Curr Res Educ (2016), 2(3) · 130-148

ARAŞTIRMA MAKALESİ

Uluslararası Dergilerde Yayımlanan Eğitim Araştırmalarının Normallik Varsayımları Açısından İncelenmesi

Ergül DEMİR* **– Özkan SAATÇİOĞLU – Fatih İMROL**

Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, Ankara, Türkiye

Öz

Bu çalışmanın amacı; istatistiksel analiz süreçleri içeren eğitim araştırmalarında normallik varsayımlarının, kullanılacak istatistiksel yöntemin belirlenmesinde ve uygulanmasında dikkate alınıp alınmadığının, bu varsayımların hangi yöntem ve tekniklerle test edildiğinin ve bu değerlendirmelere bağlı olarak kullanılacak istatistiksel yönteme yönelik ne tür kararlar verildiğinin incelenmesidir. Bu amaç doğrultusunda, seçkisiz olmayan örnekleme yöntemlerinden elverişlilik/uygunluk örneklemesine göre belirlenen uluslararası hakemli 5 dergide, 2013-2015 yılları arasında yayımlanan toplam 788 makale incelenmiştir. Temel araştırma türünde betimsel bir araştırma olarak yürütülen bu araştırmada veri toplama aracı olarak, araştırmacılar tarafından geliştirilen bir makale inceleme formu kullanılmıştır. Verilerin analizinde sıklık ve yüzde hesaplamaları ile kontenjans dağılımlarına yönelik $\chi 2$ bağımsızlık testi kullanılmıştır. Elde edilen bulgulara göre incelenen makalelerin 340'ında normallik varsayımının test edilmesi gerektiği, bu 340 araştırmanın 133'ünde normallik varsayımının test edildiği, fakat bu 133 araştırmanın 40'ında normallik varsayımının hangi yöntem ve tekniklerle test edildiğine yönelik herhangi bir açıklama yapılmadığı görülmüştür. Normallik varsayımının test edilmesinde sırasıyla betimsel yöntemler, grafiksel yöntemler ve hipotez testleri en sık kullanılan yöntemlerdir. Hipotez testleri içerisinde en sık kullanılanlar sırasıyla Kolmogorov-Smirnov ve Shapiro-Wilk testleridir. Normallik varsayımının test edilmesine bağlı olarak verilen kararlar içerisinde normallik varsayımının sağlanması durumunda parametrik yöntem ve tekniklerin kullanılması en sık gözlenen karar olmuştur. Diğer taraftan normallik varsayımı sağlanmamasına rağmen parametrik yöntem kullanılan araştırmalar da belirlenmiştir. Bu çalışma sonucunda normallik varsayımının test edilmesi gereken eğitim araştırmalarında, normallik varsayımlarının test edilmesi ve buna bağlı olarak kullanılacak istatistiksel yöntem ve tekniklere karar verilmesi açısından önemli eksiklikler ve hatalar bulunabildiği görülmüştür.

Anahtar Kelimeler: Normallik, Normal Dağılım, Varsayım, Parametrik Test, Eğitim Araştırmaları.

Makale Bilgileri:
Gönderim / Received: 10.10.2016
Kabul / Accepted: 17.10.2016

**Sorumlu Yazar / Corresponding Author:*
Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara, Türkiye
erguldemir@gmail.com

Examination of Educational Researches Published in International Journals In Terms of Normality Assumptions

Abstract

The aim of this study was to examine whether or not the normality assumptions were taken into consideration in defining which statistical methods to use in educational research articles that include statistical analysing procedures; and whether or not these assumptions were taken into account also in the application process; which methods and techniques were used towards the testing of these normality assumptions based on the evaluations on the assumptions; and finally the investigation of what kind of decisions were given in the selection of the appropriate statistical technique. In order to achieve this aim, 788 articles published from 2013 to 2015 in 5 peer-reviewed international journals which are selected according to convenience sampling, one of the non-random sampling methods, are examined. This study was carried out as a descriptive research. An article review form, developed by the researchers, was used as a data collection tool. Frequency and percentage calculations and $\chi 2$ independence test for contingency distributions were used for the analysis of the data. According to findings, it was revealed that in 340 articles normality assumption had to be tested and in 133 out of 340 normality assumption were tested and in 40 out of 133 researchers did not report how normality assumption was tested. Descriptive, graphical and hypothesis tests respectively were the most common methods in testing normality assumption. The most common hypothesis tests were Kolmogorov-Smirnov and Shapiro-Wilk, respectively. Most common decision based on normality assumption testing was to use parametric tests on the assurance of normality assumption when normality assumption was a pre-existing condition. On the other hand, there are studies that didn't meet the normality assumption yet still used parametric tests. As a result of this study, there were serious deficiencies and errors both in the testing of normality assumptions and decision-making related to the statistical methods and techniques to be used in the educational research that the normality assumption needs to be tested.

Keywords: Normality, Normal Distribution, Assumption, Parametric Test, Educational Research.

1. Giriş

Kanıta dayalı bilgi üretimi, bilimsel bilgiyi diğer bilgi kaynak ve türlerinden ayıran temel nitelik olarak karşımıza çıkmaktadır. Kanıt üretme, belli bir amaç ve hipotezler doğrultusunda sistematik gözlemler yapma, bu gözlemlerin sonuçlarını verilere dönüştürme ve bu verileri uygun istatistiksel yöntemlerle analiz ederek analiz sonuçlarını değerlendirme süreçlerini içermektedir. Bu doğrultuda istatistik, bilimsel bilgi üretiminde araştırmacıların en önemli araçlarından birisi olarak karşımıza çıkmaktadır (Uzgören, 2012; Pagano, 2009; Bakeman ve Robinson, 2005).

Veri analizinde kullanılacak istatistiksel hipotez testleri; verinin yapısına ve ölçek türüne, dağılımın biçimine ve test edilecek hipoteze göre 'parametrik testler' ve 'parametrik olmayan testler' olmak üzere sınıflandırılmaktadır. Parametrik testler; belirli bir hipotetik evren dağılımına ve bu dağılımı belirleyen ortalama, standart sapma gibi parametrelere dayalı algoritmalar içeren esnek olmayan yöntem ve tekniklerin ortak adıdır. Parametrik olmayan testler ise genellikle bir evren dağılımından bağımsız olarak, beklenen-gözlenen değer farkları, sıra ve sıra farkları gibi daha temel ve betimsel kestirimlere dayalı algoritmalar içeren esnek yöntem ve tekniklerdir. Parametrik testlerin istatistiksel gücü ve etki büyüklüğü, parametrik olmayan testlere göre daha yüksektir. Birinci ve ikinci tip hatalara karşı daha dirençlidir. Diğer taraftan parametrik olmayan testler, veri yapısı ve test edilecek hipotezin gereği olarak kullanılabildiği gibi normallik, bağımsızlık, homojenlik, küresellik, doğrusallık, çoklu bağlantılılık gibi parametrik test varsayımlarının gerekmediği ya da sağlanmadığı durumlarda da kullanılabilmektedir (Buskirk, Willoughby ve Tomazic, 2013; McKillup, 2012; Howitt ve Cramer, 2011; Ravid, 2011; Field, 2009; Stevens, 2009; Krzanowski, 2007; Foster, Barkus ve Yavorsky, 2006; Leech, Barrett ve Morgan, 2005; Sharma, 1996).

Parametrik testler, belli bir dağılım ve parametreler esas alınarak geliştirildiği için bu sınırlılıklara bağlı olarak tanımlanmış temel varsayımlar içermekte ve ancak bu varsayımlar sağlandığında

kullanılabilmektedir. Bir veya daha fazla varsayımın ihlal edilmesi durumunda test sonuçlarının hatalı ve yanıltıcı olacağı açıktır. Bu nedenle analizler öncesinde varsayımların incelenmesi önemlidir. Varsayımların sağlanıyor olması, aynı zamanda analiz sonuçlarının geçerliliğinin de bir kanıtıdır (Field, 2009; Stevens, 2009; Miles ve Banyard, 2007; Thode, 2002). Kennedy ve Bush (1985), bir veya daha fazla varsayımın ihlali durumunda analiz sonuçlarının güvenirlik ve geçerliğinin manidar düzeyde düştüğünü, benzer şekilde Bradley (1982), dağılımın normalden saptığı durumlarda istatistiksel çıkarımların hatalı olduğunu belirlemiştir.

Davranış bilimlerinde de yaygın bir şekilde kullanılan parametrik testlerin tamamına yakını, 'normal olasılık dağılımı' olarak bilinen hipotetik evren dağılımını ya da bu dağılımdan türetilmiş bir olasılık dağılımını temel almaktadır (Abbott, 2011; Stevens, 2009). Parametrik testler arasında yaygın bir şekilde kullanılan ANOVA, t testi, regresyon analizi gibi tek değişkenli; MANOVA, faktör analizi, diskriminant analizi gibi çok değişkenli teknikler, normallik varsayımlarına dayalıdır (Howitt ve Cramer, 2011; Field, 2009; Thode, 2002). Bu nedenle bu testlerin kullanılabilmesi için normallik varsayımının sağlanması bir önkoşuldur (Mertler ve Vannatta, 2005). Normallik varsayımının ihlali, analizin istatistiksel gücünü ve analiz sonuçlarının geçerliğini düşürmektedir (Wells ve Hintze, 2007).

"Gauss" veya "Gauss-Laplace" dağılımı olarak da bilinen normal dağılım, ortalama ve standart sapma ile tanımlı sürekli bir olasılık dağılımıdır (Johnson ve Bhattacharyya, 2010; Krzanowski, 2007). Normal dağılımın olasılık-yoğunluk fonksiyonu aşağıdaki gibidir:

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

Yukarıdaki formülden de görüleceği üzere sürekli bir olasılık-yoğunluk fonksiyonu olan normal dağılım iki parametre ile tanımlıdır. Bunlar evren ortalaması (μ) ve evren standart sapmasıdır (σ). Bu tanımlamaya göre normal dağılım; (-∞, +∞) aralığında X eksenine asimptotiktir, simetriktir, verilerin ortalama, mod ve medyanları simetri ekseni üzerinde çakışıktır ve eğri altındaki toplam alanın olasılığı 1'dir (Privitera, 2015; Abbott, 2011; Cohen ve Swerdlik, 2010; Lind, Marchal ve Wathen, 2006; Leech, vd. 2005; Kenny, 1987).

'Normal olasılık dağılımı' ya da kısaca 'normal dağılımın' yanı sıra sürekli değişkenlere yönelik olarak tanımlanmış, 'student t dağılımı', 'ki-kare dağılımı', 'z dağılımı', 'F dağılımı' gibi başkaca olasılık dağılımları da bulunmaktadır. Normal dağılımdan türetilmiş olan bu dağılımlar, ortalama ve standart sapmanın yanı sıra serbestlik dereceleri ile tanımlıdır ve genel olarak 'normal olasılık dağılımları ailesini' oluşturmaktadır (Field, 2009; Ross, 2009; Roussas, 2007; Lind, vd. 2006; Quinn ve Keough, 2002). Diğer taraftan süreksiz değişkenlere yönelik olarak tanımlanmış ve kullanılacak tekniği belirlemekten çok, değişkenin yapısının anlaşılmasını kolaylaştıran 'Bernoulli dağılımı', 'Binom dağılımı', 'Poisson dağılımı', 'Geometrik ve Hipergeometrik dağılımlar' gibi olasılık dağılımları da bulunmaktadır (McKillup, 2012; Krzanowski, 2007; Quinn ve Keough, 2002; Kenny, 1987; Mardia, 1972).

Birçok sürekli değişken ve davranış ölçümleri, normal dağılım ailesinde yer alan hipotetik evren dağılımlarından birine uygunluk göstermektedir. Bu nedenle günlük yaşamdaki veya doğadaki ölçülebilir pek çok değişkenin normal dağılım özelliği gösterdiği veya normal dağılım özelliğine yakın görünüm sergilediği kabul edilmektedir (Privitera, 2015; Özdamar, 2011; Ravid, 2011; Johnson ve Bhattacharyya, 2010; Johnson ve Wichern, 2007; Bakeman ve Robinson, 2005; McClave, Benson ve Sincich, 2001). Yapılan araştırmalarda öğrenci seçme sınavı puanlarının, yetişkin kadın veya erkeklerin boy uzunlukları veya ağırlıklarının, geniş bir sınıfta öğrencilerin başarı puanlarının, eğitimde zekâ testlerinden elde edilen

puanların, büyük topluluklarda yıllık hane halkı gelirlerinin normal veya normale yakın dağıldıkları belirlenmiştir (McKillup, 2012; Pagano, 2009; Kirk, 2008; Roussas, 2007; Upton ve Cook, 1996).

Normal dağılım ve normal dağılım ailesinde yer alan diğer dağılımlar hipotetik evren dağılımları oldukları için gözlem sayısı, bu dağılımların belirlenebilmesinde önemlidir. 'Merkezi limit teoremine' göre normal dağılıma sahip olan bir evrenden gelen ve bağımsız gözlemlerden oluşan yansız örneklemlerin her biri, örneklem büyüklüğü 30 ve üzerinde olmak koşuluyla normal dağılım gösterir. Ayrıca evrenden bağımsız olarak, örneklem büyüdükçe dağılımın şekli normal dağılıma yaklaşır ve böylece parametrelere yönelik olarak yapılacak çıkarımların geçerlik ve güvenirlik düzeyleri artar (Privitera, 2015; Wilcox, 2012b; Ravid, 2011; Ross, 2009; Roussas, 2007; Lind, vd. 2006; Dekking, Kraaikamp, Lopuhaä ve Meester, 2005).

Genel olarak normal dağılım, bağımlı değişkenin sayısına göre 'tek değişkenli normal dağılım' ve 'çok değişkenli normal dağılım' olmak üzere sınıflandırılabilmektedir. Bağımsız değişkenin ya da değişkenlerin süreksiz yani kategorik olması durumunda ise bağımlı değişkenin, bağımsız değişkenin her bir kategorisinde ya da bağımsız değişken kategorilerinin her bir kombinasyonunda normal dağılım göstermesi gerekmektedir. Bu durum 'tek yönlü normal dağılım've 'çok yönlü ya da çoklu normal dağılım' ayrımını da ortaya çıkarmaktadır. Bağımlı bir değişkene ait ölçümlerin gruplar içi faktörünün her bir düzeyi için ya da her bir alt grup için normal dağılımın sağlanması anlamına gelen "çoklu normallik" varsayımının incelenmesinde de tek değişkenli normallik için uygulanan yöntem ve teknikler kullanılabilmektedir (Büyüköztürk, 2013; Tabachnick ve Fidell, 2013; Stevens, 2009).

Tek değişkenli normallik varsayımı, örneklemdeki her bir sürekli değişkene ilişkin ölçümlerin normal dağılımdan geldiğini ifade etmektedir (Stevens, 2009; Hair, Anderson, Tatham ve Black, 1998). Sürekli tek bir değişkenden elde edilen puanların normal dağılıp dağılmadığı betimsel, grafiksel ve istatistiksel yöntemlerle incelenebilmektedir (Ghasemi ve Zahediasl, 2012; Abbott, 2011; Thode, 2002; Sharma, 1996).

Grafiksel yöntemlerde, puanların bir doğru üzerindeki konumu ya da çan eğrisi şeklinde dağılım gösterip göstermediği belirlenerek puanların dağılımı görsel olarak sunulmaktadır (Field, 2009; Quinn ve Keough, 2002). Bu amaçla normal dağılım eğrisinin de çizdirildiği histogram, gövde-yaprak (stem-and-leaf) grafiği, uç ya da aykırı değerlerin gözlemlendiği kutu-bıyık grafiği (box plot, box-and-whiskers), P-P (Probability-Probability) grafiği, normal Q-Q (Quantile-Quantile) grafiği veya detrended normal Q-Q grafiği sıklıkla kullanılmaktadır (McKillup, 2012; Lind, vd. 2006; Green ve Salkind, 2005; Thode, 2002). Grafiksel yöntemlerden histogram, değişkenin sürekli olduğunda çizilmekte ve puanlar normal dağılıma sahip olduğunda çan eğrisi görünümünde olmaktadır (Wilcox, 2012b; Thode, 2002). P-P ve Q-Q grafikleri ise beklenen ve gözlenen değerler arasındaki uyumu ortaya koymaktadır. Bu grafikler küçük örneklemlerde ($n<20$) yeterli bilgi verememektedir (Field, 2009; Johnson ve Wichern, 2007; Sharma, 1996). Görselliğin ön planda olduğu grafiksel yöntemler, örneklem büyüklüğünden görece bağımsız olmalarına rağmen (Çokluk, Şekercioğlu ve Büyüköztürk, 2010) büyük örneklemlerde objektif sonuçlar ortaya koyamamaktadır (McKillup, 2012). Bu nedenle grafiksel yöntemlerin diğer yöntemlerle birlikte değerlendirilmesi önerilmektedir (Stevens, 2009; Thode, 2002; Hair, vd. 1998).

Verilerin dağılımlarının, aritmetik ortalama, mod, medyan, çarpıklık ve basıklık katsayıları gibi istatistikler üzerinden incelenmesi, betimsel yöntemler olarak belirtilmektedir (Abbott, 2011; Kirk, 2008). Bu kapsamda aritmetik ortalama, mod ve medyanın eşit ya da yakın olması, çarpıklık ve basıklık katsayılarının ±1 sınırları içinde 0'a yakın olması, çarpıklık ve basıklık katsayılarının kendi standart hatalarına bölünmesi ile hesaplanan çarpıklık ve basıklık indekslerinin ±2 sınırları içinde 0'a yakın olması, standart sapma ile ortalamanın oranını yüzde olarak ifade eden bağıl değişim katsayısının 20 ile 25 aralığında olması normal dağılımın varlığına kanıt olarak değerlendirilmektedir (Tabachnick ve Fidell, 2013; McKillup, 2012; Wilcox, 2012b; Howitt ve Cramer, 2011; Lind, vd. 2006). Normallik varsayımına yönelik incelemelerde betimsel yöntemlerin de diğer yöntemlerle birlikte kullanılması ve sonuçların birlikte değerlendirilmesi önerilmektedir (McKillup, 2012; Abbott, 2011; Gnanadesikan, 1997).

Sürekli bir değişkenden elde edilen puanların normalliğini test etmek amacıyla başvurulan bir başka yöntem ise istatistiksel hipotez testlerinin kullanılmasıdır (Krzanowski, 2007). Bu kapsamda farklı yaklaşımlara göre geliştirilmiş olan normallik testlerinden yararlanılabilmektedir. Çarpıklık indeksinin istatistiksel olarak manidarlığının test edildiği 'Z testi' bu amaçla kullanılabilecek testlerden biridir. Bu testte "puanların dağılımının normal dağılımdan manidar bir şekilde farklılaşmadığı" şeklinde kurulan yokluk hipotezi test edilmektedir. Test sonucunda hesaplanan p değerinin α=.05 için 1.96 ve α=.01 için 2.58'den küçük çıkması, puanların normal dağılımdan geldiğinin kanıtı olmaktadır (Howitt ve Cramer, 2011; Field, 2009; Hair, vd. 1998). Bununla birlikte çarpıklık indeksi örneklem büyüklüğünden etkilenmekte ve büyük örneklemlerde (200 ile 400 arası) çarpıklık katsayısının standart hatası küçük çıkarken çarpıklık indeksi büyümekte, küçük örneklemlerde (30 ile 50 arası) ise çarpıklık katsayısının standart hatası büyük çıkarken çarpıklık indeksi küçülmektedir (Abbott, 2011). Anderson-Darling Testi, Shapiro-Wilk W Testi, Kolmogorov-Smirnov Testi, Pearson Ki-Kare Uygunluk Testi, Cramer-Von Mises Testi, Shapiro-Francia Testi, Lilliefors Testi, Martinez-Iglewicz Testi, Jarque-Bera Testi, D'Agostino Çarpıklık Testi, D'Agostino Basıklık Testi, D'Agostino Omnibus Testi gibi (Thode, 2002; Gnanadesikan, 1997) çok sayıdaki normallik testlerinden ise Shapiro-Wilks W Testi ve Kolmogorov-Smirnov Testi uygulamalarda en sık kullanılan testler olarak karşımıza çıkmaktadır (Hair, vd. 1998).

Örneklem büyüklüğünün 35'den büyük olması durumunda Kolmogorov-Smirnov (K-S) testi (McKillup, 2012), küçük olması durumunda ise Shapiro-Wilk testi (Shapiro ve Wilk, 1965) kullanılabilmektedir. Ancak örneklem büyüklüğünün 20 ile 50 arasında olduğu durumlarda 'Lilliefors düzeltmeli Kolmogorov-Smirnov testinin' de yeterli güce sahip olmadığı belirtilmektedir (Schoder, Himmelmann ve Wilhelm, 2006). Bu testlerde "puanların dağılımının normal dağılımdan manidar bir şekilde farklılık göstermediği" şeklinde kurulan yokluk hipotezi test edilmektedir. Test sonucunda hesaplanan p değerinin α=.05'den büyük çıkması, bu manidarlık düzeyinde puanların normal dağılımdan geldiğinin kanıtı olarak değerlendirilmektedir (Mertler ve Vannatta, 2005). Shapiro, Wilk ve Chen (1968) tarafından yapılan bir simülasyon çalışmasında Shapiro-Wilk testinin normallik varsayımını değerlendirmek için kullanılmakta olan en güçlü test olduğu ortaya konulmuştur. Kolmogorov-Smirnov testinin ise örneklem büyüklüğü arttıkça gözlenen ve beklenen dağılımlar arasındaki küçük farkların manidar çıkma eğiliminde olması nedeni ile örneklem büyüklüğünden etkilendiği ve grafiksel veya betimsel yöntemlerle birlikte kullanılmasının gerektiği belirtilmektedir (Çokluk, vd. 2010; Sprent ve Smeeton, 2007; Quinn ve Keough, 2002; Hair, vd. 1998).

Bağımlı değişkenlerin sürekli olduğu tek değişkenli ve çok değişkenli parametrik testlerin her ikisi de tek değişkenli normallik varsayımına (Thode, 2002; Mardia, Kent ve Bibby, 1979), çok değişkenli parametrik testler ise ayrıca çok değişkenli normallik varsayımına dayanmaktadır (Mertler ve Vannatta, 2005; Hair, vd. 1998). Birden fazla değişkenin aynı andaki normalliği anlamına gelen çok değişkenli normal dağılım, her bir değişkenin tek değişkenli normal dağılımı sağlamasının yanı sıra değişkenlerin doğrusal kombinasyonlarının ve her türlü ikili kombinasyonlarının da her bir grup ve alt grup için normal olduğunu varsaymaktadır. Çok değişkenli normallik sağlandığında tek değişkenli normalliğin de sağlandığı ifade edilmesine rağmen tek değişkenli normalliğin sağlanması çok değişkenli normalliğin olduğuna kesin kanıt sağlamamaktadır (Tabachnick ve Fidell, 2013; Stevens, 2009; Johnson ve Wichern, 2007; Mertler ve Vannatta, 2005; Thode, 2002; Hair, vd. 1998; Sharma, 1996).

Tek değişkenli normal dağılım varsayımında olduğu gibi çok değişkenli normalliğin incelenmesi için de betimsel, grafiksel ve istatistiksel yöntemler bulunmaktadır (Gnanadesikan, 1997). Çok değişkenli normalliğin incelenmesi için betimsel yöntemler olarak çok değişkenli çarpıklık ve basıklık katsayıları kullanılabilmektedir (Thode, 2002; Mardia, vd. 1979; Mardia, 1970). Grafiksel yöntem olarak çoklu saçılma grafiğinin (Johnson ve Wichern, 2007) yanı sıra marjinal grafikler, Kowalski'nin çizgi grafikleri ve çok değişkenli Q-Q grafikleri de kullanılmaktadır (Thode, 2002). İki değişkenli normal dağılım için bu değişkenlerin saçılma grafiklerinin elips veya elipse benzer şekilde olduğunda çok değişkenli normallik

varsayımının sağlandığı ifade edilmektedir (Stevens, 2009; Gnanadesikan, 1997; Mardia, vd. 1979). Üç değişkenli normal dağılım için ise uygun programlar yardımıyla üç boyutlu saçılma grafikleri çizilebilmesine rağmen dört ve üzeri değişken için çok boyutlu uzayda grafik çizmenin son derece zor olduğu belirtilmektedir (McKillup, 2012; Field, 2009; Johnson ve Wichern, 2007; Thode, 2002). Ayrıca tek değişkenli uyum iyiliği testlerinden Cramer-von Mises testinden türetilmiş Koziol's testi (1982), Shapiro-Wilk testinden türetilmiş The Mudholkar-Srivastava-Lin testi (1995), Anderson-Darling testinden türetilmiş The Paulson-Roohan-Sullo testi (1987), The Romeu-Ozturk testi (1993), ki-kare testi ve Kolmogorov-Smirnov testinden türetilmiş bazı özel testler (Gnanadesikan, 1997) kullanılabilmektedir.

Çok değişkenli normal dağılımı, gerçek hayattaki sürekli değişkenler için tam anlamıyla modellemek güçtür (Abbott, 2011). Bu nedenle çok değişkenli analizlerde öncelikle tek değişkenli ve çok değişkenli uç değer incelemelerinin yapılması, sonrasında her bir değişken düzeyinde aynı 'veri dönüşümleri (data transformation)' ile dağılımların normalleştirilmesi önerilmektedir. Verinin farklı birime dönüştürülmesi ve yeniden ölçeklenmesi olarak tanımlanan veri dönüştürme sürecinde, kare ya da karekök, ters çevirme, logaritmik veya arcsin, probit, logit dönüşüm işlemleri uygulanabilmektedir (Tabachnick ve Fidell, 2013; Stevens, 2009; Johnson ve Wichern, 2007; Thode, 2002; Sharma, 1996; Kenny, 1987). Veriden uç değerler çıkarıldığında veya veri yeni bir dağılıma dönüştürüldüğünde puan dağılımlarının normal dağılımdan aşırı sapma göstermesi durumunda ise bu varsayımı gerektirmeyen parametrik olmayan testlerin kullanılması tercih edilmelidir (McKillup, 2012; Howitt ve Cramer, 2011; Field, 2009; Krzanowski, 2007).

Araştırmalarda sıklıkla kullanılan parametrik testler, temelde normallik varsayımlarına dayalı olmakla birlikte araştırmacıların bu varsayımları sıklıkla göz ardı ettiği ifade edilmektedir (Wilcox, 2012a; Micceri, 1989). Gerek Türkiye dışında gerekse Türkiye'de normallik varsayımlarına yönelik olarak yürütülen araştırmalar sınırlı sayıda olmasına rağmen, bu konu ile dolaylı olarak ilişkilendirilebilecek araştırmalara rastlamak mümkündür (Doğru, Gençosman, Ataalkın ve Şeker, 2012; Ruscio ve Roche, 2012; Tavşancıl vd. 2010). Yapılan araştırmalar incelendiğinde çoğu araştırmada yöntem bölümünün açık bir şekilde raporlaştırılmadığı (Kashy, Donnellan, Ackerman ve Russell, 2009) veya bazı araştırmalarda ise sonuçların yanlış raporlaştırıldığı (Bakker ve Wicherts, 2011) bulunmuştur. Araştırmalarda parametrik testler kullanılmasına rağmen çok az sayıdaki çalışmada tek değişkenli ve çoklu normallik varsayımının kontrol edildiği sonucuna ulaşılmıştır (Ergun ve Çilingir, 2013; Delice, 2010; Tavşancıl vd. 2010; Keselman vd. 1998). Çoklu normallik ve çok değişkenli normallik varsayımlarının ise genellikle kontrol edilmediği veya bu varsayımların göz ardı edildiği ifade edilmiştir (Evrekli, İnel, Deniş ve Balım, 2011; Keselman vd. 1998). Diğer taraftan normallik varsayımının kontrol edildiği parametrik testlerde çoğunlukla istatistiksel hipotez testlerinden Kolmogorov-Smirnov testinin tercih edildiği belirtilmiştir (Gömleksiz ve Bozpolat, 2015; Kabaca ve Erdoğan, 2007).

Gerek Türkiye dışında gerekse Türkiye'de yapılan araştırmalar, veri analizlerinde parametrik testlerin kullanılması durumunda genel olarak analiz süreçlerinin raporlaştırılması, normallik varsayımı gibi temel varsayımların test edilmesi, normallik varsayımının test edilmesinde kullanılacak yöntem ve tekniğin belirlenmesi ve buna bağlı olarak uygun yöntemlere karar verilmesi açılarından önemli eksikliklerin ve hataların bulunduğunu göstermektedir. Bu tür eksiklikler içeren araştırmalarda, elde edilen sonuçların sağlam ve isabetli kararlara esas oluşturması beklenemez. Eğitim alanındaki pek çok soruna bilimsel çözümler üretme iddiasına sahip eğitim araştırmalarında da normallik varsayımının dikkate alınma düzeyi, çözümlerin sağlamlığının değerlendirilmesinde kullanılabilecek ölçütlerden birisidir. Bu bağlamda çalışmanın amacı; istatistiksel analiz süreçleri içeren eğitim araştırmalarında normallik varsayımlarının, kullanılacak istatistiksel yöntemin belirlenmesinde ve uygulanmasında dikkate alınıp alınmadığının, bu varsayımların hangi yöntem ve tekniklerle test edildiğinin ve bu değerlendirmelere bağlı olarak kullanılacak istatistiksel yönteme yönelik ne tür kararlar verildiğinin incelenmesidir. Bu amaç doğrultusunda, belirli ölçütleri karşılayan uluslararası hakemli 5 dergide, 2013-2015 yılları arasında yayımlanan makaleler incelenmiş ve aşağıdaki sorulara yanıt aranmıştır:

1. İncelenen dergilerde, normallik varsayımlarının test edilmesi gereken istatistiksel teknikleri içeren araştırmaların sıklığı ve oranı nedir? Bu sıklık ve oranlar, (a) dergilere ve (b) yıllara göre farklılık göstermekte midir?
2. Bu araştırmalar içerisinde normallik varsayımlarının test edildiği araştırmaların sıklığı ve oranı nedir? Bu sıklık ve oranlar, (a) dergilere ve (b) yıllara göre farklılık göstermekte midir?
3. Normallik varsayımlarının test edildiği araştırmalarda, bu varsayımları test etmede kullanılan yöntem ve teknikler nelerdir? Bu yöntem ve teknikler, (a) dergilere ve (b) yıllara göre farklılık göstermekte midir?
4. Normallik varsayımlarının test edildiği araştırmalarda, normalliğe yönelik ne tür bulgular elde edilmiştir ve bu bulgular doğrultusunda ne tür kararlar verilmiştir? Bu kararlar, (a) dergilere ve (b) yıllara göre farklılık göstermekte midir?

2. Yöntem

2.1. Araştırmanın Modeli

Bu çalışma, temel araştırma türünde bir betimsel araştırmadır. Temel araştırmalar, uygulamalı araştırmalardan farklı olarak saf kuramsal bilgi üretmeye yönelik olan ve mevcut kuramsal bilgilere yenilerini katmayı amaçlayan araştırmalardır. Temel araştırmalar içerisinde yer alan betimsel araştırmalar ise kuramsal olarak tanımlanmış bir durumu, belli özellikler ve değişkenler düzeyinde gözlem ve ölçme sonuçları ile destekleme ve olduğu gibi ortaya koymaya yönelik araştırmalardır (Karasar, 2015).

2.2. Örneklem

Bu çalışmanın örneklemi, seçkisiz olmayan örnekleme yöntemlerinden elverişlilik/uygunluk örneklemesine göre belirlenen beş dergide, 2013-2015 yılları arasında yayımlanan toplam 788 araştırma makalesinden oluşmaktadır. Elverişlilik/Uygunluk örneklemesi (convenience sampling); araştırma amaçları doğrultusunda elverişlilik açısından uygun ve gönüllü küme, grup ya da bireylerin seçilmesi ve araştırmaya dâhil edilmesi olarak açıklanmaktadır (Onwuegbuzie ve Collins, 2007).

Örneklemin, yani incelenecek makalelerin belirlenmesinde, kullanılan örnekleme yöntemi doğrultusunda öncelikle, tanımlanmış ölçütleri karşılayan bilimsel dergilerden ulaşılabilenler belirlenmiştir. Dergilerin belirlenmesinde; (1) eğitim alanında yayın yapıyor olmak, (2) uluslararası hakemli dergi olmak, (3) AEI, SSCI, ERIC gibi uluslararası atıf indekslerinden en az birinde listelenmek, (4) son beş yıldır uluslararası atıf indekslerinden en az birinde yer alıyor olmak, (5) elektronik ortamda açık erişim sağlıyor olmak ölçütleri dikkate alınmıştır. Bu ölçütleri karşıladığı görülen iki yüzün üzerinde dergi arasından dört tanesi, seçkisiz olarak belirlenmiş ve araştırmaya dâhil edilmiştir. Türkiye'de yayın yapmakta olan ve belirlenen ölçütleri karşıladığı görülen dergiler içerisinden ise bir tanesi yine seçkisiz olarak belirlenmiş ve araştırmaya dâhil edilmiştir. Örneklemi oluşturan makalelerin yer aldığı dergilerin bazı özellikleri ve yıllara göre makale sayılarının dağılımları Tablo 1'de verilmiştir.

Tablo 1'de görüldüğü gibi örneklemi oluşturan dergiler farklı ülkelerde, yılda en az dört sayı olarak yayımlanmaktadır. Ulusal ve uluslararası tanınmış ve yaygın indekslerde yer almaktadır. Dergiler yüksek, orta, düşük olarak tanımlanabilecek farklı etki büyüklüklerine sahiptir. Bu dergilerde 2013-2015 yılları arasında yayımlanmış olan ve bu çalışmada incelenen toplam makale sayısı 788'dir.

Tablo 1. *Örneklemde Yer Alan Makalelerin Yayımlandığı Dergilere, Dergilerin Bazı Özelliklerine ve Yıllara Göre Dağılımları*

Dergi Adı	Ülke	Etki Değeri	İndeksleri	Yıl			Toplam
				2013	2014	2015	
American Educational Research Journal	ABD	2.157	EBSCO, EBSCOhost, Higher Education Abstracts, ProQuest Education Journals, PsycINFO, SafetyLit, Scopus, Social SciSearch, SSCI, Wilson Education Index/Abstracts	43	36	36	115
British Educational Research Journal	Birleşik Krallık	0.891	British Education Index, EBSCO, ERA, ERIC, ERIH, FRANCIS, IBR, IBSS, IBZ, ISI Current Contents Social and Behavioral Sciences, PsycINFO, SCOPUS, SSCI, VOCEDplus.	51	53	58	162
European Physical Education Review	İngiltere	0.673	Academic Search Premier, British Education Index, CSA, CHID, ERIC, ERA, PEI, Political Science Index, PsycINFO, SciVal, Scopus, Social SciSearch, SSCI, Worldwide Political Science Abstracts	23	29	32	84
Research in Higher Education	Hollanda	1.160	SSCI, SCOPUS, PsycINFO, EBSCO, CSA Environmental Sciences, ERA, ERIC, ERIH PLUS, IBR, IBZ, JSTOR, MathEDUC, PSYCLINE, Research into Higher Education Abstracts, Summon by ProQuest	37	34	36	107
TED Eğitim ve Bilim	Türkiye	0.342	SSCI, Scopus, EBSCO Education Source, TÜBİTAK ULAKBİM Sosyal Bilimler Veritabanı	115	118	87	320
Toplam				269	270	249	788

2.3. Veri Toplama Aracı

Bu çalışmada veri toplama aracı olarak, araştırmacılar tarafından geliştirilmiş olan bir makale inceleme formu kullanılmıştır. Makale inceleme formu, araştırmanın amacı doğrultusunda ve araştırma sorularına yanıt oluşturacak şekilde 9 soru içermektedir. İnceleme formunun geliştirilmesinde, öncelikle araştırmacılar bir taslak form oluşturmuş ve bu formu düzenleyerek uzman görüşüne sunulmaya hazır hale getirmiştir. Öğretim üyeleri ve araştırma görevlilerinden oluşan 11 uzmanın görüş ve önerileri, bu amaçla hazırlanan bir 'Uzman Görüş Formu' ile alınmıştır. Uzmanlardan yazılı olarak alınan görüşler doğrultusunda gözden geçirme ve düzenleme çalışmaları yapılarak 'Makale İnceleme Formu'na son şekli verilmiştir.

Makale inceleme formunun kullanılabilir ve yeterli olduğuna yönelik kanıt elde edebilmek için her bir dergiden 1 tane olmak üzere rastgele belirlenen 5 makale, bu form kullanılarak üç araştırmacı tarafından değerlendirilmiş ve değerlendirmeler karşılaştırılmıştır. Üç araştırmacının da benzer değerlendirmeler yaptığı gözlenmiştir. Ayrıca araştırmacılar dışında doktora yeterliği bulunan 4 uzmandan bu 5 makaleyi, makale inceleme formunu kullanarak değerlendirmeleri istenmiş, bu değerlendirmeler de hem kendi arasında hem araştırmacıların değerlendirme sonuçları ile karşılaştırılmış ve benzer değerlendirmelerin yapıldığı gözlenmiştir.

2.4. Verilerin Toplanması ve Analizi

Bu çalışmada veriler, araştırmacılar tarafından geliştirilen makale inceleme formu kullanılarak örneklemi oluşturan 788 makalenin ayrı ayrı değerlendirilmesi ile elde edilmiştir. Elde edilen veriler, sayma ve sınıflamaya dayalı betimsel yöntemlerle analiz edilmiş, analiz sonuçları tablolarla betimlenmiştir. Dergilere ve yıllara göre farklılıklar, kontenjans dağılımları üzerinden $\chi 2$ bağımsızlık testi ile test edilmiştir.

3. Bulgular

Bu bölümde araştırma soruları doğrultusunda yürütülen analizlerden elde edilen bulgular ve bu bulgulara yönelik genel yorumlar, alt başlıklar halinde verilmiştir.

3.1. Normallik Varsayımlarının Kontrol Edilmesi Gereken Araştırmaların Dağılımı

İncelenen araştırmalarda normallik varsayımlarının kontrol edilmesi gerektiği kararının verilmesinde, araştırmanın metodolojisi ve araştırmada kullanılan istatistiksel teknikler dikkate alınmıştır.

İncelenen 788 araştırmadan yarıdan fazlasını oluşturan 462'sinde (%58,6) araştırma tür ve modeli açıkça belirtilmiştir. Diğer taraftan araştırmaların 216'sında (%27,4) araştırma tür ve modeli kısmen ya da dolaylı olarak belirtilmiş, 110'unda (%14) ise araştırma tür ve modelinin hiçbiri belirtilmemiştir. Araştırma tür ve modeli açıkça belirtilmemiş olan çalışmalarda bu özellikler, araştırmacılar tarafından araştırmanın tamamı dikkate alınarak değerlendirilmiş ve belirlenmiştir. İncelenen araştırmaların metodoloji açısından yaklaşımları ve türlerine göre dağılımları Tablo 2'de verilmiştir.

Tablo 2. *Araştırmaların Metodolojik Yaklaşımlarına ve Türlerine Göre Dağılımı*

Araştırma Yaklaşımı ve Türü	F	%	Kümülatif %
Kuramsal	34	4,3	4,3
Nitel	267	33,9	38,2
Nicel (betimleyici, tarama)	98	12,4	50,6
Nicel (ilişkisel, yordayıcı)	257	32,6	83,2
Nicel (deneysel)	56	7,1	90,4
Nicel (ölçek geliştirme)	16	2,0	92,4
Karma Model	50	6,3	98,7
Diğer*	10	1,3	100,0
Toplam	788	100,0	

*Ölçek geliştirme ve tarama modeli, ölçek geliştirme ve deneysel model, ölçek geliştirme ve ilişkisel model gibi farklı araştırma modellerin bir arada kullanıldığı araştırmalar.

Tablo 2'de görüldüğü gibi incelenen araştırmaların %38,2'si, nitel yaklaşıma dayalı veya kuramsal araştırmalardır. Bu araştırmalarda normallik varsayımlarını içeren herhangi bir istatistiksel teknik kullanılmamıştır. Bu araştırmalar dışında en sık gözlemlenen araştırmalar, nicel yaklaşıma dayalı ilişkisel ya da yordayıcı araştırmalardır (%32,6). Nicel yaklaşıma dayalı betimsel ya da tarama modelindeki araştırmalar (%12,4) bunu takip etmektedir. Bunların dışındaki deneysel araştırma tür ve modellerinin ise daha az sayıda olduğu görülmektedir.

Nitel yaklaşıma dayalı ve kuramsal araştırmalar dışında kalan 487 (%61,8) araştırmada, normallik varsayımlarının test edilmesi gereken istatistiksel tekniklerin kullanılıp kullanılmadığı incelenmiştir. Bu 487 araştırmadan 367'sinde (%75,4) kullanılan istatistiksel yöntem ve istatistiksel işlem süreci açıkça belirtilmişken, 91'inde (%18,7) kısmen ya da eksik şekilde belirtilmiştir. 29 araştırmada (%6) ise kullanılan istatistiksel yöntem ve istatistiksel işlem süreci hakkında herhangi bir bilgi verilmemiştir. Araştırmada kullanılan istatistiksel teknik ve işlem süreci açıkça belirtilmemiş çalışmalarda bu özellikler, araştırmacılar tarafından araştırmanın tamamı dikkate alınarak değerlendirilmiş ve belirlenmiştir.

Bu 487 araştırmadan 145'inde (%29,8) sadece bir istatistiksel teknik kullanılmıştır. Bir istatistiksel teknik kullanılan 145 araştırmanın 100'ünde parametrik olmayan ve diğer 45'inde ise parametrik bir teknik kullanılmıştır. Diğer 342 (%70,2) araştırmada birden fazla istatistiksel teknik bir arada kullanılmıştır. 208 araştırmada (%42,7) parametrik ve parametrik olmayan teknikler bir arada kullanılmıştır. 487 araştırmanın 145'inde (%29,8) sadece parametrik olmayan teknikler, 140'ında (%28,7) ise sadece parametrik teknikler

kullanılmıştır. İki araştırmada kullanılan tekniğin parametrik ya da parametrik olmadığı yönünde uzlaşma sağlanamamıştır. Bu iki araştırma ihmal edildiğinde geriye kalan 485 araştırmanın, kullanılan istatistiksel teknikler açısından test edilmesi gereken normallik varsayımlarına göre dağılımı Tablo 3'te verilmiştir.

Tablo 3. *Kullanılan İstatistiksel Tekniklerin Test Edilmesi Gereken Normallik Varsayımlarına Göre Dağılımı*

Normallik Varsayımları	f	%	Kümülatif %
Tek değişkenli (tek yönlü)	27	5,6	5,6
Tek değişkenli (çoklu/çok yönlü)	13	2,7	8,3
Çok değişkenli	30	6,2	14,5
Tek değişkenli (tek yönlü ve çok yönlü)	142	29,3	43,8
Tek değişkenli (tek yönlü) ve çok değişkenli	20	4,1	47,9
Tek değişkenli (çok yönlü) ve çok değişkenli	6	1,2	49,1
Tek değişkenli (tek yönlü ve çok yönlü) ve çok değişkenli	102	21,0	70,1
Gerek yok	145	29,9	100,0
Toplam	485	100,00	

Tablo 3'te görüldüğü gibi normallik varsayımlarının test edilmesi gerekmeyen araştırmaların yanı sıra, sadece bir normallik varsayımının test edilmesi gereken araştırmalar olduğu gibi birden fazla normallik varsayımının test edilmesi gereken araştırmalar da bulunmaktadır. Bu 485 araştırmanın büyük çoğunluğunda (%29,9) kullanılan istatistiksel teknik açısından herhangi bir normallik varsayımının test edilmesine gerek yoktur. 142 araştırmada (%29,3) ise tek değişkenli normallik varsayımının, hem tek yönlü yani sadece bir değişken düzeyinde hem de çok yönlü yani ilişkili bağımsız değişkenin kategorileri düzeyinde test edilmesi gerekmektedir. Toplam 340 araştırmada (%70,1) normallik varsayımının ya da varsayımlarının test edilmesi gereken istatistiksel teknik ya da teknikler kullanılmıştır.

Bu çalışmada incelenen 788 araştırma içerisinde kullanılan istatistiksel teknik açısından normallik varsayımlarının test edilmesi gereken 340 (%43,1) araştırma belirlenmiştir. Bu 340 araştırmanın yayımlandığı dergilere ve yayım yılına göre dağılımı Tablo 4'te verilmiştir.

Tablo 4. *Normallik Varsayımı Test Edilmesi Gereken Araştırmaların Yayımlandığı Dergilere ve Yayım Yılına Göre Dağılımı*

Dergi Adı		Yıl			Toplam
		2013	2014	2015	
American Educational Research Journal	n	27	17	19	63
	%	7,9	5,0	5,6	18,5
British Educational Research Journal	n	8	17	24	49
	%	2,4	5,0	7,1	14,4
European Physical Education Review	n	9	7	7	23
	%	2,6	2,1	2,1	6,8
Research in Higher Education	n	12	9	11	32
	%	3,5	2,6	3,2	9,4
TED Eğitim ve Bilim	n	64	66	43	173
	%	18,8	19,4	12,6	50,9
Toplam	n	120	116	104	340
	%	35,3	34,1	30,6	100,0

Tablo 4'te verilen kontenjans dağılımı istatistiksel olarak manidar değildir ($\chi 2=15,926$, sd=8 ve $p>0,01$). Buna göre normallik varsayımlarının test edilmesini gerektiren 340 araştırma makalesinin, dergilere ve

yıllara göre sayısal dağılımlarının beklenen ve gözlenen değerleri arasında manidar bir fark gözlenememiştir. Makalelerin yıllara göre dağılımları, her bir dergi düzeyinde ve toplamda benzerdir.

3.2. Normallik Varsayımlarının Test Edildiği Araştırmalar

Önceki analizler, bu çalışmada incelenen 788 araştırmadan 340'ında (%43,1), araştırmanın metodolojik yaklaşımı, modeli ve kullanılan istatistiksel teknikler açısından normallik varsayımlarının test edilmesi gerektiğini göstermiştir. Bu 340 araştırmanın, normallik varsayımlarının test edilmiş olma durumlarına göre dağılımı Tablo 5'te verilmiştir.

Tablo 5. *Normallik Varsayımlarının Test Edilmesi Gereken Araştırmaların Test Edilen Normallik Varsayımlarına Göre Dağılımları*

Test Edilen Normallik Varsayımları	f	%	Kümülatif %
Tek değişkenli (tek yönlü)	76	22,4	22,4
Tek değişkenli (çoklu/çok yönlü)	6	1,8	24,1
Çok değişkenli	18	5,3	29,4
Tek değişkenli (tek yönlü ve çok yönlü)	15	4,4	33,8
Tek değişkenli (tek yönlü) ve çok değişkenli	8	2,4	36,2
Tek değişkenli (çok yönlü) ve çok değişkenli	2	0,6	36,8
Tek değişkenli (tek yönlü ve çok yönlü) ve çok değişkenli	8	2,4	39,1
Hiçbiri	207	60,9	100,0
Toplam	340	100,0	

Tablo 5'te görüldüğü gibi normallik varsayımlarının test edilmesi gereken 340 araştırmanın 207'sinde (%60,9) bu varsayımların test edildiğine yönelik bir kanıt bulunamamıştır. Diğer 133 araştırmada (%39,1) ise normallik varsayımları test edilmiştir. En fazla test edilen ve dikkate alınan normallik varsayımı (%22,4), tek değişkenli ve tek yönlü normalliktir. Çok değişkenli normalliğin test edildiği toplam 36 araştırma (%10,6) ve bütün normallik varsayımlarının test edildiği 8 araştırma (%2,4) bulunmaktadır.

Bu çalışmada incelenen toplam 788 araştırma içerisinde normallik varsayımlarının test edilmesi gerekip bu varsayımların test edildiği 133 (%16,9) araştırma belirlenebilmiştir. Bu 133 araştırmanın yayımlandığı dergilere ve yayım yıllarına göre dağılımları Tablo 6'da verilmiştir.

Tablo 6. *Normallik Varsayımı Test Edilmesi Gereken ve Bu Varsayımların Test Edildiği Araştırmaların Yayımlandığı Dergilere ve Yayım Yılına Göre Dağılımı*

Dergi Adı		Yıl			Toplam
		2013	2014	2015	
American Educational Research Journal	n	5	6	8	19
	%	3,8	4,5	6,0	14,3
British Educational Research Journal	n	1	0	6	7
	%	0,8	0,0	4,5	5,3
European Physical Education Review	n	2	2	2	6
	%	1,5	1,5	1,5	4,5
Research in Higher Education	n	3	7	5	15
	%	2,3	5,3	3,8	11,3
TED Eğitim ve Bilim	n	25	32	29	86
	%	18,8	24,1	21,8	64,7
Toplam	n	36	47	50	133
	%	27,1	35,3	37,6	100,0

Tablo 6'da verilen kontenjans dağılımı istatistiksel olarak manidar değildir ($\chi 2$=8,963, sd=8 ve p>0,01). Buna göre normallik varsayımlarının test edildiği 133 araştırma makalesinin, dergilere ve yıllara göre sayısal dağılımlarının beklenen ve gözlenen değerleri arasında manidar bir fark gözlenememiştir. Makalelerin yıllara göre dağılımları, her bir dergi düzeyinde ve toplamda benzerdir.

3.3. Normallik Testlerinde Kullanılan Yöntem ve Teknikler

Önceki analizler, bu çalışmada incelenen 788 araştırmadan 133'ünde (%16,9), normallik varsayımlarının test edildiğini göstermiştir. Bu 133 araştırmanın 6'sında (%4,5) normallik testlerinde kullanılan yöntem hakkında uzlaşma sağlanamamıştır. 40'ında (%30,1) normallik varsayımının test edildiği belirtilmiş fakat hangi yöntem ve teknikle incelendiği konusunda herhangi bir bilgi verilmemiştir. 65 araştırmada (%48,9) sadece 1 yöntem, diğer 22 (%16,5) araştırmada ise birden fazla yöntem kullanılmıştır.

Evren büyüklüğü gerekçe gösterilerek normalliğin sağlandığını varsayan herhangi bir araştırma gözlenmemiştir. Örneklem büyüklüğü varsayımı ile normalliğin sağlandığını belirten 5 araştırma (%3,8) belirlenmiştir. Bunun dışında normallik varsayımının doğrudan incelenmediği fakat doğrusallık, otokorelasyon, varyansların homojenliği gibi diğer varsayımların sağlanmasına bağlı olarak normalliğin de sağlandığını belirten 18 araştırma (%13,5) gözlenmiştir.

Bu 133 araştırmanın 22'sinde (%16,5) normallik varsayımları grafiksel yöntemlerle incelenmiştir. 33 araştırmada (%24,8) ortalama, mod, medyan, basıklık ve çarpıklık katsayıları gibi betimsel istatistikler üzerinden normallik varsayımlarının incelendiği görülmüştür. Box'un M istatistiği, çok değişkenli basıklık ve çarpıklık katsayıları ve Bartlett'in çok değişkenli küresellik testi gibi çok değişkenli yöntem ve tekniklerle çok değişkenli normallik varsayımının test edildiği 15 araştırma (%11,3) bulunmaktadır.

Bu 133 araştırmanın 32'sinde (%24,1) ise Kolmogorov-Smirnov testi, Shapiro-Wilk testi gibi hipotez testleri ile normallik varsayımı test edilmiştir. Normallik varsayımının bir hipotez testi ile test edildiği 32 araştırmanın, kullanılan hipotez testine göre dağılımı Tablo 7'de verilmiştir.

Tablo 7. *Normallik Varsayımının Hipotez Testi İle Test Edildiği Araştırmaların Kullanılan Hipotez Testine Göre Dağılımları*

Hipotez Testi	f	%	Kümülatif %
Shapiro-Wilk	6	18,8	18,8
Kolmogorov-Smirnov	21	65,6	84,4
Diğer	5	15,6	100,0
Total	32	100,0	

Tablo 7'de görüldüğü gibi normallik testinde en fazla kullanılan hipotez testi (%65,6) Kolmogorov-Smirnov testidir. Shapiro-Wilk testi (%18,8) bundan sonra gelmektedir. Normallik varsayımının hipotez testi ile test edildiği araştırmaların büyük çoğunluğunda (%84,4), bu iki testten biri kullanılmıştır.

3.4. Normallik Testlerinde Elde Edilen Bulgular ve Kararlar

Önceki analizler, bu çalışmada incelenen 788 araştırmadan 133'ünde (%16,9), normallik varsayımlarının test edildiğini göstermiştir. Bu araştırmalarda, normallik varsayımına yönelik incelemelere bağlı olarak verilen kararların dağılımı Tablo 8'de verilmiştir.

Tablo 8'de görüldüğü gibi normallik varsayımlarının test edildiği 133 araştırmanın yarıdan fazlasında (%66,9), normallik varsayımının karşılandığı belirtilmiş ve buna bağlı olarak parametrik istatistiksel yöntem ve teknikler kullanılmıştır. Normallik varsayımının sağlanmadığı belirtilen 20 araştırmanın 4'ünde (%3,0) istatistiksel düzeltme yapıldıktan sonra parametrik yöntem ve teknikler kullanılmış, diğer 16 araştırmada (%12,0) ise doğrudan parametrik olmayan yöntemlerin kullanılması tercih edilmiştir. 3

araştırmada (%2,3) normallik varsayımının karşılanmadığı belirtilmesine rağmen parametrik yöntemlerin kullanıldığı görülmektedir. Normallik varsayımına yönelik incelemelerle ilişkilendirilmeyerek kullanılacak istatistiksel yöntem ve tekniklere karar verilmiş olan 15 araştırma (%11,3) bulunmaktadır. Yukarıda da belirtildiği gibi 6 araştırmada (%4,5) normallik varsayımının nasıl incelendiği açıkça belirtilmemiş ve bu kapsamda araştırmacılar tarafından yapılan değerlendirmeler sonucunda da bir uzlaşmaya varılamamıştır.

Tablo 8. *Normallik Varsayımlarının Test Edildiği Araştırmalarda Verilen Kararların Dağılımı*

Karar	f	%	Geçerli %	Kümülatif %
Normal dağılım göstermektedir ve parametrik yöntem kullanılmıştır.	89	66,9	70,1	70,1
Normal dağılım göstermemektedir ve parametrik yöntem kullanılmıştır.	3	2,3	2,4	72,4
Normal dağılım göstermemiştir ve istatistiksel düzeltme yapıldıktan sonra parametrik yöntem kullanılmıştır.	4	3,0	3,1	75,6
Normal dağılım göstermemiştir ve parametrik olmayan yöntem kullanılmıştır.	16	12,0	12,6	88,2
Diğer	15	11,3	11,8	100,0
Toplam	127	95,5	100,0	
Kayıp	6	4,5		
Genel Toplam	133	100,0		

Normallik varsayımına yönelik incelemelere bağlı olarak verilen kararlar açısından, yayımlanma yılına göre manidar fark gözlenmemiştir ($\chi2$=6,824, sd=8 ve p>0,05). Buna göre verilen kararların yıllara göre dağılımı benzerdir. Diğer taraftan verilen kararların dergilere göre kontenjans dağılımlarının ise manidar düzeyde farklılaştığı gözlenmiştir ($\chi2$=58,042, sd=16 ve p<0,01). Normallik incelemeleri ve buna bağlı olarak verilen kararlar üzerinde uzlaşılamayan 6 araştırma makalesi ihmal edildiğinde geriye kalan 127 makalede verilen kararların, dergilere göre kontenjans dağılımları Tablo 9'da verilmiştir.

Tablo 9. *Normallik Varsayımlarının Test Edildiği Araştırmalarda Verilen Kararların Dergilere Göre Dağılımı*

			Dergi Adı					
			American Educational Research Journal	British Educational Research Journal	European Psychical Education Review	Research in Higher Education	TED Eğitim ve Bilim	Toplam
Karar	Normal dağılım göstermektedir ve parametrik yöntem kullanılmıştır.	n	12	7	1	5	64	89
		%	9,4	5,5	0,8	3,9	50,4	70.1
	Normal dağılım göstermemektedir ve parametrik yöntem kullanılmıştır.	n	0	0	1	0	2	3
		%	0,0	0,0	0,8	0,0	1,6	2.4
	Normal dağılım göstermemiştir ve istatistiksel düzeltme yapıldıktan sonra parametrik yöntem kullanılmıştır.	n	0	0	1	3	0	4
		%	0,0	0,0	0,8	2,4	0,0	3.1
	Normal dağılım göstermemiştir ve parametrik olmayan yöntem kullanılmıştır.	n	1	0	0	2	13	16
		%	0,8	0,0	0,0	1,6	10,2	12.6
	Diğer	n	6	0	0	2	7	15
		%	4,7	0,0	0,0	1,6	5,5	11.8
	Toplam	n	19	7	3	12	86	127
		%	15,0	5,5	2,4	9,4	67,7	100,0

Tablo 9'da görüldüğü gibi normal dağılım gösterdiği kararına varılıp bu karara bağlı olarak parametrik yöntemlerin kullanıldığı 89 araştırmanın büyük çoğunluğu (%72) 'TED Eğitim ve Bilim Dergisinde' yayımlanmıştır. Benzer şekilde normal dağılım varsayımının sağlanmadığı kararından sonra istatistiksel düzeltme yapılarak parametrik yöntemlerin kullanıldığı 4 araştırmadan 3'ü (%75) de bu dergide yayımlanmıştır. Diğer taraftan normallik varsayımının sağlanmadığı kararına bağlı olarak doğrudan parametrik olmayan yöntemlerin kullanıldığı 16 araştırmanın 13'ü (%81,3) de bu dergide yayımlanmıştır. Bunu 6 araştırma (%40) ile 'American Educational Research Journal' takip etmektedir. Ayrıca normal dağılım varsayımını karşılamamasına rağmen parametrik yöntemlerin kullanıldığı 3 araştırmadan 2'si (%66,7) 'TED Eğitim ve Bilim Dergisinde' yayımlanmıştır. Diğer 1 araştırma ise 'European Psychical Education Review' da yayımlanmıştır.

4. Sonuç, Tartışma ve Öneriler

Bu çalışmada eğitim alanında yayın yapan ve uluslararası alan indekslerinde yer alan 5 farklı dergide 2013 ile 2015 yılları arasında yayımlanmış toplam 788 araştırma makalesi, normallik varsayımları açısından incelenmiştir. İncelenen araştırma makalelerinde, önemli raporlaştırma eksiklikleri ve yetersizlikleri bulunduğu görülmüştür. Raporlaştırmalardaki bu yetersizlikler ve eksiklikler, normallik varsayımlarına yönelik incelemeleri de güçleştirmiştir. Bu bulgu Bakker ve Wicherts (2011) ve Kashy ve diğerleri (2009) tarafından yapılan araştırma sonuçları ile örtüşmektedir.

İncelenen araştırmaların 110'unda (%14) araştırmanın metodolojik yaklaşımı, türü ya da modeli ile ilgili herhangi bir açıklama yapılmadığı görülmüştür. Bu araştırmaların metodolojik yaklaşımları ve türleri, araştırmacılar tarafından değerlendirilerek belirlenmiştir. Buna göre incelenen araştırmaların yarıdan fazlası (%61,8) nicel yaklaşıma dayalı olarak yürütülmüştür. Bu bulgu eğitim alanında nicel yaklaşıma dayalı araştırmaların daha sık yayımlanmakta olabileceğine yönelik sınırlı bir kanıt sağlamaktadır. Selçuk, Palancı, Kandemir ve Dündar (2014) yaptıkları araştırmada nicel araştırmaların daha fazla yayımlandığını (%69,5), nitel ve kuramsal araştırmalara ise nicel araştırmalara kıyasla daha az (%30,5) rastlanıldığını ortaya koymuşlardır. Benzer şekilde eğitim alanında nicel araştırmaların daha sık yayımlandığı sonucuna ulaşılmış başkaca araştırmalar da bulunmaktadır (Akaydın ve Çeçen, 2015; Bektaş, Dündar ve Ceylan, 2013; Çiltaş, 2012; Çiltaş, Güler ve Sözbilir, 2012; Göktaş vd. 2012; Erdem, 2011; Gülbahar ve Alper, 2009; Kashy vd. 2009; Şimşek vd. 2009).

Nicel yaklaşımla yürütülen araştırmaların tamamına yakınında (%94), kullanılan istatistiksel yöntem ve teknikler hakkında yeterli sayılabilecek düzeyde açıklama yapılmıştır. Bu araştırmaların büyük çoğunluğunda (%70,4) birden fazla istatistiksel teknik kullanılmıştır. Sadece bir tekniğin ve sadece parametrik veya parametrik olmayan tekniklerin kullanıldığı araştırmaların oranı %30'un altındadır. Nicel araştırmaların çeşitli özellikleri açısından incelendiği birçok araştırmada da çoğunlukla birden fazla istatistiksel tekniğin birlikte kullanıldığı sonucuna ulaşılmıştır (Akaydın ve Çeçen, 2015; Selçuk, vd. 2014; Bektaş, vd. 2013; Yağmur Şahin, Kana, Varışoğlu, 2013; Göktaş vd. 2012; Erdem, 2011). Bu durum tek değişkenli, çoklu veya çok değişkenli yöntemlerin en az ikisinin bir arada kullanıldığı araştırmaların sayıca artma eğiliminde olduğunu da ortaya koymaktadır.

Araştırmanın metodolojik yaklaşımı, türü, modeli ve kullanılan istatistiksel teknikler dikkate alındığında, bu çalışmada incelenen 788 araştırmadan 340'ında (%43) normallik varsayımlarının test edilmesi gerektiği, fakat bu 340 araştırmanın yarıdan fazlasında (%60,9) normallik varsayımlarının test edilmediği tespit edilmiştir. Ayrıca normallik varsayımının test edildiği 133 (%39,1) araştırma belirlenmiş, fakat bu 133 araştırmanın 40'ında (%30,1) normallik varsayımının hangi yöntem ve teknikle incelendiği konusunda herhangi bir bilgi verilmemiştir. Bu bulgu, eğitim alanında yürütülen ve normallik varsayımının test edilmesi gereken araştırmalarda, bu varsayımların yeterince dikkate alınmadığı ve kontrol edilmediği yönünde sınırlı bir kanıt sağlamaktadır. Benzer şekilde başka araştırmalarda da parametrik test varsayımının gerekli olduğu araştırmalarda diğer varsayımların yanı sıra normallik varsayımına yönelik

kontrollerde de önemli eksiklikler olduğu belirtilmiştir (Ergun ve Çilingir, 2013; Doğru, vd. 2012; Ruscio ve Roche, 2012; Evrekli, vd. 2011; Delice, 2010; Tavşancıl vd. 2010; Kabaca ve Erdoğan, 2007; Keselman vd. 1998). Bu bağlamda normallik varsayımının yeterince dikkate alınmayarak kontrol edilmemesinin araştırma sonuçlarının geçerliği ve güvenirliği hakkında ciddi sınırlılıklar ortaya koyduğu ifade edilebilmektedir.

Normallik varsayımına yönelik incelemelerde ortalama, mod, medyan, basıklık ve çarpıklık katsayıları gibi betimsel istatistiklere dayalı yöntemlerin diğer yöntemlerden daha sık kullanıldığı görülmüştür. İkinci olarak hipotez testleri, üçüncü olarak ise grafiksel incelemeler kullanılmaktadır. Hipotez testleri içerisinde en sık kullanılan testin Kolmogorov-Smirnov testi olduğu görülmüştür. Bu sonuçlara benzer şekilde Keselman vd. (1998)'nin yaptıkları araştırmada MANOVA testinin kullanıldığı araştırmaların yüzdelik dilim oluşturmayacak kadar az bir kısmında çok değişkenli normallik varsayımının test edildiği, betimsel istatistiklere dayalı yöntemlerin ise daha fazla kullanıldığı belirlenmiştir. Kabaca ve Erdoğan (2007), normallik varsayımının çoğunlukla test edilmediğini, test edildiğinde ise Kolmogorov-Smirnov testinin kullanıldığını ortaya koymuştur. Gömleksiz ve Bozpolat (2015), çalışmalarında Shapiro-Wilk testine nazaran daha fazla sayıdaki araştırmada Kolmogorov-Smirnov testinin kullanıldığını belirtmişlerdir. Yurtdışında yapılan araştırmalar daha çok büyük örneklemler (n>1000) üzerinde yürütülürken (Yalçın, Yavuz ve İlgün Dibek, 2015), Türkiye'de yapılmış araştırmalarda çoğunlukla 50 ve 400 kişi arası çalışma gruplarından veri toplanılması nedeniyle (Akaydın ve Çeçen, 2015; Yağmur Şahin, vd. 2013; Göktaş vd. 2012; Alper ve Gülbahar, 2009) Kolmogorov-Smirnov testinin kullanılmış olabileceği düşünülmektedir. Ancak bu testin küçük (20 ile 50 arası) ve büyük örneklemlerde (n>250) manidar sonuçlar üreterek, kullanılacak istatistiksel testin hatalı olmasına neden olacağı ifade edilmektedir (Schoder, vd. 2006; Hair, vd. 1998). Keselman, Othman ve Wilcox (2013) yaptıkları araştırmada, simülatif veriler üzerinde farklı normallik testlerini karşılaştırmıştır. Örneklem büyüklüğü 20, 40 ve 80 olarak anlamlılık düzeyi ise .05, .10, .15, .20 olarak manipüle edilmiştir. Araştırma sonuçlarına göre Anderson-Darling yönteminin Kolmogorov-Smirnov ve Cramer-von Mises testlerine nazaran normalliği belirlemede daha etkili olduğu gözlenmiştir. Schoder, vd. (2006)'nin de ifade ettiği gibi örneklemin 20 ile 50 arasında olduğu durumlarda Kolmogrov-Smirnov testinin kullanılma durumunun aksine Anderson-Darling yönteminin daha geçerli sonuçlar ortaya koyduğu bulunmuştur.

Normallik varsayımının incelendiği araştırmaların büyük çoğunluğunda (%66,9), bu varsayımın sağlandığı kararı verilerek parametrik yöntemler kullanılmıştır. Bunun yanı sıra normallik varsayımının sağlanmadığı kararına bağlı olarak önce istatistiksel düzeltme yapıp sonrasında parametrik yöntemlerin kullanıldığı ya da herhangi bir düzeltme yapılmaksızın parametrik olmayan yöntemlerin kullanıldığı araştırmalar da bulunmaktadır (%15). Diğer taraftan normallik varsayımına yönelik incelemelerle ilişkilendirilmeyerek kullanılacak istatistiksel yöntem ve tekniklere karar verilmiş olan araştırmalar (%11,3) ve normallik varsayımı sağlanmamış olmasına rağmen parametrik yöntemlerin kullanıldığı araştırmalar (%2,3) da bulunmaktadır. Bu bulgu, eğitim alanında yürütülen ve normallik varsayımının test edilmesi gereken araştırmalarda kullanılacak olan istatistiksel yöntem ve tekniğe karar vermede normallik varsayımına yönelik test sonuçlarının dikkate alınma durumları açısından önemli eksiklikler bulunabildiği ve hatalar yapılabildiği yönünde sınırlı bir kanıt sağlamaktadır. Bu bulguyu destekleyecek şekilde parametrik test varsayımlarının sıklıkla göz ardı edildiği (Ruscio ve Roche, 2012; Keselman vd. 1998; Micceri, 1989) ve normallik varsayımının birçok araştırmada hatalı veya eksik test edildiği (Doğru, vd., 2012; Tavşancıl vd., 2010), bazı araştırmalarda da bilinçli veya bilinçsiz bir şekilde istatistiksel sonuçların araştırmacıların beklentileri doğrultusunda yanlış raporlaştırıldığı (Bakker ve Wicherts, 2011) belirtilmektedir. Bu durum sıklıkla araştırmacıların istatistik alanındaki yetersizlikleri ile ilişkilendirilmektedir.

Normallik varsayımı test edilmesi gereken ve bu varsayımların test edildiği araştırmaların yayımlandığı dergilere ve yayım yılına göre dağılımları benzerdir. Diğer taraftan normallik varsayımlarının test edildiği araştırmalarda verilen kararların dergilere göre dağılımları farklılaşmaktadır. 'TED Eğitim ve Bilim

Dergisi' bu farklılığın belirgin kaynağını oluşturmaktadır. Bu dergide yayımlanan araştırma makalelerinde, normallik varsayımının sağlandığı ve buna bağlı olarak parametrik yöntemlerin kullanıldığı, ayrıca normallik varsayımının sağlanmadığı ve istatistiksel düzeltmelerden sonra parametrik yöntemlerin kullanıldığı ya da herhangi bir düzeltme yapılmadan doğrudan parametrik olmayan yöntemlerin kullanıldığı araştırmaların oranının daha yüksek olduğu görülmüştür. Diğer taraftan yine bu dergide, normallik varsayımının sağlanmamasına rağmen parametrik yöntemlerin kullanıldığı araştırmaların oranı da diğer dergilerden daha yüksektir. Bu durum, dergiler arasında normallik varsayımlarına yönelik olarak yapılacak testler ve buna bağlı olarak verilen kararlar açısından manidar farklar olabileceğine yönelik sınırlı bir kanıt sağlamaktadır. İlgili literatürde normallik varsayımının araştırıldığı makalelerin yayımlandığı dergilerin karşılaştırmalı olarak değerlendirildiği bir çalışmaya rastlanmamıştır. Bununla birlikte makalelerin metodolojik ve yöntemsel açıdan incelendiği bazı araştırmalarda, dergilere göre farklılığın olduğu görülmüştür. Bu araştırmaların bir kısmında dergiler arasında istatistiksel olarak anlamlı bir farklılığın gözlenemediği (Özsoy ve Özsoy, 2013; Ruscio ve Roche, 2012), diğer bir kısmında ise anlamlı farklılık bulunduğu (Akaydın ve Çeçen, 2015; Yalçın, vd. 2015; Bakker ve Wicherts, 2011; Meline ve Wang, 2004; Keselman vd. 1998) belirlenmiştir. Buna bilgiler doğrultusunda dergiler arasında benzerlik ve farklılıkların olabileceği sonucu çıkarılabilmektedir.

Bu çalışma sonucunda normallik varsayımının test edilmesi gereken eğitim araştırmalarında, normallik varsayımlarının test edilmesi ve buna bağlı olarak kullanılacak istatistiksel yöntem ve tekniklere karar verilmesi açısından önemli eksiklikler ve hatalar bulunabildiği görülmüştür. Bu kapsamda başta araştırmacıların sonrasında yayım sürecinde görev alan editör, hakem ve diğer uzmanların, bu konuda daha titiz ve hassas davranmaları gerekmektedir. İleri araştırmalara yönelik olarak daha geniş bir örneklem üzerinde benzer araştırmaların yapılması, örneklem büyüklüğü ile normallik varsayımlarının incelenme durumları arasındaki ilişkilerin araştırılması, araştırmacıların bu konudaki yeterlik düzeylerinin ve ilişkili değişkenlerin araştırılması önerilmektedir.

Kaynakça

Abbott, M. L. (2011). *Understanding educational statistics using Microsoft Excel and SPSS*. United States: John Wiley & Sons, Inc.

Akaydın, Ş. ve Çeçen M. A. (2015). Okuma becerisiyle ilgili makaleler üzerine bir içerik analizi. *Eğitim ve Bilim, 40*(178), 183-198.

Alper, A. ve Gülbahar, Y. (2009). Trends and issues in educational technologies: A review of recent research in TOJET. *The Turkish Online Journal of Educational Technology, 8*(2), 124-135.

Bakeman, R., & Robinson, B. F. (2005). *Understanding statistics in the behavioral sciences*. United States: Lawrence Erlbaum Associates, Inc.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior research methods, 43*(3), 666-678.

Bektaş, M., Dündar, H. ve Ceylan, A. (2013). Ulusal sınıf öğretmenliği eğitimi sempozyumu (USOS) bildirilerinin çeşitli değişkenler açısından incelenmesi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 197-222.

Bradley, J. V. (1982). The insidious L-shaped distribution. *Bulletin of the Psychonomic Society, 20*(2), 85-88.

Buskirk, T. D., Willoughby, L. M., & Tomazic, T. J. (2013). Nonparametric statistical techniques. T. D. Little, (Ed.) *The Oxford handbook of quantitative methods: Volume 2 – Statistical analysis (s. 106-142)*. United States: Oxford University Press.

Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı (18. baskı)*. Ankara: PegemA Yayıncılık.

Cohen, R. J., & Swerdlik, M. (2010). *Psychological testing and assessment: An introduction to tests and measurement*. New York: McGraw-Hill Book Co.

Çiltaş, A. (2012). 2005-2010 yılları arasında matematik eğitimi alanında Türkiye'de yapılan yüksek lisans ve doktora tez çalışmalarının içerik analizi. *The Journal of Academic Social Science Studies, 5*(7), 211-228.

Çiltaş, A., Güler, G. ve Sözbilir, M. (2012).Türkiye'de matematik eğitimi araştırmaları: bir içerik analizi çalışması. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(1), 565-580.

Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları.* Ankara: Pegem Akademi.

Delice, A. (2010). Nicel araştırmalarda örneklem sorunu. *Kuram ve Uygulamada Eğitim Bilimleri*, *10*(4) 1969-2018

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how.* United States: Springer-Verlag London Limited.

Doğru, M., Gençosman, T., Ataalkın, A. N. ve Şeker, F. (2012). Fen bilimleri eğitiminde çalışılan yüksek lisans ve doktora tezlerinin analizi. *Türk Fen Eğitimi Dergisi, 9*(1), 49-64.

Erdem, D. (2011). Türkiye'de 2005–2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 2*(1), 140-147.

Ergun, M. ve Çilingir, F. (2013). İlköğretim bölümünde yapılan lisansüstü tezlerin incelenmesi: Ondokuz Mayıs Üniversitesi Örneği. İsmail, G. (Ed.), *VI. Ulusal Lisansüstü Eğitim Sempozyumu* içinde (s.85-90). Sakarya Üniversitesi.

Evrekli, E., İnel, D., Deniş, H. ve Balım, A. G. (2011). Fen eğitimi alanındaki lisansüstü tezlerdeki yöntemsel ve istatistiksel sorunlar. *İlköğretim Online, 10*(1), 206-218. .

Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll) (Third edition).* London: SAGE Publications Ltd.

Foster, J. J., Barkus, E., & Yavorsky, C. (2006). *Understanding and using advanced statistics.* London: SAGE Publications Ltd.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489.

Gnanadesikan, R. (1997). *Methods for statistical data analysis of multivariate observations (Second edition).* United States: John Wiley & Sons, Inc.

Göktaş, Y., Hasançebi, F., Varışoğlu, B., Akçay, A., Bayrak, N., Baran, M. ve Sözbilir, M. (2012). Türkiye'deki eğitim araştırmalarında eğilimler: Bir içerik analizi. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(1), 443-460.

Gömleksiz, M. N. ve Bozpolat, E. (2015). Eğitim programları ve öğretim alanında yapılan doktora tezlerinin içerik analizi (2009-2014). *Eğitim ve Bilim,* 40(182), 29-41

Green, S. B., & Salkind, N. J. (2005). *Using SPSS for windows and macintosh: Analyzing and understanding data (Fourth edition).* United States: Pearson Prentice-Hall.

Hair, J. F., Anderson, R. E., Tatham, R., L., & Black, W. C. (1998). *Multivariate data analysis (Fifth edition).* United States: Prentice-Hall, Inc.

Howitt, D., & Cramer, D. (2011). *Introduction to SPSS statistics in psychology: For version 19 and earlier (Fifth edition).* London: Pearson Education Limited.

Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics: Principles and methods (Sixth edition)*. United States: John Wiley & Sons. Inc.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis (Sixth edition).* United States: Pearson Education, Inc.

Kabaca, T. ve Erdoğan, Y. (2007). Fen bilimleri ve matematik eğitimi alanlarındaki tez çalışmalarının istatistiksel açıdan incelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 22, 54-63.

Karasar, N. 2015). *Bilimsel araştırma yöntemi (28. Baskı).* Ankara: Nobel Yayın ve Dağıtım.

Kashy, D. A., Donnellan, M. B., Ackerman, R. A., & Russell, D. W. (2009). Reporting and interpreting research in PSPB: Practices, principles and pragmatics. *Personality and Social Psychology Bulletin, 35*(9), 1131-1142.

Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the design and analysis of experiments in behavioral research.* Lanham, MD: University Press of America.

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. United States: Library of Congress Cataloging in Publication Data.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of educational research*, *68*(3), 350-386.

Keselman, H. J., Othman, A. R., & Wilcox, R. R. (2013). Preliminary testing for normality: Is this a good practice? *Journal of Modern Applied Statistical Methods*, *12*(2), 2-19.

Kirk, R. E. (2008). *Statistics an introduction (Fifth edition).* United States: Thomson Higher Education.

Koziol, J. A. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika*, *69*(2), 423-427.

Krzanowski, W. J. (2007). *Statistical principles and techniques in scientific and social investigations*. United States: Oxford University Press.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation (second edition)*. United States: Lawrence Erlbaum Associates.

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2006). *Basic statistics for business and economics (Fifth edition)*. United States: McGraw-Hill Companies.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519-530.

Mardia, K. V. (1972). *Statistics of directional data.* United States: Academic Press, Inc.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis (Probability and mathematical statistics).* United States: Academic Press Limited.

McClave, J. T., Benson, P. G., & Sincich, T. (2001). *A first course in business statistics (Eight edition).* United States: Prentice Hall.

McKillup, S. (2012). *Statistics explained: An introductory guide for life scientists (Second edition).* United States: Cambridge University Press.

Meline, T., & Wang, B. (2004). Effect-size reporting practices in AJSLP and other ASHA journals, 1999-2003. *American Journal of Speech-Language Pathology*, 13, 202-207.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation (third edition).* United States: Pyrczak Publishing.

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin*, *105*(1), 156-166.

Miles, J., & Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction or how I came to know and love the standart error.* London: SAGE Publications Ltd.

Mudholkar, G. S., Srivastava, D. K., & Lin, C. T. (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality. *Communications in Statistics: Theory and methods*, *24*(4), 953-985.

Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report* 12(2), 281-316. http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf adresinden en son 05.03.2016 tarihinde alınmıştır.

Özdamar, K. (2011). *Paket programlar ile istatistiksel veri analizi-1 (Sekizinci baskı).* Eskişehir: Kaan Kitabevi.

Özsoy, S. ve Özsoy, G. (2013). Eğitim araştırmalarında etki büyüklüğü raporlanması. *İlköğretim Online*, *12*(2), 334-346.

Pagano, R. R. (2009). *Understanding statistics in the behavioral sciences (Ninth edition).* United States: Wadsworth, Cengage Learning.

Paulson, A. S., Roohan, P., & Sullo, P. (1987). Some empirical distribution function tests for multivariate normality. *Journal of Statistical Computation and Simulation*, *28*(1), 15-30.

Privitera, G. J. (2015). *Statistics for the behavioral sciences (Second edition).* United States: SAGE Publications.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. United States: Cambridge University Press.

Ravid, R. (2011). *Practical statistics for educators (fourth edition).* United States: Rowman & Littlefield Publishers.

Rochon, J., Gondan, M., & Kieser, M. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology*, 12: 81, doi:10.1186/1471-2288-12-81.

Romeu, J. L., & Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *Journal of Multivariate Analysis*, 46, 309-334.

Ross, S. M. (2009). *Introduction to probability and statistics for engineers and scientists (fourth edition)*. Canada: Elsevier Academic Press.

Roussas, G. (2007). *Introduction to probability (first edition).* United States: Elsevier Academic Press.

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *8*(1), 1-11.

Schoder, V., Himmelmann, A., & Wilhelm, K. P. (2006). Preliminary testing for normality: Some statistical aspects of a common concept. *Clinical and Experimental Dermatology*, *31*(6), 757-761.

Selçuk, Z., Palancı, M., Kandemir, M. ve Dündar, H. (2014). Eğitim ve bilim dergisinde yayınlanan araştırmaların eğilimleri: İçerik analizi. *Eğitim ve Bilim*, *39*(173), 430-453.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (Complete samples). *Biometrika*, 52(3/4), 591-611.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of The American Statistical Association*, *63*(324), 1343-1372.

Sharma, S. (1996). *Applied multivariate techniques*. United States: John Wiley & Sons. Inc.

Sprent, P., & Smeeton, N. C. (2007). *Applied nonparametric statistical methods (Fourth edition).* United States: CRC Press.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences (Fifth edition).* United States: Taylor and Francis Group, LLC.

Şimşek, A., Özdamar, N., Uysal, Ö., Kobak, K., Berk, C., Kılıçer, T. ve Çiğdem, H. (2009). İkibinli yıllarda Türkiye'deki eğitim teknolojisi araştırmalarında gözlenen eğilimler. *Kuram ve Uygulamada Eğitim Bilimleri, 9*(2), 115-120.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics (Sixth edition).* United States: Pearson Education.

Tavşancıl, E., Çokluk, Ö., Çıtak, G. G., Kezer, F., Yıldırım, Ö. Y. ve Bilican, S. (2010). *Eğitim bilimleri enstitülerinde tamamlanmış lisansüstü tezlerin incelenmesi (2000–2008).* Ankara Üniversitesi Bilimsel Araştırma Projesi Kesin Raporu.

Thode, H. C. (2002). *Testing for normality.* United States: Marcel Dekker, Inc.

Upton, G., & Cook, I. (1996). *Understanding statistics*. United Kingdom: Oxford University Press.

Uzgören, N. (2012). *Bilimsel araştırmalarda kullanılan temel istatistiksel yöntemler ve SPSS uygulamaları (Genişletilmiş ikinci baskı).* Bursa: Ekin Basın Yayın.

Wells, C. S., & Hintze, J. M. (2007). Dealing with assumptions underlying statistical tests. *Psychology in the schools*, *44*(5), 495-502.

Wilcox, R. R. (2012a). *Introduction to robust estimation and hypothesis testing (Third edition).* United States: Elsevier Inc.

Wilcox, R. R. (2012b). *Modern statistics for the social and behavioral sciences: A practical introduction.* United States: Chapman & Hall/CRC Press.

Yağmur Şahin, E., Kana, F. ve Varışoğlu, B. (2013). Türkçe eğitimi bölümlerinde yapılan lisansüstü tezlerin araştırmaeğilimleri. *International Journal of Human Sciences*, *10*(2), 356-378.

Yalçın, S., Yavuz, H. Ç. ve İlgün Dibek, M. (2015). En yüksek etki faktörüne sahip eğitim dergilerindeki makalelerin içerik analizi. *Eğitim ve Bilim*, *40*(182), 1-28.