

İstatistiksel Güç Bir Araştırmanın Sonuçlarına Etki Eder Mi? Örneklem Büyüklüğüne Nasıl Karar Verilmeli?^a

Burak Keskin^{b, c}

Özet

İstatistiksel güç analizi; güvenilirliği, geçerliliği yüksek bir araştırma planlamayı ve araştırma sonucunda alınacak kararların geçerliliğini, güvenilirliğini ve duyarlılığını garanti altına almayı sağlayan bir yöntemdir. İstatistiksel güç, araştırmada bir hipotezin denetlenmesi için kullanılan istatistiksel testler sonucunda varılan kararın ne kadar güvenilir ve 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 daha çok veri toplama öncesinde anlamlıdır ve hipotez kurma sürecinde önemlidir. Bu bağlamda çalışmanın amacı, sosyal bilim çalışmalarında istatistiksel güç analizinin ne olduğu konusunda farkındalık oluşturarak araştırmalara katkı sağlamak ve bir araştırmada istatistiksel güç analizinin nasıl kullanılacağı konusunda araştırmacılara rehberlik etmektir.

Anahtar Kelimeler

İstatistiksel Güç
İstatistiksel Güç Analizi
Etki Büyüklüğü

Makale Hakkında

Geliş Tarihi: 09.11.2019
Kabul Tarihi: 12.10.2020
Doi: 10.18026/cbayarsos.644692

Does Statistical Power Affect a Study's Results? How Many Sample Size?

Abstract

The statistical 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 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. Various statistical analyses begin the analysis with the existing data and focus 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 study are; to contribute researches in social sciences by creating awareness about what the power analysis means and to guide researchers about how the power analyses can be employed in a social science study.

Keywords

Statistical Power
Statistical Power Analysis
Effect Size

About Article

Received: 09.11.2019
Accepted: 12.10.2020
Doi: 10.18026/cbayarsos.644692

^a Bu çalışma, yazarın yüksek lisans tezinden üretilmiştir ve 7th International Days of Statistics and Economics isimli kongrede sunulan bildirinin genişletilmiş halidir

^b İletişim Yazarı: burakkeskin@karatekin.edu.tr; burakkeskin@gmail.com

^c Dr. Öğr. Üyesi, Çankırı Karatekin Üniversitesi İİBF İşletme Bölümü, ORCID: 0000-0002-7042-2095

Giriş

Bir araştırmanın kalitesini ve olası sonuçlarının doğruluğunu, araştırmanın tasarımı, veri toplama yöntemleri, veri analiz teknikleri ve araştırmacının bakış açısı gibi birçok faktör etkilemektedir. Bir istatistiksel anlamlılık testi yürütüldüğünde ise araştırma sonuçlarının doğruluğu açısından birkaç ilave faktör daha devreye girmektedir. Bu faktörler; anlamlılık seviyesi, örneklem büyüklüğü, etki büyüklüğü ve istatistiksel güçtür. Hipotez testinde, gerçekte yanlış olan bir sıfır hipotezinin reddedilme olasılığı istatistiksel güç olarak tanımlanmaktadır. Belirli bir anlamlılık seviyesi, örneklem büyüklüğü ve etki büyüklüğü ile çalışmanın istatistiksel gücüne Cohen (1988) tarafından hazırlanan güç tabloları, çeşitli istatistiksel güç formülleri veya geliştirilen bilgisayar yazılımları kullanılarak karar verilebilmektedir. Bu sürecin tamamı ise “*İstatistiksel Güç Analizi*” olarak adlandırılmaktadır.

İstatistiksel güç, araştırılan anakitlede, araştırmacının bulmak istediği veya bulmayı umduğu gerçek etkiyi ortaya çıkarabilme ihtimali olarak tanımlanmaktadır. 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). Araştırmacı, aradığı gerçek etkiyi bulduğunda, bu 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 araştırmacı, araştırmasında yeterli gücü sağlayamazsa bu onun anakitle üzerindeki doğru etkiyi bulmasını ve yanlış bir H_0 hipotezini reddetme şansını oldukça azaltacaktır.

Cohen’in 1962 yılındaki çalışmasından sonra bazı istatistikçiler yaptıkları çalışmalarda güç analizi konusundaki endişelerini dile getirmeye başlamıştır. Shavelson (1988) bu konuda şu şekilde görüş belirtmiştir; “Araştırmacılar, çalışmalarının güç ile ilgili olan kısmını dikkate almalı ve çalışmalarını mümkün olduğunca fazla istatistiksel güç üzerine inşa etmelidirler.”

İstatistiksel güç ve örneklem büyüklüğüne karar verme sürecindeki temel amaç, gerçekleştirilmesi planlanan araştırmanın anakitle üzerinde var olduğu düşünülen ve ortaya çıkarılmaya çalışılan etkiyi belirleyebilecek seviyede bir kurgu yapmaktır. Ancak burada etkiyi belirleme aşamasında örneklem büyüklüğüne karar verme aşamasında çok cömert davranılarak kaynakların israf edilmesinin de önüne geçilmelidir (Meyners vd., 2020).

Sosyal bilimler alanındaki çalışmalarda istatistiksel güç sürecini uygulamalı olarak anlatmayı amaçlayan bu çalışmada öncelikle istatistiksel güç kavramı detaylı bir şekilde açıklanmıştır. Keza istatistiksel güç, istatistiksel güç analizi, istatistiksel gücü etkileyen faktörler, istatistiksel güç analizi çeşitleri hakkında geniş bir bilgiye yer verilmiştir. Sonuç bölümünde ise istatistiksel güç analizi konusunda araştırmacılara bazı önerilere yer verilmiştir.

İstatistiksel Güç Kavramı

İstatistiksel güç kavramının geliştirilmesi Neyman ve Pearson'a atfedilebilir. Onlar bu kavramı ilk olarak 1928 yılında tanıtmışlardır. Neyman ve 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, R.A. Fisher tarafından yaşamı boyunca katı bir şekilde reddedildiğinden o yıllarda gelişme şansı bulamamıştır (DiLullo, 1987).

Cohen, istatistiksel gücü; "yanlış bir sıfır hipotezinin reddedilme olasılığı ($1-\beta$)" olarak tanımlamıştır. Dahası bir istatistiksel testin gücü, istatistiksel olarak anlamlı sonuçlar sağlama olasılığıdır (Cohen, 1988; O'Keefe, 2007; Balkin ve Sheperis, 2011). Rossi (2012) ise istatistiksel güç tanımını "bir testin, anakitle üzerinde gerçek bir etki olduğu varsayımı altında anlamlı bir sonuç sağlama olasılığıdır" şeklinde yapmıştır. Jones ve Sommerlund (2007) istatistiksel gücü, bir araştırmanın sağlığının ölçülmesi olarak tanımlamışlardır.

1962 yılında Jacop Cohen, istatistiksel güç konusundaki çığır açıcı çalışmasını yayımlamış ve istatistiksel güç konusunun yeniden doğmasına öncülük etmiştir. Bir dergide istatistiksel güç ile ilgili makalesinde Cohen şunları amaçlamıştır:

- Güç konusuna araştırmacıların, araştırma yaptıran müşterilerin, araştırmanın planlama ve tasarım aşamasında değerlendirme yapan kişilerin (sponsorlar, dergi editörleri vs.) dikkatini çekmek,
- Araştırmalarını güç analizi üzerine inşa etmek isteyen araştırmacılara rehberlik yapmak ve işlerini kolaylaştırmak amacıyla güç tabloları oluşturmak,
- Güç analizi üzerine inşa edilecek olan çalışmaları 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).

Bir araştırmanın tasarım aşamasındaki güç analizinin amacı, anakitle üzerinde var olduğu düşünülen etkinin mümkün olabilecek en yüksek olasılık ile belirlenebilmesinin garanti altına alınmasıdır. Eğer araştırma yapılan anakitle üzerinde herhangi bir etki varsa gerçekleştirilecek istatistiksel testler neticesinde hesaplanacak istatistiksel güç, elde edilecek sonuçların istatistiksel olarak anlamlı sonuç verip vermeyeceği olasılığını yansıtmaktadır (Tener, 2000).

İstatistiksel güç, anakitlede 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 sonuçlar 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).

Teorik olarak istatistiksel testin gücü, $1-\beta$ şeklindeki matematiksel tanımın ötesindedir. Güç, araştırma altındaki anakitlede 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 (1988) göre güç analizinin anakitle 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 çalışarak 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).

İstatistiksel Güç Analizi

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).

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

- a. **Güç**; alfa, etki büyüklüğü ve örneklem büyüklüğünün bir fonksiyonudur. Bu analiz tipi günümüz çalışmalarında en yaygın olarak kullanılan güç analizidir.
- b. **Örneklem büyüklüğü**; alfa, etki büyüklüğü ve gücün bir fonksiyonudur. Bu analiz tipi, çalışma öncesinde gerekli olan örneklem büyüklüğünün belirlenmesinde kullanılmaktadır.
- c. **Etki Büyüklüğü**; alfa, güç ve örneklem büyüklüğünün bir fonksiyonudur. Bu analiz tipi yazındaki çalışmaların karşılaştırılması için kullanılmaktadır.
- d. **Alfa**; güç, örneklem büyüklüğü ve etki büyüklüğünün bir fonksiyonudur. Bu analiz tipi, araştırma gelenekleri yüzünden çok yaygın olmayan bir güç analizi çeşididir. 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 büyüklüğü ve etki büyüklüğü kullanılarak uygulanır. Analiz 2 ise çalışma öncesinde gerekli örneklem büyüklüğünü belirlemek için (Priori Power Analysis) uygulanır (Cohen, 1988).

Güç analizinin önemini anlaşılabilmesi için aşağıdaki örnek son derece uygun olacaktır.

“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” (Ellis, 2010).

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 örneklem büyüklüğüne ihtiyaç var? Araştırma altında ortaya çıkarılmak istenin etkinin anakitle üzerinde var olduğu biliniyorsa, bu durumda etkiyi ortaya çıkarabilmek için araştırma yeterli şansa sahip mi? Bu şans nasıl artırılabilir? Eğer örneklem büyüklüğü 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.

İstatistiksel Gücü Doğrudan Etkileyen Faktörler

Bir araştırmanın istatistiksel gücünü doğrudan etkileyen faktörler;

- Anlamlılık seviyesi
- Etki büyüklüğü
- Örneklem büyüklüğüdür.

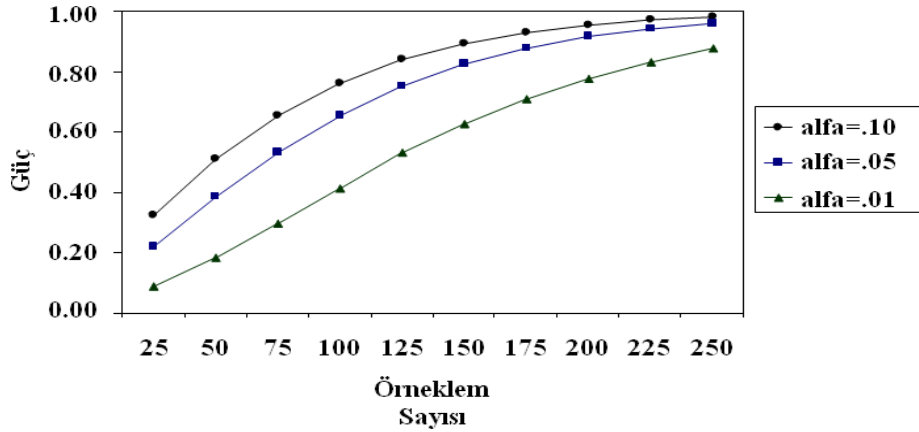
Diğer parametreler sabit olarak düşünüldüğünde bu üç parametrenin artması istatistiksel gücü artıracaktır. Bahsi geçen her bir parametre ile güç arasında doğru orantılı bir ilişki vardır (Mazen vd., 1987; O’Keefe, 2007). İstatistiksel gücü doğrudan etkileyen faktörlerin yanında dolaylı olarak etkileyen,

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

Anlamlılık Seviyesi (α)

Gücü etkileyen parametrelerden birincisi anlamlılık seviyesidir. Anlamlılık seviyesi, araştırmacı tarafından veriler toplanmadan önce belirlenmektedir. 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 edecektir (Ellis, 2010).

İlk bakışta burada gücü artırmak için sadece anlamlılık seviyesinin ve örneklem büyüklüğünün artırılması gerektiği anlamı çıkarılabilir. Fakat alfa, genellikle ya 0.01 olarak ya da 0.05 olarak kurulmaktadır ve genellikle 0.05 seviyesinden daha yukarı seviyelerde kurulmaz. İstatistiksel güç ile anlamlılık seviyesi arasındaki ilişki Şekil 1 üzerinde gösterilmiştir.



Şekil 1. Anlamlılık Seviyesinin İstatistiksel Güç Üzerindeki Etkisi
(Kaynak: Rossi, 2012)

Etki Büyüklüğü

Gücü etkileyen öncelikli parametrelerden ikincisi etki büyüklüğüdür. Etki büyüklüğü genel anlamda, incelenen olgunun, incelenen anakitle üzerinde mevcut olma derecesidir (Cohen, 1988). 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). 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.

Gücü tahmin etmek veya belirli bir güç seviyesini yakalamak amacıyla örneklem büyüklüğünü belirlemek için öncelikle yapılması gereken işlerden birisi anakitle üzerindeki etki büyüklüğünü tahmin etmektir. Bu tahmin, 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 (Mazen vd., 1987; Thomas ve Juanes, 1996; Borkowski vd., 2001; Lewis, 2006; Rose ve Bowen, 2009; Correll vd., 2020). Etki büyüklüğü, özellikle herhangi iki çalışmanın sonuçlarının karşılaştırılmasın aşamasında önemlidir (Thalheimer ve Cook, 2002).

Araştırmacıların öncelikli amacı gerçek dünyada var olan ilişkileri veya ilişkilerin gücünü veya farklılıkları tahmin etmek, bunların etkilerini ortaya çıkarmaktır. Bahsi geçen etki bir tedavi, bir deneme, bir karar, bir strateji, bir yenilik, bir icat, bir operasyon dolayısıyla her şey

olabilmektedir. Bu araştırmanın türüne bağlıdır. Küçük etkileri ortaya çıkarmak daha zordur ve daha fazla örneklem gerektirmektedir (Rotenberry ve Patterns, 1985).

Etki büyüklüğüne hayatın her alanında rastlanılabilir. Örneğin bir kilo kontrol tedavi yöntemi kişilere bir ay içinde 10 kilo zayıflama garantisi verdiğinde bu, uygulanacak tedavi yönteminin etki büyüklüğüdür. İlaveten, “servis sağlayıcınızı değiştirin ve %20 tasarruf sağlayın, botoks ile 10 yıl gençleşin” gibi ifadeler etki büyüklüğünün bir sonucu olarak ifade edilmektedir (Ellis, 2010).

Etki büyüklüğü anakitle üzerinde elde edilen bir değerdir ve sadece örneklem üzerinden tahmin edilebilir (Abraham ve Russell, 2008). Eğer, yapılan bir araştırmada anakitle 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 örneklemde elde edilen istatistikler ile anakitle parametreleri arasında anlamlı bir ilişki yok ise o zaman etki büyüklüğünün değeri sıfırdır. Eğer bazı istatistiksel anlamlılıklar elde edildiyse, o zaman anakitle üzerinde bir etki vardır ve etki büyüklüğünün değeri sıfırdan farklıdır. Bir verimlilik araştırmasında, etki büyüklüğü kavramsal olarak anakitle ortalamaları 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; McKean, 1990).

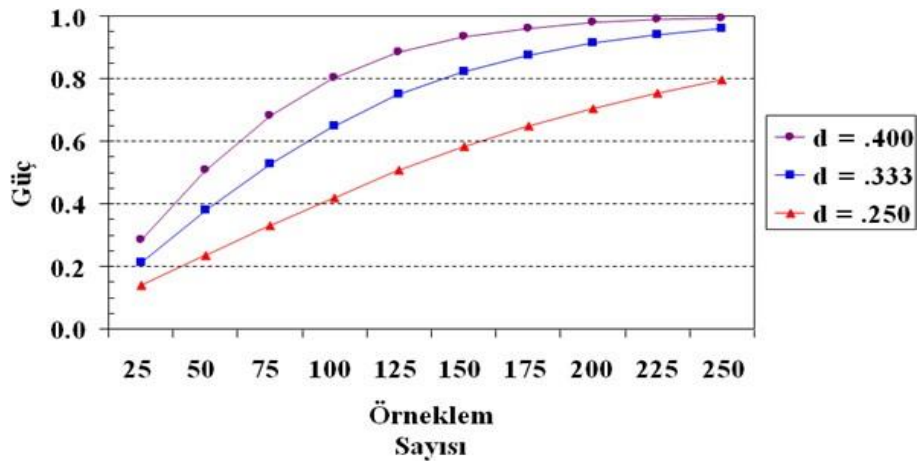
Bir etki, iki grup üzerinde uygulanan bir tedavinin sonucu olabilir veya iki ilgili değişken arasındaki ilişkinin derecesi olabilir. Etki büyüklüğünün tahmin edilmesi elde edilen sonuçların doğru yorumlanması açısından gereklidir. APA tarafından yayımlanan bir araştırmada etki büyüklüğünün rapor edilmesi konusundaki başarısızlığın editörler tarafından en çok gözlemlenen yedi kusurdan biri olduğunu belirtilmiştir (APA, 2001).

Etki büyüklüğünü belirlemenin en etkili yolu tüm anakitle üyelerini incelemektir. Fakat bu durum genellikle mümkün olmayacağı için etki büyüklüğünün tahmin edilmesi gerekmektedir. Yapılacak olan tahminin netliği kullanılan yöntemlere ve seçilen örneklem anakitleyi 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 büyüklüğünden etkilenmektedir. Nispeten daha küçük bir örneklem, gerçekleştirilen bir çalışmada daha az olasılıkla elde edilecek sonuçların istatistiksel olarak anlamlı olacağı anlamına gelmektedir. Bu durum, istatistiksel anlamlılık testinden pratik anlamlılığa dair bir sonuç çıkarılamayacağını göstermektedir (Ellis, 2010).

Etki büyüklüğü çeşitli kaynaklarda, doğru olan bir alternatif hipotezin anakitle üzerindeki boyutu ve araştırmacıya, araştırma sonucunda bulunduğu sonuçların ne kadar önemli olduğunu anlatan bir kavram olarak da tanımlanmıştır. Bu gösterge ya da belirti bir sayısal değer ile açıklanabilmektedir. Eğer sıfır hipotezi doğru ise, o zaman etki büyüklüğünün değeri "0" dır. Yani araştırılan anakitle üzerinde herhangi bir etki yoktur. Aksine, eğer sıfır hipotezi yanlışsa, o zaman etki büyüklüğünün değeri sıfırdan farklıdır ve anakitle 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 anakitle üzerinde o derecede vardır. Ayrıca aşağıdaki Şekil 2 üzerinde gösterildiği gibi, etki büyüklüğünün değeri arttıkça, anlamlılık seviyesi ve örneklem büyüklüğü sabit tutulduğunda, istatistiksel güç artmaktadır. İlaveten, daha büyük bir etki daha küçük bir örneklem gerektirmektedir (DiLullo, 1997).



Şekil 2. Etki Büyüklüğünün İstatistiksel Güç Üzerindeki Etkisi
(Kaynak: Rossi, 2012)

Cohen (1988), etki büyüklüğünün ne kadar olması gerektiğine karar verme konusunun araştırmacılar için karar vermesi en zor konulardan biri olmasına rağmen aynı zamanda en az dikkate alınan konulardan da birisi olduğunu vurgulamıştır.

Etki büyüklüğünün tahmini konusunda bir araştırmacı, çalışması hakkındaki bilgisini mümkün olduğunca çalışma öncesinde kullanmalıdır. Buna karşılık, en azından sosyal bilimlerde alanındaki araştırmalarda aşağıdaki nedenlerden ötürü yetersiz bilgiye sahip olunması nedeniyle etki büyüklüğünün doğru bir şekilde tahmin edilmesi zordur. Bu nedenler:

- Sosyal bilimlerde alanındaki teorilerin nadiren etki büyüklüğünün tahmini konusunda ipuçları vermesi,
- Güç analizi için gerekli olan bilginin genellikle yürütülen araştırmada yetersiz olmasıdır.

Bu eksikliklerden kaynaklanan zorluğu giderebilmek için Sawyer ve Ball (1981), arařtırmacılara doęru bir etki büyüklüęü tahmini yapabilmek için çalıřma öncesinde anakitleden elde edilecek bir altküme üzerinde bir ön çalıřma yapılarak etki büyüklüęünün tahmin edilmesini önermişlerdir.

İkinci olarak, eęer ön çalıřma kısmında arařtırmacı etki büyüklüęünün tahmini konusunda yeterli bir bilgiye sahip olamıyorsa, bu durumda Cohen'in etki büyüklüęü hakkında yapmış olduęu geleneksel bir sınıflandırma olan küçük, orta ve büyük seviyelerden birisini seçebilir (Gignac ve Szodorai, 2016). Fakat etki büyüklüęü, kullanılan istatistiksel teste göre deęişiklik göstermektedir. Tablo 1'de, Cohen (1988)'in sunmuş olduęu küçük, orta ve büyük seviyelerdeki etki büyüklükleri çeřitli istatistiksel testler için gösterilmektedir.

Tablo 1. Cohen'in Etki Büyüklüęü Sınıflandırmaları

Test	İlgili Etki Büyüklüęü	Etki Büyüklüęü Sınıfları		
		Küçük	Orta	Büyük
Baęımsız 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
Çoklu Regresyon	R^2	.02	.13	.26
	f^2	.02	.15	.35

Etkinin Yorumlanması

Sosyal bilimler alanındaki arařtırmacılar için anlamlılık testi sonuçlarına göre yorum yapmak nadir görülen bir durum deęildir. Kimi arařtırmacılar yüksek seviyede anlamlı çıkan sonuçlarını, orta seviyede anlamlı olan sonuçlara göre daha önemli olarak nitelendirebilirler. Kimi arařtırmacılar ise anlamlı olmayan bir sonuca ulařtıkları anda çalıřmalarını herhangi bir etkinin olmadığını belirterek sonuçlandırabilirler. Her iki durumda da yanlış yorum yapılma ihtimali oldukça fazladır. Etki büyüklüęünün yorumu, anlamlılık testinden ayrı olmalıdır. Her ikisi de birbiri ile ilişkili olmasına raęmen, istatistiksel anlamlılık örneklem büyüklüęünden etkilenmektedir. Daha büyük örneklem ile daha yüksek ihtimalle anlamlı sonuçlar elde edilebilir. Fakat $p = 0.01$ deęeri, $p = 0.05$ deęerinden daha önemli olmak zorunda deęildir. Fakat nispeten daha büyük etkilerin, küçük etkilerden daha önemli olduęu söylenebilir. Örneęin Cohen's $d = 1$ deęeri, $d = 0.2$ deęerinden daha önemlidir.

Bir çalıřma sonucunun pratik olarak deęerlendirilebilmesi için etki büyüklüęünün deęerini 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. Örneęin, bir öęrencinin herhangi bir sınavdan 140 puan aldıęına kulak misafiri olan bir kiři, bu puanın IQ test puanı mı yoksa GMAT (Graduate Management Admission Test) puanı mı olduęunu bilmeden bunu yanlış yorumlaması kaçınılmazdır. Eęer 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 (Ellis, 2010).

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

Tablo 2. Etki Büyüklüğü ve p Değeri Arasındaki Yorum Farkının Anlaşılması için Bir Örnek

	<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 $p<0.05$ çıktığı 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 büyüklüğünün 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). Sosyal bilimler alanında araştırmacılar kendilerini kesinlikle p değeri ile sınırlamamalı, daha da öteye bakmalıdırlar.

Örneklem Büyüklüğü

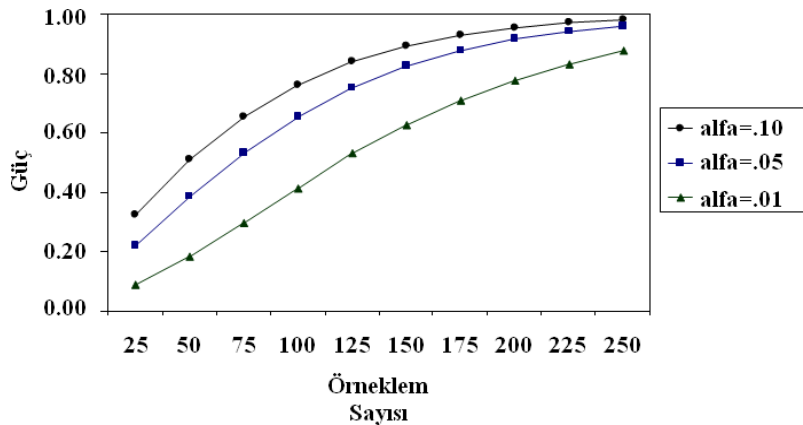
Örneklem büyüklüğü, gücü doğrudan etkileyen bir diğer parametredir. Örneklem büyüklüğü, çalışma için kullanılan denek sayısıdır. Örneklem büyüklüğü, bir istatistiksel testin gücünü etkileyen öncelikli parametredir. DiLullo'nun (1997) aktardığına göre araştırmacıların maliyet ve ulaşılabilirlik, araştırmanın tasarım aşamasındaki kısıtlar ve yeterli istatistiksel güç açısından uygun bir örneklem büyüklüğünü belirlemeleri gerekmektedir.

Araştırmanın tasarım aşamasındaki önemli konulardan birisi de belirlenen güç seviyesini elde edebilmek için gerekli örneklem büyüklüğünün ne olması gerektiğinin bilinmesidir (MacCallum vd., 1996). Araştırılan konu ile ilgili gözlem sayısı arttıkça örneklemde elde edilen verilerin güvenilirliği de artacaktır. Artan bu güvenilirliğin bir sonucu olarak araştırmacı yüksek olasılıkla yanlış olan sıfır hipotezini doğru bir şekilde reddedebilir. Bu

durumda örneklem büyüklüğünün artması gücü olumlu yönde etkilemektedir. İdeal olarak bir araştırmacı alfa, etki büyüklüğü ve arzuladığı güç seviyesini kullanarak çalışması için gerekli en uygun örneklem büyüklüğüne ulaşabilir (Cashen ve Geiger, 2004). Araştırmalarda anlamlı sonuçlar elde etme olasılığı büyük örneklem sayıları ile birlikte artarken, bu olasılık daha küçük örneklem sayıları ile birlikte azalır. Çünkü daha büyük örneklem, hedef anaktile parametreleri hakkında daha net tahminler yapılmasını sağlar (Balkin ve Sheperis, 2011).

Örneklem büyüklüğü artarsa standart hata azalacağı için güç artmaktadır. Küçük sayıdaki örneklem ile gerçekleştirilen araştırmalarda, etki büyüklüğü gerçekte büyük olmasına rağmen araştırmacı sıfır hipotezini reddetme konusunda başarısız olabilir ve mevcut olan önemli farklılıkları belirleyemeyebilir. Aksine, örneklem büyüklüğü olması gerekenden çok fazla olan çalışmalarda gerçekte anlamsız olan etkiler örneklem büyüklüğünün çok fazla olması nedeniyle yanlış bir şekilde anlamlı olarak bulunabilir. O nedenle mümkün olduğu müddetçe optimum örneklem büyüklüğü ile çalışılmalıdır (Sacco, 1982; Mazen vd., 1987; Coblick, 1998; Jones ve Sommerlund, 2007; Balkin ve Sheperis, 2011).

Anlamlılık seviyesi kurulduğunda ve etki büyüklüğüne bir teori veya yapılan bir ön çalışma ile karar verildiğinde, örneklem büyüklüğü istatistiksel çalışmanın uygulama aşamasından önce belirli bir güç elde etmek için kullanılır. Diğer parametreler sabit iken örneklem büyüklüğü ile güç arasında doğru orantılı bir ilişki vardır. Bu durum Şekil 3'te gösterilmiştir (DiLullo, 1997).



Şekil 3. İstatistiksel Güç ve Örneklem Sayısı Arasındaki İlişkinin Çift Yönlü ve Farklı Alfa Seviyelerinde Gösterilmesi
(Kaynak: Rossi, 2012)

Aslında burada sorulması gereken soru hipotezin doğru bir şekilde test edilmesi için ne kadar örnekleme ihtiyaç olduğudur. Bu sorunun cevabı, güç analizini kullanmadan, genellikle mantıklı olmayan esaslara dayanmaktadır (Ellis, 2010). Örneklem büyüklüğünü ayarlamak, araştırmacılar tarafından gücü ayarlamaktan daha kolay bir iş olarak görülmektedir. Örneğin, genellikle örneklem büyüklüğü 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 büyüklüğünün belirlenmesinde araştırmacılar böyle basit yöntemleri kullanarak genellikle gereğinden çok daha fazla örneklem ile çalışırlar ki bu da daha fazla zaman ve mali açıdan külfet demektir. Ya da olması gerekenden daha az bir örneklem 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 büyüklüğünü belirlemelidirler (McKean, 1990; DiLullo, 1997).

Sosyal bilimler alanındaki araştırmacılar, araştırmaları için gerekli olan örneklem büyüklüğüne karar verme konusunda net bir bilgiye sahip değillerdir. Genellikle örneklem büyüklüğü konusunda “ulaşabildiğin kadar örnekleme kullan” yöntemi kullanılmaktadır. Ancak bu yöntem çoğu zaman işe yaramamaktadır. Bazen ulaşılan örneklem büyüklüğü sıfır hipotezini reddetmek için yeterli olmayabilir. Bazen de normalde ulaşılan örneklem büyüklüğünden çok daha az sayıdaki örneklem ile araştırma tamamlanabilmekte ve harcanan emek ve para açısından savurganlık durumun söz konusu olmaktadır (Olejnik, 1984).

Güç konusu, örneklem büyüklüğüne karar verme aşamasında dolaylı olarak dikkate alınmaktadır. Araştırmacılar genellikle örneklem büyüklüğüne karar verme aşamasında etkin bir yöntem olan güç analizi yerine geçmişteki tecrübelerle, verilerin ulaşılabilirliğine veya ulaşabildiğin kadar örneklem metoduyla çok da verimli olmayan yöntemler kullanılmaktadır. Bu şekilde güç analizi üzerine inşa edilmeyen araştırmalar maliyet ve zaman açısından etkin değildir. Bir güç analizi, araştırmacıya yeterli sayıda örneklem ile çalışmayı ve anakitle üzerindeki eğer varsa gerçek etkiyi belirleme şansı vermektedir (Coblick, 1998). Çalışmalarda uygulanacak istatistiksel test çeşitlerine göre kullanılması gereken örneklem büyüklüğüne (Cohen’in tabloları aracılığı ile veya G-Power programı ile) kolaylıkla karar verilebilmektedir. Bu sayede araştırmacılar hem var olduğunu düşündükleri etkiyi ortaya çıkarabilmek için yeterli istatistiksel güce sahip olacaklar hem de olması gereken büyüklükte örneklem ile çalışarak zaman, maliyet ve çaba anlamında oldukça büyük tasarruflarda bulunmuş olacaklardır.

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).

Teorik Güç Analizi (Priori 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 en uygun örneklem büyüklüğüne ulaşmak mümkündür (Erdfelder vd., 1996).

Güç analizleri genellikle çalışmanın tasarım aşamasında işe yaramaktadır (Flynn vd., 2017). Teorik güç analizi, temel olan etki büyüklüğü, örneklem büyüklüğü, güç ve alfa seviyesi gibi parametrelerden örneklem büyüklüğüne 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, araştırmanın başında terk edilebilmekte ve yeni bir tasarım yapılabilmektedir (Lewis, 2006).

Deneysel Güç Analizi (Post-Hoc Power Analysis)

Güç analizleri genellikle araştırmaların tasarım aşamalarında faydalıdır. Fakat bazen özellikle

istatistiksel olarak anlamlı sonuçlar elde edilemeyen arařtırmalarda, alıřma tamamlandıktan sonra bir g analizini gerekleřtirilmesi gerekmektedir. Arařtırma ncesinde yapılacak bir etki byklę tahminine dayanan teorik g analizinden farklı olarak deneysel g analizini, arařtırma sonucunda gzlemlenen etkiye ve gzlemlenen istatistiksel gce dayanmaktadır. Bu tr bir g analizini yapmanın avantajı, elde edilecek anlamlı olmayan sonuçların dikkate deęer olmaları ihtimalinin bulunmasıdır. nk anlamlı olmayan sonuçlar yetersiz gten veya yetersiz rneklem byklęnden kaynaklanabilmektedir. Var olan kk bir etkiyi belirleyememe nedeniyle anlamsız olarak sonulandırılan bir alıřma, yeterli rneklem byklę kullanarak anlamlı olarak bulunabilir (Balkin ve Sheperis: 2011).

Bir alıřmada, istatistiksel olarak anlamlı olmayan sonuçlar elde edildięinde arařtırmacı alıřmasını yeterli bir istatistiksel g seviyesinde kurup kurmadıęından řphe etmeli ve “aradıęım etkiyi bulmak iin alıřmam yeterli istatistiksel gce sahip mi? Acaba ne kadar rneklem kullanmalıydım?” gibi soruları kendisine sormalıdır.

Anlamlı olmayan sonuçlar bir arařtırmacı iin sorundur ve teorik g analizini uygulanması her zaman sonuçların beklenildięi gibi geleceęi garantisini vermez. Teorik g analizini doęru bir etki byklę tahminine dayanmaktadır. Eęer var olan gerek etki, arařtırmanın tasarım ařamasında tahmin edilenden dřk ise sahip olunan mevcut g seviyesinden daha fazlasına ihtiya olacaktır. alıřma sonucunda elde edilecek etkiye gre deneysel g analizini yapmak, tahmin edilen g ile elde edilen g arasında bir kıyaslama yapma olanaęı saęlar. Eęer arařtırmanın tasarım ařamasında tahmin edilen g dřk ıkarsa, arařtırmacı alıřmasını řu řekilde sonulandırabilir: “Elde edilen sonuçlar anlamlı deęil fakat bu, testlerin yeterince gce sahip olmamasından kaynaklanabilir. Dolayısıyla arařtırmanın tekrarlanmaya ihtiyaı vardır.” Eęer elde edilen gcn yeterli seviyede olduęu kanısına varılırsa “sonular kesinlikle negatiftir” denilebilir (Ellis, 2010).

G analizini, bazen anlamlı olmayan sonuçlar elde edildięinde yapılmıř olan testlerin gcn lmek iin kullanılmaktadır. Bu řekilde gerekleřtirilen g analizlerinin amacı yanlış olan bir sıfır hipotezinin yanlışlıkla da olsa doęru olarak aıklanmasını nlemektir. Deneysel g analizleri sadece istatistiksel olarak anlamlı sonuçlar retmeyen alıřmalara mahsustur. Ancak yine de bazı arařtırmacılar, istatistiksel olarak anlamlı sonuçlara rneklem rastlantısı olarak ulařılmıř olabileceęi ihtimaline vurgu yaparak, istatistiksel olarak anlamlı sonuçlar elde edilse bile zellikle kk sayıdaki rneklem ile gerekleřtirilen alıřmalarda deneysel g analizinin uygulanması gerektięini belirtmiřlerdir (Ertrk, 2005).

Yetersiz ve Ařırı Gcn Yol Atıęı Sorunlar

McKean’in (1990) aktardıęına gre istatistiksel g konusunda yazında yapılan bazı incelemelerde, arařtırmacıların g seviyelerini ok fazla řiřirdikleri sonucuna varılmıř, 2011 yılında yapılan farklı bir arařtırmada ise hesaplanan gcn, gerek gcn yaklaşık olarak iki katı olduęu sonucuna ulařılmıřtır. Arařtırmacılar yeterli gce ulařamadıkları iin elde ettikleri gc şiřirerek olması gereken makul seviyeye ekmiřlerdir. Bu řekilde gerekleřtirilen bir alıřmadan elde edilecek sonuçların nemlilięi ise tartıřılacak bir konudur.

Yeterli istatistiksel g ile gerekleřtirilen alıřmalar arařtırmacılara bařarılı olmak iin yeterli řansı saęlamaktadır. Dřk g ile yapılan alıřmalarda řu rnek olaydaki gibi bir sorun vardır:

Ton balıęı avlamak iin kullanılan aę ile kk balıkları avlamaya alıřan bir kiři, yanlış tercih yaptıęı iin muhtemelen hi balık tutamayacaktır. Fakat bu kiři bařarısız olduktan sonra “bu

gölde hiç balık yok şeklinde işini sonlandırırorsa 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).

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'den daha fazla gözlem içeren örnekleme sahip 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).

Tener'in (2000) aktardığına göre Davlin vd. 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 büyüklüğünün olması gerekenden çok daha az olmasıydı. Daha olumlu sonuçlar elde edebilmek için mevcut örneklem büyüklüğünün 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.

Düşük gücün getirdiği negatif sonuçların yanında aşırı yüksek gücün de getirdiği olumsuz sonuçlar vardır. Aşırı güç, gereğinden çok fazla örneklem büyüklüğü demektir ve bu da farazi bir etkinin belirlenmesi olasılığını artırır. 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).

Medikal çalışmalar, β/α oranını değerlendirmek için oldukça uygundur. Birçok tıbbi test neredeyse hiç II. Tip hata yapılmayacak şekilde tasarlanır. Bu durum I. Tip hata yapma olasılığını artırır. Tıp alanındaki bu testleri tasarlayanlar, sağlıklı bir kişiye yapılan testler sonucunda "hastalık olduğuna dair bulgular tespit edildi, daha fazla test yapılması gerekir"

demenin, gerçekte hasta olan kişiye “her şey çok iyi durumda, herhangi bir sorununuz yok” demekten daha iyi olduğunu söylemektedirler (Ellis 2010).

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:

Tablo 3. Aşırı Güç Seviyesinin Neden Olduğu Olumsuz Duruma Bir Örnek

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 araştırmacı, normalde anlamlı olmayan bir farklılığı bu çalışmada anlamlı bir farklılıkmiş gibi gösterecek ve bu şekilde yorumlayacaktır.

Sonuç ve Öneriler

İlgili yazında, istatistiksel güç analizini içeren bilimsel çalışmalar oldukça sınırlı sayıdadır. Bu az sayıdaki bilimsel çalışmalarda kullanılan istatistiksel testlerde güç analizinin ihmal edildiği anlaşılmaktadır. Bu bağlamda bu çalışma, istatistiksel güç analizinin önemine dikkat çekmekte ve istatistiksel güç analizi sürecine odaklanmaktadır.

“İdeal güç seviyesi ne olmalı?” sorusu, cevabı kolay olan bir soru değildir. Çünkü bu durum risk ve getiri açısından bir mübadele gerektirmektedir. Eğer %50 güç seviyesi ile bir çalışma gerçekleştirilirse bu, yanlış bir sıfır hipotezini reddetmek için araştırmacının sanki yazı tura atıyormuş gibi %50 şansa sahip olduğu anlamına gelmektedir. Birçok emek harcanarak gerçekleştirilen bir çalışmanın başarılı olma durumunun yazı tura atma gibi bir şansa bağlanması açıkçası sorgulanması gereken bir durumdur. Eğer güç seviyesi %90 gibi bir seviyede kurulursa, o zaman anlamlı bir etki belirlenme şansı oldukça artacak ve II. Tip hata yapma olasılığı %10’a düşecektir. Bu nedenle, anlamlılık testi içeren çalışmalar Cohen’in önerdiği en az %80 güç seviyesinde olacak şekilde tasarlanmalıdır.

İstatistiksel güç analizleri bireysel çalışmalarda uygulanabileceği gibi bir dergide herhangi bir alanda yayımlanmış olan çalışmalar veya belirli bir alanda yazılmış tez çalışmaları 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 büyüklüğünü ve alfa seviyesini kullanmaktadır. 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.

Özet olarak, araştırmacının tasarım aşamasında güç analizini dikkate almak, araştırmacıya gerekli örneklem büyüklüğünün belirlenmesi konusunda yardımcı olmakla kalmaz, aynı zamanda;

- a. H_0 hipotezi ile ilgili olası yanlış kararları engeller.
- b. Anlamlı olmayan sonuçların doğru yorumlanmasına olanak sağlar.
- c. Araştırmanın yeniden yapılması ihtiyacını azaltarak zaman, çaba ve maliyet açısından araştırmacıya tasarruf sağlar.
- d. Kritik öneme sahip araştırmaların yayımlanmasına olanak sağlar.
- e. Büyük örneklem sayısı nedeniyle belirlenebilecek farazi etki olasılığını azaltır.
- f. II. Tip hatadan daha ciddi olan I. Tip hata yapma olasılığını araştırmacıların kontrol etmelerine olanak sağlar.
- g. Araştırmacılara H_0 hipotezlerini kanıtlama şansını sağlar.

Buradan hareketle şu açıkça söylenebilir ki teorik güç analizi, araştırmanın tasarım aşamasında kesinlikle dikkate alınmalıdır.

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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Chase, L. J. ve Chase, R. B. (1976). A Statistical Power Analysis of Applied Psychological Research. *Journal of Applied Psychology* 61(2): 234-237.
- 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.
- Correll, J. vd. (2020). Avoid Cohen’s Small, Medium and Large for Power Analysis. *Trends in Cognitive Sciences*, in press.
- 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.
- Erdfelder, E., Faul, F. ve Buchner, A. (1996). GPower: A General Power Analysis Program. *Behavior Research Methods, Instruments & Computers* 28(1): 1-11.
- Ertürk, M. Ş. (2005). Retrospective Power Analysis: When?. *Radiology* 237(2): 743-744.
- Flynn, K., Swintek, J. & Johnson, R. (2017). The Influence of Control Group Reproduction on the Statistical Power of the Environmental Protection Agency's Medaka Extended One Generation Reproduction Test (MEOGRT). *Ecotoxicology and Environmental Safety*, 136: 8-13.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect Size Guidelines For Individual Differences Researchers. *Personality and Individual Differences*, 102, 74–78.
- 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.
- Lewis, K. P. (2006). Statistical Power, Sample Sizes and the Software to Calculate Them Easily. *Bioscience* 56(7): 607-612.
- 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.
- 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.
- Meyners, M. vd. (2020). To replicate or not to Replicate, or When Did We Start to Ignore the Concept of Statistical Power. *Food Quality and Preference* 79.
- McKean, K. E. (1990). *Statistical Power Analysis of Doctoral Dissertation Research in Educational Psychology*. Doctoral Dissertation, Oklahoma State University, Oklahoma, USA.
- 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 to Necessary Sample Size. *Journal of Experimental Education 53(1)*: 40-48.
- 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. (2012). 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. *The American Naturalist, 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 XVIII*: 275-290.
- Shavelson R. J. (1988). *Statistical Reasoning for the Behavioral Sciences, 2nd Edition*, Boston, MA: Allyn & Bacon.
- 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, Erişim Tarihi: 15.05.12.
- Thomas, L. ve Juanes, F. (1996). The Importance of Statistical Power Analysis: An Example from Animal Behaviour. *Animal Behaviour 52*: 856-859.