



Bilimsel Makalelerde Raporlanması Önerilen Çıkarımsal İstatistikler

Ömer AKBULUT[✉]

Giresun Üniversitesi, Fen Bilimleri Enstitüsü, Biyosüreç Mühendisliği Anabilimdalı, 28200, Giresun, Türkiye.
<https://orcid.org/0000-0002-8860-3513>, e-posta: omer.akbulut@giresun.edu.tr

ÖZET

Bilimsel arařtırmaların planlanması yürütülmesi ve sonuçlandırılmasında yaygın olarak çıkarımsal istatistik tekniklerden yararlanılmakta ve istatistiksel bulgular raporlanmaktadır. Bu çalışmada, arařtırmalarda p-değerinin yanında raporlanması gereken diğeri istatistikler irdelenmiştir. Arařtırmalarda en sık kullanılan çıkarımsal istatistik yöntemi "Yokluk Hipotezi Anlamlılık Testi"dir. Bu yöntemin son çıkarımsal istatistiğı ise p- değeridir. İstatistik testler sonucu elde edilen bu değer gerçek değeriyle ve üç ondalık basamakla rapor edilmelidir. P- değerinin tek başına sunumundan kaçınılmalıdır. Ayrıca ifade ettiğinden daha fazla anlam yüklenilmemelidir. Bu değer test istatistiğı (t, z, χ^2 , F vb.) ile birlikte yazılmalıdır. Ayrıca arařtırma makalelerinde örneklem büyüklüğü, mutlaka belirtilmelidir. Bunun yanında, testin gücü, güven aralığı ve etki büyüklüğü istatistiklerine yer verilmesinde yarar vardır. Çünkü p-istatistiğı ve testin gücü örneklem büyüklüğünden önemli ölçüde etkilenir. Örneklem büyük olması, p- değerinin küçük, testin gücünün yüksek olmasına neden olmaktadır. Etki büyüklüğü ise örneklem büyüklüğünden etkilenmemektedir. Bu nedenle istatistiksel yorumlar özellikle etki büyüklüğü ve güven aralığı kullanılarak yapılmalıdır.

Çıkarımsal İstatistikler

Teknik Not

Makale Tariğesi

Geliş Tarihi : 13.04.2023

Kabul Tarihi : 06.07.2023

Anahtar Kelimeler

Çıkarımsal istatistik

P-değeri

Güven aralığı

İstatistiksel güç

Etki büyüklüğü

Inferential Statistics Suggested to Report in Scientific Articles

ABSTRACT

Inferential statistical techniques are widely used in the planning, execution and conclusion of scientific research and statistical findings are reported. In this study, other statistics that should be reported in addition to the p-value in studies were examined. The most frequently used inferential statistical method in research is the "Null Hypothesis Significance Test". The final inferential statistic of this method is the p-value. This value obtained as a result of statistical tests should be reported with its true value and three decimals in scientific works. P-value should not be given more meaning than it expresses and should be reported together with the test statistic (t, z, χ^2 , F etc.). In addition, the sample size should be specified in research articles. In addition, it would be beneficial to include the power of the test, confidence interval and effect size statistics. Because the p-value and the power of the test are significantly affected by the sample size. The large sample size causes the p-value to be small and the power of the test to be high. The effect size is not affected by the sample size. Therefore, statistical inference should be inferred using, especially effect size and confidence interval.

Inferential Statistics

Technical Note

Article History

Received : 13.04.2023

Accepted : 06.07.2023

Keywords

Inferential statistics

p-value

Confidence interval

Statistical power

Effect size

Atf İçin : Akbulut, Ö.,(2024). Bilimsel Makalelerde Raporlanması Önerilen Çıkarımsal İstatistikler. *KSÜ Tarım ve Doğa Derg* 27 (1), 238-247. DOI: 10.18016/ksutarimdog.vi.1282850

To Cite: Akbulut, Ö.,(2024). Inferential Statistics Suggested to Report in Scientific Articles. *KSU J. Agric Nat* 27 (1), 238-247. DOI: 10.18016/ksutarimdog.vi.1282850.

GİRİŞ

Bilimsel çalışmalarda istatistiksel düşünme ve istatistiksel analiz önemli bir yere sahiptir. Ekonomi,

tarım, mühendislik, sağlık, eğitim, vd. tüm alanlarda yapılan bilimsel çalışmalarda istatistiksel analiz tekniklerinden yararlanılmaktadır. Bilimsel çalışma

sürecinin; hipotez kurma, değişken belirleme, örnekleme, ölçme, veri toplama, veri analizi ve yorumlama vd. tüm aşamalarında istatistik teknikler yoğunlukla kullanılmaktadır. Bu teknikler içinde yaygın kullanılan yöntem “Yokluk Hipotezi Anlamlılık Testi, (YHAT)” yöntemidir (Yıldırım & Yıldırım, 2011; Erkuş 2017; Akbulut, 2022). YHAT kısaca “Hipotez Testi” (HT) olarak ifade edilebilir. YHAT'nin son çıkarımsal istatistiği “p-değeridir”. Leek & Peng (2015)'in ifadesi ile “p-değeri bilimsel araştırma ve veri analizi sürecinde buzdağının görülen ucudur”. Araştırmaların çoğunda bulgular p-değerine odaklanarak çoğu kez de p-değeri tek başına yorumlanarak sunulmaktadır. P-değerinin tanımı (Cohen, 1994; Yıldırım & Yıldırım, 2011; Aschwanden, 2016) yorumlanması (Cohen, 1994; Cohen, 2011; Kılıç, 2011; Işık 2014; Lu & Belitskaya–Levy, 2015; Halsey et al., 2015) ve önemi (Hojat & Xu, 2004; Dahiru, 2008; Lu & Belitskaya–Levy, 2015; Halsey et al., 2015) üzerinde son yıllarda yoğun değerlendirme ve tartışmalar yapılmaktadır. P-değerinin hatalı yorumlanması konusunda (Nuzzo, 2014; Lu & Belitskaya–Levy, 2015; Greenland et al., 2016; O'Leary, 2021) ve kötüye kullanımı (Vidgen & Yasseri, 2016) konusunda da makaleler yayınlanmıştır. Bu bağlamda yayınlanan makalelerin genel sonucu p-değerinin hatalı yorumundan kaçınılması ve p-değeri ile birlikte diğer çıkarımsal istatistik ölçülerin raporlanması şeklindedir (Goddman, 2008; Wasserstein & Lazar 2016; Mark et al., 2016; Wasserstein, et al., 2019; Gao, 2020; Akbulut, 2022). Bu bağlamda önerilen istatistik ölçüler etki büyüklüğü (EB), güven aralığı (GA) ve testin gücü (TG) ölçüleridir (Whitley & Ball, 2002; APA 2010; Kul, 2014; Işık, 2014; Greenland et al., 2016; Solla et al., 2018; Balkin & Lenz, 2021).

Bu makalede p-değeri ile birlikte sunulması gereken EB, GA ve TG çıkarımsal ölçüleri ele alınmış ve bu ölçülerin hesaplanması ve raporlanması sayısal örnekler ile açıklanmıştır.

α DEĞERİ ve P-DEĞERİ

İstatistiksel analizlerde α değeri, araştırmanın planlanması, örneklem büyüklüğü ve beklenen gücün hesaplanması aşamasında karar verilen I. Tip hata olasılığıdır. I. Tip hata gerçekte yokluk/sıfır hipotezi doğru olduğu halde, test sonucunda karşıt hipotezin kabul edilmesi olasılığıdır. Bu olasılık genellikle 0.05 veya 0.01 olarak alınır.

İstatistiksel anlamlılığı ifade eden p-değeri ise, örneklem verilere YHAT veya diğer istatistiksel testlerin uygulanmasının bir sonucudur. Bilimsel makalelerde istatistiksel anlamlılık p-değeri “P” veya “p” ile sembolize edilmektedir. Kısaca ifade etmek gerekirse α , veri analizi öncesinde karar verilen olasılık değeri, p ise istatistiksel analiz ile hesaplanan olasılık değeridir. Bilimsel araştırmaların

raporlanmasında en yaygın kullanılan α ve p olasılıklarının önemi, anlamı, doğru raporlanması ve doğru yorumlanması konusunda Akbulut (2022) tarafından kapsamlı bir çalışma yapılmıştır. Ayrıca bu konuda Türkçe literatürde başka çalışmalar da mevcuttur (Kılıç, 2011; Yıldırım & Yıldırım 2011; Işık, 2014; Kul, 2014; Erkuş 2017; Ünal, 2021).

Anlamlılık durumunu ifade eden p-değeri gerçek değeri ile ve üç ondalıkla ($p=0.028$ veya $p=0.002$ gibi) yazılmalıdır. İstatistik analizlerde söz konusu olasılık 0'dan büyük 1'den küçük ($0 < p < 1$) bir değer alır ve hiçbir zaman sıfır olamaz. Ancak bazı istatistiksel analiz yazılımları (SPSS gibi) çok küçük p-değerini üç ondalıkla yuvarlayarak $p=0.000$ şeklinde sunmaktadır. Bu durumda $p < 0.001$ şeklinde yazılmalıdır. Bulguların yorumlanmasında p değerine gereğinden fazla anlam yüklenmemeli ve p tek başına verilmemelidir. P değeri, örneklem büyüklüğü (n) veya serbestlik derecesi ve test istatistiği ile birlikte (z, t, χ^2 , F, gibi) rapor edilmelidir.

ETKİ BÜYÜKLÜĞÜ

Etki büyüklüğü (EB), örneklem verilerinden hesaplanan istatistiklerin yokluk/sıfır hipotezi ile tanımlanan değerden sapma düzeyini gösteren istatistiksel bir ölçüdür. (Cohen, 1994; Sullivan & Feinn 2012; Özsoy & Özsoy, 2013, Çapık, 2014; Ünal, 2021). Bu ölçü bir fark veya bir katsayıdır. En basit tanımıyla EB, yeni yöntemin eskisine kıyasla ne büyüklükte bir fark oluşturduğunu ifade eder (Kılıç, 2014). Kalaycıoğlu & Akhanlı (2020) EB'yi sağlık alanı için gruplar arasında klinik (diğer alanlarda teknik, ekonomik) anlamlı olan en küçük fark olarak tanımlamıştır. Bu fark değeri standartize edilerek, örneğin ortalamalar arası fark ortak standart sapmaya oranlanarak, karşılaştırılabilir bir şekle dönüştürülür (Yıldırım & Yıldırım, 2011; Çapık, 2014; Ünal, 2021).

EB hesaplama yöntemleri araştırma tasarımı ve kullanılan istatistik test yöntemine göre farklılık gösterir. Araştırmanın tasarımı aşamasında örneklem büyüklüğünün belirlenmesinde EB ihtiyaç duyulan bir ölçüdür. EB hedef populasyondan ön çalışma örnekleme kullanılarak belirlenebilir. Eğer ön çalışma ile yeterli güvenilirlikte etki büyüklüğünün belirlenememesi durumunda EB için düzenlenmiş standart değerlerden yararlanılarak örneklem büyüklüğü belirlenir. Çizelge 1'de yaygın kullanılan bazı istatistik testler için standart EB sınıflarının (küçük, orta, büyük) sınırları verilmiştir. Örneklem büyüklüğünü hesaplamada bu sınıfların genellikle alt sınırları kullanılır.

EB belirlenirken küçük etkilerin ortaya çıkartılmasının çok güç olması nedeniyle daha büyük örneklem gerektirdiği unutulmamalıdır.

Aslında EB kavramı günlük hayatta yaygın olarak

kullanılmaktadır. Örneğin “Yeni diyet haftada 1500 gr. zayıflama garantisi vermektedir, binada ısı yalıtımı en az %40 yakıt tasarrufu sağlamaktadır, ilave yemleme koyunlarda ikizlik oranını %25

artırmaktadır” gibi ifadeler ile ifade edilen EB’dir. EB’nin rapor edilmesi ve yorumlanması, beş önemli nedene dayandırılmaktadır. Bunlar;

Çizelge 1. İstatistiksel testlere göre etki büyüklüğü sınıfları ve sınır değerleri
Table 1. Effect size classes and its limit values according to statistical tests

| İstatistiksel testin amacı | İstatistiksel test | Etki büyüklüğü ve sembolü | Etki büyüklüğü sınıfları | | |
|---|--|--------------------------------------|--------------------------|-----------|-------|
| | | | Küçük | Orta | Büyük |
| Gruplar arası fark; d ailesi etki büyüklüğü | | | | | |
| İki oran farkı | Ki-Kare | Odds Oranı | 1.68-3.46 | 3.47-6.70 | ≥6.71 |
| İki ortalama farkı | t Test, | Cohen d , veya Hedge g | 0.20-0.49 | 0.50-0.79 | ≥0.80 |
| Oranlar arasındaki fark | Z test veya Ki-Kare | Cohen’in g istatistiği | 0.05-0.14 | 0.15-0.19 | ≥0.20 |
| Değişkenler arası ilişki; r ailesi etki büyüklüğü | | | | | |
| İki sürekli değişken arası ilişki | Korelasyon | Perarson’un r değeri | 0.10-0.29 | 0.30-0.49 | ≥0.50 |
| Çapraz tablo analizi | Ki-Kare | Cramer’in V , φ | 0.10-0.29 | 0.30-0.49 | ≥0.50 |
| Sürekli ve sıralı değişkenler arası ilişki | Pearson veya Spearman Rank korelasyonu | r ve ρ | 0.10-0.29 | 0.30-0.49 | ≥0.50 |
| İkiden fazla bağımsız grup veya bağımlı grupların arasındaki fark | ANOVA veya tekrarlı ölçümlerde ANOVA | η ² | 0.01-0.05 | 0.06-0.13 | ≥0.14 |
| | | f | 0.10-0.24 | 0.25-0.39 | ≥0.40 |
| Fonksiyonel bağıntı analizi | Çoklu Regresyon | R ² | 0.02-0.12 | 0.13-0.25 | ≥0.26 |

Kaynak: Özçomak ve Çebi (2017); Akbulut, (2021); Balkin ve Lenz, (2021)

- i) Araştırma bulgularının pratik olarak “düşük” “orta” ve “yüksek” şeklinde değerlendirilmesini sağlar ve kullanım açısından yorumlanmasını kolaylaştırır (Hojat & Xu, 2004).
- ii) Aynı konuda farklı çalışmaların rapor edilen sonuçlarının karşılaştırılmasına imkân sağlar.
- iii) Daha sonra aynı kapsamda yapılacak araştırmalarda raporlanan EB değerleri dikkate alınarak gerekli ve yeteri örneklem büyüklüğünün belirlenmesini sağlar (Téllez et al., (2015).
- iv) Klinik, teknik veya ekonomik yüksek EB ile sonuçlanan çalışmaların istatistiksel anlamsız bulunması durumunda, araştırmanın daha büyük örneklem ile tekrarlanarak gerçek anlamlı etkinin belirlenmesine ışık tutar.
- v) EB istatistiksel anlamlılık ve diğer çıkarımsal istatistiklere göre iki noktada avantaja sahiptir. Bunlar a) Diğer çıkarımsal istatistikler örneklem büyüklüğünden etkilenirken EB örneklem büyüklüğünden etkilenmez. b) EB skalası serbest bir indekse sahiptir (Hojat & Xu, 2004; Nelson et al., 2015).
- vi) Verilerin ön görülen dağılıma uygunluğu

(Örneğin normal dağılım) varsayımı altında istatistiksel anlamlılık, güven aralığı ve istatistiksel güç için doğru tahminler yapılabilir. Dağılım geçerli değil ise tahminler de hatalı olacaktır. Hâlbuki EB teorik dağılımlardan bağımsız bir ölçüdür.

Bu nedenlerle araştırma raporlarının metot bölümünde ön görülen ve kullanılan EB, bulgular kısmında ise gözlenen EB rapor edilmelidir.

EB ve TG istatistiklerinin araştırma makalelerde raporlanması gerekliliği belirtilmesine rağmen Türkçe literatürde bu ölçülerin raporlanması henüz istenilen düzeyde yaygınlaşmamıştır. Eğitim bilimleri alanında Türkiye kaynaklı yayımlanan ve SSCİ’da taranan bilimsel dergilerdeki makalelerin sadece %7.2’sinde EB’nin raporlandığı bildirilmektedir (Özsoy & Özsoy, 2013). Türkçe literatürde hipotez testlerinin yaygın kullanıldığı tarım bilimleri ve mühendislik alanındaki çalışmalarda p-değeri sıklıkla raporlanırken EB’yi raporlayan çalışmaya rastlanılmamıştır. Halbuki Sullivan & Feinn (2012). Cohen’e atfen EB’nin önemini “Bir araştırma sorusunun birincil ürünü p değerleri değil, bir veya daha fazla EB ölçüsüdür” şeklinde özgün bir ifade ile vurgulamıştır. Kılıç (2014) istatistiksel anlamlılıkla birlikte EB ve EB’nin güven

sınırlarının raporlanmasını önermektedir. Aynı şekilde Téllez et al., (2015) APA editörlerine atfen istatistiksel anlamlılık testlerinin bazı noktalardaki zayıflığı nedeniyle, EB ve EB'nin güven aralığının raporlanmasına dikkat çekmişlerdir.

GÜVEN ARALIĞI

Güven aralığı (GA) belirli bir olasılıkla populasyon parametrelerinin hangi değerler arasında (alt ve üst sınırlar) bulunabileceğini ifade eder. GA'nın bu değerler arasında bulunma olasılığı güven katsayısı $1-\alpha$ ile ifade edilir. Bu değer yüzle çarpımına ise $100(1-\alpha)$ güven düzeyi denir. GA'nın genel hesaplama formülü,

$$GA = \text{istatistik} \pm (\text{standart hata}) \times \left(\text{dağılımın } \frac{\alpha}{2} \text{ olasılığındaki değeri} \right)$$

şeklindedir. GA'nın doğru tahmini için örneklemin rastgele yapılması, örnekleme giren birimlerin birbirinden bağımsız olması gerekir. Değişkenin normal dağılıma uyum göstermemesi durumunda uygun tekniklerle GA tahmini yapılmakla birlikte (Cebeci, 2020), uygulamada daha çok değişkenin normal dağılım göstermesi şartı altında yapılır.

Örneğin varyansı $\sigma^2 = 100$ olan bir populasyondan alınan 25 birimlik bir örneklemin ortalaması 40 ise, %95 güvenle ile ($\alpha=0.05$) bu populasyonun ortalamasına ait güven aralığı hesaplanmak istensin. Bu örnek için, standart hata, $\sigma_{\bar{x}} = \sqrt{100/25}=2$ 'dir. Ayrıca $\alpha/2=0.025$ olup Z dağılımının sağ kuyruğunda

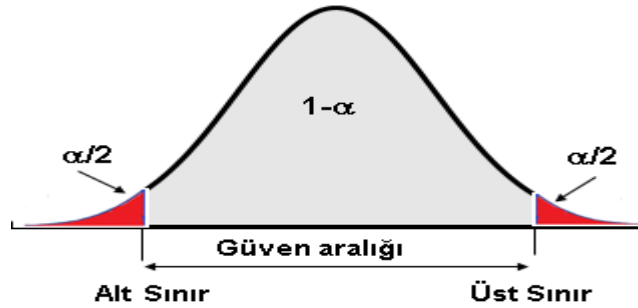
$Z_{0.025}$ olasılığı için kesme değeri 1.96'dır. Buradan, $\bar{X} \mp \sigma_{\bar{x}} Z_{\alpha/2}$ formülü ile söz konusu populasyonun ortalaması %95 güvenle $(40 \pm 2 \cdot 1.96)$ 36.08 ile 43.92 arasında olacaktır.

Bu sonuç şu şekilde de yorumlanabilir: Bu populasyondan 100 kez örneklem alınması halinde bunların %95'inde \bar{X} , 36.08 ile 43.92 arasında değer alacaktır. Bu tanımlı genelleştirmek gerekirse, güven aralığı tahmini, tekrarlanan örneklemler yapıldığında, ön görülen güven düzeyinde populasyon ortalamasının hangi değerler arasında olabileceğinin bir tahminini sunar.

Uygulamada populasyon varyansı σ^2 genellikle bilinmez ve örneklem de çoğu kez sınırlı ($n < 30$) büyüklüktedir. Bu durumlarda GA tahmini için örneklem varyansı S^2 kullanılır.

Örnekleme varyansının (S^2) kullanılması durumunda yukarıdaki eşitlik $\bar{X} \mp s_{\bar{x}} t_{\alpha/2(n-1)}$ şekline dönüşür (Walpole, 1969). Burada örneklem büyüklüğüne bağlı t dağılımına ait kritik değerin ve örneklem ortalamasının standart hatası $s_{\bar{x}} = \sqrt{S^2/n}$ istatistiğinin kullanıldığına dikkat edilmelidir.

GA belirli bir güven düzeyinde populasyon parametresinin içinde bulunabileceği alt ve üst sınırları belirlenmiş sayılar kümesidir. Güven aralığını ve güven sınırlarını Şekil 1'deki gibi göstermek mümkündür.



Şekil 1: Normal dağılım varsayımı altında güven aralığı ve güven sınırları

Figure 1: Confidence interval and confidence limits under the assumption of normal distribution

GA belirlenen bir güven düzeyinde populasyon parametresinin alabileceği değerleri belirleyen örneklemden populasyona doğru yapılan bir tahmindir. Güven sınırları veya güven aralığı uygulamada genellikle %95 veya %90 güven düzeylerinde bildirilir. (Yıldırım & Yıldırım 2011). Güven aralığını yine %5'lik anlamlılık düzeyinde ($\alpha=0.05$) reddedilemeyecek yokluk/sıfır hipotezi değerler kümesi olarak tanımlamak mümkündür. Yani yokluk hipotezi anlamlılık testi ile ulaşılan sonuçlara güven aralığı ile de ulaşılabilir (Yıldırım & Yıldırım 2011). Aynı durum oran testleri ve ilişki katsayılarının testi için de geçerlidir.

Dahiru (2008) p-değerinin doğru yorumlanması ve p-değerini destekleyecek diğer ölçüt olarak güven

aralığının (GA) p-değeri ile birlikte sunumunun faydalı olacağına dikkat çekmektedir. Güven aralığının hesaplanmasının sağladığı avantajlar Dahiru (2008) ve Yıldırım & Yıldırım (2011) tarafından, aşağıdaki gibi ifade edilmiştir.

- Çalışma öncesi yokluk hipotezi kurulmasını gerektirmez. Güven aralıkları hipotez testinin anlamlı veya anlamsız sonuç vermesinden daha çok bilgilendiricidir.
- Araştırmalardan elde edilen ortalama, oran vb. ölçülerin mutlak olarak anlaşılmasını önler. Güven aralığının genişliği, tahminin güvenilirliği veya kesinliği örneklemden elde edilen değerlerin bu ölçü için hesaplanacak değerlerden biri olduğunun anlaşılmasını

kolaylaştırır.

- iii- Güven aralıkları, istatistiksel anlamlılığın aksine, bir bulgunun önemli olup olmadığı yorumlanmasını (örneğin, klinik önem) daha kolay hale getirir.
- iv- İstatistiksel anlamlılık testlerinin I. tip hataya karşı önlem almadığı halde, güven aralığı için böyle bir durum söz konusu değildir.
- v- Güven aralığının genişliği, çalışmanın yenilenmesi durumunda benzer bulgulara ulaşılma ihtimali hakkında bilgi verir.
- vi- Güven aralıkları bilgisi meta analiz çalışmalarında farklı çalışmalardan elde edilen bilgilerin birleştirilmesini kolaylaştırır.

Bunlara ilave olarak güven aralıkları kullanılarak benzerlik veya farklılıklar karşılaştırmalı olarak daha objektif değerlendirilebilir. Bu bağlamda karşılaştırılan gruplara ait istatistiklerin tartışılmasında daha aydınlatıcı olduğu gibi kolaylık sağlar. Ayrıca GA istatistiği, araştırma sonuçlarının literatür bildirimleri ile karşılaştırılmasında benzerlik ve farklılıkları belirlemede daha tutarlı yorum yapmaya imkân verir. Yani; ortalama, oran vb. nokta tahminlerine ait bulgular literatür ile karşılaştırmalı olarak tartışılırken tarım bilimlerinde genellikle "...bildirimleri ile benzer, ...bildirimlerinden küçük veya daha az, ...bildirimlerinden daha yüksek, büyük" gibi göreceli ifadeler kullanılmaktadır. Güven sınırları tahmini yapıldığında güven aralığındaki bildirimler benzer, alt sınırdan küçükler daha düşük, üst sınırdan büyükler daha yüksek olarak yorumlanabilir. Böylece güven sınırları kullanılarak yapılacak tartışma ve yorumlar ile daha objektif çıkarımlar yapılabilecektir.

Ancak aynı α hatası düzeyi veya $1-\alpha$ güven düzeyinde örneklem büyüklüğü arttıkça güven aralığı daralır. Yani parametre için birbirine daha yakın sınırlar tahmin edilir. Yani güven aralığı p-değerinde olduğu gibi örneklem büyüklüğü arttıkça küçülür. Bu nedenle aynı değişken için farklı çalışmalarda yapılan güven aralığı tahminlerinin karşılaştırılmasında araştırmaların örneklem büyüklükleri de dikkate alınarak yorumlanmalıdır.

İSTATİSTİKSEL GÜÇ, TESTİN GÜCÜ

YHAT ile çözümlenen araştırmalarda rapor edilmesi gereken bir diğer çıkarımsal istatistik testin gücü (TG) istatistikidir. TG; $1-\beta$ olup burada β , istatistiksel analizlerdeki II. Tip hata düzeyidir. II. Tip hata gerçekte karşıt hipotez doğru olduğu halde, test sonucunda yokluk hipotezinin kabul edilmesi olasılığıdır. II. Tip hatanın β genellikle en fazla 0.20 olması istenir. Bu durumda yapılacak YHAT sürecinde hedeflenen testin gücü en az 0.80 olacaktır.

TG, testin gerçekte yanlış olan yokluk hipotezinin ret edilme olasılığıdır. Diğer bir ifade ile testin gücü bir

testin gerçekte var olan fark veya etkiyi bulabilme olasılığıdır (Kalaycıoğlu ve Akhanlı, 2019; Ünalın, 2021). TG tıpkı p-değeri gibi şartlı bir olasılıktır. Olasılık terminolojisi ile

$TG=1-\beta = P(H_0 \text{ Ret} | H_0 \text{ Yanlış})$ şeklinde ifade edilir (Sun et al., 2011).

Bir çalışmada TG arttıkça hem doğru etkiyi bulma şansı hem de yokluk hipotezini reddetme şansı artmaktadır. Diğer taraftan bir araştırmada TG yetersiz ise hem doğru etkiyi bulma şansı hem de yokluk hipotezini reddetme şansı azalır.

TG, anlamlılık düzeyi α , örneklem büyüklüğü n ve EB'nin bir fonksiyonudur. Yani hipotez testi sürecinde $TG=f(\alpha, n, EB)$ şeklinde ifade edilebilir (Özçomak & Çebi, 2017; Keskin, 2020). Bunların dışında TG, istatistiksel testin çeşidi, araştırmadaki grup sayısı, gruplardaki gözlem sayısının farklı olması, örneklem hatası, üzerinde çalışılan değişkenin varyasyonu ve testin yönü (tek yanlı veya iki yanlı) tarafından da etkilenir (Keskin, 2020). TG, anlamlılık seviyesi α , EB ve örneklem büyüklüğü ile doğru orantılı, standart sapma ile ters orantılı değişir. Yani α , EB ve örneklem büyüklüğü arttıkça TG artarken, değişkenlik ölçüsü standart sapma arttıkça TG azalır. Ayrıca TG, tek yönlü testlerde iki yönlü testlere göre daha yüksektir.

TG, teorik güç (prospective power, a priori power,), deneysel güç (retrospective power, observed power, post hoc power, archived power) ve karşılaştırmalı güç (compromise power) olmak üzere üç çeşittir (O'Kafee, 2007; Özçomak & Çebi, 2017; Keskin, 2020). Ancak karşılaştırmalı güç diğer güç analizlerine göre daha az bilinen ve uygulanan bir güç türüdür (Özçomak & Çebi, 2017).

Teorik güç; araştırmanın planlanması aşamasında örneklem büyüklüğü belirlenirken öngörülen α (genellikle 0.05), EB ve diğer etkenler dikkate alınarak örneklem büyüklüğünü hesaplamada kullanılır. Bu aşamada kullanılan teorik güç genellikle 0.80 olarak alınır. Bununla birlikte II. Tip hatayı daha küçük ($\beta=0.10$) alarak daha büyük bir örneklem ile daha güvenilir sonuçlar elde edebilmek için testin gücü 0.90 olarak da alınabilmektedir (Kalaycıoğlu & Akhanlı, 2020; Ünalın, 2021).

Keskin, (2020) teorik güç analizinin yapılmasını güç analizi için ideal kabul etmektedir. Teorik güç analizini araştırmanın tasarımı aşamasında başlangıçta yapmak oldukça yararlıdır. Araştırmalar planlanırken %80 güç, öngörülen anlamlılık düzeyi α ve klinik, teknik veya ekonomik farklılık veya ilişkiyi (EB) test edebilecek yeterli ve gerekli EB'nin belirlenmesi, araştırma tasarımının doğru yapılmasına imkân verir. Araştırmanın başında yeterli güç sağlamayan çalışmalar öngörülen araştırma tasarımı yerine yeni bir tasarım yapılabilmektedir (Lewis, 2006). Teorik güç analizinin yapılması, bir çalışmada II. Tip hata yapma

olasılığının belirlenmesi için de kullanılmaktadır (Gürkan 2007).

Bir diğer güç analizi yöntemi karşılaştırmalı güç analizidir. Karşılaştırmalı güç analizinin yapılması yaygın değildir. Ancak TG türlerinden biri olarak ifade edilmektedir. (Özçomak & Çebi 2017; Keskin 2020). Karşılaştırmalı güç analizi ölçüsü üzerine örneklem büyüklüğü, EB, α ve β/α şeklinde ifade edilen hata oranlarının bir sonucudur (Özçomak & Çebi 2017). Yani karşılaştırmalı güç değerinin belirlenmesinde teorik güç parametrelerine ilave olarak hata oranı parametresi β/α da dikkate alınmaktadır. Rutin hata düzeyleri $\alpha=0.05$, $\beta=0.20$ alındığında, hatalar oranı $\beta/\alpha = 4$ standart değer olarak kullanılmaktadır.

Araştırma raporunda teorik güç değeri ve karşılaştırmalı güç kriterleri araştırmanın metot bölümünde istatistiksel analizler alt başlığında α olasılığı, EB ile birlikte belirtilir.

Deneysel güç veya gözlenen güç, YHAT analizlerinden sonra yapılan güç analizidir. Bazı yazarlar bu güç analizinin yapılmasını doğru bulmamaktadır (Plate, et al., 2018). Fakat bazen istatistiksel olarak anlamlı sonuçlar elde edilemeyen araştırmalarda, deneysel güç analizi yapmak yararlı olmaktadır (Özçomak & Çebi 2017). Çünkü istatistiksel olarak anlamlı olmayan sonuçlar yetersiz örneklem büyüklüğünden kaynaklanabilmektedir. Bu durumda var olan etkiyi kullanarak, çalışma yeterli örneklem büyüklüğüne ulaştırılarak anlamlı sonuçlar bulunabilir (Keskin, 2020). Ancak Sun et al. (2011) istatistiksel olarak anlamlı bulunmayan testler için güç analizi yapmanın bulgulara bir katkısının olmadığını bildirmişlerdir. Bu yazarlar, istatistiksel anlamlı sonuçlar bulunmayan durumlarda deneysel güç yerine GA ve gözlenen EB'nin raporlanmasını önermişlerdir. Ayrıca istatistiksel anlamlı bulunmayan ancak yüksek güç ile sonuçlanan araştırmalarda yokluk hipotezinin ret edilmesinde tereddüt edilmemelidir. Diğer yandan istatistiksel anlamlı bulunan sonuçlar için yokluk hipotezinin hangi güç ile ret edildiğinin bilinmesi araştırma bulgularına önemli katkı sağlamaktadır (Işık, 2014; Keskin, 2020). Bazı araştırmacılar, istatistiksel olarak anlamlı sonuçların örneklemin rastlantısal bir sonucu olabileceği ihtimalini dikkate alarak, istatistiksel anlamlı sonuçlar için özellikle küçük örneklemlemler ile gerçekleştirilen çalışmalarda deneysel güç analizinin gerekli olduğunu belirtmişlerdir (Ertürk, 2005). Keza APA yazım standardını esas alan dergilerde ve son yıllarda özellikle sağlık ve psikoloji alanında yayınlanan makalelerde deneysel güç analizi bulgularının da rapor edilmesi istenmektedir. Deneysel güç araştırmanın bulgular bölümünde ve genellikle veri analiz tablolarında sunulur.

Son yıllarda Türkçe literatürde yayınlanmış

araştırmaların gücünü inceleyen araştırmalar yaygınlık kazanmıştır (Çapık,2014; Özçomak & Çebi 2017; Şevgin & Çetin 2017). Bu kapsamda analitik istatistik yöntemlerinin (test) uygulandığı yayınlanmış araştırmalar incelenerek Çapık (2014) hemşirelik araştırmalarında, Özçomak & Çebi (2017) işletme-iktisat alanlarında Şevgin & Çetin (2017) ise eğitim araştırmalarında güç ve etki büyüklüğünü hesaplayarak bulguları yorumlamışlardır. Çapık (2014) 61 araştırmada sonuçları verilen 725 test sonucunu incelemiş ve bu makalelerin sadece birinde istatistiksel gücün hesaplandığı ve makale metninde rapor edildiğini bildirilmiştir. Şevgin & Çetin (2017) 25 makalenin sadece birinde güç analizinin rapor edildiğini, Özçomak & Çebi (2017) inceledikleri 95 makalenin hiçbirinde güç hesaplaması yapılmadığı ve rapor edilmediğini tespit etmişlerdir. Işık (2014) ise psikoloji alanında yayın yapan Türk Psikoloji Dergisinde 1995-2013 yılları arasını 5'er yıllık periyotlarla EB raporlama oranı bakımından incelemiştir. Yazar yıllar itibariyle başlangıçta %1 düzeyinde olan oranın son 5 yılda %18'e ulaştığını tespit etmiştir. Bu araştırmaların genel sonucu olarak, analitik istatistiksel metotların yer aldığı makalelerde güç analizinin önemini yeterince kavranmadığını vurgulamışlardır. Araştırmacılar istatistik analiz yapılan çalışmalarda EB ve güç analizi sonuçlarının rapor edilmesi gerektiğine dikkat çekmişlerdir.

Anlamlılık Testi, EB, TG ilişkisini şekilsel olarak aşağıdaki gibi göstermek mümkündür (Şekil 2).

HİPOTEZ TESTİ VE GÜVEN SINIRLARI İÇİN SAYISAL BAZI ÖRNEKLER (1)

Bağımsız İki Grup Ortalamasının Karşılaştırılması (Sayısal Örnek 1)

A ve B gibi iki bağımsız grupta ölçülen sistolik kan basıncı değerlerine ait veriler analiz edilmiştir. Ön analizde verilerin normal dağılımlı ve varyansların homojen olduğu tespit edilmiştir. Araştırmanın tasarımı ve ön analiz bulguları doğrultusunda, veriler bağımsız gruplarda t testi ile analiz edilmiştir. İki yönlü test için bulgular aşağıdaki çizelgede özetlenmiştir (Çizelge 2). Çizelge 2'de görüldüğü gibi tanımlayıcı istatistiklerin yanında sadece p değeri değil, GA, EB ve TG çıkarımsal istatistikleri de verilmiştir. Bu istatistiklerden yararlanarak araştırma bulgularının yorumlanması ve tartışılması daha fazla ve daha doğru bilgi elde edilmesine imkân verecektir.

Bu analiz sonuçları şu şekilde yorumlanabilir: Gruplar arasında 11.9 birimlik fark tespit edilmiştir. Bu fark istatistiksel olarak anlamlı ($t_{(30)}=3.12$; $p=0.004$) bulunmuştur. Testin gücü 0.87 ve etki büyüklüğü $d=1.12$ olarak her iki ölçünün yüksek düzeyde olduğu

¹ : Örnek çalışmalar sanal olup veriler simülasyonla üretilmiştir.

tespit edilmiştir.

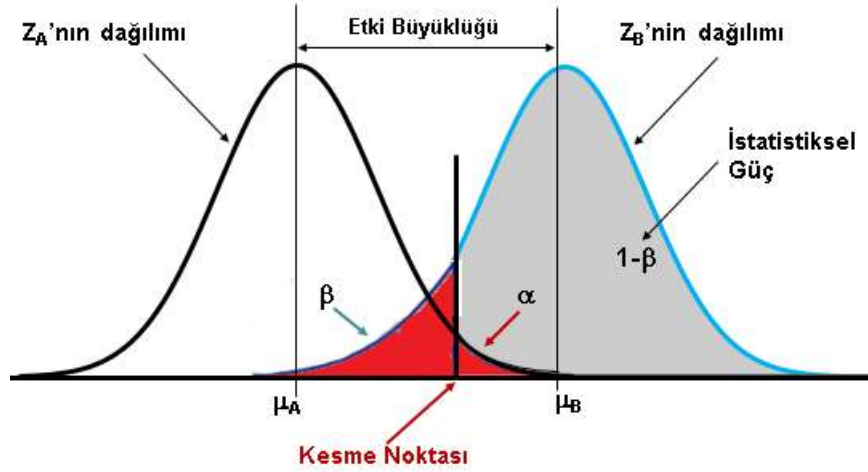
Çizelge 2’de özetlenen bulgular şu şekilde hesaplanmıştır: Ortalamalar arası fark, $\Delta = 122.8 - 110.9 = 11.9$ olarak hesaplanır.

GA, EB ve daha ileri testler için ortak standart sapma

(S_o) ve ortak standart hata ($S_{\bar{x}_o}$) gereklidir. $S_o = n_A$ ve n_B eşit olduğu durumda

$$S_o = \frac{S_A + S_B}{2} \quad (1)$$

şeklinde dir. Buradan $S_o = \frac{9.0 + 12.3}{2} = 10.65$ ’dir.



Şekil 2. Anlamlılık testi, etki büyüklüğü ve istatistiksel güç
Figure 2. Significance test, effect size and statistical power

Çizelge 2. Sayısal örnek 1 için tanımlayıcı ve çıkarımsal istatistikler

Table 2. Descriptive and inferential statistics for the numerical example 1

| Grup | n | Ortalama | Standart sapma | Fark, GA, EB ve GA _(EB) | İstatistik anlamlılık ve güç |
|------|----|----------|----------------|--|--------------------------------|
| A | 16 | 122,8 | 9,0 | $\Delta = 11.9$; GA = [4.1-19.7] | $t_{(30)} = 3.12$; p = 0.004; |
| B | 16 | 110,9 | 12,3 | d = 1.12; GA _(EB) = 0.34-1.90 | Güç = 0.87 |

GA: Güven aralığı, EB: Etki büyüklüğü

Örneklem büyüklükleri $n_A = n_B$ eşit olduğu durumda ortak standart hata:

$$S_{\bar{x}_o} = \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \quad (2)$$

eşitliği kullanılarak $S_{\bar{x}_o} = \sqrt{\frac{9^2}{16} + \frac{12.3^2}{16}} = 3.81$ olarak hesaplanır.

GA genel eşitliği ise;

$$GA = \text{İstatistik} \pm \text{standart hata} \left(\text{dağılımın } \frac{\alpha}{2} \text{ olasılığındaki değeri} \right) \quad (3)$$

şeklinde olup, bu örnek için t dağılımının kritik tablo değeri yani $t_{(30)0.025} = 2.042$ ’dir.

Ortalama fark 11.9 ve standart hata 3.81 değerleri Eşitlik 3’te yerlerine yazılarak,

$GA = 11.9 \pm 3.81 \times 2.042$ eşitliğinde gerekli hesaplamalar yapıldığında bu ortalama fark için alt sınır 4.1, üst sınır ise 19.7 olarak bulunur.

EB olarak Cohen’in d istatistiği $d = \Delta / S_o$ eşitliğinden yararlanılarak $11.9 / 10.65 = 1.12$ olarak hesaplanır.

Etki büyüklüğünün standart hatası ise,

$$Sh_{(EB)} = \sqrt{\left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right) + d^2 / (2(n_1 + n_2))} \quad (4)$$

şeklinde dir. (G*Power Manuel, 2017). Bu eşitliğe göre etki büyüklüğünün standart hatası

$$Sh_{(EB)} = \sqrt{\left(\frac{16+16}{16 \cdot 16} \right) + 1.12^2 / (2(16 + 16))} = 0.38 \text{’dir.}$$

Etki büyüklüğü için %95 olasılıkla güven sınırları Z dağılımından yararlanarak hesaplanabileceği gibi, bu örnek için t dağılımının $t_{(30)0.025} = 2.042$ değeri kullanılarak Eşitlik 3’e göre $GA_{(EB)} = 1.12 \pm 0.38 \times 2.042 = 0.34$ ile 1.90 aralığında olabileceği bulunur.

YHAT için t dağılımına göre test istatistiği,

$$t = \frac{\bar{X}_A - \bar{X}_B}{S_{\bar{x}_o}} \quad (5)$$

olup, bu eşitliğe göre $t = \frac{122.8 - 110.9}{3.81} = \frac{11.9}{3.81} = 3.12$ ’dir. Bu test istatistiğinin t dağılımının iki tarafındaki olasılık değeri, yani $t_{(30)}; 3.12$ için olasılık değeri $p = 0.004$ ’dür.

Bu örnek için iki yönlü test yapılması durumuna göre $\alpha = 0.05$ ve EB 1.12 alındığında TG = 0.87 yani %87 olarak hesaplanır. Güç değeri hesaplama işlemleri hacimli olduğu için makale boyutu dikkate alınarak GPower 3.1 programı ile hesaplanmıştır (Faul ve ark., 2007).

Bu bulgular kısaca özetlenecek olursa; Çizelge 2’de verilen bulgular değerlendirildiğinde A ve B gruplarının ortalamaları arasındaki fark 11.9 olup, bu

fark %95 güvenle 4.1 ile 19.7 arasında değişebilmektedir. Bu fark ve %87 güç düzeyinde istatistiksel olarak anlamlıdır ($t_{(30)}=3.12$; $p=0.004$).

İkiden Fazla Bağımsız Grup Ortalamasının Karşılaştırılması (ANOVA) (Sayısal Örnek 2)

İncelenen bir X değişkeninin dağılımı normal dağılıma uygun ve varyansları homojen olan A, B ve C gibi üç bağımsız grupta X değişkenine ANOVA testi uygulanmıştır. Elde edilen analiz bulguları Çizelge 3 ve Çizelge 4’de sunulmuştur. Bu çizelgelerdeki çıkarımsal istatistiklerin hesaplanması ve yorumlanması aşağıda detaylı olarak açıklanmıştır.

Bu örnekte (Örnek 2) araştırma tasarımı verilerine herhangi bir istatistiksel yazılım kullanılarak (SPSS,

MINITAB vb.) ANOVA uygulandığında Çizelge 3 ve Çizelge 4’deki özet istatistikler elde edilebilir. Burada Çizelgelerde sunulan istatistiklerin nasıl hesaplandığı ve nasıl yorumlanması gerektiği örneklendirilmiştir.

GA değerleri şu şekilde hesaplanmıştır: Örneğin A metodunun ortalaması 122.8, standart sapması 9.0 ve standart hatası ise $S_x=9/\sqrt{16} = 2.25'$ dir. Ayrıca $\alpha=0.05$ için 15 serbestlik dereceli ve çift taraflı t dağılımının tablo değeri, $t_{(15) 0.025} = 2.131'$ dir. Bu ön hesaplamalar Eşitlik 3’te yerlerine yazılarak A grubu için GA’nın alt ve üst sınırları; $122.8 \pm 2.25 \times 2.131 = 118.0$ ile 127.6 olarak hesaplanır. Diğer gruplar için güven aralıkları benzer şekilde bulunmuş ve Çizelge 3’te sunulmuştur.

Çizelge 3. Sayısal örnek 2 için tanımlayıcı istatistikler
Table 3. Descriptive statistics for numerical example 2

| | n | Ortalama | Standart sapma | %95 Güven Aralığı | |
|-------|----|--------------------|----------------|-------------------|-----------|
| | | | | Alt sınır | Üst sınır |
| A | 16 | 122.8 ^a | 9.0 | 118.0 | 127.6 |
| B | 16 | 110.9 ^b | 12.3 | 104.4 | 117.5 |
| C | 16 | 110.8 ^b | 8.2 | 106.4 | 115.2 |
| Genel | 48 | 114.8 | 11.3 | 111.6 | 118.1 |

a, b: Farklı harfle gösterilen ortalamalar farklıdır ($p<0.01$).

Çizelge 4. Örnek 2 için ANOVA bulguları ve bazı çıkarımsal istatistikler

Table 4. ANOVA findings and some inferential statistics for the example 2

| | Sd | (#)Kareler toplamı | Kareler ortalaması | F | Önemlilik | EB | Güç |
|---------------|----|--------------------|--------------------|-------|-----------|-------------------------------|-------|
| Gruplar Arası | 2 | 1529.2 | 764.6 | 7.639 | 0.001 | $f = 0.50$ $\eta^2 = 0.25$ | 0.858 |
| Gruplar İçi | 45 | 4503.9 | 100.1 | | | | |
| Toplam | 47 | 6033.1 | | | | | |

(#) Analizlerin raporlanmasında genellikle yazılmaz.

Bu çalışma için tek yönlü ANOVA ile veriler değerlendirildiğinde (Çizelge 4) $p=0.001$ hata ile en az iki ortalama arasında fark olduğu bulgusuna ulaşılmıştır. Çizelge 4 izlendiğinde istatistiksel anlamlılık p değerinin yanında EB ve güç istatistikleri de verilmiştir. Bu istatistikler analiz hakkında daha açıklayıcı bilgiler sunmaktadır.

Farklı grupları belirlemek için yapılan LSD (Least Significant Difference) çoklu karşılaştırma testi sonucuna göre A grubu, B ve C gruplarından farklı ($p<0.01$) B ve C grupları ise benzerdir. Grupların güven aralıkları incelendiğinde bu sonucu görmek mümkündür. Örneğin A grubunun güven sınırları, B ve C gruplarının güven sınırları ile çakışmayıp, sayı doğrusunda daha büyük değer almaktadır. B ve C gruplarının ise güven aralıkları ise çakışmaktadır. Yani sınırlar ayrıktır.

Bu analize ait etki büyüklüğü EB, η^2 (eta kare) Eşitlik (6) yardımı ile aşağıdaki gibi hesaplanır:

$$\eta^2 = \frac{\text{Gruplar arası kareler toplamı}}{\text{Genel kareler toplamı}} \quad (6)$$

$$\eta^2 = \frac{1529.2}{6033.1} = 0.25 \text{ olarak hesaplanır.}$$

ANOVA için etki büyüklüğü " f " Eşitlik 7 kullanılarak hesaplanır.

$$f = \sqrt{\frac{\sum(\bar{x}_i - \bar{x})^2}{k(S_0^2)}} \quad (7)$$

Eşitlik 7’de; \bar{x}_i : i. grup ortalamasını, \bar{x} : genel ortalamayı, k: grup sayısını, S_0 : genel (ortak) standart sapmayı göstermektedir.

Eşitlik 7 kullanıldığında ANOVA için etki büyüklüğü

$$f = \sqrt{\frac{(122.8-114.8)^2 + (110.9-114.8)^2 + (110.8-114.8)^2}{3(11.3)^2}} = 0.50 \text{ olur.}$$

G*Power yazılımı ile bu etki büyüklüğünün 0.05 hata ile 0.858 güce sahip olduğu hesaplanmıştır. Yani bu analiz gerek f istatistiği gerekse η^2 istatistiği bakımından büyük etkiye (sırasıyla; $0.50 > 0.40$ ve $0.25 > 0.14$) sahiptir. Ayrıca bu testin gerçekleşen gücü $\alpha=0.05$ hata düzeyinde yaklaşık olarak %86’dır.

SONUÇ ve ÖNERİLER

Sonuç olarak istatistiksel analiz uygulanan araştırmalar rapor edilirken, aşağıdaki hususlara dikkat edilmelidir.

- İstatistiksel sonuçlar sunulurken anlamlılık düzeyi “p” ile birlikte test istatistiği değeri de (t, z, χ^2 , F, vb.) yazılmalıdır. Anlamlılık düzeyi p değeri, tercihen üç ondalıkla, örneğin p=0.018 gibi, yazılmalıdır.
- P-değerinin belirlenen α değerinden (genellikle 0.05) küçük çıkması durumunda karşıt/alternatif hipoteze çok yüksek bir anlam yüklenmemelidir. Bu sonucunda hatalı olabileceği dikkate alınmalıdır. Ayrıca istatistiksel olarak anlamlı bulunan etki, fark veya ilişki için ekonomik, teknik, klinik anlamlılığı belirlemek için etki büyüklüğü de hesaplanmalıdır. Ayrıca test sonucu çıkan anlamlılık düzeyini daha güvenilir hale getirmek için gözlenen güç değeri de hesaplanmalıdır.
- P-değerinin α değerinden büyük çıkması durumunda da sonucun anlamlı olabileceği veya anlamsız çıkmasının istatistik test metodu veya örneklem büyüklüğünden kaynaklanıp kaynaklanmadığı da irdelenmelidir. Örneklem büyüklüğünün yetersiz olup olmadığını belirlemek için de gözlenen güç hesaplanmalıdır.
- Araştırmanın metot bölümünde öngörülen istatistik test, bu teste bağlı etki büyüklüğü, α ve güç değeri (1- β) rapor edilmelidir. Bu varsayımlar altında belirlenen örneklem büyüklüğü “n” de belirtilmelidir.
- Araştırma bulguları rapor edilirken test istatistiği, n ve p'nin yanında EB ile gözlenen güç değeri (TG) de rapor edilmelidir.
- Ortalama, oran veya katsayıların yorumları; nokta tahminlerinden çok güven aralığı değerleri ve güven sınırları dikkate alınarak yapılmalıdır. Bunun içinde tanımlayıcı istatistiklerin verildiği tablolara güven sınırları değerleri de eklenmelidir.

KAYNAKLAR

- Akbulut, Ö., (2021). Çok Değişkenli ve Farklı Ölçekli Araştırmalarda Örneklem Büyüklüğünün Tespiti. *Journal of Animal Sciences and Products*. 4 (2), 199-215,
- Akbulut, Ö., (2022). Bilimsel Araştırmalarda İstatistiksel Anlamlılığın Raporlanmasında Güncel Yaklaşımlar: Hatalar ve Doğrular. *International Journal of Eastern Mediterranean Agricultural Research*, 4(1), 68-87.
- APA, (2010). American Psychological Association. (2010). Publication Manual of the American

- Psychological Association (6th Ed.). Washington, DC
- Aschwanden, C. 2016. Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing P-Values. Five Thirty Eight. 17 Haziran 2016 tarihinde kaynağından arşivlendi. Erişim tarihi: 25 Kasım 2022.
- Balkin, R.S. & Lenz S.A. (2021). Contemporary Issues in Reporting Statistical, Practical, and Clinical Significance in Counseling Research. *Journal of Counseling & Development* 99(2), 227-237
- Cebeci, Z. (2020). R Uygulamalı Yeniden Örnekleme Teknikleri. Pegem Akademi. ISBN:978-625-7052-60-3
- Cohen, H, W. (2011). P values: use and misuse in medical literature. *Ame. J. Hypertension*. 24(1), 18-23
- Cohen, J., (1994). The earth is round (p<0.05). *American Psychologist*, 49(12), 997-1003
- Çapık, C., (2013). Bir Hemşirelik Dergisinde Yayınlanan Makalelerde İstatistiksel Güçlerin İncelenmesi. *Anadolu Hemşirelik ve Sağlık Bilimleri Dergisi* 16(3), 170-5.
- Dahiru T. (2008). P – value, a true test of statistical significance? A cautionary note. *Annals of Ibadan Postgraduate Medicine*. 6(1), 21-26.
- Erkuş, A., (2017). Denence Testi ve H₀ Denencesinin Reddedilememesinin Dayanılmaz Ağırlığı, Düşünce yazısı-Opinion paper İlköğretim Online, 16(4), 12-16. [Online]: <http://ilkogretim-online.org.tr>
- Ertürk, M. Ş., (2005). Retrospective power analysis: When? *Radiology* 237 (2), 743-749
- Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3.1 A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Gao, J., (2020). P-values – a chronic conundrum. *Gao BMC Medical Research Methodology*. 20, 167. <https://doi.org/10.1186/s12874-020-01051->
- Greenland, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman S.N.& Altman, D.G., (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 31, 337–350.
- Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, 45(3), 135-140
- Gürkan, A., (2007). Klinik Peridontolji araştırmalarında bağımsız iki grup ortalamasının karşılaştırılmasında örnek genişliği, istatistiksel güç ve anlamlılık. *EÜ Diş Hekimliği Fak. Derg.* 28, 123-134.
- G*Power Manual. (2017). GPower Manual.pdf . [https://www.psychologie.hhu.de Mathematisch-Naturwissenschaftliche Fakultaet/Psychologie/](https://www.psychologie.hhu.de/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/) Erişim: 01/05/2022.

- Halsey, L., Curran-Everett, D., Vowler, S., & Drummond, G. (2015). The fickle P value generates irreproducible results. *Nat. Methods* 12, 179–185.
- Hojat M & Xu, G., (2004). A Visitor's Guide to Effect Sizes Statistical Significance Versus Practical (Clinical) Importance of Research Findings *Advances in Health Sciences Education* 9, 241–249
- Işık, İ., (2014). Yokluk hipotezi anlamlılık testi ve etki büyüklüğü tartışmalarının psikoloji araştırmalarına yansımaları. *Eleştirel Psikoloji Bülteni*. 5, 55-80
- Kalaycıoğlu, O. & Akhanlı, S.E., (2020). Sağlık araştırmalarında güç analizinin önemi ve temel prensipleri: Tıbbi çalışmalar üzerinde uygulamalı örnekler. *Turk J Public Health* 18(1), 103-112
- Keskin, B., (2020). İstatistiksel güç bir araştırmann sonuçlarına etki eder mi? örneklem büyüklüğüne nasıl karar verilmeli? *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi* 18 (Armağan Sayısı), 157-174.
- Kılıç, S., (2014). Etki büyüklüğü. *Journal of Mood Disorders* 4(1), 44-46
- Kılıç, S., (2011). Neyin Peşindeyiz? Kutsal p-değerinin mi (İstatistiksel Önemlilik) Yoksa Klinik Önemliliğin mi? *Journal of Mood Disorders*.11, 1:46-48.
- Kul, S., (2014); İstatistik Sonuçlarının Yorumu: P-değeri ve Güven Aralığı Nedir? *Türk Toraks Derneği Derg*, 8(1), 11-13. DOI:10.5152/ pb.2014. 003.
- Leek, J.T.& Peng, R.D., (2015). Statistics: P values are just the tip of the iceberg. *Nature*. 520(7549), 612. doi: 10.1038/520612a.
- Lewis, K. P., (2006). Statistical power, sample sizes and the software to calculate them easily. *Bioscience* 56(7), 607-612.
- Lu, Y.& Belitskaya-Levy I., (2015). The debate about p-values, *Shanghai Arch Psychiatry*. 27 (6), 381-385.
- Mark, D.B., Lee, K.L., & Jr Harrell, F.E., (2016). Understanding the Role of P Values and Hypothesis Tests in Clinical Research. *JAMA Cardiol*. 1(9), 1048-1054. doi:10.1001/jamacardio.2016.3312
- Nelson, M. S., Wooditch, A., & Dario, L. M. (2015). Sample size, effect size, and statistical power: a replication study of Weisburd's paradox . *J Exp Criminol* 11, 141–163 DOI 10.1007/s11292-014-9212-9
- Nuzzo, R., (2014) Scientific method: Statistical errors. *Nature*, 506, 150–152.
- 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 Analyses. *Communication Methods and Measures*, 4, 291-299.
- O'Leary, T.J., (2021) Rigor, Reproducibility, and the P value (Commentary). *The American Journal of Pathology*, 191(5), 806-808.
- Özçomak, M.S., & Çebi, K., (2017). İstatistiksel Güç Analizi: Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi Üzerine Bir Uygulama. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 31(2) 413-425
- Özsoy, S. & Özsoy, G., (2013). Eğitim Araştırmalarında Etki Büyüklüğü Raporlanması. *İlköğretim Online*, 12(2), 334-346.
- Plate, J.D.J., Borggreve, A.S., Hillegersberg, R. & Peelen, L.M., (2018). Post Hoc Power Calculation: Observing the Expected. *Annals of Surgery* 269(1), e11. DOI: 10.1097/SLA.0000000000002910
- Solla, F., Tran, A., Bertonecelli, D., Musoff, C. & Bertonecelli, C.M., (2018) Why a P-Value is Not Enough. *Clin. Spine. Surg*. 31, 385–388.
- Sullivan, G. M.& Feinn (2012). Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <http://doi.org/10.4300/JGME-D-12-00156.1>
- Sun, S., Pan W. & Wang, L.L., (2011). Rethinking Observed Power. *Methodology*, 7(3), 81–87
- Şevgin, H. & Çetin B., (2017). Eğitim Araştırmalarında Güç Analizi ve Bir Uygulama *YYÜ Eğitim Fakültesi Dergisi*; 14(1), 1462-1480, doi.org/10.23891/efdyu.2017.52
- Téllez, A., García, C.H. & Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia. State of the Art*, 8(3), 27-47.
- Ünalın, A., 2021. Sample Size in Clinical Researches: Power of the Test and Effect Size. *BSJ Health Sci*, 4(3), 221-227.
- Vidgen, B., & Yasseri, T. (2016). P-Values: Misunderstood and Misused. *Frontiers in Physics*, 4(6), 1-5.
- Walpole, R.E, (1969).Introduction to Statistics, Third Printing. The McMillan Company, New York p:365.
- Wasserstein, R.L. & Lazar, N.A. (2016). Editorial. The ASA's statement on P-values: context, process, and purpose. <http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108>.
- Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. (2019). Moving to a world beyond “p< 0.05”. *The American Statistician* 73, 1–19.
- Whitley, E. & Ball, J., (2002). Review Statistics review 3: Hypothesis testing and P values *Critical Care* 6, 222-225
- Yıldırım, H. H. & Yıldırım, S. (2011). Hipotez Testi, Güven Aralığı, Etki Büyüklüğü ve Merkezi Olmayan Olasılık Dağılımları Üzerine. *İlköğretim Online* , 10 (3) , 1112-1123 .Retrieved from <https://dergipark.org.tr/tr/pub/ilkonline/issue/8591/106798>.