaktır (s. 33-34).

1.15 Araştırmacıları İstatistiksel Gücü İhmal Etmeye İten Nedenler

Cohen'in 1962 yılındaki çalışmasından sonra güç analizi üzerine yapılan çalışmalar, yayımlanan kitaplar araştırmacıların bu konuya çok fazla önem vermediklerini göstermektedir. Ancak bunda sosyal bilimler ve davranış bilimleri alanlarındaki araştırmacıların matematik ve istatistik bilgilerinin zayıf olmasının etkisi çok fazladır. Fakat bu durum son yıllarda güç, etki büyüklüğü, güven aralıkları, anlamlılık testleri ile ilgili kitap sayısının artmasıyla değişmiştir. Bu kitaplardan bazıları; Aberson (2010), Murphy ve Myers (2009), Ellis (2010), Cumming (2012), Davey ve Salva (2010), Dattalo (2008) şeklindedir (Rossi, 2012, s. 16).

İstatistiksel gücün neden ihmal edildiğini açıklamak için birkaç neden öne sürülmektedir. Tener'in (2000) aktardığına göre, Kraemer ve Thienmann (1987), işin eğitimsel probleminden bahsetmektedir. Onlar, istatistiksel gücün matematiksel karışıklığının araştırmacılara öğretilmesinin zorluğundan bahsetmişler ve araştırmacılara genellikle anlamlılık seviyesi ile başa çıkmanın yollarının öğretilmediğini, ancak nadiren de olsa istatistiksel güç sorunu ile başa çıkmanın yollarının öğretilmediğini belirtmişlerdir. Kraemer ve Thienmann (1987), istatistiksel gücü dikkate almanın, araştırmanın tasarım aşamasının en önemli aşamalarından biri olduğunu ve bu sürecin araştırmanın özel bir bölümü olacak şekilde resmi bir eğitim ve deneyim gerektirdiğini belirtmektedirler. Gereken tüm bu eğitim ve deneyimin sadece bir kurs ile öğretilmesi kolay bir iş değildir (s. 46).

Sedlmeier ve Gigerenzer, istatistiksel gücü ihmal etme nedenlerinden birinin de H_0 hipotez testi konusundaki teori farklılıklarından kaynaklandığını belirtmişlerdir. Ronald Aymler Fisher'in teorisi, katı bir şekilde H_0 hipotezinin reddine dayanmakta ve genellikle anlamlılık seviyesi 0,05 olarak kurulmaktadır. Bu teori H_1 hipotezini dikkate almaz. Jerzy Neyman ve Egon Pearson's H_0 Hipotez Testi Teorisi ise H_1 hipotezinin de açıklanmasını içermekte ve istatistiksel gücü dikkate almaktadır. İşte bu iki temel teori arasındaki farklılığın neden olduğu kafa karışıklığı, istatistiksel gücün yıllarca araştırmacılar tarafından ihmal edilmesine neden olmuştur (Sedlmeier ve Gigerenzer, 1989).

Sedlmeier ve Gigerenzer, istatistiksel gücün önceleri ihmal edilmesinde tarihsel bir sebep olduğunu belirtmektedirler. Fisher'in istatistiksel teorisi 1935'li yıllarda geliştirilmiştir. Buna karşın Neyman ve Pearson istatistiksel teorisi ise II. Dünya Savaşı yıllarında ortaya atılmıştır. Bu nedenle tarihsel anlamda daha önceleri ortaya çıkmış olan ve istatistiksel gücü dikkate almayan Fisher modeline araştırmacılar, Neyman ve Pearson modelinden daha aşındır.

Daha sonraki yıllarda bu iki teorinin Hibrit Teori H_0 Hipotez Testi adı altında birleştirildiğini ve bunun sonucunda istatistiksel gücün ihmal edilmeye devam edildiğini belirtmişlerdir. Bu iki teorinin birleştirilmesinden sonra bir kavram karmaşası meydana gelmiş ve bunun doğrudan sonucu olarak istatistiksel güç ihmal edilmeye devam edilmiştir (Sedlmeier ve Gigerenzer, 1989)

İstatistiksel gücün ihmal edilme nedenlerinden biri de belki araştırmacıların örneklem sayısına karar vermelerinde etkili olan güç, etki büyüklüğü ve anlamlılık seviyesi gibi faktörlerin yanlış anlaşılmasından kaynaklanmaktadır. Yapılan bir araştırmada telefon ile ulaşılan yazarlara örneklem sayısına nasıl karar verdikleri konusunda açık uçlu sorular sorulmuştur. Bütçe kısıtlamaları, ulaşılabilecek mevcut veri sayısı örneklem sayısı belirlemede en çok kullanılan yöntemler olarak öne çıkmıştır. 28 yazardan sadece beşi örneklem sayısına istatistiksel kurallara göre veya beklenen etki büyüklüğüne göre karar verdiklerini açıklamışlardır. İstatistiksel gücün ihmal edilmesindeki bir başka neden de dergi editörlerinin güç konusunun çok da önemli olmadığını ve güç konusunun çalışmalarda kullanılmasının gerekmediğini düşünmeleridir (Tener, 2000, s. 48).

Sedlmeier ve Gigerenzer, 1989 yılında güç analizi ile ilgili çalışmaların ve kitapların sosyal bilimler alanındaki etkisini gözlemlemek amacıyla bir çalışma gerçekleştirmiş ve yaptıkları çalışmada 1962 yılında Cohen'in çalışmasından sonra herhangi bir olumlu artış olmadığı sonucuna ulaşmışlardır. Yapılan bu çalışmalardan sonra Rossi (1990) ise 1982 yılında yayımlanmış çalışmaların gücü ile ilgili bir araştırma yapmış ve Cohen'in 1962 yılındaki çalışmasında elde ettiği değerlerden biraz daha fazlasına ulaşmıştır. Fakat yine de bu artan değerler, olması gereken güç seviyesine ulaşmak için kayda değer bir artış değildir (Rossi, 1990, s. 650).

Yapılan tüm bu çalışmalardaki düşük gücün nedeni olarak Cohen, sosyal bilimler alanındaki düşük metodolojik gelişmeleri görmüştür. Bu durumun nedeni ise Cohen'e göre sosyal bilimler alanında yazılmış kitaplardaki hatalar ve yazındaki mevcut belirsizliklerdir. Daha önce de bahsedildiği gibi bu konudaki mevcut belirsizliklerin temel nedeni Sedlmeier ve Gigerenzer (1989)'e göre Fisher ve Neyman-Pearson Teorilerinin hibrit edilmesidir.

İstatistiksel gücün ihmal edilmesindeki bir başka olası neden ise araştırmacıların güç analizi yaparak zaman harcamak istememelerinden kaynaklanmaktadır. 1996 yılında yazarlar üzerinde yapılan bir araştırmada 169 yazardan 108 tanesi çalışmalarında hiç güç analizi yapmadıklarını belirtmişlerdir. İşin daha da vahim yanı çalışmalarında güç analizi yapmayan

108 yazardan %67'sinin güç analizinin ne olduğu konusunda hiçbir fikrinin olmamasıdır. Çoğu yazar ise güç analizinden haberdar olup, çalışmalarda kullanmadığını açıklamıştır (Tener, 2000, s. 49).

1.16 İstatistiksel Güç Hesaplamaya Yönelik Bilgisayar Programları

1.16.1 G-POWER: Genel Bir Güç Analizi Programı

Araştırmacılar, geçmişte araştırmaları için gerekli örneklem sayısına karar vermek için çok net sonuçlar vermeyen araçlar kullanmışlardır. Bunlar çoğu zaman araştırmacının bütçesine, zamanına, konuda uzman olan kişilerin tavsiyelerine vs. gibi yöntemler olmuştur. Ancak ne yazık ki araştırmacıların, güç analizi gibi etkili bir yöntemi, araştırma için veri toplama aşamasından önce uygulayarak gerekli örneklem sayısına karar verme imkânları varken çok mantıklı olmayan yollara başvurmuşlardır. Cohen, araştırmacıların güç analizini çok karmaşık bulup, ondan kolayca kaçmak isteyeceklerini belirtmiştir. İşte bu nedenle Cohen tarafından oluşturulan güç tabloları ve çoğu uzman tarafından oluşturulan bilgisayar programları araştırmacılara oldukça kolaylık sağlamıştır. Bahsi geçen programlardan birisi de G-Power 3.1.3 programıdır. G-Power, her türlü işletim sistemi için kullanılabilen ücretsiz bir güç analizi programıdır (Balkin ve Sheperis, 2011, s. 269). Bu program ücretsiz olarak <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register> internet sitesinden indirilebilmektedir.

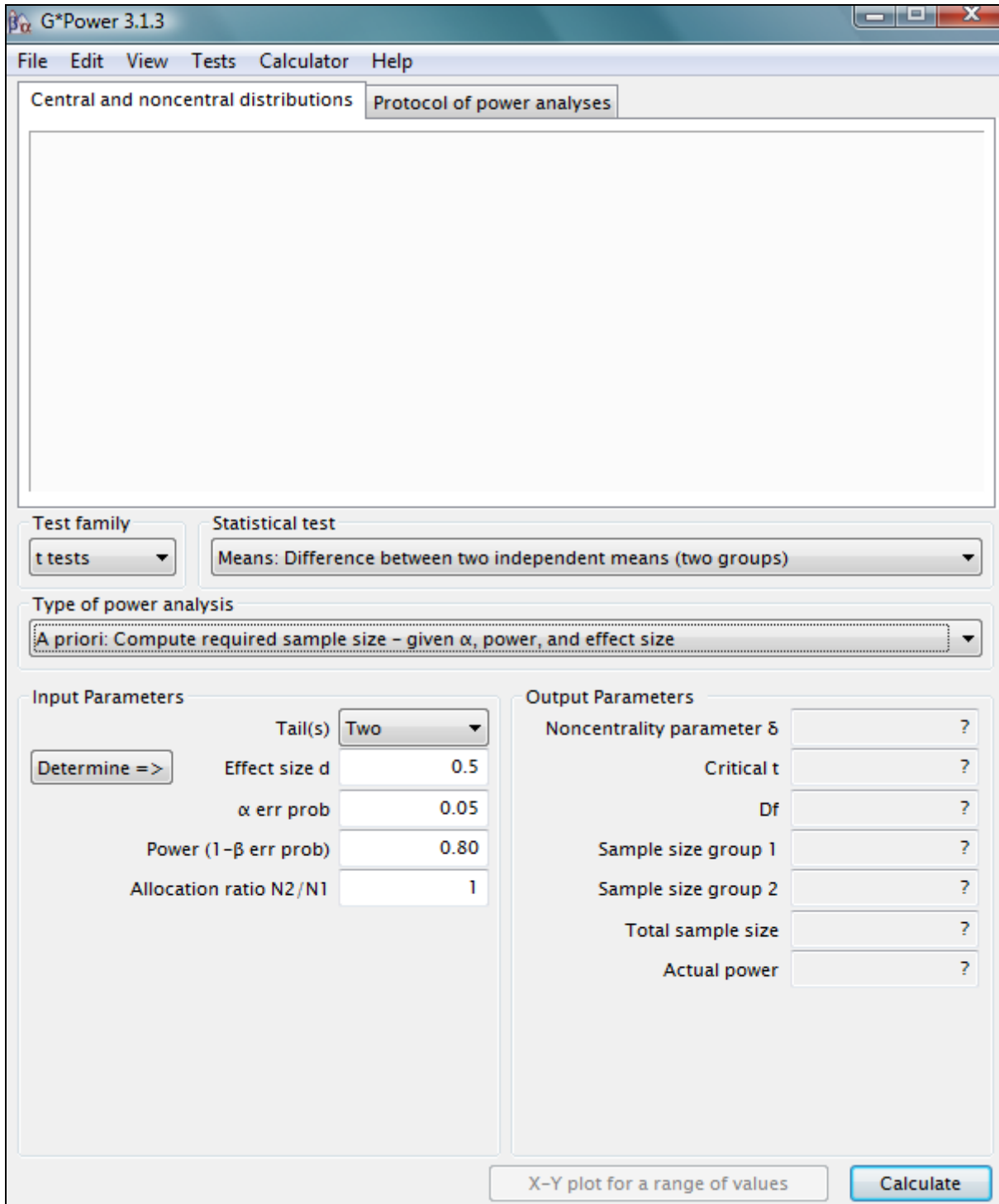
G-Power, tamamen interaktiftir ve menüler yardımı ile kullanılan bir programdır. Bu program sosyal bilimlerde kullanılan t-testi, F testi, ki kare testi gibi birçok yaygın istatistiksel testin istatistiksel güç analizlerinin hesaplanmasını yüksek kesinlik oranında gerçekleştirmektedir.

G-Power;

1. Verilen örneklem sayısı, etki büyüklüğü ve alfa seviyesi ile (post hoc güç analizi-deneysel güç analizi) bir istatistiksel testin güç değerini hesaplar.
2. Verilen etki büyüklüğü, alfa seviyesi ve güç değerleri ile (priori güç analizi-teorik güç analizi) gerekli örneklem sayısını hesaplar.
3. Verilen örneklem sayısı, etki büyüklüğü ve beta/alfa oranı ile (compromise güç analizi) gerekli alfa ve beta değerlerini hesaplar(Erdfelder, Faul, Buchner, 1996, s. 3).

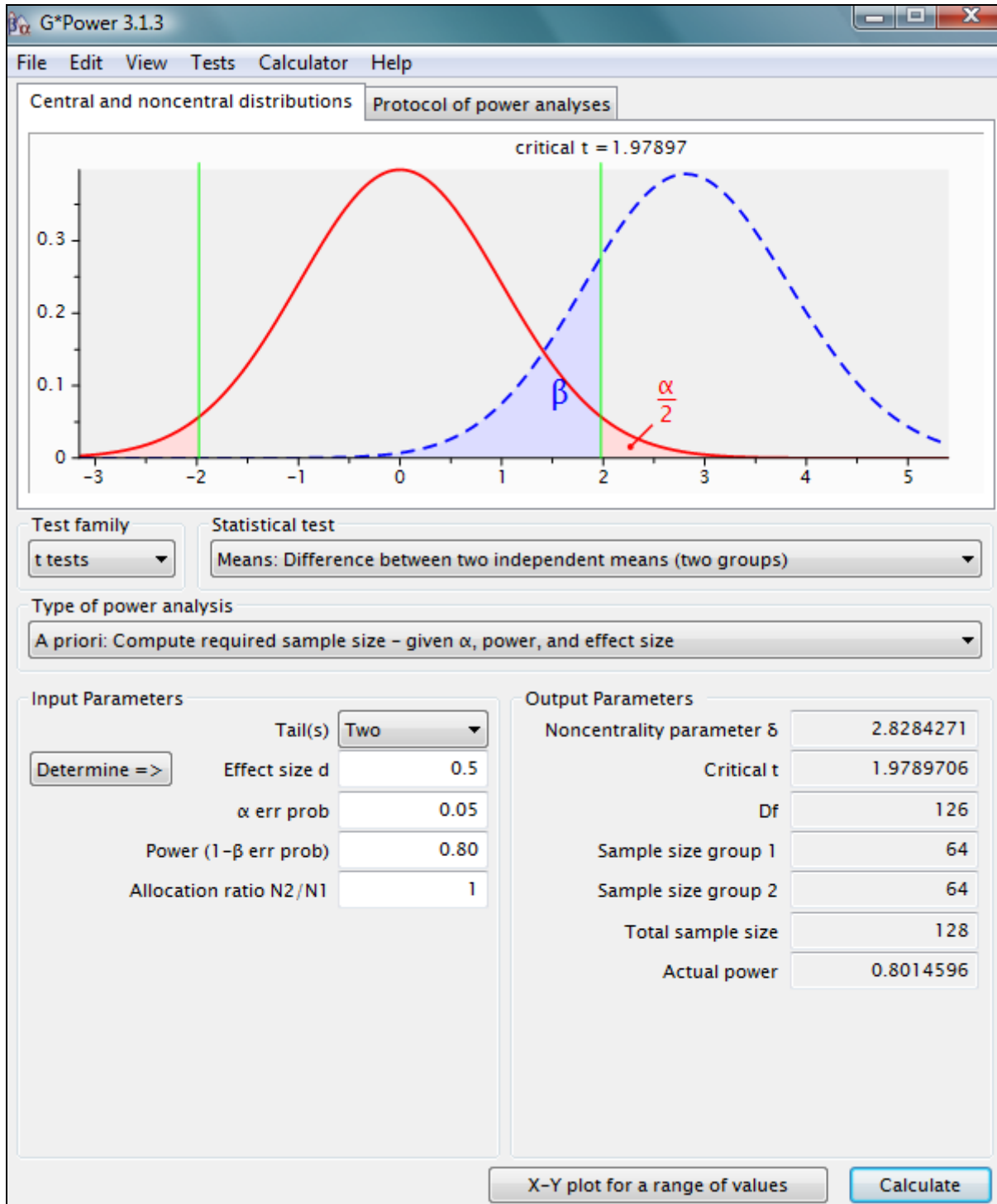
4. Verilen bir güç, alfa ve örneklem sayısı değerleri ile belirlenebilecek en küçük etkiyi hesaplar(sensitivity power analysis)

Eğer bir araştırmacı çalışması için etki büyüklüğü tahminini yapmış, arzu ettiği güç seviyesini ve anlamlılık seviyesini belirlemiş ise G-Power programını kullanarak ihtiyacı olan optimum örneklem sayısını hesaplayabilir. Bazı testler için bu değerlere ilaveten analizde kullanılacak grup sayısı, bağımsız değişken sayısı vs. gibi ilave bilgilere de ihtiyaç olabilir. Gerekli örneklem sayısını belirlemek için program açtıktan sonra sonra “*test family*” menüsü kullanılarak hangi tür test kullanılacak ise (t-testi, F testi, Z testi, X^2 vs. gibi) seçilir. Daha sonra “*statistical test*” menüsünden hangi test ailesi seçildiyse o test ailesine ait ilgili testlerden birisi seçilir. Son olarak ise “*type of power analysis*” menüsünden eğer örneklem sayısına karar verilecek ise “*A priori: Compute required sample size – given alpha, power and effect size*” seçilir. Daha önceden belirlenmiş olan alfa, etki büyüklüğü ve güç seviyeleri ekrandaki ilgili yerlere girilerek analiz gerçekleştirilir. Eğer çalışma sonucunda elde edilen güç seviyesi belirlenmek isteniyorsa “*type of power analysis*” menüsünden “*Post Hoc: Compute achieved power – given alpha, sample size and effect size*” seçilir ve “*compute*” kısmına tıklanarak analiz gerçekleştirilir (Balkin ve Sheperis, 2011, s. 270-271). Şekil 1.10’da değerler girilmiş ve analize hazır G-Power programından bir kesit verilmiştir.



Şekil 1.10 G-Power 3.1.3 Programından Güç Analizi Öncesi Bir Kesit

Şekil 1.11’de ise gerekli değerler girildikten sonra ihtiyaç olan örneklem sayısının gösterildiği program çıktısı verilmiştir.

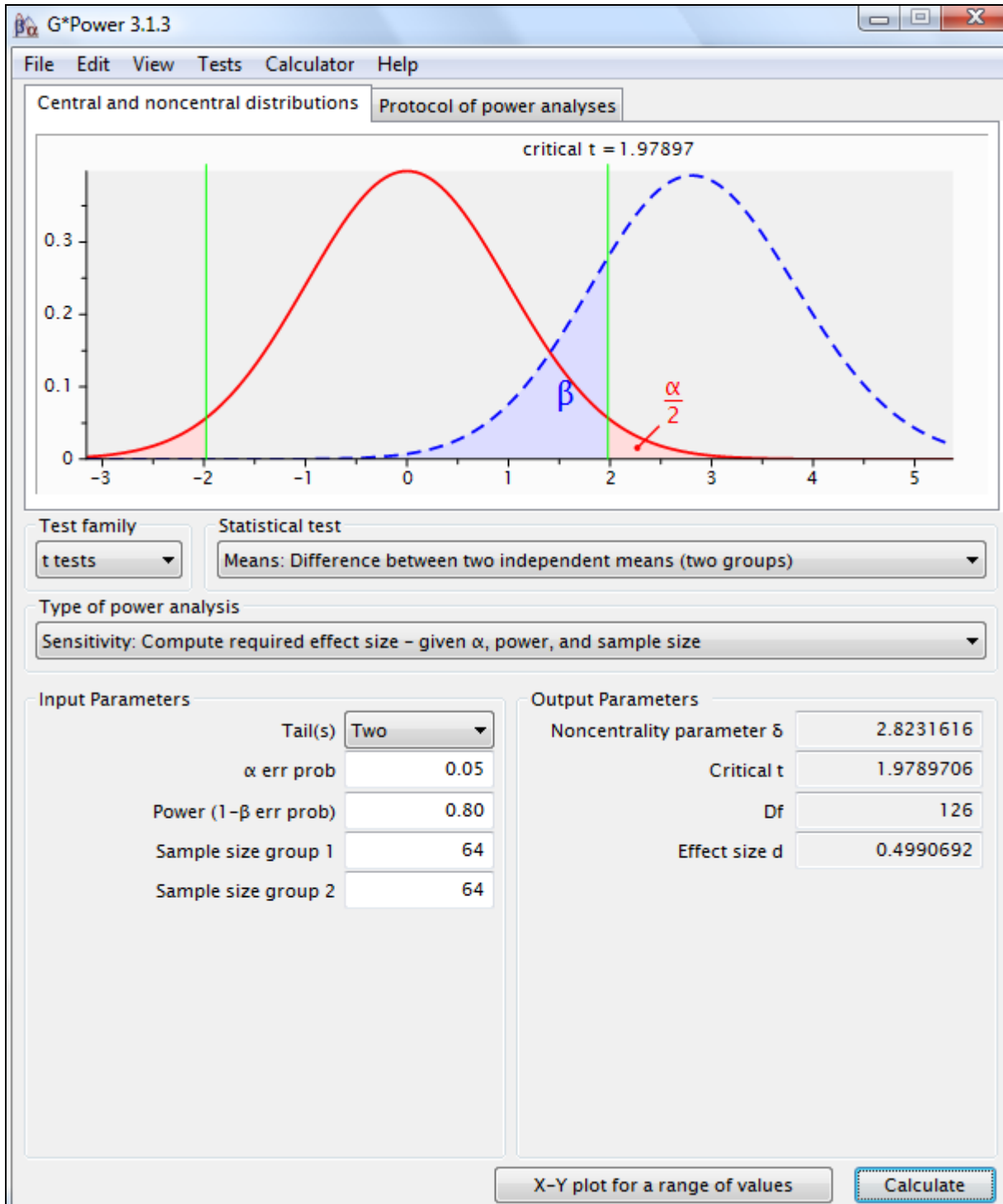


Şekil 1.11 G-Power 3.1.3 Programından Güç Analizi Sonrası Bir Kesit

Yapılan işlemler sonucunda her bir grup için gerekli olan örneklem sayısı 64 olarak, toplamda ise 128 olarak hesaplanmıştır. Program, üst kısımda ise kritik test değerinin ve sıfır hipotezi ret bölgelerinin gösterildiği grafiği çizmiştir.

G-Power programının bir diğer işlevi, verilen güç, alfa ve örneklem sayısı değerleri ile belirlenebilecek en küçük etkiyi hesaplamasıdır. Bu tür bir güç analizi “*sensitivity power*”

analysis” olarak adlandırılmaktadır. Bu analizi gerçekleştirmek için test ailesi ve kullanılacak istatistiksel test seçildikten sonra “*type of power analysis*” menüsünden “*Sensitivity: Compute required effect size – given alpha, power and sample size*” seçilir ve *compute* butonuna tıklanır. Gerçekleştirilen bu tür bir güç analizinin program çıktısı Şekil 1.12’de gösterilmiştir.



Şekil 1.12 G-Power 3.1.3 Programından Belirnelebilecek Etkiye Dair Bir Kesit

1.16.2 PASS ve STATISTICA

Son yıllarda internet üzerinden satın alınabilecek veya indirilebilecek bazı yeni güç analizi araçları geliştirilmiştir. Bunlardan bazıları PASS (Power Analysis and Sample Size) ve STATISTICA'dır. PASS, güç analizi yapmaya ve örneklem sayısı hesaplamaya yarayan bir programdır. Bu program t-testi, F testi, Z testi, çoklu regresyon, varyans analizi gibi çok geniş yelpazede istatistiksel analizler gerçekleştirmektedir.

STATISTICA ise kapsamlı, kullanıcı dostu, güç analizi ve örneklem sayısı hesaplama işlemini gerçekleştiren bir programdır (Deng, 2000,s. 45-46).

İKİNCİ BÖLÜM

İSTATİSTİKSEL GÜÇ BELİRLEMeye İLİŞKİN

SOSYAL BİLİMLER ALANINDA BİR UYGULAMA

2.1 Araştırmanın Amacı, Kapsamı ve Önemi

Bu araştırmanın asıl amacı, sosyal bilimler çalışmalarında istatistiksel güç analizi konusunun ne olduğu konusunda farkındalık oluşturarak ilgili yazına katkı sağlamaktır. Araştırmanın bir diğer amacı ise bir sosyal bilim çalışmasında istatistiksel güç analizinin nasıl kullanılacağı konusunda araştırmacılara rehberlik etmektir. Türkçe sosyal bilimler yazınında böyle bir çalışmaya rastlanılmadığı için bu çalışma ile yazındaki mevcut istatistiksel güç analizi bilgisine önemli ölçüde katkı yapılacağı düşünülmektedir.

Bu bağlamda “İnsan Kaynakları” alanı seçilmiş, YÖK’ün Ulusal Tez Merkezi’ne tezlerini kayıt eden yüksek lisans ve doktora öğrencilerinin istatistiksel güç analizi konusuna tezlerinde ne derece yer verdikleri irdelenerek bu konuda bir farkındalık oluşturmak amaçlanmaktadır.

Yapılan araştırma, Yükseköğretim Kurulu (YÖK) Ulusal Tez Merkezi’ne kayıtlı olan ve “İnsan Kaynakları” alanında başarıyla tamamlanmış olan yüksek lisans ve doktora tezlerini kapsamaktadır.

2.2 Araştırmanın Yöntemi

Bu bölümde araştırmanın örnekleme, veri toplama araçları ve veri analiz tekniklerinden bahsedilecektir.

2.2.1 Araştırmanın Örnekleme

Araştırmanın evrenini YÖK’ün Ulusal Tez Merkezi’ne kayıtlı, “İnsan Kaynakları” alanında 2005-2011 yılları arasında başarıyla tamamlanmış tüm yüksek lisans ve doktora tezleri oluşturmaktadır. YÖK’ün Ulusal Tez Merkezi’ne kayıtlı tezlere 2005 yılından itibaren elektronik ortamda erişim mümkün olduğundan 2005-2011 yılları arasındaki yüksek lisans ve doktora tezleri seçilmiştir. Erişme izni bulunan toplam 254 adet tez araştırmanın evrenini oluşturmaktadır. Ancak elde edilen 254 tezin istatistiksel güç analizi açısından incelenebilmesi için bazı kriterleri sağlaması gerekmektedir. Tezlerin araştırmaya dâhil edilip edilmeyeceğine karar vermek için aşağıdaki kriterler kullanılmıştır:

- i) Araştırmaya sadece herhangi bir istatistiksel anlamlılık testi uygulanan tezler dâhil edilmiştir. Herhangi bir istatistiksel anlamlılık testi uygulanmayan tezler araştırmaya dâhil edilmemiştir.
- ii) Sadece Cohen'in güç tablolarında ve G-Power programında yer alan testler ile anlamlılık testi gerçekleştiren tezler araştırmaya dâhil edilmiştir. Bu testler: Bağımsız Gruplar t-testi, Pearson Product Moment Korelasyon Katsayısı, Mann-Whitney U Testi, Kruskal Wallis Testi, Ki-Kare Testi, ANOVA, Çoklu Regresyon Analizi, Basit Regresyon Analizi ve Wilcoxon Testi' dir.
- iii) Eğer tezin yazarı araştırması için herhangi bir anlamlılık seviyesi belirlemişse istatistiksel gücü hesaplamak için o anlamlılık seviyesi kullanılmıştır.
- iv) Herhangi bir anlamlılık seviyesi belirtilmeyen tezlerde istatistiksel gücü hesaplamak için anlamlılık seviyesi 0.05 olarak alınmıştır.
- v) Eğer tez yazarı alternatif hipotezin tek yönlü mü yoksa iki yönlü mü olduğunu belirtmemişse, alternatif hipotezin iki yönlü olduğu varsayılmıştır.
- vi) Eğer tezde örneklem sayısı ile ilgili herhangi bir bilgiye yer verilmemişse o tez analize dâhil edilmemiştir.
- vii) Küçük, orta ve büyük seviyelerdeki etki büyüklükleri Cohen'in (1988) *Statistical Power Analysis for the Behavioral Sciences* isimli kitabından alınmıştır. Etki büyüklüğü tahminleri her bir istatistiksel test için farklılık göstermektedir.
- viii) Tezlerde kullanılan anlamlılık testlerinden Wilcoxon Testi, örneklem ve istatistiksel güç ortalamalarını aşırı değerde etkilediği için analizde dâhil edilmemiştir.

Yukarıda bahsi geçen kriterleri karşılamayan 161 adet yüksek lisans ve doktora tezi araştırmaya dâhil edilmemiştir. Kriterlere uyan 93 tez araştırmanın örneklemini oluşturmuştur. Tablo 2.1 örneklemin dağılımını göstermektedir.

Tablo 2.1 Araştırmaya Dâhil Edilen Tezlerin Yüksek Lisans ve Doktora Göre Dağılımı

TEZ	SAYI	%
YÜKSEK LİSANS	63	67,74
DOKTORA	30	32,26
TOPLAM	93	100,00

2.2.2 Veri Toplama Araçları

Araştırma kapsamında oluşturulan YÖK Ulusal Tez Merkezi'ne kayıtlı 2005-2011 yılları arasında başarıyla tamamlanmış olan yüksek lisans ve doktora tez listesi YÖK'ün resmi internet sitesi olan www.yok.gov.tr üzerinden elde edilmiştir. Güç analizi için araştırma yapmak üzere belirlenen tezlerin listeye eklenmesi konusunda tez yazarlarının tezlerini internet üzerinden erişime açma durumuna göre hareket edilmiştir.

2.2.3 Veri Analiz Teknikleri

Hesaplanan istatistiksel gücün doğruluğu, analiz edilmek üzere seçilen tezlerin yazarları tarafından gerçekleştirilen anlamlılık testlerinin doğruluğuna bağlıdır. Her bir istatistiksel test için belirlenen istatistiksel güç seviyesi doğrudan Jacop Cohen'in 1988 yılında yayımladığı "*Statistical Power Analysis for the Behavioral Sciences*" kitabında yer verdiği güç tablolarından ve G-Power 3.1.3 programından yararlanılarak oluşturulmuştur. Elde edilen verilerin düzenlemesi için Microsoft Office 2007 Excell programından yararlanılmıştır. Fakat bu tablolar, araştırmanın tasarımı, araştırmanın doğası, bağımlı ve bağımsız değişkenin seçilmesi, denekler arasındaki çeşitlilik, test edilen gruplar arasındaki varyans gibi bir istatistiksel testin gücüne karar verme sürecinde önemli rol oynayan etkenleri yansıtmamaktadır. Bu nedenle, Cohen'in kitabında yer verdiği güç tablolarına göre hesaplanan güç seviyeleri sadece birer tahmindir ve bu değerler, analize dâhil edilen tezlerin veya analiz edilen istatistiksel testlerin gerçek gücü olarak ifade edilemez.

İstatistiksel güç analizini gerçekleştirebilmek için öncelikle her bir yüksek lisans ve doktora tezi için kayıt edilen örneklem sayısı, anlamlılık seviyesi alternatif hipotezin yönü kayıt edilmiştir. Daha sonra kayıt edilen bu değerler kullanılarak Cohen'in tanımlamış olduğu küçük, orta ve büyük seviyedeki etki büyüklükleri için her bir anlamlılık testine post-hoc

(deneysel) istatistiksel güç analizi uygulanmıştır. İstatistiksel güç değerleri doğrudan Cohen'in güç tablolarından ve G-Power 3.1.3 programından elde edilmiştir. Böylece her bir yüksek lisans ve doktora tezi için istatistiksel güç seviyeleri hesaplanmıştır. Daha sonra tez yazarlarının güç konusuna olan genel yaklaşımlarını belirlemek adına istatistiksel güç seviyelerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için ortalama bir değeri hesaplanmıştır.

Bu çalışma, anlamlılık testlerini gerçekleştiren ve güç analizi için gerekli olan verileri sağlayan tezler ile sınırlıdır. Netice olarak, çok sayıda yüksek lisans ve doktora tezine ulaşılmıştır. Fakat yapılan analizler sonucunda elde edilen bulgular daha geniş anakütlelere genelleştirilemez.

2.2.4 Araştırmanın Bulguları ve Yorumları

Araştırmaya dâhil edilen 93 adet yüksek lisans ve doktora tezinde toplamda 2970 adet istatistiksel anlamlılık testi uygulanmıştır. Ancak parametrik olmayan Kruskal Wallis testinin hem Cohen'in güç tablolarında hem de G-Power 3.1.3 programında olmaması nedeniyle bu testi kullanan testlerin istatistiksel güçleri, onun parametrik karşılığı olan ANOVA testi uygulanmış gibi düşünülerek hesaplanmıştır. İstatistiksel anlamlılık testlerinin dağılımı Tablo 2.2' de gösterilmiştir.

Tablo 2.2 İstatistiksel Anlamlılık Testlerinin Dağılımı

İstatistiksel Anlamlılık Testi	Frekans	%
ANOVA	1099	37,00
Bağımsız Gruplar t-Testi	617	20,77
Çoklu Regresyon Analizi	281	9,46
Basit Regresyon Analizi	45	1,52
Kruskal Wallis Testi	128	4,31
Mann Whitney U Testi	151	5,08
Pearson Korelasyon Analizi	162	5,45
Ki Kare Testi	87	16,40
Toplam	2970	100,00

Tablo 2.2'den de görüldüğü gibi çalışmaya dâhil edilen yüksek lisans ve doktora tezleri içinde en fazla kullanılan istatistiksel anlamlılık testi ANOVA (%37)'dir. Bir sonraki en fazla kullanılan test ise Bağımsız Gruplar t-Testi (%20,77)'dir.

Cohen (1988), eğer araştırmacının elinde istatistiksel güç seviyesinin ne olması ile ilgili bir dayanağı yoksa bu seviyenin en az %80 olması gerektiğini belirtmiştir. Eğer dikkate alınması gereken başka önemli nedenler yoksa bu tablolarda önerilen optimum örneklem sayısı araştırmacıya en uygun örneklem sayısına karar verme konusunda yardımcı olmaktadır. Aksi takdirde araştırmacı bulmayı umduğu etkiyi ortaya çıkarmak için gerekli olan örneklemden çok daha az sayıda örneklem kullanabilmekte veya gerekenden çok daha fazla örneklem kullanarak kaynaklarını israf etmiş olabilmektedir. Araştırmaya dahil edilen tezlerde kullanılan ortalama örneklem sayısı ve orta seviyedeki etki büyüklüğünü belirleyebilmek için her bir istatistiksel anlamlılık testi için gereken optimum örneklem sayıları Tablo 2.3'te verilmiştir.

Tablo 2.3 Araştırmaya Dâhil Edilen Tezlerde Kullanılan Gerçek Örneklem Sayıları ve Optimum Örneklem Sayıları

	Test Sayısı	Ortalama
Örneklem Sayısı	2970	173,15
Optimum Örneklem Sayısı	2970	75,62

Tablo 2.3'ten açıkça görülmektedir ki optimum örneklem sayısı 75,62 olmasına rağmen tezlerde kullanılan gerçek örneklem sayısı 173,15'tür, yani optimum örneklem sayısının yaklaşık 2,3 katıdır.

Tezlerde kullanılan gerçek örneklem sayılarının ve optimum örneklem sayılarının ortalamalarının test başına dağılımları Tablo 2.4'te gösterilmiştir.

Tablo 2.4 Gerçek ve Optimum Örneklem Sayılarının Test Başına Dağılımları

İstatistiksel Anlamlılık Testi	Frekans	Ortalama Örneklem Sayısı	Optimum Örneklem Sayısı
ANOVA	1099	184,44	64
Bağımsız Gruplar t- Testi	617	206,06	64
Çoklu Regresyon Analizi	281	180,13	85
Basit Regresyon Analizi	45	205,12	55
Kruskal Wallis Testi	128	202,54	64
Mann Whitney U Testi	151	158,66	67
Pearson Korelasyon Analizi	162	96,04	84
Ki Kare Testi	487	152,26	122
Toplam/Ortalama	2970	173,15	75,62

Tezlerde kullanılan örneklem sayılarının ortalamasına (173,15) en yakın olan istatistiksel anlamlılık testi ortalama 180,13 örneklem sayısı ile Çoklu Regresyon Analizi'dir. Ortalama örneklem sayısına en uzak ortalamaya sahip olan test ise Pearson Korelasyon Analizi (%96,04)'dir.

2.2.4.1 İstatistiksel Güç Analizi

Araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için hesaplanan istatistiksel güçleri Tablo 2.5'te gösterilmiştir.

Tablo 2.5 Araştırmaya Dâhil Edilen Tezlerin Küçük, Orta ve Büyük Seviyedeki Etki Büyüklüklerine Göre Hesaplanan Ortalama İstatistiksel Güçleri

	Küçük Etki	Orta Etki	Büyük Etki
Ortalama İstatistiksel Güç	0,23	0,72	0,92

Ortalama istatistiksel güç seviyelerinin frekans dağılımı ise Tablo 2.6'da gösterilmiştir.

Tablo 2.6 Tezlerde Uygulanan İstatistiksel Anlamlılık Testlerinin Güç Ortalamalarının Frekans ve Kümülatif Yüzdeleri

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	474	15,96	62	2,09	-	-
0,10 - 0,19	1051	51,35	53	3,87	63	2,12
0,20 - 0,29	306	61,65	125	8,08	-	2,12
0,30 - 0,39	471	77,51	174	13,94	52	3,87
0,40 - 0,49	374	90,10	83	16,73	1	3,91
0,50 - 0,59	42	94,39	273	25,93	25	4,75
0,60 - 0,69	94	91,52	240	34,01	18	5,35
0,70 - 0,79	106	94,68	441	48,86	270	14,44
0,80 - 0,89	51	98,28	132	53,30	109	18,11
0,90 - 0,99	1	100,00	1386	99,97	1603	72,09
1,00	-	100,00	1	0,03	829	100,00
Toplam	2970	---	2970	---	2970	---

Tablo 2.6'dan da görüldüğü gibi, yüksek lisans ve doktora tezlerinin %94,68'inin istatistiksel gücü küçük bir etkiyi ortaya çıkarabilmek için Cohen'in tavsiye ettiği istatistiksel güç seviyesi olan %80'den daha aşağıdadır. Buna karşılık, tezlerin %48,86'sı orta seviyedeki bir etkiyi ortaya çıkarabilmek için gerekli %80 istatistiksel güç seviyesinden daha aşağıda iken, tezlerin %14,44'ü büyük bir etkiyi ortaya çıkarabilmek için gerekli istatistiksel güce sahip değildir.

Ayrıca araştırmaya dâhil edilen 93 adet tezin her birindeki istatistiksel anlamlılık testlerinin istatistiksel güçleri ayrı ayrı hesaplanmıştır. Bu sonuçlar ise Tablo 2.7'de gösterilmektedir.

Tablo 2.7 Her Bir İstatistiksel Anlamlılık Testi İçin Ortalama İstatistiksel Güç Seviyesi

İstatistiksel Anlamlılık Testi	Frekans	Küçük Etki	Orta Etki	Büyük Etki
ANOVA	1099	0,17	0,63	0,87
Bağımsız Gruplar t- Testi	617	0,29	0,78	0,94
Çoklu Regresyon Analizi	281	0,26	0,87	0,95
Basit Regresyon Analizi	45	0,43	0,91	0,99
Kruskal Wallis Testi	128	0,19	0,67	0,98
Mann Whitney U Testi	151	0,23	0,68	0,86
Pearson Korelasyon Analizi	162	0,16	0,75	0,97
Ki Kare Testi	487	0,12	0,62	0,89
Toplam/Ortalama	2970	0,23	0,72	0,92

Bahsi geçen anlamlılık testlerinden Mann-Whitney U Testi, Kruskal Wallis Testi ve Ki Kare Testi parametrik olmayan istatistiksel testlerdir. Cohen'in 1962 yılında gerçekleştirdiği ilk sistematik istatistiksel güç analizi çalışmasındaki tavsiyesine göre Mann-Whitney U Testi, Bağımsız Gruplar t-Testi ile, Kruskal Wallis Testi ise Anova ile aynı özellikleri gösteren birbirine çok yakın testlerdir. Cohen (1962), kendi çalışmasında bu parametrik olmayan testlerin yerine onların parametrik karşılıkları olan t-testi ve Anova'yı kullanmıştır. Ancak G-Power 3.1.3 programının parametrik olmayan testlerden Mann Whitney U testinin istatistiksel gücünü hesaplamaya imkân vermesi nedeniyle bu çalışmada parametrik olmayan Mann Whitney U testlerinin parametrik karşılığı olan t-testi kullanılmamış, her bir testin kendi istatistiksel gücü ayrı ayrı hesaplanmıştır. Ancak Kruskal Wallis testinin G-Power 3.1.3. programında da olmaması nedeniyle bu test yerine onun parametrik karşılığı olan ANOVA kullanılmıştır.

Tablo 2.7'den de görüldüğü gibi küçük, orta ve büyük seviyedeki etkileri belirleyebilmek için en iyi ortalama güç seviyesine sahip olan istatistiksel test sırasıyla 0,43, 0,91 ve 0,99 değerleri ile basit regresyon analizidir. En düşük güç ortalamasına sahip olan istatistiksel test ise sırasıyla 0.12, 0.62 ve 0.89 değerleri ile ki- kare testidir.

Tablo 2.8’de arařtırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan ANOVA testlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.8 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin ANOVA Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	265	24,11	62	5,64	-	-
0,10 - 0,19	321	53,32	14	6,92	6	0,55
0,20 - 0,29	46	57,51	18	8,55	-	0,55
0,30 - 0,39	228	78,25	159	23,02	30	3,28
0,40 - 0,49	175	94,18	17	16,73	-	3,28
0,50 - 0,59	-	94,18	98	24,57	73	9,92
0,60 - 0,69	64	100,00	150	33,48	117	20,56
0,70 - 0,79	-	100,00	68	47,13	25	22,84
0,80 - 0,89	-	100,00	4	53,32	94	31,39
0,90 - 0,99	-	100,00	509	53,69	733	98,09
1,00	-	100,00	-	100,00	21	100,00
Toplam	1099	---	1099	---	1099	---

Tablo 2.8’den de görüldüğü gibi arařtırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan ANOVA testlerinin hiçbirisi küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %47,13’ü orta seviyedeki bir etkiyi, %22,84’ü ise büyük seviyedeki bir etkiyi belirlemek için gerekli %80 güç seviyesi kriterini karşılayamamaktadır.

Tablo 2.9’da arařtırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan bağımsız gruplar t-testlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.9 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Bağımsız Gruplar t-Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	1	0,16	-	-	-	-
0,10 - 0,19	163	26,58	-	-	-	-
0,20 - 0,29	79	39,38	-	-	-	-
0,30 - 0,39	28	43,92	162	26,26	-	-
0,40 - 0,49	153	68,72	34	31,77	-	-
0,50 - 0,59	37	74,72	4	32,41	-	-
0,60 - 0,69	24	8,61	42	39,22	-	-
0,70 - 0,79	87	92,71	94	54,46	23	3,73
0,80 - 0,89	45	100,00	106	71,64	144	27,07
0,90 - 0,99	-	100,00	175	100,00	324	79,58
1,00	-	100,00	-	100,00	126	100,00
Toplam	617	---	617	---	617	---

Tablo 2.9'dan görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan bağımsız gruplar t-testlerinin %92,71'i küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %54,46'sı orta seviyedeki bir etkiyi, %3,73'ü ise büyük seviyedeki bir etkiyi belirlemek için gerekli %80 güç seviyesi kriterini karşılayamamaktadır.

Tablo 2.10'da araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan çoklu regresyon analizlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.10 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Çoklu Regresyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	22	7,83	4	1,42	-	-
0,10 - 0,19	35	20,28	-	1,42	-	-
0,20 - 0,29	122	63,70	-	1,42	-	-
0,30 - 0,39	57	83,99	-	1,42	4	1,42
0,40 - 0,49	34	96,09	18	7,83	-	1,42
0,50 - 0,59	5	97,86	-	7,83	-	1,42
0,60 - 0,69	-	97,86	-	7,83	-	1,42
0,70 - 0,79	-	97,86	20	14,95	-	1,42
0,80 - 0,89	-	100,00	5	16,73	18	7,83
0,90 - 0,99	-	100,00	228	97,86	150	61,21
1,00	-	100,00	6	100,00	109	100,00
Toplam	281	---	281	---	281	---

Tablo 2.10'dan görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan çoklu regresyon analizlerinin %97,86'sı küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %14,95'i orta seviyedeki bir etkiyi, %1,42'si ise büyük seviyedeki bir etkiyi belirlemek için gerekli %80 güç seviyesi kriterini karşılayamamaktadır.

Tablo 2.11'de araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan basit regresyon analizlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.11 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Basit Regresyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	-	-	-	-	-	-
0,10 - 0,19	22	48,89	-	-	-	-
0,20 - 0,29	4	57,78	-	-	-	-
0,30 - 0,39	5	68,89	-	-	-	-
0,40 - 0,49	-	68,89	-	-	-	-
0,50 - 0,59	-	68,89	-	-	-	-
0,60 - 0,69	6	82,22	-	-	-	-
0,70 - 0,79	7	97,78	6	13,33	-	-
0,80 - 0,89	-	97,78	25	68,89	-	-
0,90 - 0,99	1	100,00	14	100,00	26	57,78
1,00	-	100,00	-	100,00	19	100,00
Toplam	45	---	45	---	45	---

Tablo 2.11’den görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan basit regresyon analizlerinin %97,78’i küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %13,33’ü orta seviyedeki bir etkiyi belirlemek için gerekli %80 güç seviyesi kriterini karşılayamamaktadır. Tezlerde uygulanan tüm basit regresyon analizleri, büyük seviyedeki bir etkiyi belirleyebilmek için ise gereken %80 istatistiksel güç seviyesinden daha fazlası ile gerçekleştirilmişlerdir.

Tablo 2.12’de araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan Kruskal Wallis testlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.12 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Kruskal Wallis Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	5	3,91	-	-	-	-
0,10 - 0,19	64	53,91	1	0,78	-	-
0,20 - 0,29	13	64,06	4	3,91	-	-
0,30 - 0,39	34	90,63	-	3,91	1	0,78
0,40 - 0,49	12	100,00	-	3,91	4	3,91
0,50 - 0,59	-	100,00	62	52,34	-	3,91
0,60 - 0,69	-	100,00	2	53,91	-	3,91
0,70 - 0,79	-	100,00	-	53,91	-	3,91
0,80 - 0,89	-	100,00	-	53,91	-	3,91
0,90 - 0,99	-	100,00	59	100,00	111	90,63
1,00	-	100,00	-	100,00	12	100,00
Toplam	128	---	128	---	128	---

Araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan Kruskal Wallis testlerinin hiçbiri küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %53,91'i orta seviyedeki bir etkiyi, %3,91'i ise büyük seviyedeki bir etkiyi belirleyebilmek için gereken %80 istatistiksel güç seviyesi kriterini karşılayamamaktadır.

Tablo 2.13'te araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan Mann-Whitney U testlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.13 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Mann-Whitney U Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	32	21,19	-	-	-	-
0,10 - 0,19	65	64,24	30	19,87	-	-
0,20 - 0,29	5	67,55	11	27,15	-	-
0,30 - 0,39	34	92,05	1	27,81	30	19,87
0,40 - 0,49	-	92,05	-	27,81	-	19,87
0,50 - 0,59	-	92,05	10	34,44	10	26,49
0,60 - 0,69	-	92,05	15	44,37	1	27,15
0,70 - 0,79	12	100,00	30	64,24	1	27,81
0,80 - 0,89	-	100,00	5	67,55	10	34,44
0,90 - 0,99	-	100,00	49	100,00	86	91,39
1,00	-	100,00	-	100,00	13	100,00
Toplam	151	---	151	---	151	---

Tablo 2.13'ten görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan Mann-Whitney U testlerinin hiçbiri küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %64,24'ü orta seviyedeki bir etkiyi, %27,81'i ise büyük seviyedeki bir etkiyi belirleyebilmek için gereken %80 istatistiksel güç seviyesi kriterini karşılayamamaktadır.

Tablo 2.14'ten araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan Pearson korelasyon analizlerinin hiçbirinin küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamadığı görülmektedir.

Tablo 2.14 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Pearson Korelasyon Analizlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	21	12,96	-	-	-	-
0,10 - 0,19	92	69,75	-	-	-	-
0,20 - 0,29	37	92,59	-	-	-	-
0,30 - 0,39	12	100,00	8	4,94	-	-
0,40 - 0,49	-	100,00	3	6,79	-	-
0,50 - 0,59	-	100,00	63	45,68	-	-
0,60 - 0,69	-	100,00	4	48,15	-	-
0,70 - 0,79	-	100,00	10	54,32	-	-
0,80 - 0,89	-	100,00	28	71,60	8	4,94
0,90 - 0,99	-	100,00	46	100,00	150	97,53
1,00	-	100,00	-	100,00	4	100,00
Toplam	162	---	162	---	162	---

Buna karşılık, korelasyon analizlerinin %54,32'lik kısmı orta seviyedeki bir etkiyi belirlemek için gerekli olan istatistiksel güçten yoksun iken analizlerin tamamı büyük seviyedeki etki büyüklüklerini belirleyebilecek yeterli istatistiksel güce sahiptir.

Tablo 2.15'te araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde kullanılan Ki-kare testlerinin istatistiksel güçlerinin küçük, orta ve büyük seviyedeki etki büyüklükleri için frekans dağılımları gösterilmiştir.

Tablo 2.15 Küçük, Orta ve Büyük Seviyedeki Etki Büyüklükleri İçin Ki-kare Testlerinin İstatistiksel Güçlerinin Frekans Dağılımı

Güç	Küçük Etki		Orta Etki		Büyük Etki	
	f	Kümülatif %	f	Kümülatif %	f	Kümülatif %
0,00 - 0,09	128	26,28	-	-	-	-
0,10 - 0,19	289	85,63	4	0,82	-	-
0,20 - 0,29	-	85,63	92	19,71	-	-
0,30 - 0,39	70	100,00	-	19,71	4	0,82
0,40 - 0,49	-	100,00	7	21,15	-	0,82
0,50 - 0,59	-	100,00	-	21,15	4	1,64
0,60 - 0,69	-	100,00	52	31,83	-	1,64
0,70 - 0,79	-	100,00	249	82,96	88	19,71
0,80 - 0,89	-	100,00	12	85,42	-	19,71
0,90 - 0,99	-	100,00	71	100,00	321	85,63
1,00	-	100,00	-	100,00	70	100,00
Toplam	487	---	487	---	487	---

Tablo 2.15'ten görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan ki-kare testlerinin hiçbiri küçük bir etkiyi belirleyebilmek için gereken %80 güç seviyesi kriterini karşılayamamaktadır. Buna karşılık, testlerin %82,96'sı orta seviyedeki bir etkiyi, %19,71'i ise büyük seviyedeki bir etkiyi belirleyebilmek için gereken %80 istatistiksel güç seviyesi kriterini karşılayamamaktadır.

Tablo 1.11'de 58 adet farklı dergide yayımlanan makaleler için 1960 - 2004 yılları arasında gerçekleştirilen 25 farklı istatistiksel güç analizi çalışması sonucunda küçük, orta ve büyük seviyedeki etki büyüklükleri için ortalama istatistiksel güç seviyelerinin sırasıyla 0.24, 0.64 ve 0.86 olduğu görülmektedir. Bu değerler, 2005-2011 yılları arasında insan kaynakları alanında başarıyla tamamlanmış yüksek lisans ve doktora tezlerinin sırasıyla küçük, orta ve büyük seviyedeki etki büyüklükleri için sahip oldukları ortalama güç seviyeleri olan 0.23, 0.72 ve 0.92 ile karşılaştırıldığında, küçük seviyedeki etki büyüklüğü dışında diğer güç seviyelerinin daha aşağıda olduğu görülmektedir.

2.3 İstatistiksel Güç ile Anlamlılık Testlerinden Elde Edilen Sonuçlar Arasındaki İlişki

Bir çalışmanın istatistiksel gücünün düşük olması o çalışmadaki yanlış bir sıfır hipotezinin doğru bir şekilde reddedilme şansının düşük olması anlamına gelmektedir. Bu kısımda, araştırmaya dâhil edilen 93 adet yüksek lisans ve doktora tezinde uygulanan toplam 2970 adet istatistiksel anlamlılık testinin anlamlı sonuç verip vermemesine göre bir dağılımı verilmiştir. Tablo 2.16, bu dağılımı göstermektedir.

Tablo 2.16 İstatistiksel Anlamlılık Testlerinin Sonuç Dağılımları

Anlamlılık Testi	Anlamlı Sonuç Sayısı	Anlamsız Sonuç Sayısı	Toplam
ANOVA	359	740	1099
Bağımsız Gruplar t-Testi	309	308	617
Çoklu Regresyon Analizi	143	138	281
Basit Regresyon Analizi	44	1	45
Korelasyon Analizi	110	52	162
Ki Kare Testi	231	256	487
Mann-Whitney U Testi	41	110	151
Kruskal Wallis Testi	48	80	128
Toplam	1285	1685	2970

Tablo 2.16'dan da görüldüğü gibi araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinde uygulanan toplam 2970 adet anlamlılık testinden 1685'i (%56,73) anlamsız olarak sonuçlanırken, tezlerin 1285'i (%43,27) anlamlı olarak sonuçlanmıştır. Anlamsız sonuç elde edilen bir çalışmada, sonuçların kesinlik kazanması için istatistiksel gücün en az %80 seviyesinde olması gerekmektedir. Aksi halde çalışma kesinlikle “sonuçlar anlamsızdır” veya “ilişki yoktur” şeklinde sonlandırılmamalıdır. Düşük güce sahip olan çalışmalar, yeterli güç elde edilip mutlaka tekrarlanmalıdır.

Tablo 2.16’da dikkat çeken bir nokta da en fazla anlamlı sonuç veren anlamlılık testinin basit regresyon analizi olmasıdır. Çünkü, Tablo 2.7’de verilen ortalama istatistiksel güç seviyesi en yüksek olan test basit regresyon analizidir. İstatistiksel güç seviyesi en yüksek olan basit regresyon analizinin en fazla anlamlı sonuç veren test olması tesadüf ile açıklanacak bir durum değildir.

SONUÇ

İlgili yazında, istatistiksel güç analizini içeren bilimsel çalışmalar sınırlı sayıdadır ve bu az sayıdaki bilimsel çalışmalarda kullanılan istatistiksel testlerde güç analizinin ihmal edildiği anlaşılmaktadır. Bu bağlamda bu tez çalışması istatistiksel güç analizinin önemine dikkat çekmekte, istatistiksel güç analizi sürecine odaklanmakta ve sosyal bilimlerde istatistiksel güç analizinin nasıl yapılacağı uygulamalı olarak anlatılmaktadır. Uygulama için YÖK'ün Ulusal Tez Merkezi'ne kayıtlı ve "İnsan Kaynakları" alanında 2005 – 2011 yılları arasında başarıyla tamamlanmış olan ve erişimine izin verilen 254 adet yüksek lisans ve doktora tezine ulaşılmıştır. Ancak gerekli kriterleri karşılayan 93 adet tez araştırmaya dâhil edilmiştir. Araştırmaya dâhil edilen 93 adet tezin hiçbirinde istatistiksel güçten, II. Tip hatadan ve etki büyüklüğünden bahsedilmemiştir. Ayrıca tezlerin hiçbirinde örneklem sayılarına nasıl karar verildiğine dair herhangi bir formül veya tabloya yer verilmemiştir.

Araştırmaya dâhil edilen yüksek lisans ve doktora tezlerinin istatistiksel güçleri sırasıyla küçük, orta ve büyük seviyedeki etki büyüklükleri için 0.23, 0.72 ve 0.92 şeklinde bulunmuştur. Eğer bir tez yazarı, tezinde küçük bir etkiyi belirlemek isteseydi, yanlış bir sıfır hipotezini doğru bir şekilde reddedebilmek için sadece %23 oranında şansa sahip olacaktı. Aynı yazarın orta seviyedeki bir etkiyi belirlemek için elde edeceği şans %72, büyük seviyedeki bir etkiyi belirlemek için elde edeceği şans ise %92 olacaktı. İstatistiksel güç değerleri, büyük seviyedeki etkileri ortaya çıkarma durumu dışında Cohen'in tavsiye ettiği %80 seviyesinden daha aşağıdadır.

Tezlerde uygulanan 2970 adet anlamlılık testinden 1685'i (%56,73) anlamsız sonuçlar verirken, 1285'i (%43,27) anlamlı sonuçlar vermiştir. Ortalama istatistiksel gücün düşük olması nedeniyle anlamsız sonuçlar elde edilen çalışmaların, yeterli istatistiksel güç sağlandıktan sonra mutlaka tekrar edilmeleri gerekmektedir. Buna karşılık elde edilen değerler çeşitli alanlarda gerçekleştirilmiş istatistiksel güç analizlerinden elde edilen değerlerden nispeten daha yüksek çıkmıştır.

Tezlerin istatistiksel güç seviyelerinin ortalaması, yazında yer alan çalışmalar ile kıyaslandığında nispeten daha yüksektir. Ancak 1962–2004 yılları arasında yapılan çalışmaların istatistiksel güçlerinin ortalamaları düşük olmasına rağmen son yıllarda yapılan çalışmalarda ortalama güç seviyelerinin yükseldiği açıkça görülmektedir. Örneğin, sadece 2001–2004 yılları arasında 8 dergide yayımlanan makaleler üzerinde gerçekleştirilen 5 adet

istatistiksel güç analizi çalışmasında küçük, orta ve büyük seviyedeki etki büyüklükleri için ortalama istatistiksel güçler sırasıyla 0.32, 0.75 ve 0.93 şeklindedir. Bu durum Cohen (1982)'in ilk sistematik istatistiksel güç analizi çalışmasını gerçekleştirdiği 1962 yılından sonra istatistiksel güç konusunun o yıllarda tam anlamıyla anlaşılabilmesi ile açıklanabilir. Son zamanlarda gerçekleştirilen istatistiksel güç analizi çalışmalarında ortalama istatistiksel güç seviyelerinin daha yüksek olması, konunun gittikçe daha da dikkat çektiğinin ve nispeten istatistiksel güç seviyelerinin tatmin edici bir noktaya geldiğinin bir göstergesi olarak görülebilir.

Araştırmaya dâhil edilen 93 adet yüksek lisans ve doktora tezinde toplamda 2970 adet istatistiksel anlamlılık testi gerçekleştirilmiştir. Tez başına kullanılan ortalama test sayısı 31,93'tür. Tezlerde toplamda 14,503 örneklem kullanılmış ve tez başına ortalama örneklem sayısı 155,94 olmuştur. Test başına kullanılan ortalama örneklem sayısı 173,15 iken kullanılması gereken optimum örneklem sayısı 75,62 olarak bulunmuştur. Tezlerde en fazla kullanılan test 1099 adet ile ANOVA (%37), bir sonraki en fazla kullanılan test 617 adet ile Bağımsız Gruplar t-Testi (%20,77) olurken en az kullanılan ise 45 adet ile Basit Regresyon Analizi (%1,52) olmuştur. İstatistiksel güç seviyesi açısından en yüksek güç ortalamasına sahip olan test küçük, orta ve büyük seviyedeki etkileri belirleyebilmek için sırasıyla 0.43, 0.91 ve 0.99 değerleri ile Basit Regresyon Analizi ve en düşük güç ortalamasına sahip olan testler ise sırasıyla 0.12, 0.62 ve 0.89 değerleri ile Ki -Kare Testi ve 0.17, 0.63 ve 0.87 değerleri ile ANOVA olarak bulunmuştur.

Yukarıda bahsedilen sonuçlar, yüksek lisans ve doktora öğrencilerinin tezlerinde ne kadar örneklem ile çalışmaları gerektiği konusunda net bir fikirlerinin olmadığını göstermektedir. Bu nedenle, araştırmacıların örneklem seçimi konusuna ve dolaylı olarak istatistiksel güç konusuna biraz daha fazla önem vermeleri gerekmektedir. Bu bağlamda aşağıdaki öneriler yapılabilir:

- i) Yüksek lisans ve doktora öğrencilerine ders aşamasında, araştırmalarda örneklem seçimini en iyi şekilde nasıl yapacakları öğretilmelidir. Cohen (1988)'in örneklem tablolarından ve kullanılması gereken optimum örneklem sayısını gösteren istatistiksel güç analizi programlarından tez aşamasına geçecek öğrencilere bahsedilmelidir.
- ii) Bu çalışma 2–3 yılda bir tekrarlanarak tezlerin istatistiksel güç seviyelerindeki değişim incelenmelidir.

iii) Sadece “İnsan Kaynakları” alanındaki tezler kapsamında gerçekleştirilen bu çalışma, diğer alanlarda yazılmış tezler üzerinde yapılacak istatistiksel güç analizi çalışmaları ile karşılaştırılabilir. Bu sayede istatistiksel güç konusu hakkında Türkiye’de eğitim gören yüksek lisans ve doktora öğrencilerinin ne derece bilgi sahibi olduklarının değerlendirilmesi yapılabilir.

KAYNAKÇA

- ABERSON, C. L. (2010), *Applied power analysis for the behavioral sciences*, New York: Routledge.
- ABRAHAM, W. T. ve RUSSELL, D. W. (2008), “Statistical Power Analysis in Psychological Research”, *Social and Personality Psychology Compass*, 2/1, 283–301.
- AERA (2006), “Standards for reporting on empirical social science research in AERA publications,” American Educational Research Association, *Educational Researcher*, 35;6, 33–40.
- APA (2001), “Publication Manual of the American Psychological Association, *5th Edition*”, Washington, DC: American Psychological Association.
- BALKIN, R. S. ve SHEPERIS, K. J. (2011), “Evaluating and Reporting Statistical Power in Counseling Research”, *Journal of Counseling and Development: JCD*; Summer 2011, 89, 268-272.
- BECKER, L. A. (2000), “Effect Size”, <http://web.uccs.edu/lbecker/Psy590/es.htm> (alıntı tarihi Mayıs, 2012).
- BORKOWSKI, S. C., WELSH, M. J. ve ZHANG, Q., (2001), “An Analysis of Statistical Power in Behavioral Accounting Research”, *Behavioral Research in Accounting*, 13, 63-84.
- BÜYÜKÖZTÜRK, Ş. (2010), *Sosyal Bilimler İçin İstatistik*, 6. Baskı, Pegem Akademi, Ankara.
- CAFRI, G., KROMREY, J. D. ve BRANNICK, M. T. (2009), “A Sas Macro for Statistical Power Analysis in Meta-Analysis”, *Behavior Research Methods*, 41;1, 35-46.
- CARTER, D. C. (1997), “The Account Taken of Statistical Power in Research Published in the British Journal of Psychology”, *British Journal of Psychology*, 88, 71-83.
- CASHEN, L. H. ve GEIGER S. W. (2004), “Statistical Power and the Testing of Null Hypothesis: A Review of Contemporary Management Research and Recommendations for Future Studies”, *Organizational Research Methods*, 7;2, 151-167.
- CHASE, L. J. ve CHASE, R. B. (1976), “A Statistical Power Analysis of Applied Psychological Research”, *Journal of Applied Psychology*, 61;2, 234-237.

CLARK, D. (1996), "Statistical Power as a Contributing Factor Affecting Significance Among Dissertations in the School of Religious Education at Southwestern Baptist Theological Seminary", Doctoral Dissertation, USA.

COBLICK, G. E. (1998), "Statistical Power Analysis in Nursing Research", Doctoral Dissertation, Auburn University, Alabama, USA.

COHEN, J. (1962), "The statistical power of abnormal-social psychological research: A review, *Journal of Abnormal and Social Psychology*, 65, 145–153.

COHEN, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum.

COHEN, J. (1990), "Things I Have Learned (so far)", *American Psychologist*, 45, 1304-1312.

COHEN, J. (1992), "A power primer," *Psychological Bulletin*, 112;1, 155–159.

COHEN, J. (1994), "The earth is round ($p < .05$). *American Psychologist*", 49, 997-1003.

DENG, H. (2000), "Statistical Power Analysis of Dissertations Completed by Students Majoring in Educational Leadership at Tennessee Universities", Doctoral Dissertation, East Tennessee State University, Tennessee, USA.

DILULLO, L. K. (1997), "A Post Hoc Power Analysis of Inferential Research Examining The Relationship Between Mathematic Anxiety and Mathematic Performance", Doctoral Dissertation, Auburn University, Alabama, USA.

ELLIS, P. D. (2010), *The Essential Guide to Effect Size, Statistical Power, Meta-Analysis and Interpretation Research Results*, Cambridge University Press.

ERTÜRK, M. Ş. (2005), "Retrospective Power Analysis: When?", *Radiology*, 237;2, 743-744.

FAGLEY, N. S. (1985), "Applied Statistical Power Analysis and the Interpretation of Nonsignificant Results by Research Consumers", *Journal of Counseling Psychology*, 32;3, 391-396.

FAUL, F., ERDFELDER, E. ve BUCHNER, A. (1996), "GPower: A General Power Analysis Program", *Behavior Research Methods, Instruments & Computers*, 28;1, 1-11.

FERN, E. F. ve MONROE, K. B. (1996), "Effect-Size Estimates: Issues and Problems in Interpretation", *Journal of Consumer Research*, 23;2, 89-105.

GÜRKAN, A. (2007), "Klinik Periodontoloji Araştırmalarında Bağımsız İki Grup Ortalamalarının Karşılaştırılmasında Örnek Genişliği, İstatistiksel Güç ve Anlamlılık", *Ege Üniversitesi Diş Hekimliği Fakülte Dergisi*, 28: 123-134.

HALLAHAN, M. ve ROSENTHAL, R. (1995), "Statistical Power: Procedures and Applications", *Behav. Res. Ther.*, 34;5/6, 489-499.

HOENIG, J. ve HEISEY, D. W. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis", *The American Statistician*, 55;1, 1-6.

JONES, A. ve SOMMERLUND, B. (2007), "A Critical Discussion of Null Hypothesis Significance Testing and Statistical Power Analysis within Psychological Research", *Nordic Psychology*, 59;3, 223-230.

KOSCIULEK, J. F. ve SZYMANSKI, E. M. (1993), "Statistical Power Analysis of Rehabilitation Counseling Research", 36;4.

LEWIS, K. P. (2006), "Statistical Power, Sample Sizes and the Software to Calculate Them Easily", *Bioscience*, 56;7, 607-612.

LIEBER, R. L. (1990), "Statistical Significance and Statistical Power in Hypothesis Testing", *Journal of Orthopaedic Research*, 8, 304-309.

MACCALLUM, R. C., BROWNE, M. ve SUGAWARA, H. M. (1996), "Power Analysis and Determination of Sample Size for Covariance Structure Modeling", *Psychological Methods*, 1;2, 130-149.

MAXWELL, S. E. (2004), "The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences and Remedies", *Psychological Methods*, 9;2, 147-163.

MAXWELL, S. E., KELLEY, K. ve RAUSCH, J. R. (2008), "Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation", *Annual Review of Psychology*, 59, 537-563.

MAZEN, A. M., GRAF, L., A., KELLOG, K. E. ve HEMMASI, M. (1987), "Statistical Power in Contemporary Management Research", *The Academy of Management Journal*, 30;2, 369-380.

MCKEAN, K. E. (1990), "Statistical Power Analysis of Doctoral Dissertation Research in Educational Psychology", Doctoral Dissertation, Oklahoma State University, Oklahoma, USA.

MONÉ, M. A., MUELLER, G. C. ve MAULAND, W. (1996) "The Perceptions and Usage of Statistical Power in Applied Psychology and Management Research", *Personnel Psychology*, 49;1, 103-120.

MURPHY, K., MYORS, B. ve WOLACH, A. (2004), *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, 2nd. Edition.

NAKAGAWA, S. (2004), "A Farewell to Bonferroni: The Problems of Low Statistical Power and Publication Bias", *Behavioral Ecology*, 15;6, 1044-1045.

NAKAGAWA, S. ve FOSTER, M. T. (2004), "The Case Against Retrospective Statistical Power Analyses with an Introduction to Power Analysis", *Acta Ethol*, 7: 103-108.

NICKERSON, R. S. (2000), "Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy", *Psychological Methods*, 5;2, 241-301.

O'KEEFE, D. J. (2007), "Post Hoc Power, Observed Power, A Priori Power, Retrospective Power, Prospective Power, Achieved Power: Sorting Out Appropriate Uses of Statistical Power Analyses", *Communication Methods and Measures*, Lawrence Erlbaum Associates, Inc, I(4), 291-299.

OLEJNIK, S. F. (1984), "Planning Educational Research: Determining Necessary Sample Size", *Journal of Experimental Education*, 53;1, 40-48.

ONWUEGBUZIE, A. J. ve LEECH, N. L. (2004), "Post Hoc Power: A Concept Whose Time Has Come", *Understanding Statistics*, 3;4, 201-230.

OTTENBACHER, K. (1982), "Statistical Power and Research in Occupational Therapy", *OTJR*, 2;1, 13-25.

ÖZDAMAR, K. (2003), *Modern Bilimsel Araştırma Yöntemleri*, Kaan Kitabevi, Eskişehir.

PETERMAN, R. M. (1990), "The Importance of Reporting Statistical Power: The Forest Decline and Acidic Deposition Example", *Ecology*, 71;5, 2024-2027.

- ROGERS, T. ve HOPKINS, K. (1988), "Power Estimates in the Presence of a Covariate and Measurement Error", *Educational and Psychological Measurement*, 48, 647-656.
- ROSE, R. A. ve BOWEN, G. L. (2009), "Power Analysis in Intervention Research: Designing Cluster-Randomized Trials", *Social Work Research*, 33;1, 43-52.
- ROSSI, J. S. (1990), "Statistical Power of Psychological Research: What We Have Gained in 20 Years ?", *Journal of Consulting and Clinical Psychology*, 58;5, 646-656.
- ROSSI, J. S. (2012, in Press), *Statistical Power Analysis*. In J. A. Schinka & W. F. Welicer(Eds.), *Handbook of Psychology. Volume 2: Research Methods in Psychology* (2nd. Ed.). John Wiley & Sons.
- ROTENBERRY, J. T. ve WIENS, J. A. (1985), "Statistical Power Analysis and Community-Wide Patterns", 125;1, 164-168.
- SACCO, W. (1982), "Statistical Power Considerations in the Use of Cost-Effectiveness Analysis", *Professional Psychology*, 13;5, 752, 758.
- SAWYER, A. G. ve BALL, D. A. (1981), "Statistical Power and Effect Size in Marketing Research", *Journal of Marketing Research*, Volume: XVIII, 275-290.
- SEDLMEIER, P. ve GIGERENZER, G. (1989), "Do Studies of Statistical Power Have an Effect on the Power of Studies?", *Psychological Bulletin*, 105;2, 309-316.
- SİPAHİ, B., YURTKORU S. ve ÇİNKO, M. (2008), *Sosyal Bilimlerde SPSS'le Veri Analizi*, Beta Yayınları, İstanbul.
- STEIDL, R. J., HAYES, J., P. ve SCHAUBER, E. (1997), "Statistical Power Analysis in Wildlife Research", *The Journal of Wildlife Management*, 61;2, 270-279.
- STEPHANO, J. (2003), "How Much Power is Enough? Against the Development of an Arbitrary Convention for Statistical Power Calculations", *Functional Ecology*, 17;5, 707-709.
- TENER, Mark, A. (2000), "A Post Hoc Statistical Power Analysis a Survey of the Research Published in the Journal of Athletic Training", *Doctoral Dissertation*, Middle Tennessee State University, Tennessee, USA.

THALHEIMER, W. ve COOK, S. (2002), "How to Calculate Effect Sizes From Published Research Articles: A Simplified Methodology", http://education.gsu.edu/coshima/EPRS8530/Effect_Sizes_pdf4.pdf, Eriřim Tarihi: 15.05.12.

THOMAS, L. (1997), "Retrospective Power Analysis", *Conservation Biology*, 11;1, 276-280.

THOMAS, L. ve JUANES, F. (1996), "The Importance of Statistical Power Analysis: An Example from Animal Behaviour", *Animal Behaviour*, 52, 856-859.

EK-1

ÇALIŞMADA KULLANILAN TERİMLER VE ANLAMLARI

Alfa (α -Anlamlılık Seviyesi): Doğru bir sıfır hipotezinin reddedilme olasılığıdır (I. Tip Hata).

Beta (β): Yanlış olan bir sıfır hipotezinin reddedilememesi olasılığıdır (II. Tip Hata).

Etki Büyüklüğü (Effect Size): Araştırılan ilişkilerin, farklılıkların anakütle üzerinde ne derece var olduğudur (Cohen, 1988, s. 8). Özellikle H_0 ve H_1 hipotezleri arasındaki farktır.

İnference (Tümevarım): Araştırmacının örneklemden elde ettiği sonuçları tüm anakütle üzerinde genelleştirmesi durumudur (Sipahi vd., 2008, s. 6).

Gözlemlenen Etki Büyüklüğü: Araştırmadan elde edilen veriler analiz edildikten sonra hesaplanan etkinin boyutudur (Cohen, 1988, s. 9).

Gözlemlenen Güç: Araştırmadan elde edilen veriler analiz edildikten sonra hesaplanan güçtür (Balkin ve Sheperis, 2011, s. 271).

Anakütle: Araştırmacı tarafından incelenmek istenen birimlerin tümünün oluşturduğu topluluktur (Sipahi vd., 2008, s. 4).

Post-Hoc (Retrospective) Power Analysis (Deneysel Güç Analizi): İstatistiksel gücün, veriler toplanıp analiz edildikten sonra hesaplanmasıdır (Özdamar, 2003, s. 109).

Priori (Prospective) Power Analysis (Teorik Güç Analizi): Verileri toplama aşamasından önce, araştırmanın planlama aşamasında gerekli olan istatistiksel gücün hesaplanmasıdır (Özdamar, 2003, s. 109).

İstatistiksel Güç: Yanlış olan bir sıfır hipotezinin reddedilmesi için gerekli olan doğru etkinin bulunabilmesi olasılığıdır (Cohen, 1988, s. 4).

Güç Analizi: Bir çalışmanın istatistiksel gücünün belirlenmesi, anlamlılık seviyesi, etki büyüklüğü ve örneklem sayısı ile gerçekleştirilen veya örneklem sayısının belirlenmesi için etki büyüklüğü, anlamlılık seviyesi ve istatistiksel güç ile gerçekleştirilen bir ön süreçtir (Cohen, 1988, s. 4).

Meta Analiz: Bir meta analiz, araştırma yapmak istenen alanda daha önce gerçekleştirilmiş benzer çalışmaların sonuçlarını ortak bir havuzda toplayarak o alan ile ilgili ortalama değerlere ulaşmaktır (Becker, 2000, s. 11)

Örneklem (Sample): Anakütleden alınan deneklerin alt kümesidir.

Örneklem Sayısı (Sample Size): Örneklemde dâhil olduğu deneklerin toplam sayısıdır.

İstatistik: Gözlemlenen değerlerin yorumlanması ve değerlendirilmesi için objektif bir yöntemdir (Sipahi vd., 2008, s. 3).

I. Tip Hata (False Positive): Gerçekte doğru olan bir H_0 hipotezinin reddedilmesi hatasıdır (Balkin ve Sheperis, 2011, s. 269; Murphy ve Myors, 2004, s. vii).

II. Tip Hata (False Negative): Gerçekte yanlış olan bir H_0 hipotezinin reddedilememesi hatasıdır (Rossi, 2012, s. 8; Murphy ve Myors, 2004, s. vii).

Ö Z G E Ç M İ Ş

Adı ve Soyadı : Burak KESKİN

Doğum Tarihi ve Yeri : 23.10.1986 / ADANA

Medeni Durumu : Bekar

Eğitim Durumu

Mezun Olduğu Lise : İbrahim Atalı Çok Programlı Lisesi/ADANA

Lisans Diploması : Akdeniz Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümü

Lisans Bitirme Projesi : Kartezyen Dönüşümler

Yabancı Dil : İngilizce

İş Deneyimi

Çalıştığı Kurumlar : Çankırı Karatekin Üniversitesi/Araştırma Görevlisi (2010- Devam)

Çelebi Ground Handling AYT / 2010

USAŞ AYT / 2008-2010

Riva Diva Hotel Antalya / 2007

E-mail : burakkeskin@karatekin.edu.tr