SPORMETRE
The Journal of Physical Education and Sport Sciences
Beden Eğitimi ve Spor Bilimleri Dergisi

DOI: 10.33689/spormetre.794015

Geliş Tarihi (Received): 12.09.2020 Kabul Tarihi (Accepted): 22.03.2021 Online Yayın Tarihi (Published): 30.03.2021

SPOR BİLİMLERİNDE ETKİ BÜYÜKLÜĞÜ VE ALTERNATİF İSTATİSTİK YAKLAŞIMLARI

Süleyman ULUPINAR[1], İzzet İNCE[2]
[1]Milli Eğitim Bakanlığı, Ermenek İlçe Milli Eğitim Müdürlüğü, KARAMAN
[2]Ankara Yıldırım Beyazıt Üniversitesi, Sağlık Bilimleri Fakültesi, Egzersiz ve Spor Bölümü, ANKARA

Öz: Diğer disiplinlerde olduğu gibi spor bilimleri alanında da bilimsel çalışmaların sonuçlarının sahaya aktarılmasında bazı sınırlılıklar olduğu bilinmektedir. Birçok araştırmanın sadece yokluk hipotezinin test edildiği analiz sonuçları ve buna bağlı *p-değeri* ile rapor edilmesi, spor bilimlerinde araştırma ve uygulama arasındaki boşluğun başlıca sebebi olarak gösterilmektedir. Bazı araştırmacılar gerçek dünyada yokluk hipotezinin daima yanlış olduğunu ileri sürerek müdahale etkisinin pratikteki büyüklüğüne daha fazla odaklanılması gerektiğini savunmaktadır. Son yıllarda mevcut istatistik yaklaşımının kullanıldığı araştırma sonuçlarının etki büyüklüğü gibi pratiğe hitap eden yöntemler ile desteklenmesine sıkça vurgu yapılmaktadır. Bu çalışmanın amacı bilimsel çalışmalarda sıklıkla kullanılan genel istatistik yaklaşımına alternatif olabileceği ileri sürülen modelleri ve destekleyici nitelik taşıyan bazı yöntemleri incelemek ve faydalı olduğu düşünülen yöntemlerin spor bilimleri alanında yaygınlaşmasına katkıda bulunmaktır. Literatürde özellikle etki büyüklüğünün rapor edilmesine ilişkin bir fikir birliği sağlanmış olsa da etki büyüklüğünün sınıflandırılmasına ilişkin bazı görüş ayrılıkları vardır. Büyüklük temelli çıkarım modeli ve Bayesci istatistik modelinin de pratikte sağladığı bazı avantajlara rağmen mevcut şartlarda kullanılan istatistik yaklaşımına bir alternatif olamayacağı daha fazla kabul görmektedir. Diğer taraftan pratikte önemli kabul edilen en küçük değişim miktarının ve hata terimlerinin vurgulandığı destekleyici yöntemlerin spor bilimlerinde pratik bir fayda sağlayacağı ileri sürülmektedir. Sonuç olarak literatürde mevcut istatistik yaklaşımına alternatif olabilecek modeller ile ilgili tartışmaların devam etmesine rağmen, yokluk hipotezinin test edildiği sonuçların (*p-değeri*, istatistiksel anlamlılık analizleri) pratik kullanıma ilişkin destekleyici yöntemler ile birlikte sunulmasının önemli yararlar sağladığı konusunda yaygın bir kabul oluşmuştur.
Anahtar Kelimeler: Cohen's *d*, p değeri, pratik anlamlılık, minimum anlamlı değişim, tipik hata.

EFFECT SIZE AND ALTERNATIVE STATISTICAL APPROACHES IN SPORTS SCIENCES

Abstract: It is known that there are some limitations in reflection the results of scientific studies on the practice in the sports sciences as in other disciplines. Main reason for the gap between research and practice in sports sciences is considered as the reporting approach of only the analysis results and *p-value* in many studies using null hypothesis. Some researchers have suggested that practical importance of the intervention effect should be focused, arguing that the null hypothesis is always false in the real world. In recent years, it has been emphasized that the research results using the current statistical approach should be supported with methods that appeal to practice, such as the effect size. The aim of this study is to examine models that are suggested to be an alternative to the current statistical approach commonly used in scientific studies and some supportive methods, and to increase to the use of methods considered useful in the field of sports sciences. Although there is a consensus in the literature regarding the reporting of effect size, there are some disagreements regarding the classification of the effect size. Despite some practical advantages of the magnitude-based inferences and the Bayesian statistical models, it is more accepted that it cannot be an alternative to valid statistical approach used in the current conditions. On the other hand, it is stated that the supportive methods that emphasize the smallest important change amount and error terms, which are considered critical in practice, will provide practical benefits in sports sciences. Consequently, although the discussions about the models that can be an alternative to the current statistical approach in the literature continue, there is a widespread consensus that supporting analyzes of the null hypothesis (*p-value*, statistical significance) with methods for practical use provide important benefits.
Key Words: Cohen's *d*, p value, practical importance, smallest worthwhile change, typical error.

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

GİRİŞ

Spor bilimleri alanında yapılan akademik çalışmalardan elde edilen sonuçların sahaya aktarılmasında bazı sınırlılıklar olduğu belirtilmektedir (Bernards ve ark., 2017; Hopkins, 2002; Sullivan ve Feinn, 2012). Tüm dünyada geçerli olduğu vurgulanan, akademik çalışmalar ile saha uygulamaları arasındaki boşluğun en az aynı oranda ülkemizde de var olduğu düşünüldüğünde araştırma ve uygulama arasında işlevsel bir köprünün gerekli olduğu açıktır. Akademik araştırma sonuçlarının saha profesyonelleri tarafından anlaşılması, yorumlanması ve uygulamada kullanılmasını sağlamak için bazı öneriler sunulmaktadır (Hopkins ve ark., 2009; Sullivan ve Feinn, 2012). Yapılan deneysel çalışmalardaki antrenman programı, ısınma protokolü, egzersiz müdahalesi, besin desteği veya benzeri bir müdahalenin etkisinin rapor edildiği çalışmaların sonuçlarını karşılaştırabilmek ve pratikteki etkisine karar verebilmek için bazı destekleyici yöntemlere ihtiyaç duyulduğu savunulmaktadır (Cumming, 2014; Fritz ve ark., 2012; Hopkins, 2019; Welsh ve Knight, 2015).

Araştırma ve uygulama birimleri arasındaki boşluğun giderilmesine ilişkin en temel öneri bilimsel verilerin pratiğe ve saha profesyonellerinin kullanımına uygun olarak analiz edilmesi ve raporlanmasıdır (Rhea, 2004). Mevcut istatistik yaklaşımında hipotez doğrulama süreci, genellikle bir olasılık değeri belirlemeye (*p-değeri*) ve yokluk hipotezinin reddedilip reddedilmeyeceğine karar vermeye (istatistiksel anlamlılık testleri) dayanmaktadır (Alpar, 2010; Sullivan ve Feinn, 2012). Yani bir araştırmacı kendi çalışması için geçerli olan kriterlere dayanarak yokluk hipotezini reddettiğinde alternatif hipotezi kabul etmiş olur. Ancak sadece yokluk hipotezini test etmek ve *p-değeri* sunmakla yetinilmeyip pratiğe yönelik destekleyici bilgilerin raporlanmasına ilişkin tavsiyelere rağmen spor bilimlerindeki araştırmaların çoğunun önerilen yöntemleri içermediği görülmektedir (Fröhlich ve ark., 2009). Spor bilimlerindeki araştırmaların birçoğunun *p-değeri*ni hesaplamaktan daha ileri gitmediği, dolayısıyla uygun şekilde raporlama ve yorumlama gibi temel nitelikleri göz ardı ettiği vurgulanmaktadır (Tomczak ve Tomczak, 2014). Bu çalışmanın amacı, genel istatistik yaklaşımına alternatif olabileceği ileri sürülen modelleri ve destekleyici nitelik taşıyan bazı yöntemleri incelemek ve bilimsel çalışmalarda faydalı olduğu düşünülen bu yöntemlerin spor bilimleri alanında yaygınlaşmasına katkı sunmaktır.

YÖNTEM

Bu derleme çalışmasında spor bilimlerinde güncel istatistiki yaklaşımlar, alternatif ve destekleyici nitelik taşıyan hesaplamaları değerlendiren çalışmalar incelenmiştir. Pub Med, Web of Science, Medline, Cochrane Library, Science Direct, Google Scholar ve ULAKBİM elektronik veri tabanları "effect size", "practical importance", "smallest worthwhile change", "minimal detectable change", "standard error of measurement", "typical error", "Bayesian statistic", magnitud-based inferences" anahtar kelimeleri kullanılarak taranmıştır. Elektronik arama ile ulaşılan ilgili tüm yazıların başlık ve özetleri araştırmacılar tarafından gözden geçirilmiştir. Konu açısından uygun olduğuna karar verilen çalışmalardan meta-analiz araştırmaları, sistematik derlemeler ve deneysel çalışmaların tam metni okunmuştur. Ayrıca konu ile ilgili İngilizce ve Türkçe dillerinde yazılmış kitaplar ve ilgili konuya öncülük eden web siteleri incelenerek (Örneğin, https://sportscience.sportsci.org/) konu ile ilgili kapsamlı bir bütünlük oluşturmak amaçlanmıştır.

Mevcut istatistik yaklaşımında hipotez testleri ve "*p-değeri*" eleştirisi

Diğer disiplinlerde olduğu gibi spor bilimlerinde de bilimsel araştırmaların önceden belirlenen bir hipoteze uygunluğu ölçülerek sonuçların istatistiksel açıdan anlamlı olup olmadığını

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

değerlendirir. Bir başka deyişle mevcut araştırma yapısı olarak kabul edilen Neyman-Pearson istatistiksel yaklaşımı, sonuçların ne kadar doğru olduğunu değil ne derecede şansa bağlı olarak ortaya çıktığını gösterir (Cohen, 1988; Fröhlich ve ark., 2009; Rosnow ve Rosenthal, 2003). Örneğin bir çalışmadan elde edilen bir analizin sonucunun genelde anlamlılık derecesi olarak kabul edilen 0,05'ten küçük olması, bu araştırmanın 100 tekrarının en az 95'inde sonuçların aynı sınırlar içinde olacağı anlamına gelmektedir. Bu nedenle *p-değeri,* bilimsel araştırmaların bulgularına ilişkin güven aralıklarını değerlendirmede önemliyken, çalışma tasarımında kullanılan müdahalenin etkisinin ne kadar büyük olduğuna dair yeterli bilgi sağlamamaktadır (Hopkins ve ark., 2009; Hopkins, 2002).

Hipotezlerin önceden kurulup test edildiği mevcut istatistik modellerinde gerçekte var olmayan bir etkinin istatistiksel eşiğe ulaşıp anlamlı kabul edilmesi için manipülasyon niteliği taşıyan birçok faktör olduğu belirtilmektedir (Bernards ve ark., 2017). Örneğin bilimsel çalışmalarda kullanılan istatistik hesaplamalarının, örneklem büyüklüğü ve ölçümlerin ortalamaya olan uzaklık dağılımından oldukça etkilenmektedir. Yani istatistiksel açıdan anlamlı olmayan bir sonuç, diğer değişkenler sabit kalmak kaydıyla örneklem sayısının fazla olması veya ölçüm sonuçlarının ortalamaya yakın bir aralıkta yığılması durumunda 0,05 düzeyine ulaşabilir ve anlamlı olarak kabul edilir (Cohen, 1988; Sullivan ve Feinn, 2012). Özetle, örneklem sayısı ve standart sapma üzerinden hesaplanan standart hata örneklem sayısı arttıkça küçülürken *p-değeri* de buna bağlı olarak düşme eğilimindedir. Anlamlılık olarak kabul edilen bu yargının örneklem sayısına olan yüksek duyarlılığı, bu yaklaşımının manipüle edilebilir olduğuna ilişkin eleştirilere sebep olmaktadır (Altman ve Bland, 2005; Hopkins, 2019; Tomczak ve Tomczak, 2014). Bu sebeplerle *"p-değeri*" üzerinden değerlendirilerek rapor edilen sonuçların, müdahale etkisinin büyüklüğüne dair destekleyici bilgiler ile birlikte sunulmasının bulguların karşılaştırılmasını, yorumlanmasını ve sahada uygulanmasını kolaylaştıracağı savunulmaktadır (Cumming, 2014; Rhea, 2004; Sullivan ve Feinn, 2012).

Akademik ortamda bilimsel yayınların hacmini arttırmak için yoğun bir baskı kurulmuş olsa da spor bilimleri ile ilgili araştırmalardan elde edilen sonuçların pratiğe uygun ve ön yargılardan uzak olması beklenmektedir (Bernards ve ark., 2017). Birçok alanda olduğu gibi spor bilimlerinde de çoğu araştırmanın sonucu *p-değeri*, yani istatistiksel olasılığı üzerinden rapor edilmektedir. Bazı araştırmacılar karar vermek için *p-değeri*ni temel alan Neyman-Pearson istatistiksel yaklaşımında yokluk (boş hipotez, sıfır hipotezi) hipotezinin gerçekte daima hatalı olduğunu, bu yaklaşımdan elde edilen sonuçların bir etkinin büyüklüğü konusunda bilgi sağlamadığını ve bu etkinin pratikteki önemi konusunda kanıt sayılamayacağını savunmaktadır (Bernards ve ark., 2017; Greenland ve ark., 2016; Mengersen, Drovandi, Robert, Pyne ve Gore, 2016). Pratikte daha geçerli ve uygulanabilir çıkarımlar elde edilebilmesi mevcut istatistik bakış açısına alternatif olabileceği düşünülen etki büyüklüğü temelli bir istatistik modeline veya Bayesci yöntemler gibi bir çerçeveye geçişin ya da sonuçların birlikte rapor edilmesinin çözüm olabileceği savunulmaktadır (Bernards ve ark., 2017; Hopkins, 2019).

Mevcut istatistik yaklaşımı üzerindeki en temel eleştirilerden birisi de gerçekliği "siyah ya da beyaz" olarak görmeye sevk ettiği yönündedir (Cumming, 2014). Bir başka deyişle analiz sonuçlarının mutlak terimler ile "anlamlı – anlamlı değil" veya "fark var – fark yok" şeklindeki ikili kategorize etme çabası spor bilimlerindeki araştırma sonuçlarını pratikten uzaklaştırmaktadır (Bernards ve ark., 2017; Greenland ve ark., 2016; Hopkins, 2019). Diğer taraftan gerçek dünyada yokluk hipotezinin her zaman yanlış olduğu, yeterli sayıda ve hassasiyette ölçüm yapılmaya devam edildiği sürece daima bir farka ulaşılacağı savunulmaktadır (Cohen, 2013). Ayrıca birçok disiplinden farklı olarak spor bilimlerinde sahadaki gelişmeler birçok açıdan bilimsel faaliyetlerin daha önünde gelmektedir. Dolayısıyla

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

spor bilimlerinde bilimsel bir çalışmanın sonucu istatistiksel açıdan bir anlamlılığa işaret etse bile, bu etkinin büyüklüğüne ve pratikteki önemine odaklanılmaya ihtiyaç olduğu açıktır (Bernards ve ark., 2017; Mengersen ve ark., 2016; Rhea, 2004).

Spor bilimlerinde mevcut istatistik yaklaşımının göz ardı ettiği bir diğer faktör de katılımcıların antrenman veya yarışma statüsü gibi niteliksel özellikleridir (Peterson, Rhea ve Alvar, 2004; Rhea, 2004). Örneğin katılımcı sayısının eşit olduğu iki araştırmanın birinde elit sporcular, diğerinde yeni başlayan sporcuların tercih edildiğini ve her iki grubun skuat egzersizi ile ölçülen 1 tekrar maksimum kuvvet artışlarının 10,0 ± 2,0 (ort ± ss) kg olduğunu varsayalım. Bu durumda iki araştırmanın istatistiki açıdan benzer şekilde sonuçlandığını söylemek mümkün olsa da uygulamada bu farkın elit sporcular için çok daha değerli olduğu tartışılmazdır (Hopkins, 2019). Özetle, performans gelişimi açısından elit sporculardaki 10 kg, yeni başlayanlardaki 10 kg'dan daha büyüktür ancak pratikte kesin olarak kabul edilen bu gerçeğin mevcut istatistiksel modeller ile ortaya koyulması mümkün değildir.

Bilimsel çalışmalarda hakim olan mevcut istatistik yaklaşımının sebep olduğu bir diğer olumsuzluk ise literatüre yön veren önemli yayın merkezlerinin ve editörlerin anlamlı sonuçların, yani bir etkinin ya da bir farkın ortaya koyulduğu çalışmaları tercih etme eğilimidir. Bilimsel araştırmalarda yapılan istatistiksel analizlerin anlamlı ya da anlamlı olmayan farklar ile sonuçlanmasının eşit derecede bilimsel değere sahip olduğu kabul edilse de anlamlı sonuçların ortaya koyulduğu çalışmaların bilimsel dergilerde daha fazla kabul gördüğü bilinmektedir. Ancak *p-değerinin* büyük oranda standart hatadan, standart hatanın da örneklem sayısından etkilendiği düşünüldüğünde anlamlılık eşiği, 'özellikle az sayıda sporcu üzerinde yapılan araştırmalar için' önemli bir sınırlılık oluşturmaktadır (Altman ve Bland, 2005). Diğer taraftan bir araştırma sonucunda bir farkın anlamlı olduğuna kanıt olarak sunulan *p-değeri,* önemli bir farktan ziyade aslında araştırmanın çok büyük sayıdaki bir örneklem üzerinde yapıldığına işaret ediyor olabilir (Sullivan ve Feinn, 2012). Bu sebeple mevcut şartlardaki yaygın kabul, etki büyüklüğünün *p-değerinin* yanında sunulması gereken ve pratikte kullanımı kolaylaştıran destekleyici bir yöntem olduğudur (Tomczak ve Tomczak, 2014).

Etki büyüklüğü (*Effect Size*)

Yokluk hipotezinin sınandığı anlamlılık testleri araştırmanın örnekleminden elde edilen sonuçlara şans faktörüyle ulaşılma ihtimalini değerlendirirken, etki büyüklüğü ise pratik anlamlılığın bir göstergesi olarak kullanılmaktadır (Cohen, 1969; Hopkins, 2002). Yani bir araştırmanın sonucunda ortaya çıkan farkın pratikte ne kadar önemli olduğuna karar vermede kullanılan bir ölçüttür. Etki büyüklüğü hem aynı araştırma içinde uygulanan müdahalelerin hem de farklı zamanlarda yapılmış araştırmaların ortaya koyduğu etkilerin karşılaştırılmasına imkan sağlar (Fritz ve ark., 2012; Rosnow ve Rosenthal, 2003; Sawilowsky, 2009). Etki büyüklüğü ayrıca istatistiksel bir testin gücünü hesaplamak için kullanılabilir, böylece araştırmacıların bir çalışma için gerekli örneklem sayısını belirlemelerine yardımcı olur (Tomczak ve Tomczak, 2014). Ayrıca pilot çalışmalar veya küçük örneklem sayısını içeren diğer ön çalışmalardan elde edilen etki büyüklükleri gelecekteki çalışmaların sonuçlarından ulaşılacak beklentilerin bir göstergesi olarak da kullanılabilir (Fritz ve ark., 2012; Tomczak ve Tomczak, 2014). Dolayısıyla farklı araştırma sonuçlarını ortak bir birim üzerinden yorumlayabilmek ve sonuçların saha profesyonelleri tarafından kolayca anlaşılabilmesi için özellikle spor bilimleri alanındaki çalışmalarda etki büyüklüğünün rapor edilmesinin gerekliliğine dikkat çekilmektedir (Cohen, 1969, 1988; Hopkins, 2019).

Etki büyüklüğü raporlanması, araştırmadan elde edilen sonuçların büyüklüğünün ve pratik öneminin değerlendirilmesine izin verir (Hopkins ve ark., 2009; Tomczak ve Tomczak, 2014).

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

Örneğin iki grubun karşılaştırıldığı bir araştırmada gruplardan birine Olimpik Halter antrenman müdahalesinin, diğerine de geleneksel kuvvet antrenman müdahalesinin uygulandığını varsayalım. Araştırma periyodunun sonunda her iki grubun da kuvvet gelişiminin anlamlı derecede artmış olması bu müdahalelerin her ikisinin de etkili olduğu konusunda önemli bir bilgi sağlarken, hangi müdahalenin etkisinin daha büyük olduğu sorusuna cevap vermekte yetersiz kalmaktadır. Ya da farklı zamanlarda uygulanan bağımsız iki araştırmanın her ikisinin de kullandığı müdahalenin performansı anlamlı derecede arttırdığını varsayalım. Bu iki araştırmanın ortak bir birimde karşılaştırılması için yüzde değerlerinin sunulması örneklemin ortalama etrafındaki dağılımından bağımsız olarak hesaplandığı için sonuçların doğrudan karşılaştırılması yanıltıcı olabilmektedir. Etki büyüklüğünün hesaplanmasında standart sapma değerleri de kullanıldığından ortak bir birim üzerinden araştırma sonuçlarını kıyaslama şansı vermektedir (Cohen, 1969; Rosenthal ve Rosnow, 1985; Rosnow ve Rosenthal, 2003). Etki büyüklüğü bu gibi durumlarda farkların büyüklüğünü mukayese etmek için oldukça işlevsel bir yöntem olarak kullanılabilmektedir.

Etki büyüklüğü aynı zamanda *"p-değeri"* anlamlı olmasa bile pratikte önemli sayılabilecek sonuçların fark edilmesine imkan verir (Cohen, 2013; Rhea, 2004). Örneğin bir antrenman protokolünün bir grup üzerindeki etkisine dair *p-değeri:* 0,08 olarak hesaplanır ve sadece bu şekilde rapor edilirse bu protokolün etkisinin önemsiz olduğuna karar verilmiş olur. Ancak 0,08 *p-değeri*, bahsi geçen denemenin aynı şartlarda 100 kez yapılması durumunda 92 kez beklenebileceği anlamını taşır ve örneklem sayısının artmasıyla anlamlılık eşiğine ulaşması mümkündür. Aynı çalışmanın "p = 0,08; EB = 0,70 (orta düzeyde etki)" şeklinde rapor edilmesi hem çalışma bileşenlerini revize ederek araştırmanın tekrar edilmeye değer olup olmadığına karar vermek için araştırmacılara fayda sağlarken, hem de saha profesyonellerine sonuçların pratikte uygulanabilirliği konusunda bilgi vermektedir (Rhea, 2004).

Etki büyüklüğünün raporlanmasına daha çok ortalamaların karşılaştırıldığı hipotez testlerinde (*t*-testler ve ANOVA testleri) ihtiyaç var gibi gözükmektedir. Çünkü regresyon ve korelasyon gibi analizleri içeren çalışmalarda *p-değeri* ile birlikte verilen bu katsayılar ilişkinin veya bağımlı değişkendeki varyansın açıklanma düzeyinin ne kadar büyük olduğuna dair bir bilgi sunmaktadır. Ortalamalar arasındaki farklara ilişkin etki büyüklüğünün sunulmasında en sık tercih edilen yöntemler eta kare (η^2), epsilon kare ($\varepsilon 2$), omega kare (ω^2), Hedges'g ve Cohen's *d* katsayısıdır. Bağımsız iki grup arasındaki farkın veya ön test − son test ölçümleri arasındaki farkın pratikteki büyüklüğünü belirlemek için Cohen aşağıdaki formülleri önermektedir (Cohen, 1969).

$$EB_{\text{ilişkisiz gruplar}} = (Ort_{\text{deney\_gr}} - Ort_{\text{kontrol\_gr}}) / SS_{\text{control\_gr}}$$

$$EB_{\text{ilişkili gruplar}} = (Ort_{\text{son\_test}} - Ort_{\text{ön\_test}}) / SS_{\text{ön\_test}}$$

Bununla birlikte, formülden de anlaşılabileceği gibi Cohen's *d* formülünün ilk formu (Cohen, 1969), hem ilişkisiz grupların hem de ilişkili grupların (ön test − son test) sınandığı durumlarda standart sapma değerlerinden birini göz ardı etmektedir. Örneğin son testten elde edilen ortalama hesaplamada kullanılırken, bu teste ilişkin standart sapma hesaplamaya dahil edilmez. Bu durumun çözümü için Bortz ve Döring (2007), Cohen's *d* formülüne ortak varyansı kullanarak küçük bir düzeltme önerisi sunmuştur. Her iki ölçümün standart sapmasının yer aldığı formül aşağıdaki gibidir (Bortz ve Döring, 2007).

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*

$$d = \frac{\text{Ort1} - \text{Ort2}}{\sqrt{\frac{SS_1^2 + SS_2^2}{\text{n1} + \text{n2} - 2}}}$$

Daha sonra Hedges ve Olkin (2014), harmanlanmış standart sapmanın (*pooled standard deviation*) kullanılmasını önererek benzer bir formül geliştirmişlerdir. Temelde Cohen's *d* altyapısıyla hesaplanan ve *g* değeri olarak sunulan bu değer aşağıdaki şekilde hesaplanmaktadır (Hedges ve Olkin, 2014).

$$g = \frac{\text{Ort1} - \text{Ort2}}{\sqrt{\frac{(\text{n1} - 1)SS^2 + (\text{n2} - 1)SS^2}{\text{n1} + \text{n2} - 2}}}$$

İkiden fazla ortalamanın karşılaştırıldığı ANOVA analizlerinin kullanıldığı durumlarda ise etki büyüklüğü genelde eta kare (η^2), epsilon kare ($\varepsilon 2$) ve omega kare (ω^2) değerleri kullanılarak rapor edilmektedir. Eta kare hesaplanmasının aşağıdaki formül ile yapılabilmektedir (Cohen, 1973; Fritz ve ark., 2012; Lakens, 2013).

$$\eta 2 = \frac{\text{Gruplar arası kareler toplamı}}{\text{Toplam kareler toplamı}}$$

Formülde yer alan "gruplar arası kareler toplamı" ve "toplam kareler toplamı" istatistik programlarında yer alan ANOVA tablosunda sunulduğu için kolayca hesaplanabilmektedir. Diğer taraftan, bir bağımsız değişkenin olduğu durumlarda eta karenin kullanılmasının uygun olduğu belirtilmektedir ancak birden fazla bağımsız değişken varsa incelenmek istenen değişkenin etkisi, diğer değişkenlerin etkisine bağlı olarak değişebilmektedir. Bu durumda kısmi eta karenin (η_p^2) hesaplamasının daha uygun olduğu düşünülmektedir (Fritz ve ark., 2012; Lakens, 2013; Rosenthal ve Rosnow, 1985).

$$\eta_p^2 = \frac{\text{Gruplar arası kareler toplamı}}{\text{Gruplar arası kareler toplamı} + \text{Standart hatalar toplamı}}$$

Bunun yanında epsilon kare ve omega kare hesaplamaları da ANOVA analizlerinde etki büyüklüğünün bir göstergesi olarak kullanılmakla birlikte bu hesaplamaların çoğu istatistik programlarında otomatik olarak hesaplanmamaktadır. Özellikle en çok kullanılan istatistik programı olan SPSS (*Statistical Package for the Social Sciences*), istenildiği takdirde kısmi eta kare katsayısını ANOVA tablosunda sunmaktadır. Bu nedenle araştırmalarda ANOVA için sunulan etki büyüklüklerinin çok yüksek bir çoğunluğu kısmi eta kare katsayısıdır. Kısmi eta kare katsayısı küçük (0,0099), orta (0,0588) ve büyük (0,1379) şeklinde sınıflandırılsa da (Cohen, 1988), böyle bir sınıflandırma olmadan da sonuçları kolaylıkla yorumlamak mümkündür. Kısmi eta kare katsayısı 0 ve 1 arasında bir değer alırken, 100 ile çarpıldığında bağımlı değişkene ait varyansın ne kadarının bağımsız değişken tarafından açıklandığını gösterir. Örneğin kısmi eta kare katsayısının 0,74 olması, bağımlı değişkene ait varyansın %74'ünün bağımsız değişken tarafından açıklandığı anlamına gelir. η^2, η_p^2, ε^2 ve ω^2 değerlerinin karşılaştırıldığı bir çalışmada örneklem sayısının büyük olduğu durumlarda η^2, ε^2 ve ω^2 değerlerinin birbiri ile oldukça benzer, ancak η_p^2 değerinin diğerlerinden daha büyük olduğu gösterilirken örneklem sayısının küçük olduğu durumlarda ise dört değerin de birbirine yakın olduğu gösterilmiştir (Levine ve Hullett, 2002). Aynı çalışmada ayrıca tek-yönlü ANOVA tasarımında dört değerin birbirine yakın olduğu ancak iki-yönlü ANOVA tasarımında

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

η^2, ε^2 ve ω^2 değerlerinin birbiri ile oldukça benzer, η_p^2 değerinin ise diğerlerinden daha büyük olduğu rapor edilmiştir (Levine ve Hullett, 2002).

Bununla birlikte korelasyon katsayısı olarak kullanılan *r* değeri üzerinden etki büyüklüğü hesaplama işlemi birçok istatistik programıyla yapılan testler sonucunda sunulan *t* değeri yardımıyla kolayca yapılabilmektedir (Rosnow ve Rosenthal, 2003). Örneğin bağımsız örneklem *t*-testi (*Student's t-test*) birbiri ile ilişkili olmayan iki grubun ortalamasını karşılaştırmak için kullanılan parametrik bir testtir. İstatistik programlarının sunmuş olduğu *t* değerleri kullanılarak aşağıdaki formül ile etki büyüklüğü hesaplanabilmektedir (Fritz ve ark., 2012; Rosnow ve Rosenthal, 2003).

$$r = \sqrt{\frac{t^2}{t^2 + sd}}$$

Formülde yer alan "sd" serbestlik derecesi anlamına gelmektedir ve iki gruptaki örneklem sayısının her ikisinin de 1 eksiği toplanarak hesaplanır ($n_1 - 1 + n_2 - 1$). Formülde yer alan "r" nokta çift serili korelasyon katsayısını (*point-biserial correlation coefficient)* ifade eder ve bu sayının karesi (r^2) 100 ile çarpıldığında elde edilen değer, bağımlı değişkendeki varyansın ne kadarının bağımsız değişken tarafından açıklandığını göstermektedir (Tomczak ve Tomczak, 2014).

Tablo 1. Hopkins'e göre korelasyon (r) değerlerine ilişkin etki büyüklüğü sınıflandırması

Önemsiz (trivial)	Küçük (small)	Orta (moderate)	Büyük (large)	Çok büyük (very large)	Mükemmele yakın (nearly perfect)
0	0,1	0,3	0,5	0,7	0,9

Bununla birlikte, iki ortalama arasındaki farkın etki büyüklüğünün raporlanması konusunda büyük ölçüde ortak bir görüş olsa da sınıflandırılması konusunda yeni arayışlar devam etmektedir. Cohen ilk olarak 1969 yılında etki büyüklüğü sınıflandırmasını küçük (0,2), orta (0,5) ve büyük (0,8) şeklinde sınıflandırırken (Cohen, 1969), daha sonra bu sınıflandırma üzerinde bir modifikasyon yaparak $d \leq 0,40$ ise küçük; $d = 0,41 - 0,70$ ise orta; $d > 0,70$ ise büyük şeklinde tanımlamıştır (Cohen, 1988). Daha sonra Cohen tarafından davranış bilimleri ve sosyal bilimlerde kullanılmak üzere yapılan bu sınıflandırmalar Sawilowsky (2009) tarafından genişletilmiş olsa da bu değer aralıklarının spor bilimleri ve uygulamalı bilimlerde geçerli olmadığına dair birçok karşıt görüş ortaya atılmıştır (Hopkins, 2002, 2019; Peterson ve ark., 2004; Rhea, 2004; Welsh ve Knight, 2015).

Tablo 2. Cohen'e göre iki ortalama arasındaki farka ilişkin etki büyüklüğü sınıflandırmaları (Cohen, 1969, 1988).

	Küçük (*small*)	Orta (*moderate*)	Büyük (*large*)
Cohen (1969)	0,2	0,5	0,8
Cohen (1988)	< 0,4	0,41 – 0,70	> 0,70

Tablo 3. Sawilowsky tarafından genişletilen iki ortalama arasındaki farka ilişkin etki büyüklüğü sınıflandırması (Sawilowsky, 2009).

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*

Çok küçük (*trivial*)	Küçük (*small*)	Orta (*moderate*)	Büyük (*large*)	Çok büyük (*very large*)	Muazzam (*huge*)
0,01	0,2	0,5	0,8	1,2	2

Cohen etki büyüklüğünü sınıflandırırken küçük etki büyüklüğünü tesadüfi olarak ortaya çıkamayacak büyüklükteki asgari ölçüt şeklinde tanımlarken, büyük etki büyüklüğünü zor ama ulaşılması mümkün olabilecek bir kriter olarak tanımlamıştır (Cohen, 1969). Daha sonra Rhea (2004) kuvvet antrenmanlarında etki büyüklüğünü sınıflandırmak için bir dizi meta-analiz çalışmasını (Peterson ve ark., 2004; Rhea ve Alderman, 2004; Rhea, Alvar ve Burkett, 2002; Rhea, Alvar, Burkett ve Ball, 2003) incelemiş ve 400'den fazla araştırmadan elde edilen yaklaşık 3000 etki büyüklüğünün ortalamasını 1,25 (± 1,0 standart sapma) olarak bulmuştur. Böylece Rhea, bu değerin Cohen'in çalışmalarında (Cohen, 1969, 1988) "büyük" olarak tanımlanan değerlerden (0,8 ve 0,7) çok yüksek olduğunu vurgulayarak katılımcıların antrenman statüsünü de göz önünde bulundurarak yeni bir sınıflandırma önermiştir (Rhea, 2004).

Tablo 4. Rhea'ya göre katılımcıların antrenman durumuna göre etki büyüklüğü sınıflandırması (Rhea, 2004).

Etkinin büyüklüğü	Düşük seviyede antrenmanlı bireyler (< 1yıl)	Orta seviyede antrenmanlı bireyler (1-5 yıl)	Yüksek seviyede antrenmanlı bireyler (> 5 yıl)
Önemsiz (*Trivial*)	< 0,50	< 0,35	< 0,25
Küçük (*Small*)	0,50 – 1,25	0,35 – 0,80	0,25 – 0,50
Orta (*Moderate*)	1,25 – 1,90	0,80 – 1,50	0,50 – 1,0
Büyük (*Large*)	> 2,0	> 1,5	> 1,0

Rhea sınıflandırmasının hem deneklerin antrenman statüsüne göre düzenlenmesi hem de etki büyüklüğü aralıklarının spor bilimlerine uygun olarak belirlenmesi alanda kabul görse de (Bernards ve ark., 2017; Dankel ve ark., 2017; Flanagan, 2013), sadece kuvvet antrenmanlarına özel olması sebebiyle çok fazla kapsayıcı olamamıştır. Spor bilimlerinde daha çok kabul gören sınıflandırma Hopkins tarafından yapılmıştır (Hopkins, 2002) ve güncel çalışmalarda da yaygın olarak kullanılmaktadır (Bernards ve ark., 2017; Hazir, Kose ve Kin-Isler, 2018; Panissa ve ark., 2018).

Tablo 5. Hopkins'e göre iki ortalama arasındaki farka ilişkin etki büyüklüğü sınıflandırması (Hopkins, 2002).

Önemsiz (*trivial*)	Küçük (*small*)	Orta (*moderate*)	Büyük (*large*)	Çok büyük (*very large*)	Mükemmele yakın (*nearly perfect*)
< 0,2	0,2 – 0,59	0,60 – 1,19	1,20 – 1,99	2,0 – 3,99	> 4,0

Özetle yokluk hipotezinin test edilmesine dayalı *p-değeri* temelde bir farkın (veya bir etkinin, bir ilişkinin) varlığı ya da yokluğu hakkında bilgi sağlarken, bu bilginin derecelendirilmesine izin veren etki büyüklüğüdür (Tomczak ve Tomczak, 2014). Ancak araştırmacılar için etki büyüklüğünün sınıflandırmasına ilişkin farklı seçenekler bulunmaktadır (Cohen, 1969, 1988; Hopkins, 2002; Rhea, 2004; Sawilowsky, 2009). Bu seçeneklerin tam olarak hangisinin doğru ya da yanlış olduğunu söylemek mümkün değildir. Ancak yaygın olarak kabul edildiği şekliyle deneklerin antrenman statüsünü de göz önüne alarak kuvvet antrenmanın etkisinin incelendiği çalışmalarda Rhea (2004) sınıflandırması, herhangi bir antrenman müdahalesi, besin takviyesi

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*

veya ısınma protokolü gibi bir denemenin yapıldığı araştırmalarda Hopkins (2002) sınıflandırması daha fazla kabul görmektedir. Cohen (1969) ve Sawilowsky (2009) sınıflandırmaları ise davranış bilimleri ve sosyal bilimler alanları temel alınarak geliştirildiği için bu disiplinler ile ilişkili olan spor ve sağlık bilimleri çalışmalarında kullanılması uygun gözükmektedir.

En küçük değerli değişim *(Smallest worthwhile change, SWC).*

Spor bilimlerinde, diğer disiplinlerden farklı olarak istatistik kullanımındaki bir diğer paradoks da örneklem sayısı ve gözlenebilir fark arasındaki etkileşimdir. Özellikle elit seviyedeki sporcular, uluslararası şampiyonalarda madalya alan sporcular veya nadir görülen hastalıklara sahip kişiler gibi az sayıdaki gruplarda çalışmak isteyen araştırmacıların bu değerli çabasının istatistiksel açıdan anlamlı olmayan sonuçlar ortaya çıkarması muhtemeldir. Çünkü anlamlılık tanımı için kullanılan *p-değerinin* bir farkı ortaya koyabilmesinde örneklem sayısının fazla ya da farkın oldukça büyük olması avantaj sağlamaktadır (Cohen, 1988; Sullivan ve Feinn, 2012; Welsh ve Knight, 2015). Ancak hem az sayıdaki elit sporculara ulaşmanın güç olması hem de bu sporcuların mevcut performans çıktılarının birbirine yakın olması sebebiyle yapılan herhangi bir müdahale diğer gruplara göre sayısal açıdan daha küçük gelişimler sağlayacağından istatistiki açıdan anlamlı sonuçlara ulaşmayı zorlaştırmaktadır. Bununla birlikte uygulamada ise istatistiki anlamlılıktan bağımsız olarak özellikle elit sporcuların performanslarını çok küçük düzeyde artışlar sağlayabilecek yöntemler bile oldukça değerlidir (Bernards ve ark., 2017; Mengersen ve ark., 2016; Rhea, 2004).

Özellikle, bir uzun vadeli sporcu gelişim programında fiziksel performansların belli periyotlar ile ölçülmesinin gerekliliği açıktır. Uygulanan beslenme, antrenman, psikolojik destek veya benzeri bir müdahalenin etkisinin hangi boyuta ulaştığında anlamlı kabul edildiğini bilmek oldukça önemlidir. Bununla birlikte belli aralıklarda ölçülen test sonuçlarındaki değişimin performansın yanı sıra küçük miktarlarda da olsa yorgunluk seviyesi, uyku düzeni veya bir besin takviyesinin geçici etkisi gibi sebeplerle çeşitlilik göstermesi kaçınılmazdır. Dolayısıyla ölçümlerdeki değişkenliğin de hesaplamaya dahil edildiği en küçük değerli değişim (*smallest worthwhile change, SWC*) bu tür durumlara uygun bir çözüm olarak sunulmuştur (Bernards ve ark., 2017; Coyne ve ark., 2015; Özbay ve Ulupınar, 2019; Pyne, 2003). SWC uygulamada önemli olarak kabul edilen en küçük değişim miktarı olarak tanımlanabilir (Bernards ve ark., 2017; Pyne, 2003). SWC aşağıdaki formül ile kolayca hesaplanabilmektedir (Pyne, 2003).

$$SWC = 0{,}2 \times \text{denekler arası standart sapma}$$

Formülden anlaşılacağı gibi SWC bir grubun bir kez ölçülmesi durumunda hesaplanabilmektedir. Ancak SWC değerinin uygulamada kabul edilebilir olması için ölçümlerin standart hatasının (*standart error of measurements*) ve tipik hatasının (*typical error of the measurements*) SWC değerinden daha küçük olmasının gerekli olduğu belirtilmektedir (Bernards ve ark., 2017). Ancak bu hata terimlerinin hesaplanabilmesi için en az iki ölçümün gerçekleştirilmesi gerekmektedir. Tipik hata, özellikle test-tekrar test güvenirlik analizlerini içeren çalışmalarda iki deneme arasındaki varyasyonun (değişimin) bir ölçütü olarak kullanılmaktadır (do Nascimento ve ark., 2017; Hopkins, 2000; Özbay ve Ulupınar, 2019; Özbay ve ark., 2019). Örneğin 12 haltercinin katıldığı bir çalışmada ikişer kez koparma kaldırışı sonunda, katılımcıların iki ölçümü arasındaki farklarının sırasıyla 2, 3, 5, 0, 4, 0, 2, 1, 0, 2, 3, 1 kg olduğunu varsayalım. Bu durumda iki ölçüm arasındaki farklardan elde edilen 12 değerin standart sapması (ölçümler arası farkların standart sapma) 1,62 kg; tipik hatası da 1,15 kg olacaktır. Bu çalışmada, katılımcılar arası standart sapma değerinin 0,2 ile çarpılması ile elde edilen SWC, tipik hatadan daha büyük olması durumunda pratikte kullanılabilir bir ölçüt

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

sağlayabilmektedir (Hopkins, 2000). Aksi takdirde SWC, küçük etki boyutunun eşik değeri olan 0,2 yerine, orta düzey etki boyutunun eşik değeri olan 0,6 ile çarpılarak kullanılabilir (Ferreira da Silva Santos, Lopes-Silva, Loturco ve Franchini, 2020).

Tipik hata = Ölçümler arasındaki farkların standart sapması / √2

Tipik hata formülünde √2 değerinin kullanılması her iki ölçümün ayrı ayrı hata varyansı içermesinden kaynaklanmaktadır (Weir, 2005). Başka bir deyişle, bir gruptan iki ölçüm alındığı zaman her iki ölçümün de kendi içinde bir miktar hata içerdiği kabul edilir. Tipik hatanın küçük olması, birinci ölçümdeki sonuçların ikinci ölçümde büyük oranda tekrar edildiğini gösterir böylece güvenilir sonuçların üretildiğine dair bir çıkarım yapmayı mümkün kılar.

Ayrıca literatürde yokluk hipotez testlerinin analiz edildiği yaklaşıma dayanarak en küçük gerçek fark (*smallest real difference*) veya tespit edilebilen minimal değişim (*smallest detectable difference, minimal detectable change, MDC*) olarak adlandırılan ve benzer bir amaca hizmet eden hesaplama formülü de aşağıda sunulmaktadır (Guyatt, Kirshner ve Jaeschke, 1992). Ancak bu formülde kullanılan SEM (ölçümlerin standart hatası) değeri en az iki ölçümün yapılmasını gerektirmektedir.

$$\text{MDC} = 1{,}96 \times \sqrt{2} \times SEM$$

Formülde yer alan 1,96 değeri %95 güven aralığı sınırları ile ilişkili z skorudur. Dolayısıyla formülde yer alan 1,96 değeri çalışmalarda en sık tercih edilen % 95 güven aralığı sınırları içindir, ancak araştırmacının daha muhafazakâr ya da daha liberal bir aralık tercihine göre bu sayı değişebilir. Formülde yer alan SEM ölçümlerdeki tesadüfi (rastgele, random) hataların standartlaştırılmış bir değeridir ve standart sapmadan ve güvenirlikten oldukça etkilenmektedir. SEM; sınıf-içi korelasyon katsayısı (*ICC*) değeri (r ile gösterilebilir) ve standart sapma kullanılarak aşağıdaki şekilde hesaplanır.

$$SEM = \text{ss} \times \sqrt{(1 - \text{r})}$$

Formülden de anlaşılacağı gibi SEM'in düşmesinin iki temel sebebi denekler arası standart sapmanın düşük olması ve/veya güvenirlik değerinin yüksek olmasıdır. Örneğin 40 m sprint sürelerinin ölçüldüğü bir takımda denekler arası standart sapma 0,30 saniye; r değeri (örneğin test − tekrar test sonucu arasındaki ICC) 0,90 ise; $SEM = 0{,}30 \times \sqrt{(1 - 0{,}90)}$ = 0,0949 olarak hesaplanır. Böylece en küçük gerçek fark = 1,96 × √2 × 0,0949 = 0,26 s olarak bulunur. Yani 40 m sprint süresi 5,40 s olan bu grubun ortalamasının en az 5,14 s (5,40 – 0,26) olması durumunda pratikte önemli bir performans gelişimi sağlanmış olur.

Tipik hata ve SEM bazı çalışmalarda aynı isim altında değerlendirilse de (do Nascimento ve ark., 2017; Hopkins, 2000; Weir, 2005) tipik hata daha çok test – tekrar test arasındaki varyasyonu vurgulamak için uygundur. Tipik hata temelde katılımcıların iki ölçüm arasındaki tutarlılığını vurguladığı için hesaplamasında ölçümler arasındaki farkın standart sapması (ölçümler-arası standart sapma) kullanılır. SEM hesaplanırken ise bir grubun denekler-arası standart sapma değerleri kullanılır. SEM, ölçme işlemindeki hataların miktarını göstermez ancak sonuçların hatadan arınıklık derecesinin standartlaştırılmış bir göstergesi olarak kullanılır. Örneğin bir çalışmadaki tüm katılımcıların ikinci ölçümdeki maksimum koparma kaldırışı skorları ilk ölçümdekinden 2 kg daha fazla olursa tipik hata 0 olacaktır (0 / √2 = 0). Bu örnek bize aynı zamanda tipik hatanın ölçümlerdeki sabit hataları değil, sistematik ve rastgele hatalara duyarlı olduğunu da göstermektedir. Dolayısıyla ikinci ölçümde tüm

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

değerlerin 3 birim yükseldiği veya 5 birim düştüğü iki durum için de tipik hata 0 olacaktır. SEM değerinin 0 olabilmesi için ise katılımcı grubun tüm üyelerinin aynı değere sahip olması $0 \times \sqrt{(1-\mathrm{r})=0}$ veya güvenirlik değeri olan ICC'nin maksimum değere ulaşması (1 olması) gerekir ($\mathrm{SS} \times \sqrt{(1-1)=0}$). Her iki hesaplama için de standart sapma değerinin küçülmesi (güvenirliğin yükselmesi, ölçümler arasındaki tutarlılığın yüksek olması) hata miktarlarının daha küçük bir değere sahip olması anlamına gelmektedir. Böylece daha küçük SEM ve tipik hata değerleri, ölçümlerin hatadan arınmış olduğunu ve daha küçük farkların araştırma müdahalesinin bir etkisi olarak kabul edilebileceğini göstermektedir.

Test–tekrar test tasarımları içeren çalışmalarda sınıf-içi korelasyon katsayısının katılımcıların heterojenliğine aşırı duyarlı olduğundan ölçümlerin güvenilirliğinden emin olmak için tek başına yeterli olmayacağı belirtilmektedir (Hopkins, 2000). Bu sebeple, pratikte işlevsel olabilecek tipik hata, SEM, SWC ve MDC gibi analizlerin de sunulması gerektiği vurgulanmaktadır (Hopkins, 2000; Wilkinson ve ark., 2019). MDC ve SWC düşük varyasyon içeren çalışma tasarımlarında kullanmak için oldukça elverişlidir. Örneğin test–tekrar test güvenirliği yüksek olan yöntemlerde ölçümler arasında düşük varyasyon ürettiği için verilerin güvenilir ve sistematik hatalardan arınmış olduğu kabul edilir (Bernards ve ark., 2017; Coyne ve ark., 2015; Özbay ve Ulupınar, 2019). Dolayısıyla ölçümler arasındaki varyasyonu azaltmak performansta önemli kabul edilebilecek küçük miktardaki gerçek bir değişimi tespit etme fırsatı sunar (Bernards ve ark., 2017; Coyne ve ark., 2015). Böylece elit sporcuların birbirine yakın skorları sebebiyle varyasyonun daha küçük olması beklendiği için, elit sporcuların kullanıldığı çalışmalarda *p-değeri* ile karar verilen anlamlılık yargısının oluşturduğu dezavantaja bir çözüm sunabilir. Bir başka deyişle elit sporcular üzerinde yapılan çalışmalarda denek sayısı az olduğu için *p-değerinin* anlamsız çıkma olasılığı artarken, grup içi homojenliğin yüksek olması sebebiyle SWC daha düşük olacaktır. Örneğin 40 m sürelerinin ölçüldüğü bir çalışmada elit sporcuların standart sapması 0,30 saniye iken, amatör sporcuların standart sapmasının 0,80 saniye olduğunu varsayalım. Bu durumda önemli bir performans gelişimi sayılabilmesi için elit sporcuların 0,06 saniye (0,2 × 0,30), amatör sporcuların ise 0,16 saniye (0,2 × 0,8) daha hızlı koşmaları gerekecektir.

SWC hesaplanırken standart sapmanın 0,2 ile çarpılması Cohen (1969)'un etki büyüklüğü hesaplanmasına dayanmaktadır. Literatürde 0,2 değeri pratikte küçük ama önemli sayılabilecek en küçük eşik değeri olarak kabul edilirken hem Cohen hem de Hopkins tarafından yapılan sınıflandırmalardaki tek ortak değerdir. Ancak daha önce de belirtildiği gibi SWC değerinin uygulamada kullanılabilmesi için ölçümlerin mümkün olduğunca hatadan arınmış olması gerekmektedir. Bunun için ölçümleri standartlaştırmak ve uyku düzeni, beslenme, ölçüm cihazı, test ortamı ve yorgunluk gibi faktörlerin etkisini minimum düzeye indirmek gerekmektedir (Peltola, 2005).

Bayesci İstatistiksel Yaklaşım

İngiliz matematikçi Thomas Bayes'in (1702-1761) ismiyle anılan Bayes teoreminin temelini bir olayın gerçekleşmiş olmasının başka bir olayın gerçekleşme olasılığına etkisi oluşturmaktadır (Bayes, 1991). Bayesci istatistik modeli, model parametreleri için verinin içermiş olduğu objektif bilgi ile kişisel, bilimsel ya da geçmişten gelen önsel bilgilerin birleştirilmesine ve parametrelere ilişkin çıkarımların yapılmasına olanak sağlayan bir yöntemdir (Bernards ve ark., 2017; Mengersen ve ark., 2016). Parametrelere ilişkin önsel bilgiler öznel (subjektif) bir nitelik taşır ve uygun dağılımlarla modelleme sürecine dahil edilirler. Bu nedenle Bayesci istatistik modeli, önsel bilgilerden yola çıkarak önsel dağılımlara, sonsal dağılımlardan yola çıkarak sonsal bilgilere ulaşma çabasını içerir (Bayes, 1991; Mengersen ve ark., 2016).

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

Bayesci istatistik modelinin mevcut istatistiksel yaklaşıma kıyasla eleştirildiği en temel nokta bir bilginin tahmininde kesinlik içermeyen gözlemleri ve subjektif görüşleri kullanmasıdır. Bayes teoreminin detayları oldukça karmaşık ve halen tartışmaların devam ettiği bir istatistik modelidir. Bununla birlikte Bayes teoreminin temel prensipleri kısaca şu şekilde sıralanmaktadır (Bayes, 1991; Bernards ve ark., 2017; Mengersen ve ark., 2016).
1. Bir olayın kesin olma ihtimal 1, asla olmama ihtimali 0 olarak kabul edildiğinde, her olay 0 ve 1 arasında bir olasılığa eşittir.
2. Bir olayın gerçekleşme ve gerçekleşmeme olasılığının toplamı daima 1'e eşittir.
3. İki olayın birlikte gerçekleşme ihtimali, birincinin olma ihtimali ile ikincinin birinci ile birlikte olma ihtimalinin çarpımına eşittir.

Bu teoremin temel mantığını anlatmak için oldukça basit örnekler kullanılsa da bilimsel çalışmalarda bir istatistik modeli olarak kullanılmasında kapsamlı hesaplamalar gerektiren ve oldukça farklı görüşler içeren bir modeldir. Basit bir örnekten yola çıkacak olursak, aynı anda 5 adet bozuk paranın havaya atıldığında en az 3'ünün yazı gelme olasılığını objektif olarak kolaylıkla hesaplayabiliriz. Ancak 1 numaralı paranın tura geldiği bilindiğinde bu olasılık değişecektir. Ya da paralardan ikisinin hileli olduğu ve tura gelme olasılıklarının %50 yerine % 80 olduğu bilindiğinde hesaplanan olasılık yine değişecektir. Dolayısıyla bir olasılığın bir olayın olmasından veya bilinen bir bilgiden etkilenmesi Bayes yaklaşımının temelini oluşturmaktadır. Ancak burada verilen örnek objektif niteliktedir ve sayısal hesaplamalara dayanmaktadır. Bilimsel bir araştırma tasarımında ise bir olasılığın etkilendiği bilgi ve gözlemin genelde subjektif, yani yoruma dayalı olması bu yaklaşımın temel eleştirisidir. Bayesci yaklaşımda objektif ve subjektif analizler konusunda tartışmalar devam ederken yakın zamanda bütüncül bir istatistik yaklaşımı için bir fikir birliğinin oluşmasının beklenmediği belirtilmektedir (Çelebi, 2019; Mengersen ve ark., 2016).

Mevcut istatistik yaklaşımından farklı olarak Bayesci yaklaşımda temel mantık tümevarım olasılığından yola çıkarak denemeler yoluyla en yüksek kesinliğe ulaşmaya çalışmaktır. Bazı araştırmacılar tarafından olasılıklara ilişkin doğrulamalar yaparak ilerleyen bir yapı olan Bayesci yaklaşımın daha geniş kapsamlı bir çerçeve sunduğu savunulmaktadır (Bayes, 1991; Mengersen ve ark., 2016). Başka bir deyişle mevcut istatistik yaklaşımındaki bir hipotezin doğru ya da yanlış olmasına karar vermekten ziyade Bayes modelinde hipotezin sahip olduğu düşünülen olasılık belirlenmeye çalışılır (Çelebi, 2019; Mengersen ve ark., 2016). Bayesci istatiksel modelinin sonuçlar üzerinde doğrudan bir olasılık karşılaştırmasına imkan verdiği ve örneklem sayısının az olduğu durumlarda küçük etki büyüklüklerini tespit etmek için elverişli olduğu savunulmaktadır (Bayes, 1991; Mengersen ve ark., 2016). Mevcut istatistiksel yaklaşımlar ile Bayesci modelin karşılaştırıldığı bir çalışmada, Bayesci modelin daha küçük farkların saptanması ve daha kullanışlı olasılık yorumları sağladığı ortaya koyulmuştur (Mengersen ve ark., 2016).

Bilimsel araştırmalarda bir istatistik modeli olarak Bayes teoreminde subjektif bilgi, henüz veriyi gözlemlemeden parametre hakkında araştırmacının tecrübeleri, önceki çalışmaların sonuçları veya uzman kişilerin görüşlerinden elde edilen ve önsel olarak adlandırılan bir bilgidir (Bayes, 1991). Önsel bilgi kullanılarak önsel bir olasılık ile örneklemden elde edilen olabilirlik fonksiyonu ile gelen objektif bilginin birleştirilmesi sonucu sonsal olasılık ortaya çıkar. Sonsal bilgi ve ona karşılık gelen sonsal olasılık, veriyi gözlemledikten sonra tahmin edilmesi istenen parametre hakkındaki olasılık değeridir (Bayes, 1991; Çelebi, 2019; Mengersen ve ark., 2016). Sonuç olarak elde edilen subjektif ve objektif bilgi bütünü parametre hakkında çıkarsama yapmak ve karar vermek için kullanılmaktadır. Giriş düzeyinde değinilen Bayesci yaklaşımın

Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

derinliği bu çalışmada bahsedilenden çok daha fazladır. Önsel dağılım türlerinin yanı sıra objektif ve subjektif analiz yaklaşımı gibi kendi içinde büyük bir kapsama dahil olan bu yaklaşımın sadece temel bilgileri bu çalışmanın konusu olarak incelenmiştir.

Büyüklük Temelli Çıkarımlar Yöntemi (*Magnitude based inferences, MBI*)

Batterham ve Hopkins (2006) yokluk hipotezi anlamlılık testlerinin uygulamadaki sınırlılıklarından dolayı Büyüklük Temelli Çıkarım (*Magnitude Based inferences, MBI*) ismiyle yeni bir istatistiksel çıkarım yöntemi geliştirmişlerdir. Bu yöntemde güven aralıkları, en küçük değerli değişim ile ilişkili olarak yorumlanmaktadır (Batterham ve Hopkins, 2006). MBI yöntemi geliştirildikten hemen sonra spor bilimi topluluğunda büyük bir ilgi görmüştür ve bilimsel araştırmalarda kullanılmaya başlanmıştır (Hopkins ve ark., 2009; Hopkins, 2017; Lohse ve ark., 2020). Bu yöntem, etki büyüklüklerine daha fazla vurgu yapmayı temel alarak pratikteki geçerliliği arttırma arzusundan doğmuştur. MBI, elde edilen sonuçları zararlı (negatif), önemsiz ve faydalı (pozitif) olarak üç etki büyüklüğü kategorisine ayırmaktadır. Standartlaştırılmış etki büyüklüğü birimleri ise − 0,2 ile + 0,2 arasındaki bir değer alabilmektedir.

Bir araştırmanın raporladığı değerler genellikle örneklemden elde edilen değeri kullanarak temsil ettiği ana kütlenin gerçek değeri hakkında bir çıkarım yapabilmeyi sağlamaktadır (Barker ve Schofield, 2008). Araştırmacılar geleneksel olarak, yokluk hipotez testinden türetilen bir *p-değeri* temelinde istatistiksel olarak anlamlı veya anlamlı olmayan bir değer beyan ederek çıkarımda bulunur (Cumming, 2014; Sullivan ve Feinn, 2012). Bu yaklaşımın yeterince açık olmadığı ve grubun dağılımına, ölçüm hatalarına ve örneklem büyüklüğüne bağlı olarak yanıltıcı olabileceği savunulmaktadır (Hopkins ve ark., 2009; Hopkins, 2017, 2019; Welsh ve Knight, 2015). MBI, istatistiğin gerçek değerindeki belirsizliğe dayalı daha sezgisel ve pratik bir yaklaşım olarak ileri sürülmüştür (Batterham ve Hopkins, 2006). İlk olarak MBI'de, "belirsizlik" gerçek değerin olası aralığını tanımlayan güven sınırları olarak ifade edilmektedir, daha sonra "yararlı" ve "zararlı" gibi bazı olumlu veya olumsuz anlamda pratikte kullanılabilir istatistik tanımlamaları dikkate alarak sonuçların sahaya aktarılması ile ilgilenmektedir (Hopkins, 2019). Örneğin, "çok büyük olasılıkla yararlıdır" gibi gerçek değerin gözlemlenen büyüklüğe sahip olma olasılığının nitel olarak açıklanmasının çıkarımı daha açık hale getirdiği kabul edilmektedir (Hopkins ve ark., 2009; Welsh ve Knight, 2015). Ayrıca gerçek değerin derecelendirilmiş büyüklüklere (önemsiz, küçük, orta ve büyük gibi) sahip olduğuna dair nicel olasılıkların da birlikte sunulabilmesi sonucun faydası hakkında karar vermeye rehberlik edebilmektedir (Batterham ve Hopkins, 2006).

MBI ilk olarak 2006 yılında uluslararası hakemli bir dergide tanıtılmıştır. Daha sonra aynı dergide, Barker ve Schofield (2008) tarafından eleştirilmiştir. Yazarlara göre, "Müdahalenin faydalı olma ihtimali % 90'dır" gibi bir ifade, tamamen Bayesci bir yaklaşımı benimsemektir (Barker ve Schofield, 2008). Diğer bir eleştiri ise MBI'nin güven aralıkları ile yanlış ve aşırı iyimser çıkarımlar yapıldığı yönündedir (Sainani, 2018). Yakın tarihli bir yorumda Sainani (2018) ise klasik *p-değeri* ve güven aralıklarının ötesine geçme fikrinin iyi bir fikir olmasına rağmen, MBI'nin hata kontrolü açısından sorunları olduğuna işaret etmektedir. Sainani (2018) MBI'nin örneklem büyüklüğüne ve yarar/zarar eşiklerine bağlı olarak, geleneksel hipotez testlerine göre 2 ila 6 kat daha büyük olabilen pozitif oranlar ürettiğini ileri sürmektedir. Buna ilaveten Sainani'ye (2018) göre MBI olasılık terimleri Bayesci değildir ve keyfi bir bakış açısı üzerine temellendirilmiştir. Sainani'nin (2018) gündeme getirilen bu eleştirileri ilk değildir ancak alandaki etkisi diğerlerine göre daha büyük olmuştur. Bu çalışma sonrası spor bilimleri alanındaki önemli dergiler, sadece MBI analizi kullanılarak gerçekleştirilmiş araştırmaları reddetmektedir veya hipotez testleri ile desteklenmesini istemektedir. Batterham ve Hopkins

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

(2018) ise bu eleştirilere cevap olarak MBI'nin olasılık tanımlarının Bayesçi yaklaşımla örtüştüğünü ileri sürmüşlerdir. Ayrıca diğer araştırmacılar tarafından kullanılan çıkarımsal hata oranlarını yeniden analiz ederek Sainani'nin (2018) iddialarını reddetmişlerdir (Hopkins ve Batterham, 2018). Bir başka çalışmada ise MBI'nin, hipotez testleri olarak "Büyüklük temelli kararlar" (MBD) isim değişikliği ile sunulması durumunda, istatistik topluluğunun en azından bazı üyeleri için daha kabul edilebilir olabileceği öne sürülmüştür (Hopkins 2019). Sainani'nin yazar olarak yer aldığı yakın tarihli bir derlemede ise benzer olarak MBI'nin, aşırı iyimser sonuçlar nedeniyle spor bilimi ve tıp literatürüne zarar verdiği ve ayrıca MBI'nin araştırmacıları küçük örneklemli çalışmalara teşvik ettiği vurgulanmıştır (Lohse ve ark., 2020). Sonuç olarak MBI veya yeni ismiyle MBD'nin *p*-değerlerine uygun bir alternatif olup olmadığı konusundaki tartışmaların bir süre daha devam etmesi muhtemel görünmektedir.

TARTIŞMA VESONUÇ

Diğer disiplinlerde olduğu gibi spor bilimleri alanında da akademik araştırmaların sonuçlarının saha uygulamalarına aktarılması önemlidir. Ancak birçok çalışmada sadece yokluk hipotezinin test edildiği anlamlılık sonuçları ve *p-değerinin* sunulması sonuçların pratikte yorumlanmasını ve kullanılmasını zorlaştırmaktadır. Mevcut istatistik yaklaşımının araştırma sonuçlarını anlamlı−anlamsız veya fark var−yok şeklinde ikili kategorize ederek açıklaması alternatif istatistik yaklaşımlarının ortaya çıkmasına sebep olmuştur. Bayesci istatistik modeli ve büyüklük temelli çıkarımlar modeli (MBI) belli bir periyot için benimsenmişlerse de bu modellerin yeterli olmadığı konusundaki görüşler daha yaygındır. Diğer taraftan, *p-değerinin* etki büyüklüğü (Cohen's d, Hedges'g), en küçük değerli değişim (SWC), tespit edilebilen minimal değişim (MDC), ölçümlerin standart hatası (SEM) ve tipik hata (TE) gibi pratikte kullanılabilecek yöntemler ile desteklenmesi genel bir kabul haline gelmiştir. İkiden fazla grup karşılaştırmasını veya bir grubun ikiden fazla ölçümünü içeren çalışmalarda ise kısmi eta kare katsayısı (η_p^2), istatistiksel programlarda da sunulduğu için yaygın olarak tercih edilmektedir. Korelasyon ve regresyon katsayıları ise doğrudan bir ilişki veya çıkarım hakkında bilgi verdiği için herhangi bir dönüşüme ihtiyaç duyulmadan kullanılabilmektedir.

Diğer taraftan konu ile ilgili tartışmalar sürerken mevcut durumda, performans araştırmalarında veya fiziksel ölçümlerin kullanıldığı tasarımlarda Cohen's d ya da Hedges'g değerlerinin hesaplanarak Hopkins'e göre sınıflandırılması daha uygun gözükmektedir. Ancak Cohen's d değeri hesaplanırken ilk ortaya çıktığı şekliyle değil, her iki standart sapmanın da hesaplamaya dahil edildiği revize edilmiş formunun kullanılması daha uygundur. Diğer taraftan bir kuvvet antrenmanının etkisine odaklanan çalışmalarda aynı zamanda katılımcıların antrenman seviyeleri de göz önünde bulundurulmak isteniyorsa Rhea sınıflandırması geçerli bir alternatiftir. Cohen ve Sawilowsky tarafından yapılan sınıflandırmalar ise daha çok davranışsal değişimlerin değerlendirildiği veya tutum ölçekleri gibi radikal değişimlerin beklenmediği ölçme araçlarının kullanıldığı çalışmalarda tercih edilebilmektedir.

Etki büyüklüğü sınıflandırmaları arasındaki çeşitlilik temelde ölçülen özelliklerin değişim potansiyeli ile ilgilidir. Örneğin, ortalama 60 kg ağırlık ile skuat yapabilen bir grup altı aylık bir antrenman protokolünün ardından bu ağırlığı yaklaşık iki katına çıkarabilmektedir. Ancak bir duruma özgü tutum ölçümünde bu süre içinde bu büyük farkla değişim gerçekleşmesi çok muhtemel değildir. Bu sebeple araştırmalarda odaklanılan değişkenlere uygun etki büyüklüğü sınıflandırmasını tercih etmek önemlidir.

*Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com

KAYNAKLAR

Alpar, R. (2010). *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik*: Detay yayıncılık.

Altman, D. G., Bland, J. M. (2005). Standard deviations and standard errors. *Bmj, 331*(7521), 903.

Barker, R. J., Schofield, M. R. (2008). Inference about magnitudes of effects. *International journal of sports physiology and performance, 3*(4), 547-557.

Batterham, A. M., Hopkins, W. G. (2006). Making meaningful inferences about magnitudes. *International journal of sports physiology and performance, 1*(1), 50-57.

Bayes, T. (1991). An essay towards solving a problem in the doctrine of chances. 1763. *MD computing: computers in medical practice, 8*(3), 157.

Bernards, J. R., Sato, K., Haff, G. G., Bazyler, C. D. (2017). Current research and statistical practices in sport science and a need for change. *Sports, 5*(4), 87.

Bortz, J., Döring, N. (2007). *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler: Limitierte Sonderausgabe*: Springer-Verlag.

Cohen, J. (1969). Statistical Power Analysis for the Behavioral Science: New York: Academic Press.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and psychological measurement, 33*(1), 107-112.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, 2nd edn. Á/L: Erbaum Press, Hillsdale, NJ, USA.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Academic press.

Coyne, J. O., Tran, T. T., Secomb, J. L., Lundgren, L., Farley, O. R., Newton, R. U., Sheppard, J. M. (2015). Reliability of pull up ve dip maximal strength tests. *J Aust Strength Cond, 23*, 21-27.

Cumming, G. (2014). The new statistics: Why and how. *Psychological science, 25*(1), 7-29.

Çelebi, V. (2019). Bayes teoremi bağlamında olasılıkçı Bayes epistemolojisinin kapsamı üzerine bir inceleme. *Felsefe ve Sosyal Bilimler Dergisi (FLSF)*(28).

Dankel, S. J., Mouser, J. G., Mattocks, K. T., Counts, B. R., Jessee, M. B., Buckner, S. L., . . . Loenneke, J. P. (2017). The widespread misuse of effect sizes. *Journal of Science and Medicine in Sport, 20*(5), 446-450.

do Nascimento, M. A., Ribeiro, A. S., de Souza Padilha, C., da Silva, D. R. P., Mayhew, J. L., do Amaral Campos Filho, M. G., Cyrino, E. S. (2017). Reliability and smallest worthwhile difference in 1RM tests according to previous resistance training experience in young women. *Biology of Sport, 34*(3), 279-285.

Ferreira da Silva Santos, J., Lopes-Silva, J. P., Loturco, I., Franchini, E. (2020). Test-retest reliability, sensibility and construct validity of the frequency speed of kick test in male black-belt taekwondo athletes. *Ido Movement for Culture. Journal of Martial Arts Anthropology, 20*(3), 38-46.

Flanagan, E. P. (2013). The effect size statistic—Applications for the strength and conditioning coach. *Strength ve Conditioning Journal, 35*(5), 37-40.

Fritz, C. O., Morris, P. E., Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General, 141*(1), 2.

Fröhlich, M., Emrich, E., Pieter, A., Stark, R. (2009). Outcome effects and effects sizes in sport sciences. *International Journal of Sports Science and Engineering, 3*(3), 175-179.

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*

Greenland, S., Senn, 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. *European journal of epidemiology, 31*(4), 337-350.

Guyatt, G. H., Kirshner, B., Jaeschke, R. (1992). Measuring health status: what are the necessary measurement properties? *Journal of clinical epidemiology, 45*(12), 1341-1345.

Hazir, T., Kose, M. G., Kin-Isler, A. (2018). The validity of running anaerobic sprint test to assess anaerobic power in young soccer players. *Isokinetics and Exercise Science, 26*(3), 201-209.

Hedges, L. V. ve Olkin, I. (2014). *Statistical methods for meta-analysis*: Academic press.

Hopkins, W. G., Batterham, A. (2018). The vindication of magnitude-based inference. *Sportscience, 22*, 19-29.
Hopkins, W. G., Marshall, S., Batterham, A., Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine+ Science in Sports+ Exercise, 41*(1), 3.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports medicine, 30*(1), 1-15.

Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. *A new view of statistics, 502*, 411.

Hopkins, W. G. (2017). Estimating Sample Size for Magnitude-Based Inferences. *Sportscience, 21*.

Hopkins, W. G. (2019). Compatibility intervals and magnitude-based decisions for standardized differences and changes in means. *Sportscience*.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863.

Levine, T. R., Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625.

Lohse, K., Sainani, K., Taylor, J. A., Butson, M. L., Knight, E., Vickers, A. (2020). Systematic Review of the use of "Magnitude-Based Inference" in Sports Science and Medicine.

Mengersen, K. L., Drovandi, C. C., Robert, C. P., Pyne, D. B., Gore, C. J. (2016). Bayesian estimation of small effects in exercise and sports science. *PloS one, 11*(4), e0147311.

Özbay, S., Ulupınar, S. (2019). Reliability of 1RM, 5RM and 10RM Tests in Upper Body Resistance Exercises. *The Journal of Turkish Sport Sciences for Health, 2*(1), 1-7.

Özbay, S., Ulupınar, S., Çınar, V., Akbulut, T. (2019). Reliability of Easily Applicable Non-Laboratory Methods Used for Determination of the Upper Body Strength. *Turkiye Klinikleri Journal of Sports Sciences, 11*(2).

Panissa, V. L., Fukuda, D. H., Caldeira, R. S., Gerosa-Neto, J., Lira, F. S., Zagatto, A. M., Franchini, E. (2018). Is oxygen uptake measurement enough to estimate energy expenditure during high-intensity intermittent exercise? Quantification of anaerobic contribution by different methods. *Frontiers in Physiology, 9*, 868.

Peltola, E. (2005). Competitive performance of elite track-and-field athletes: variability and smallest worthwhile enhancements. *Sportscience, 9*, 17-21.

Peterson, M. D., Rhea, M. R., Alvar, B. A. (2004). Maximizing strength development in athletes: a meta-analysis to determine the dose-response relationship. *The Journal of Strength ve Conditioning Research, 18*(2), 377-382.
Pyne, D. B. (2003). *Interpreting the results of fitness testing.* Paper presented at the International science and football symposium.

Rhea, M. R. (2004). Determining the magnitude of treatment effects in strength training research through the use of the effect size. *Journal of strength and conditioning research, 18*, 918-920.

Rhea, M. R., Alderman, B. L. (2004). A meta-analysis of periodized versus nonperiodized strength and power training programs. *Research quarterly for exercise and sport, 75*(4), 413-422.

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*

Rhea, M. R., Alvar, B. A.,Burkett, L. N. (2002). Single versus multiple sets for strength: a meta-analysis to address the controversy. *Research quarterly for exercise and sport, 73*(4), 485-488.

Rhea, M. R., Alvar, B. A., Burkett, L. N.,Ball, S. D. (2003). A meta-analysis to determine the dose response for strength development.

Rosenthal, R., Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*: CUP Archive.

Rosnow, R. L., Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 57*(3), 221.

Sainani, K. L. (2018). The Problem with" Magnitude-based Inference". *Medicine and science in sports and exercise, 50*(10), 2166-2176.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 26.

Sullivan, G. M., Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education, 4*(3), 279-282.

Tomczak, M., Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength ve Conditioning Research, 19*(1), 231-240.

Welsh, A. H., Knight, E. J. (2015). "Magnitude-based inference": a statistical review. *Medicine and science in sports and exercise, 47*(4), 874.

Wilkinson, T. J., Xenophontos, S., Gould, D. W., Vogt, B. P., Viana, J. L., Smith, A. C., Watson, E. L. (2019). Test–retest reliability, validation, and "minimal detectable change" scores for frequently reported tests of objective physical function in patients with non-dialysis chronic kidney disease. *Physiotherapy theory and practice, 35*(6), 565-576.

**Sorumlu Yazar: İzzet İnce, Doç. Dr., E-mail: izzetince43@gmail.com*