

## BİLİMSEL ARAŞTIRMALARDA YAPILAN İSTATİSTİKSEL VE YÖNTEMBİLİMSEL HATALAR-II: GRAFİK, TABLO VE GÖSTERİM HATALARI

Saime SAYIN\*

### Öz

*Bu çalışmada, eğitim alanında yapılan, ulusal hakemli dergilerde yayımlanmış makalelerde karşılaşılan grafik, \*tablo ve yorumlamalardaki açıklama ve gösterim hataları üzerinde durulmuştur. Bu amaçla 1999-2006 yılları arasında yayımlanmış olan makale ve bildiri metinleri incelenmiştir. Yapılmış olan çeşitli istatistiksel hatalara örnek teşkil edebilecek olanlardan kesitler alınıp ayrıntılı olarak incelenmiştir. İncelemeler sonucunda, istatistiksel kavramların doğru kullanılmadığı ya da birbirlerinin yerine hatalı olarak kullanıldığı, amaca uygun grafik hazırlanamadığı ve yapılan analizlere yönelik özet tabloların hazırlanmasında güçlük çekildiği, bağımlı ve bağımsız grup t testlerinin nasıl yapılması ve hipotezlerin hangi koşullar altında kabul/ret edilmesi gerektiğinin bilinmediği, test sonuçlarının metin içerisinde gösterimlerinde hatalar yapıldığı gözlenmiştir.*

**Anahtar Sözcükler:** Bilimsel araştırma, yöntembilimsel hata, istatistiksel hata

### Abstract

*Focus of this study is mistakes made on evaluation and representation of graphics and tables within education articles published in national journals with referees. Articles and declarations published within 1999-2006 have been investigated. Samples representing various statistical mistakes were collected and examined. Conclusions are; wrong usage of statistical concepts, graphics aren't satisfying the purpose, inappropriate evaluation of graphics, insufficient knowledge in how to conduct t tests for dependent and independent groups, difficulty in preparing summary tables of analysis made, not knowing under what conditions hypotheses are to be accepted or rejected, mistakes in representation of test results within the text.*

**Keywords:** Scientific research, methodological mistake, statistical mistake

Her bilim bilgi üretir ama her bilgi bilimsel değildir (Güvenç, 1994). Hempel, Reischer ve Yıldırım'a göre, "her bilim dalı ya da disiplini geçerli ve güvenilir (doğruluğu irdelenebilir, kanıtlanabilir) bilgi üretmeye çalışır. Ama her geçerli ve güvenilir bilgi, bilimsel olmayabilir. Bilimsel bilgiyi diğer bilgi türlerinden ayıran ölçüt, metot veya yöntem (bilginin nasıl/ hangi yoldan elde edildiği) bilgisidir" (Akt. Güvenç, 1994). "Araştırmanın değerini ve güvenilirliğini azaltan her türlü girişim ise, bilimsel yanılma (scientific misconduct) olarak ifade edilmektedir. Araştırmacı araştırma planlamasını, uygun yöntem seçimini, yöntemlerin uygulanmasını, sonuçların analizini ve yorumunu bilmediği için farkında olmaksızın güvenilir olamayan bilgiler üretebilir. Disiplinsiz araştırma (sloopy research) olarak adlandırılan böyle durumlarda araştırmacılara araştırma eğitimi verilerek, araştırma disiplini öğretilir ve böylece bu türdeki bilimsel yanılmalar düzeltilerek araştırmacılar bilime kazandırılabilirler" (Kansu, 1994). Ancak öncelikle bu durumun saptanması ve eksikliklerin belirlenmesi gerekmektedir. Oğuz (1999) günümüzde her bilim insanının bilimsel çalışmaların denetlenmesinde sorumluluk taşıması, ilgi alanına giren konularda yayınları izlemesi, ortaya konulan bilim dışı savlara ya da çalışmalara ilişkin yanlışlıkları saptayacak biçimde özenle incelemesi ve bu tür saptamaları bilimsel yayın ortamında ortaya koyması gerektiğini ifade etmektedir.

Dolayısıyla araştırmalardaki hataların tespit edilmesi ve sunulması hem bilimsel çalışmaların niteliğini artırmak, hem de bu tür hataları yapan araştırmacıları Kansu'nun (1994) deyiimiyle "bilime kazandırmak" adına önemli olacaktır. Aksi takdirde araştırma sayısı arttığı hâlde bilimsel bilgilerin artmayacağı bir gerçektir.

Yapılan istatistiksel ve yöntemsel hatalara yönelik olarak özellikle tıp alanında birkaç çalışmaya rastlanmaktadır. Örneğin, Erkuş (2004) çalışmasında, bazı tıp dergilerinin son sayılarında yer alan çalışmaları yöntemsel ve istatistiksel açıdan ele almış ve çeşitli değerlendirmelerde bulunmuştur. Fakat, Erkuş (2004) yaptığı çalışmada tespit ettiği hatalardan ciddi olarak gördüğü bir kaçına kısaca yer vermiş, ancak hatanın ne olduğunu ve nasıl düzeltilebileceğini belirtmemiştir. Karaağaoğlu (2005), klinik çalışmalarda dikkat edilmesi gereken istatistiksel işlemleri ve aksi durumda olabilecek hataları ortaya koymaya çalışmıştır. Demirbağ (2007) ise, 4 tıbbi dergiye yayımlanmak üzere gönderilen, Epidemiyoloji ve İstatistiksel açıdan yazardan değerlendirilmesi istenen, ardışık 222 medikal

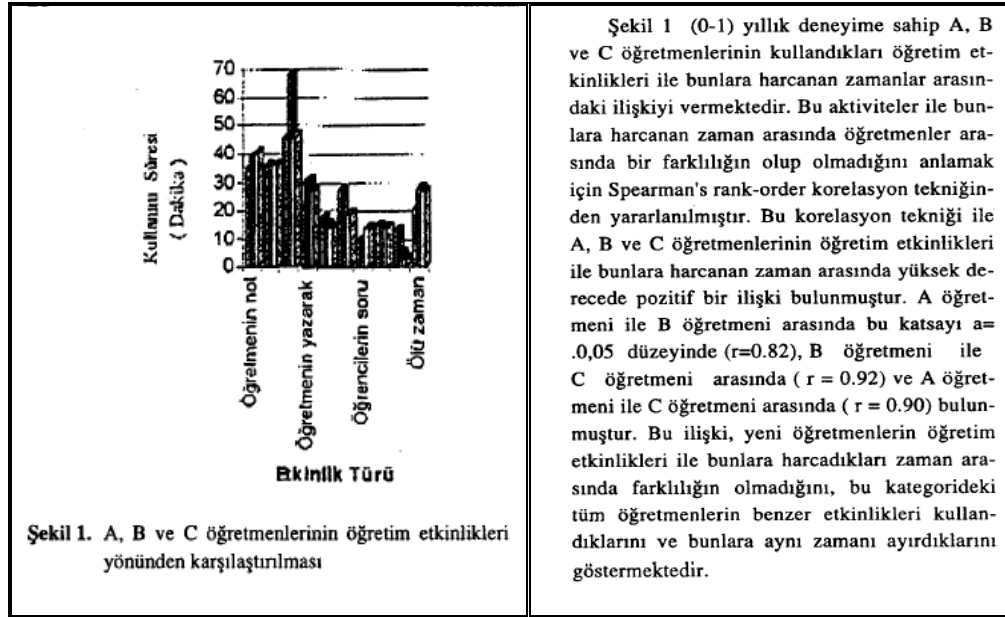
çalışmanın sonuçlarını incelemiş ve özellikle hangi konularda hatalar yapıldığını tespit etmeye çalışmıştır. Demirbağ (2007)' da sadece yapılan hataları vurgulamıştır. Eğitim alanında ise, Bademci'nin (2005a, 2005b, 2006a, 2006b ve 2007) ölçme ve yöntembilimsel hatalara yönelik yaptığı ayrıntılı çalışmaların dışında, arařtırmalarda yapılan istatistiksel hatalara ilişkin çalışmalara rastlanılmamaktadır. Bu açıdan bakıldığında Bademci'nin (2005a, 2005b, 2006a, 2006b ve 2007) çalışmaları eğitim bilim alanında ilk örnekleri oluşturmaktadır: Bu çalışmalarda arařtırmacı, ölçme ve arařtırma yöntembilimine ilişkin yapılan bazı hataları açıkça ortaya koyarak, bu hatalara yönelik çözüm ve düzeltilme yollarını ayrıntılı biçimde sunmuştur. Erdoğan ve Kabaca (2006) ise fen bilimleri ve matematik eğitimi alanlarında yapılan istatistiksel hataları incelemek amacıyla, 1996-2006 yılları arasında Marmara Üniversitesi Eğitim Bilimleri Enstitüsünde, fen bilimleri ve matematik eğitimi alanlarında yapılan yüksek lisans ve doktora tezlerinden 100 adet tez çalışmasını random (rastgele) olarak seçmişler ve bu tezlerde veri toplama araçları, geçerlik ve güvenilirlik çalışmaları, örneklem, betimsel istatistiklerin kullanımı, normal dağılım tespiti, parametrik testlerin seçimi, çok değişkenli istatistiklerin kullanımı, parametrik olmayan testlerin kullanımı ve biçimsel hatalar (tabloların oluşturulmasına yönelik hatalar vb.) olmak üzere yapılan hataları incelemişlerdir.

Bu çalışmada da, eğitim alanında yapılan arařtırmalardaki istatistiksel hataların tespit edilmesi ve düzeltilmesi amaçlanmıştır. Bu amaçla, kapsam daraltılarak sadece bir alt alanda, *-fen eğitimi* alanında- 1999-2007 yılları arasında yapılmış, ulusal hakemli dergilerde yayımlanmış ve internette ulaşılabilen makaleler taranmıştır. Bu makalelerde istatistiksel yöntemlerin kullanımına yönelik olarak varsayımların karşılanıp karşılanmadığına, veri türlerinin kullanılan istatistiksel yöntemlere uygun olup olmadığına bakılmayarak, sadece istatistiksel analiz sonuçlarının grafik ve tablolar hâlinde sunulması ve yorumlanması aşamasında gözlenen hatalar ele alınmıştır. Gözlenen istatistiksel hatalara yönelik olarak arařtırmalardan kesitler alınmış ve ayrıntılı olarak incelenmiştir. Yapılan istatistiksel hataların neler olduğunun somut olarak görülmesinin, bu tür hataların giderilmesine olanak sağlayabileceği düşünülmektedir. Dolayısı ile bu çalışmada yapılan istatistiksel hataların miktarlarının sayısal olarak incelenmesi değil, yapılan ve tekrarlanan hataların gözler önüne serilmesi amaçlanmıştır. Böylelikle arařtırmacıların -iyi niyetli olarak- farkında olmadan güvenilir olmayan sonuçlar ortaya koyan arařtırmalar yapmalarının ortadan kalkacağı ve dolayısıyla bilimsel çalışmaların niteliğinin artacağı umulmaktadır.

### Araştırmalarda Görülen Bazı Ciddi İstatistiksel Hatalar

1999-2007 yılları arasında ulusal hakemli dergilerde yayımlanan makalelerden özellikle grafik, tablo, yorum ve gösterim hataları içerenlerden her yıl için bir adet olmak üzere dokuz adet örnek makale seçilmiştir. Bu makalelerden kesitler alınmış ve istatistiksel hatalara yönelik tespitler ve eleştiriler tarih sırasına göre aşağıda verilmiştir.

**Örnek 1 :** Azar ve Çepni'nin (1999) ulusal hakemli bir dergi olan *H.Ü. Eğitim Fakültesi Dergisi*'nde yayımlanmış "Fizik Öğretmenlerinin Kullandıkları Öğretim Etkinliklerinin Mesleki Deneyime Göre Değişimi" isimli çalışmaları ve çalışmalarında kullandıkları grafik ve ilgili yorumları (sayfa 27 ve 28).



**Eleştiri ve Bilimsel Katkı 1:** Yukarıda yer alan açıklamalar incelendiğinde araştırmacıların fark ve ilişki kavramını hatalı olarak birbirlerinin yerine kullandıkları görülmektedir. Araştırmacılar Şekil 1'in öğretmenlerin kullandıkları öğretim etkinlikleri ile etkinliklere harcanan zamanlar arasındaki ilişkiyi verdiği ifade etmişler, ancak Şekil 1'i "A, B ve C öğretmenlerinin öğretim etkinlikleri

yönünden karşılaştırılması” olarak adlandırmışlardır. Üstelik Şekil 1’de yer alan grafik, araştırmacıların belirttiği gibi bir ilişki (saçılım) grafiği (Guilford ve Fruchter, 1978: 79; Çelik, 1999: 172-173) olmamasına rağmen şeklin açıklamasında ilişkiden bahsedilmektedir. Bunun dışında araştırmacılar “Bu aktiviteler ile bunlara harcanan zaman arasında öğretmenler arasında bir farklılığın olup olmadığını anlamak için Spearman’ın rank-order korelasyon tekniğinden yararlanılmıştır.” şeklinde farklılığı belirlemek üzere ilişki katsayısı hesapladıklarını belirtmişlerdir. Spearman’ın rank-order korelasyon katsayısı (Spearman sıra korelasyonu katsayısı) Pearson ilişki katsayısının parametrik olmayan karşılığı olup *iki ortalama arasındaki farkı değil, sadece iki değişken arasındaki ilişkiyi* verir (bakınız, Gönen ve Tatlıdil, 1985: 150; Çelik, 1999: 247). Araştırmacıların hesapladıkları ilişki katsayısına ilişkin olarak “Bu ilişki, yeni öğretmenlerin öğretim etkinlikleri ile bunları harcadıkları zaman arasında farklılığın olmadığını, bu kategorideki tüm öğretmenlerin benzer etkinlikleri kullandıklarını ve bunlara zaman ayırdıklarını göstermektedir” şeklindeki yorumları ise istatistiksel olarak “ilişki” ve “fark” kavramlarının anlamlarını karıştırdıklarını ortaya koymaktadır.

**Örnek 2:** *Aydoğdu’nun (2000) ulusal hakemli dergi olan H.Ü. Eğitim Fakültesi Dergisi’nde yayımlanmış “Kimya Öğretiminde Deneylerle Zenginleştirilmiş Öğretim ve Geleneksel Problem Çözme Etkinliklerinin Kimya Ders Başarısı Açısından Karşılaştırılması” isimli çalışması ve çalışmasında hazırladığı tablo ve yorumları (sayfa 29-31).*

mından faydalanılmıştır(9). Analizlerden elde edilen sonuçlar başlangıçta deney ve kontrol grupları arasında kimya başarısı açısından istatistiksel olarak anlamlı bir fark olmadığını ortaya çıkarmıştır ( $t=0.34$ ,  $P>0.05$ ). Hesaplanan  $t$  değeri (hesaplanan  $=0.34$ ), tablodaki  $t$  değerinden daha küçüktür (tablo= 1.96). Uygulama sonunda deneylerle zenginleştirilmiş kimya öğretiminden faydalanan grubun Kimya Başarı Testinde daha başarılı olduğu gözlenmiştir ( $t=2.21$ ,  $P>0.05$ ). Hesaplanan  $t$  değeri tablodaki  $t$  değerinden daha büyüktür. Her iki grup için ön ve son test ortalamaları, standart sapma ve  $t$  değerleri aşağıdaki tablo 1’de verilmiştir.

Tablo 1. Grupların ön test ve son test ortalamaları ve  $t$  değerleri.

Testler	Grup no	Öğr.sayısı	Ortalama	Std.sapma	t-testi
Ön test	Grup 1	56	6.05	2.09	0.34
	Grup 2	54	6.21	2.31	
Son test	Grup 1	56	9.00	3.29	2.21
	Grup 2	54	7.58	2.67	

**Eleştiri ve Bilimsel Katkı 2:** Araştırmacının t testi ile ilgili açıklamaları incelendiğinde, araştırmacının deney ve kontrol grupları arasındaki kimya başarıları açısından anlamlı bir fark olmadığını belirttiği ve bu durumu ( $t=0.34$ ,  $P>0.05$ ) şeklinde gösterdiği görülmektedir. Ancak araştırmacı zenginleştirilmiş kimya öğretiminden faydalanan grubun Kimya Başarı Testinde daha başarılı olduğunu belirterek, “hesaplanan t değeri tablodaki t değerinden daha büyüktür” şeklinde bir açıklama yapmış ve ( $t=2.21$ ,  $P>0.05$ ) şeklinde göstermiştir. Oysa araştırmacının anlamlı farkı belirtmek üzere ( $t(108) = 2.21$ ,  $P<0.05$ ) gösterimini kullanması gerekmektedir.

**Örnek 3:** Buluş’un (2001) ulusal hakemli dergi olan *P.Ü. Eğitim Fakültesi Dergisi*’nde yayımlanan “Öğretmen Adaylarında Bilişsel Tutarlılık Tercih” isimli çalışması ve çalışmasında hazırladığı tablo ve ilgili yorumları (sayfa 26-33).

**Tablo 2. Öğretmen Adaylarının Bilişsel Tutarlılık Tercih ve Akademik Başarı Puanlarının Ortalama ve Standart Sapmaları**

Değişkenler	n	$\bar{x}$	s
Tutarlılık Terc.	332	111.3	23.4
Akadem. Başarı	297	71.8	6.5

### 1. Bilişsel Tutarlılık Tercihinin Akademik Başarı ile İlişkisi

Bunun için, öncelikle denencede belirtilen değişkenler üzerinden bütün grubun aldığı puanlar karşılaştırılmış ( $r$ ), sonra bilişsel tutarlılığı tercih etme düzeyi düşük ve yüksek olanların akademik başarı puanlarının ilişkileri t testi ile araştırılmıştır.

Ortalama ve standart sapmaları Tablo 2’de verilen bilişsel tutarlılık tercihi ve akademik başarı değişkenlerinin ilişkileri Pearson Korelasyon Katsayısı ile analiz edilmiş ve  $r = -.005$ ;  $p > .05$  düzeyinde istatistiksel olarak anlamsız bir ilişki bulunmuştur. Bu sonuç, bilişsel tutarlılık tercihinin akademik başarı düzeyi ile anlamlı bir ilişki içinde olmadığını göstermektedir.

-----

Tutarlılık tercihinin cinsiyet ile iliřkisi t testi ile, diđer deęiřkenlerle iliřkisi ise varyans analizi ile arařtırılmıřtır. Elde edilen bulgulardan, anlamlı iliřki ierenleri ařaęıda tablolafıtılarak aıklanmıřtır.

**Tablo 4. ğretmen Adaylarının Kendilerini Zeka ve Yetenek Aısından Yeterli Grme Derecelerine Gre Tutarlılıęı Tercih Etme Dzeyinin Ortalaması ve Standart Sapması**

Zeka Yet.Der.	n	$\bar{x}$	ss
Biraz yeterli	13	96.6	36.2
Orta dzeyde	161	114.8	21.6
Olduka yeterli	158	109.0	23.3
Toplam	332	111.3	23.4

Tablo 4 incelendięinde, kendisini zeka ve yetenek aısından biraz yeterli gren adayların en dřük ( $\bar{x} = 96.6$ ), orta dzeyde yeterli bulanların ise en yksek ( $\bar{x} = 114.8$ ) biliřsel tutarlılık tercihi dzeyi ortalamasına sahip oldukları grlmektedir.

Ortalamalar arasındaki farkların anlamlı olup olmadıęı varyans analizi ile test edilmiř ve sonular ( $F(5.255)$ ;  $p = .006 < .01$ ) ortalamalar arasındaki farkların .01 dzeyinde nemli olduęunu gstermiřtir. Buna gre, biliřsel tutarlılıęı tercih etme dzeyi kendisini zeka ve yetenek aısından yeterli bulma dzeyine gre deęiřmektedir. Farkın kaynaęı iin yapılan Tukey Testi'ne gre, anlamlı fark, zeka ve yetenek aısından orta dzeyde yeterli bulanların ortalaması ile biraz ve oldukça yeterli bulanların ortalamaları arasındaki farkların yksek olmasından (iliřkilerin nem deęeri sırasıyla  $p = .017 < .05$  ve  $p = .050 < .05$ ) kaynaklanmıřtır.

**Eleřtiri ve Bilimsel Katkı 3:** Arařtırmacı ğretmen adaylarının Biliřsel Tutarlılık Tercihleri ile akademik bařarı puanları arasındaki iliřkiyi Pearson korelasyon katsayısı ile analiz ettięini belirtmiřtir. Ancak Tablo 2'de grldęi zere ğrenci sayıları farklıdır. Arařtırmacının ka ğrenci iin biliřsel tutarlık tercihi ve akademik bařarı arasındaki korelasyonu hesapladıęı anlařılmamaktadır. Ayrıca arařtırmacı tutarlılık tercihinin cinsiyet ile iliřkisini t testi ile, diđer deęiřkenlerle iliřkisini ise varyans analizi ile arařtırdıęını ve anlamlı iliřki ierenleri tablolafıtılarak aıkladıęını rapor etmiř, ama tablo ile ilgili aıklamada ortalamalar arasındaki farkların anlamlı olup olmadıęını varyans analizi ile test ettięini belirtmiřtir. Arařtırmacı daha sonra farkın kaynaęını bulmak iin Tukey testi yaptıęını belirtmiř ve "... farkların yksek olmasından (iliřkilerin nem deęeri sırasıyla  $p=.017<.05$  ve  $p=.050<.05$ ) kaynaklanmıřtır." řeklinde aıklama yapmıřtır. Bu durum arařtırmacının kullandıęı testleri hangi amala kullanması gerektięi

konusunda bir karmaşa yaşadığını ya da *ilişki* ve *fark* kavramlarını birbirlerinin yerine hatalı olarak kullandığını göstermektedir.

**Örnek 4:** Kanlı ve Yağbasan'ın (2002) ulusal hakemli bir dergi olan *Milî Eğitim Dergisi*'nde yayımlanmış "2000 Yılında Ankara'da Fizik Öğretmenleri İçin Düzenlenen Hizmetiçi Eğitim Yaz Kursunun Etkinliği" isimli çalışmaları ve çalışmalarında hazırladıkları tablolar ve ilgili yorumları (sayı 150, sayfasız).

Bu araştırmanın yapılmasındaki temel amaç, fizik öğretmenleri için düzenlenen hizmetiçi eğitim yaz kurslarının, kursa katılan öğretmenlerin meslekî bilgi ve becerilerini geliştirmede katkısının olup olmadığını ve hizmetiçi eğitim kurslarının etkinliği hakkında öğretmenlerin görüşlerini tespit etmektir. Bu bağlamda M.E.B.Hizmetiçi Eğitim Dairesi Başkanlığı tarafından 2000 yılı Ağustos ayında Ankara FenLisesi'nde düzenlenen "Fizik Laboratuvar Araçlarını Kullanma ve Formatörlük Kursu"na araştırmacı-gözlemci olarak katılarak, düzenlenen kursun; öğretmenlerin meslekî bilgi ve becerilerini geliştirmedeki etkinliği, kursun başında ve sonunda uygulanan başarı testi ile belirlenmiştir. Elde edilen veriler istatistik işlemlerle incelenerek ortaya çıkan sonuçların sebepleri araştırılmıştır.

-----

..... Ortalamalardaki bu artışın anlamlılığını test etmek için null hipotezi ve t-testi kullanılmıştır. Kursta, test puanları arasındaki ilişkiyi ifade eden t-testleri hesaplanmış, Tablo-1'de belirtilmiştir.

	Ortalama X	Standart sapma S	Kişi sayısı N	t-testi	Olasılık değeri P
Ön-test	12,419	3,0062	29	6,361	0,000
Son-test	17,931	3,5750	29		

Tablo-1. Düzenlenen Kursta Ait Verilerin Değerlendirilmesi

$t=6,361$  değeri  $=0,05$  anlamlılık düzeyi ve Serbestlik derecesi  $SD=56$ 'a ait tablo değeri olan  $t=2,021$  değerinden büyük olduğu için veya bulunan  $p$  değeri  $0,025$ 'ten küçük olduğundan; Ankara'daki kursta uygulanan ön-test ve son-test sonuçları arasında "anlamlı bir ilişki vardır" sonucuna ulaşılır."



**Eleřtiri ve Bilimsel Katkı 4:** Yukarıda yer alan tabloya iliřkin aıklama incelendiđinde, arařtırmacıların ortalamalardaki artışın anlamlılıđını test etmek için null hipotez ve t-testi kullandıklarını ifade ettikleri grlmektedir. ncelikle “null hipotez” (yokluk hipotezi, sıfır hipotezi, istatistiksel hipotez), iki deđer arasında anlamlı bir fark olmadığını ifade eden hipotezdir (İnal ve Gnay, 1993: 386; Ergn Blbl, 2001: 212). Dolayısıyla “null hipotez” ile anlamlılık test edilmez, aksine test edilen hipotezin kendisidir. Ayrıca, arařtırmacılar “...test puanları arasındaki iliřkiyi ifade eden t testleri hesaplanmış...” řeklinde bir ifade kullanmışlardır. t testi arařtırmacıların iddia ettikleri gibi (tablo incelendiđinde iki ortalama arasındaki farka iliřkin t testi olduđu grlmektedir) iliřkiyi ifade eden bir test deđildir ve ayrıca, testler hesaplanmazlar, yalnızca t deđeri (t istatistiđi) hesaplanır ve tablo deđerleri ile karřılařtırılarak yokluk hipotezinin kabul ya da red edilmesine karar vermede kullanılır (bakınız, Ergn Blbl, 2001: 211-220; zdamar, 2004: 316). Ancak arařtırmacılar bu yanlış ifadeyi kullanmakla kalmayıp tablo zerinde de *t* yerine *t testi* yazmışlardır. Bunların dıřında, arařtırmacılar kursa katılanların n-son test puan ortalamalarını karřılařtırmışlar (bađımlı grup ya da bađımlı iki rneklem) ancak serbestlik derecesinin 56 olduđunu ifade etmişlerdir. Oysa, arařtırmanın rneklemini 29 kiřiden oluřmaktadır ve serbestlik derecesi 28 olmalıdır (bakınız, zdamar, 2004: 321). Son olarak da, arařtırmacılar t testi tablosunu yorumlarken “n-test ve son-test sonuları arasında anlamlı bir iliřki vardır” sonucuna ulařtıklarını ifade etmişlerdir. Syledikleri gibi n-test ve son-test puanları arasında anlamlı bir iliřki sz konusu ise, bu durum, n test puanları ile son test puanlarının birlikte deđiřtiđi anlamına gelmektedir. Bir bařka deyiřle n test ve son test puanlarının birlikte artacađı ya da azalacađı veya n test puanları artarken son test puanlarının azalacađı ya da tam tersinin olacađı anlamına gelmektedir. Sonu olarak, arařtırmacıların istatistiksel yntemler hakkındaki bilgilerinin yeterli olmadığını ifade etmek pek de yanlış olmayacaktır.

**rnek 5: Tezcan ve Gnay’ın (2003) ulusal hakemli dergi olan *Milli Eđitim Dergisi*’nde yayımlanan “Lise Kimya đretiminde Laboratuvar Kullanımına İliřkin đretmen Grřleri” isimli alıřmaları ve alıřmalarında hazırladıkları tablolar ve ilgili yorumları (sayı 159)**

Öğretmenlere yöneltilen 22 adet anket sorusunun cevapları ile cinsiyetlerinin ve okul türlerinin ilişkileri Chi-Square Testi ile değerlendirildi. Değerlendirmede bayan öğretmenler 1, erkekler 2 ile soru şıkları a=1, b=2, c=3, d=4 rakamı ile belirtildi. Örnek olarak 1 sorunun değerlendirilmesi Tablo 1’de verildi. 1. soru: Kimya öğretiminde laboratuvar kullanıyor musunuz? a: evet, b: hayır.

**Tablo 1.** Cinsiyet ile 1. sorunun Chi-Square test sonuçları

Soru ve şıkları	Soru 1.		
	1	2	TOPLAM
<b>CİNSİYET</b>	-	-	-
1	14	7	21
2	12	1	13
<b>Toplam</b>	26	8	34

	-	-	Asymp.
	Value	-	Sig.
	-	df	(2-sidee)
Pearson Chi-Square	38.020	4	,000
Likelihood Ratio	12.398	4	,015
N of Valid Cases	35		

Bu sorunun değerlendirilmesine göre öğretmenler kimya öğretiminde laboratuvar kullanmaktadır ( $P > 0.005$ ). Diğer soruların analizi şöyledir: Öğretmenler laboratuvar destekli öğretimin gereğine inanmaktadır ( $P < 0.005$ ). Ancak laboratuvar istenen düzeyde uygulanamamaktadır.

**Bilgi:** Bu çalışma bir “Yüksek Lisans Tez” çalışmasıdır.

Eleřtiri ve Bilimsel Katkı 5: Arařtırmacılar 2x2'lik Khi-Kare tablosu hazırlamıřlardır. Ama arařtırmacıların bu tabloya iliřkin verdikleri analiz sonuçları (bilgisayar ıktısı aynen kullanılmıř) incelendiđinde, serbestlik derecesinin 4 olduđu grlmektedir. Halbuki 2x2'lik bir tablo iin serbestlik derecesi 1'dir (bakınız, elik, 1999:206; Spiegel ve Stephens, 1999: 263). Ayrıca sz konusu 2x2'lik tablo incelendiđinde gzelerden birinde denek sayısının 5 'ten az olduđu grlmektedir. Oysaki "Khi-Kare dađılıřında herhangi bir sıra ya da stunda 5'ten az sayıda denek olmamalıdır." (Kutsal ve Muluk, 1978). Bunlara ilaveten arařtırmacılar cinsiyet ile 1. soruya (1. Soru: Kimya đretiminde laboratuvar kullanıyor musunuz?) verilen yanıtları incelemiř olduklarını belirtmelerine rađmen yorumlarını "Bu sorunun deđerlendirilmesine gre đretmenler kimya đretiminde laboratuvar kullanmaktadırlar." řeklinde hatalı olarak yapmıřlardır. Sz konusu alıřmanın bir yksek lisans tez alıřmasından hazırlandıđı da gz nne alındıđında yapılan istatistiksel hataların nemi daha iyi anlařılacaktır.

**rnek 6:** Gzel'in (2004) ulusal hakemli bir dergi olan *Trk Fen Eđitimi Dergisi*'nde yayımlanan "Genel Fizik ve Matematik Derslerindeki Bařarı ile Matematiđe Karřı Olan Tutum Arasındaki İliřki" isimli alıřması ve alıřmasında hazırladıđı tablolar ile ilgili yorumları (sayfa 49-57).

Öğrencilerin matematik başarıları ile matematik tutum puanları arasındaki ilişki Tablo 5’ de görülmektedir.

*Tablo 5. Matematik Başarıları İle Matematik Tutum Puanının Karşılaştırılması*

Matematik Başarı Puanı	Ort.	N	Ss	F	P
0-40	154,14	22	26,99	3,71	0,06
40-60	162,65	34	25,23		
60-70	164,98	46	26,90		
70-80	169,16	37	27,76		
80-100	177,09	64	28,03		
Toplam	168,00	203	27,88		

Tablo 5’den F değerinin 3,71 olduğu görülmektedir. Bu değer, 0,05 manidarlık düzeyinde anlamlı farklılığı ifade etmektedir. Farkın hangi başarı grubu lehine olduğunu tespit etmek için Tukey testi uygulanmıştır. Tukey testinin sonuçları Tablo 6’ da verilmiştir.

*Tablo6. Tukey Testi Sonuçları*

Matematik Başarı Puanı.	Ort. Farkı	P
0-40 :40-60	-8,51	0,782
0-40 :60-70	-10,84	0,536
0-40 :70-80	-15,03	0,240
0-40 :80-100	-22,96	0,06*
40-60 :60-70	-2,33	0,996
40-60 :70-80	-6,52	0,851
40-60 :80-100	-14,45	0,089
60-70 :70-80	-4,18	0,957
60-70 :80-100	-12,12	0,142
70-80 :80-100	-7,93	0,619

Tablo 6 incelendiğinde matematik başarı notları (0-40) ile (80-100) olanların arasında anlamlı bir farkın olduğu, yani matematik başarı düzeyi yüksek olanların tutumlarının başarı düzeyi düşük olanlara kıyasla olumlu yönde olduğu görülmektedir. Diğerleri arasında anlamlı bir fark görülmemektedir.

**Eleřtiri ve Bilimsel Katkı 6:** Yukarıda yer alan tablo ve aıklamalar incelendiğinde, arařtırmacının matematik bařarı ve matematik tutum puanları arasındaki iliřkiyi belirlemek iin varyans analizi yapmıř olduėu, ama tabloyu ‘‘Matematik Bařarı ile Matematik Tutum Puanlarının Karřılařtırılması’’ řeklinde adlandırdıėı grlmektedir. Arařtırmacı iliřki ve fark kavramlarını birbirlerinin yerine kullanmaktadır. Bunun dıřında, arařtırmacı nem dzeyi olarak .05 almasına raėmen .06 řeklinde olan p deėerini yorumlarken ‘‘anlamlı fark vardır’’ řeklinde bir aıklama yapmıřtır. Oysaki p deėeri 0.05’ten byk olduėu iin istatistiksel olarak anlamlı fark yoktur. Bu durumun basit bir yazım hatası olmadığı, arařtırmacının ‘‘...farkın hangi bařarı grubu lehine olduėunu tespit etmek iin Tukey Testi uygulanmıřtır’’ řeklindeki aıklaması ve aynı yorumu Tukey testi sonucunda da yapması ile ortaya konulmuřtur. Ayrıca varyans analizi sonucunda yapılan Tukey Testi ile *farkın hangi grubun lehine olduėu deėil, sadece farkın hangi gruplar arasında olduėu* tespit edilebilir. Sonu olarak arařtırmacının, kurduėu hipotezi hangi kořullar altında kabul/ret etmesi ve nasıl yorumlaması gerektiėini ve testlerin kullanım amalarını iyi bilmediėini sylemek yanlıř olmayacaktır.

**rnek 7:** zkan ve Azar’ın (2005) ulusal hakemli bir dergi olan *Milli Eėitim Dergisi*’nde yayımlanmıř ‘‘rnek Olaya Dayalı ğretim Ynteminin Dokuzuncu Sınıf ėrencilerinin Ders Bařarı ve Derse Karřı Tutumlarına Olan Etkisinin İncelenmesi’’ isimli alıřmaları ve alıřmalarında hazırladıkları tablolar ve ilgili yorumları (sayı 168, sayfasız).

**izelge 1. Kontrol Grubunda Cinsiyete Gre Fizik Dersi Bařarısına İliřkin Eřli Gruplar t Testi Sonuları**

Faktrler	N	X	S	t	sd	P
Kız (kontrol)	22	50,54	24,46	0,70	58	0,487
Erkek (kontrol)	38	40,30	24,39			

\*p>0.05

**Çizelge 2. Deney Grubunda Cinsiyete Göre Fizik Dersi Başarısına İlişkin Eşli Gruplar t Testi Sonuçları**

	N	X	S	t	sd	P
Kız (deney)	26	49,69	25,93	-0,892	58	0,376
Erkek (deney)	34	55,64	25,39			

\*p>0.05

Çizelge 1 ve Çizelge 2'de görüldüğü gibi deney ve kontrol grubu kız ve erkek öğrencilerin başarı düzeyleri arasında anlamlı bir fark yoktur. ....

**Eleştiri ve Bilimsel Katkı 7:** Yukarıda yer alan tablolar incelendiğinde, araştırmacıların kontrol ve deney grupları için ayrı ayrı olmak üzere kız ve erkeklerin puanlarını “Eşli Gruplar t Testi” (paired/ related samples t test, yani bağımlı iki örneklem t testi) ile kontrol ettikleri görülmektedir. Araştırmacıların söz konusu testin hangi durumlarda kullanıldığına ilişkin olarak bilgilerini yeniden gözden geçirmelerinin gerektiği, kız ve erkek öğrencilerin karşılaştırmalarını yaparken bağımlı iki örneklem t testi kullandıklarını rapor etmiş olmalarından anlaşılmaktadır. Çünkü söz konusu test, “n birimlik gruptan iki farklı uygulamaya ilişkin iki veri seti elde edilmiş ise bu veri setleri bağımlıdır ve farkların ortalaması sıfır olan bir toplumun rasgele örneğidir” varsayımının test edilmesinde kullanılır (Özdamar, 2004: 332). Araştırmacıların eşli gruplar t testi (bağımlı iki örneklem t testi) uyguladıklarını belirtmelerine rağmen serbestlik derecesi (n-1 olmalı) olarak bağımsız iki örneklem t testinin serbestlik derecesini (n<sub>1</sub>+n<sub>2</sub>- 2) kullanmışlardır. Bu durumda araştırmacıların gerçekte belirttikleri gibi eşli gruplar t testi (bağımlı iki örneklem t testi) mi, yoksa bağımsız iki örneklem t testi mi yaptıkları anlaşılamamaktadır. Üstelik araştırmacıların yaptıkları bu hata, çok daha ciddi soru/sorunları gündeme getirmektedir. Zira araştırmacılar SPSS programını kullanmışlardır. SPSS programında bağımlı ve bağımsız iki örneklem t-testi için veri girişleri farklı yapılmaktadır. Dolayısıyla araştırmacıların veri girişlerini uygun bir şekilde yapıp yapmadıkları sorusu/sorunu da akla gelmektedir. Bütün bunlara ilaveten tabloların altında verilen “\*p > 0.05” ifadesi yer almaktadır. Ancak bu tür gösterimler istatistiksel olarak anlamlı fark bulunduğu, bu farkın hangi önem düzeyinde olduğunu göstermek amacıyla kullanılır (bakınız, Özdamar, 2004: 316). Araştırmacıların bu gösterimi niçin kullandıklarına yönelik makul bir açıklama bulunmamaktadır. Hem istatistiksel yöntemin hatalı kullanılması ya da hatalı rapor

edilmesi, hem de veri giriři ile ilgili ortaya ıkan sorulardan dolayı sz konusu alıřmanın bilimsel yeterlilięi ciddi bir Őekilde sarsılmaktadır.

**rnek 8:** Bozdoęan, Tařdemir ve Demirbař'ın (2006) ulusal hakemli bir dergi olan *İnn Üniversitesi Eęitim Fakltesi Dergisi*'nde yayımlanmıř "Fen Bilgisi ęretiminde İřbirlikli ęrenme Ynteminin ęrencilerin Bilimsel Sre Becerilerini Geliřtirmeye Ynelik Etkisi" isimli alıřmaları ve alıřmalarında hazırladıkları tablolar ve ilgili yorumları (sayfa 23-36).

#### Verilerin Analizi

Arařtırma verilerinin analizinde SPSS 10.0 programı kullanılmıřtır. Deney ve Kontrol gruplarının bilimsel sre becerileri testleri arasında anlamlı bir iliřkinin olup olmadıęını belirlemek iin aritmetik ortalama, yzde-frekans ve t-Testi analizleri kullanılmıřtır.

Tablo 1. Deney ve Kontrol Grubu ęrencilerinin n test Puanlarına İliřkin t-Testi Sonuları

GRUP	N	$\bar{X}$	S	sd	t	P
Deney	34	2.88	1.93	64	1.859	.068
Kontrol	32	2.12	1.28			

Tablo 1 incelendięinde deney grubu ęrencilerinin bilimsel sre beceri n test puanlarının ( $\bar{X} = 2.88$ ) kontrol grubu ęrencilerinin bilimsel sre beceri n test puanlarından ( $\bar{X} = 2.12$ ) daha yksek olduęu grlmektedir. Gruplar arasındaki bu puan farkı istatistiksel olarak anlamlı bulunmamıřtır ( $t_{(64)} = 1.859, p > .05$ ). Bu sonuca gre, arařtırma ncesinde grupların denk dzeyde olduęu grlmektedir.

Deney ve kontrol grubunda yer alan öğrencilerin bilimsel süreç beceri son test puanları cinsiyetlerine göre incelenmiş, yapılan t-Testi analiz sonuçları Tablo 5 ve Tablo 6' da verilmiştir.

**Tablo 5. Deney Grubu Öğrencilerinin Bilimsel Süreç Son Test Puanlarının Cinsiyetle Olan İlişkisi**

CİNSİYET	N	$\bar{X}$	S	sd	t	p
Erkek	16	3.68	1.81	32	1.345	.188
Kadın	18	4.44	1.46			

Tablo 5 incelendiğinde, deney grubu öğrencilerinin bilimsel süreç son test puanlarının cinsiyetle olan ilişkisi anlamlı çıkmamıştır ( $t_{(30)} = 1.345$ ,  $p > .05$ ).

**Eleştiri ve Bilimsel Katkı 8:** Yukarıda verilen örnek ile ilgili olarak öncelikle, araştırmacıların “deney ve kontrol gruplarında bilimsel süreç becerileri testleri arasında anlamlı bir ilişki olup olmadığını belirlemek için aritmetik ortalama, yüzde-frekans ve t testi kullanılmıştır” şeklindeki açıklamalarının uygun olmadığı söylenebilir. Çünkü ortalama ve yüzde, sıklık dağılımlarının birbirlerine göre yerlerini, uzaklıklarını kısacası konumlarını belirten ölçülerdir (Saraçbaşı ve Kutsal, 1987: 41-43) ve ilişkiyi belirlemek için kullanılmazlar. Ayrıca örnek incelendiğinde, araştırmacıların t testini iki farklı amaçla yaptıkları görülmektedir. Araştırmacılar ilk tabloda t testini, deney ve kontrol gruplarının ön test puan ortalamaları arasındaki farkı kontrol etmek için, ikinci tabloda ise t testini, son test puanları ile cinsiyet ilişkisini kontrol etmek için kullanmışlardır. Araştırmacıların t testi'nin (bu hâli ile) kullanım amacını yeterince bilmedikleri söylenebilir, çünkü t testi ilişkiyi tespit etmek amacı ile kullanılamaz.



**Örnek 9:** Ergin, Kanlı ve Tan'ın (2007) ulusal hakemli dergi olan *Gazi Eđitim Fakóltesi Dergisi*'nde yayımlanmış "Fizik Eđitiminde 5E Modeli'nin Öđrencilerin Akademik Başarısına Etkisinin İncelenmesi" isimli çalıřmaları ve çalıřmalarında hazırladıkları tablo ve ilgili yorumları (191-209).

Tablo 1. Deney ve Kontrol Gruplarında Yer Alan Öđrencilerin "Mantıksal Düşünme Yeteneđi Testi (MDYT)" Puanlarına İliřkin Bađımsız Gruplar t-Testi Sonuçları

Test	Grup	N	$\bar{X}$	S.S.	s.d.	t	p
MDYT	Kontrol	40	17.750	1.904	82	.467	.642
	Deney	44	17.909	1.157			

Tablo 1 incelendiđinde; iki ortalama puan deđeri birbirine oldukça yakın olduđu görülecektir. Bu ortalamalar arasındaki farkın istatistiksel olarak anlamlı olup olmadıđı, bađımsız gruplar için t-testi ile yoklanmıř, hesaplanan t deđeri ve anlamlılık düzeyine ( $p < 0.05$ ) göre, gruplar arasında anlamlı bir fark gözlenmemiřtir.

Tablo 4. Deney ve Kontrol Gruplarında Yer Alan Öđrencilerin YAHBT Son Test Puanlarına İliřkin Bađımsız Gruplar t-Testi Sonuçları

Ölçüm	N	$\bar{X}$	S.S.	s.d.	t	p
Son Test (Kontrol)	40	15.725	2.374	82	7.319	.000
Son Test (Deney)	44	19.000	1.613			

Tablo 4 incelendiđinde; iki ortalama puan deđerinin birbirinden oldukça farklı olduđu görülür. Bu ortalamalar arasındaki farkın istatistiksel olarak anlamlı olup olmadıđı, bađımsız gruplar için t-testi ile yoklanmıř, hesaplanan t deđeri ve anlamlılık düzeyine ( $p < 0.05$ ) göre, gruplar arasında anlamlı bir fark gözlenmiřtir. Bu farkın deney grubu lehine olduđu görülmektedir.

**Eleřtiri ve Bilimsel Katkı 9:** Arařtırmacılar, çalıřmalarına uygun řekilde t testlerini yapmıřlar ve tablolařtırmıřlardır. Ancak tabloların yorumlanması esnasında "... Hesaplanan t deđeri ve anlamlılık düzeyine ( $p < 0.05$ ) göre, gruplar

arasında anlamlı bir fark gözlenmemiştir.” şeklindeki yorumlarını yaparken ciddi bir hata yapmışlardır. Çünkü ( $p < .05$ ) gösterimi hesaplanan test istatistiğinin .05 düzeyinde istatistiksel olarak anlamlı olduğunu belirtmek üzere kullanılır. Araştırmacılar çalışmada yer alan tüm tabloların altında, sonuç anlamlı olsa da olmasa da aynı gösterimi kullanmışlardır. Bu durum araştırmacıların bu gösterimi hangi amaçla kullanmaları gerektiği konusunda yeterli bilgiye sahip olmadıkları olasılığını akla getirmektedir.

### Sonuç ve Öneriler

Sargut’a (2006) göre; “Bilimsel makalelerin en temel işlevi, bilim kişilerinin alanlarında neler olup bittiğini izlemelerine olanak vermesidir. Bilim kişileri, birbirlerini okumak, birbirleriyle iletişim kurmak ve tartışmak zorundadırlar ve bu çerçevede bilim yapma sürecinin ortamını makaleler oluşturmaktadır. Bilimsel makalelerin niteliği de güçlü doktora programları ile bağlantılıdır.” Ayrıca Sargut (2006) “sosyal bilimlerde nitelikli makale üretimi konusunda sorunlar yaşandığını ifade etmektedir. Nitelsiz makale üretimi sorununun çözümüne yönelik olarak da; öncelikle güçlü doktora programlarına gereksinim olduğunu, bu programlarda bilim kişisi adaylarına doğru okuma alışkanlıkları verilmesi ve bilimsel değer ve normların öğretilmesi gerektiğini ifade etmiştir. Böylece kişiler bilim yapmanın vazgeçilmez ögesi olan buluşsal arayış ve savın önemini doktora süreci içerisinde öğreneceklerdir.” Sargut’un (2006) sosyal bilimlerle ilgili bu tespit ve önerilerine katılmamak mümkün değildir. Ancak sosyal bilimlerde ortaya çıkan nitelsiz makale sorununun, sosyal bilimlerle birlikte diğer alanlarda da çıktığını ve bu çalışmada da tipik örneklerinin verildiği türde pek çok makalenin dergilerde yer aldığını; ayrıca, ölçme ve araştırmalarda istatistiksel ve yöntembilimsel hataların yapıldığı, değişik disiplinlerde ve çok sayıda kitap, makale ve tez çalışmalarının bulunduğunu da ilave etmek gerekmektedir (geniş bilgi için bakınız, Bademci, 2004, 2005a, 2005b, 2005c, 2006a, 2006b, 2007; Erdoğan ve Kabaca, 2006; Sayın, 2008, baskıda). Yapılan bazı istatistiksel hatalar, *İstatistik Mezunları Derneği, 4. İstatistik Kongresi, Çalışma Grubu Raporu*’nda da (2005) ele alınmıştır. Çalışma Grubunca belirlenen sorunlar; istatistik alanında PC yazılım/paket programlarının yapılmakta olan hataları artırıcı bir zemin hazırladığı ve programların bilinçsiz ve hatalı uygulanmalarının gözlemlendiği, değişik alan araştırmalarında uygun olmayan analiz tekniklerinin kullanıldığı, veri derleme yöntem/lerinin doğru olmayan biçimde kullanıldığı, daha önce yapılan ve yayımlanmış olan ancak daha sonra hatalı oldukları saptanan çalışmaların düzeltilmediği ve bunları referans alan araştırmacıların yanlış yönlendirildiği, farklı sektörlerde yapılan araştırmalarda

tasarım, veri deęerlendirme, analiz ve sonuca varma konularında eksik ve yetersiz bilgi düzeyi ve uygulamaların olduęu vb. řeklinde özetlenebilir. alıřma Grubunca da dikkate alındıęı üzere çeřitli sektörlerde ya da alanlarda benzer istatistiksel ve yöntembilimsel hatalar yapılmaktadır.

Yurt dıřında yapılan alıřmalara bakıldıęında da benzer konulara dikkat çekilmeye alıřıldıęı görölmektedir. Lang ve Lantz (2003) ve Lang (2004) biomedical arařtırma makalelerinde bulunan betimsel istatistiklerin yanlış kullanımı, sonuçlar için yalnızca p deęerlerinin rapor edilmesi, istatistiksel testlerin sayıltularının kontrol edilmeden analizlerde kullanılması vd. olmak üzere 20 istatistiksel hatadan bahsetmişlerdir. Benzer řekilde Strasak ve dięerleri (2007) tıp alanında yapılan istatistiksel hataları tespit ettikleri alıřmalarında, alıřmanın deseni, verilerin analizi, uygulanan istatistiksel yöntemin belgelenmesi, alıřma verilerinin sunumu, alıřma bulgularının yorumu ile ilgili olarak 47 potansiyel istatistiksel hata ve eksiklięi rapor etmişlerdir. Bu alıřmaların dıřında özellikle tıp alanında istatistik ile etik arasında bir baę kurulduęu ve istatistięin yanlış kullanımının etik kural ihlali olarak ele alındıęı da görölmektedir (bakınız, Altman, 1980; Demirbaę, 2007). Tüm bu alıřmalar bilimsel arařtırmalarda istatistik ve yöntembilimin yeri ve önemini vurgulamaktadır.

Bu alıřmada da bilimsel bilgi üretme sürecinde önemli bir yeri olan istatistięin doęru anlaşılması ve uygulanmasına katkı saęlamak için fen eęitimi alanındaki yayınlarda gözlenen istatistiksel hatalardan (özellikle grafik, tablo ve yorumlardaki açıklama ve gösterim hataları) kesitler alınmış ve incelenmiştir. Yapılan incelemelerin sonucunda, arařtırmacıların istatistiksel kavramları iyi bilmedikleri için bazı kavramları hatalı olarak birbirlerinin yerine kullandıkları, arařtırmalarında kullandıkları testlerin kullanım amaçlarını yeterli düzeyde bilmedikleri, istatistiksel analiz sonuçlarını tablolařtırmada güçlükler yaşadıkları ve tabloları yorumlamada zorlandıkları, yapılan analizler sonucunda elde edilen deęerlerin anlamlılıęını göstermekte de problem yaşadıkları, arařtırmacıların yeterli istatistik bilgisine sahip olmadıkları görölmektedir. Bunların dıřında, arařtırmacıların istatistiksel veri analizi için bilgisayar kullanımlarında sorunlar yaşadıkları da gözlenmektedir. Özellikle veri girişlerini kurallarına uygun olarak yapıp yapamadıkları, ciddi bir soru ve sorun olarak ortaya çıkmaktadır. Sonuç olarak, yapılan istatistiksel hataların ortaya konulması ve düzeltilmesi gerekmektedir. Kansu'nun (1994) deyimiyile, arařtırmacılara arařtırma eęitimi verilmeli, arařtırma disiplini öğretilmeli ve arařtırmacılar bilime yeniden kazandırılmalıdır.

### İncelenen Çalışmalar

1. Azar, A. ve Çepni, S. (1999). Fizik öğretmenlerinin kullandıkları öğretim etkinliklerinin mesleki deneyime göre değişimi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 16–17, 24 – 33. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

[www.egitimdergisi.hacettepe.edu.tr/199916ALI%20AZAR.pdf](http://www.egitimdergisi.hacettepe.edu.tr/199916ALI%20AZAR.pdf)

2. Aydoğdu, C. (2000). Kimya öğretiminde deneylerle zenginleştirilmiş öğretim ve geleneksel problem çözme etkinliklerinin kimya ders başarısı açısından karşılaştırılması. *H.Ü. Eğitim Fakültesi Dergisi*, 19, 29–31. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://www.egitimdergisi.hacettepe.edu.tr/200019CEM%C4%B0L%20AYDO%C4%9EDU.pdf>.

3. Buluş, M. (2001). Öğretmen adaylarında bilişsel tutarlılık tercihi. *P.Ü. Eğitim Fakültesi Dergisi*, 10, 26–33. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://egitimdergi.pamukkale.edu.tr/makale/say%C4%B110/3-M.Bulu%C5%9F.pdf>

4. Kanlı, U. ve Yağbasan, R. (2002). 2000 Yılında Ankara'da Fizik Öğretmenleri İçin düzenlenen hizmet içi eğitim yaz kursunun etkinliği. *Millî Eğitim Dergisi*, 153–154. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir:

<http://yayim.meb.gov.tr/dergiler/153-154/kanli.htm>

5. Tezcan, H. ve Günay, S. (2003). Lise Kimya Öğretiminde laboratuvar kullanımına ilişkin öğretmen görüşleri. *Millî Eğitim Dergisi*, 159. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://yayim.meb.gov.tr/dergiler/159/tezcan-gunay.htm>

6. Güzel, H. (2004). Genel Fizik ve Matematik derslerindeki başarı ile Matematiğe karşı olan tutum arasındaki ilişki. *Türk Fen Eğitimi Dergisi*, 1, 49–56. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://www.tused.org/internet/tufed/tufedmain.htm>

7. Özkan, M. ve Azar, A. (2005). Örnek olaya dayalı öğretim yönteminin dokuzuncu sınıf öğrencilerinin ders başarısı ve derse karşı tutumlarına olan etkisinin incelenmesi. *Millî Eğitim Dergisi*, 168. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://yayim.meb.gov.tr/dergiler/168/index3-ozkan.htm>

8. Bozdođan, A.E., Tařdemir, A., Demirbař, M. (2006). Fen Bilgisi Öğretiminde işbirlikli öğrenme yönteminin öğrencilerin bilimsel süreç becerilerini geliřtirmeye yönelik etkisi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 7(11), 23-36. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiřtir. <http://web.inonu.edu.tr/~efdergi/arsiv/Bozdogan.pdf>
9. Ergin, İ., Kanlı, U. ve Tan, M. (2007). Fizik Eğitiminde 5E modeli'nin öğrencilerin akademik başarısına etkisinin incelenmesi. *Gazi Eğitim Fakültesi Dergisi*, 27(2), 191–209. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiřtir. <http://www.gefad.gazi.edu.tr/2007-2/11.pdf>

#### Kaynaklar

- Altman, D. G. (1980). Statistics and ethics in medical research: misuse of statistics is unethical. *British Medical Journal*, 281, 1182–1184. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiřtir <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1714517>.
- Bademci, V. (2004). “Testin güvenilirliđi” veya “test güvenilirlidir” diye ifade etmek doğru deđildir. *Türk Eğitim Bilimleri Dergisi*, Cilt 2, 367–373.
- Bademci, V. (2005a). Arařtırmalarda ölçme ile ilgili bazı büyük hataları düzeltmek ve bir reformu başlatmak: güvenilirlik, testlerin bir özelliđi deđildir. *Eğitim Fakültelerinde Yeniden Yapılandırmanın Sonuçları ve Öğretmen Yetiřtirme Sempozyumu'nda Sunulmuř Bildiri*. Ankara: Gazi Üniversitesi, Gazi Eğitim Fakültesi, 22–23–24 Eylül.
- Bademci, V. (2005b). Hakemlerin deđerlendirmelerindeki hatalar üzerine: Fisher'in z dönüşümü ve güvenilirlik çalışmaları için örneklem büyüklüđü. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, 17, 46–75. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiřtir. [http://www.esef.gazi.edu.tr/html/yayinlar/17\\_pdf/17\\_d.pdf](http://www.esef.gazi.edu.tr/html/yayinlar/17_pdf/17_d.pdf)
- Bademci, V. (2005c). Testler güvenilir deđildir: ölçüm güvenilirliđine yeterli dikkat ve güvenilirlik çalışmaları için örneklem büyüklüđü. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, 17, 33–45. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiřtir. [http://www.esef.gazi.edu.tr/html/yayinlar/17\\_pdf/17\\_c.pdf](http://www.esef.gazi.edu.tr/html/yayinlar/17_pdf/17_c.pdf)

Bademci, V. (2006a). Güvenirliđi dođru anlamak ve bazı klişeleri yıkmak: bilinenlerin aksine, cronbach'ın alfa katsayısı, negatif ve -1'den küçük olabilir. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 7(12), 3-26. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

<http://web.inonu.edu.tr/~efdergi/arsiv/bademci.pdf>

Bademci, V. (2006b). Tartışmayı sonlandırmak: cronbach'ın alfa katsayısı, iki değerli [0,1] ölçümlenmiş maddeler ile kullanılabilir. *Kazım Karabekir Eğitim Fakültesi Dergisi*, 13, 438-446.

Bademci, V. (2007). *Ölçme ve araştırma yöntem biliminde paradigma deđişikliği: testler güvenilir deđildir*. Ankara: Yenyap Yayınları.

Çelik, M.Y. (1999). *Biyoistatistik araştırma ilkeleri, yeni bir yaklaşım*. Diyarbakır: Dicle Üniversitesi Rektörlüğü Basımevi.

Demirbađ, A. E. (2007). Tıbbi yayınların epidemiyolojik ve istatistiksel analizindeki sorunlar ve yapıcı eleştiri yaklaşımının yayın ve dergi kalitesi üzerine etkisi. *Sađlık Bilimlerinde Süreli Yayıncılık- 5. Ulusal Sempozyumunda Sunulmuş Bildiri*. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.ulakbim.gov.tr/cabim/vt/uvt/tip/sempozyum5/page171-180.pdf>

Erdoğan, Y. ve Kabaca, T. (2006). Fen Bilimleri ve Matematik Eğitimi alanlarındaki tez çalışmalarında yapılan istatistiksel hatalar ve çözüm önerileri. *7. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi'nde Sunulmuş Bildiri*. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.fenmat.gazi.edu.tr/tam%20metin.pdf>

Ergün Bülbül, S. (2001). *Çözümsel istatistik*. İstanbul: Alfa Yayınları.

Erkuş, A. (2004). Bazı tıp dergilerinin son sayılarındaki makalelerin yönetsel ve istatistiksel açıdan incelenmesi. *Mersin Üniversitesi Tıp Fakültesi Dergisi* 2, 175-181. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.

[ftp://tip.mersin.edu.tr/pub/tipdergisi/cilt5\(2004\)/cilt5isayi2inisan2004/makale27\(175-181\).pdf](ftp://tip.mersin.edu.tr/pub/tipdergisi/cilt5(2004)/cilt5isayi2inisan2004/makale27(175-181).pdf)

Gönen, S. ve Tatlıdil, H. (1985). *Parametrik olmayan istatistiksel yöntemler*. Yayınlanmamış Ders Notu, Ankara: Hacettepe Üniversitesi.

Guilford, J.P. ve Fruchter, B. (1978). *Fundamental statistics in psychology and education* (Sixth Edition). McGraw-Hill International Editions.

- Güvenç, B. (1994) Türkiye’de sosyal bilimler gelişmeler ve süreklilikler. İçinde: *Dünya’da ve Türkiye’de Bilim, Etik ve Üniversite*. Ankara: Türkiye Bilimler Akademisi (TÜBA), Bilimsel Toplantı Serileri: 1, 23–35.
- İnal, C. ve Günay, S. (1993). *Olasılık ve matematiksel istatistik*. Ankara: H.Ü. Fen Fakültesi Yay. Ders Kitapları Dizisi: 16.
- İstatistik Mezunları Derneği, 4. İstatistik Kongresi, Çalışma Grubu Raporu* (2005). İstatistik alanında yapılan çalışmalarda çok sık rastlanan hatalar ve bilimsel etik. (10 Mayıs, 2005). Antalya: İstatistik Mezunları Derneği, 4. İstatistik Kongresi. İnternette (son olarak) 10 Aralık 2007 tarihinde elde edilmiştir. <http://www.imd.org.tr/gerekli/workgroupreport.pdf>
- Kansu, E. (1994). Bilimsel yanıtma ve önlenmesi. İçinde: *Dünya’da ve Türkiye’de Bilim, Etik ve Üniversite*, Ankara: Türkiye Bilimler Akademisi (TÜBA), Bilimsel Toplantı Serileri: 1, 71–75.
- Karaağaoğlu, E. (2005). Bilimsel makalelerde istatistiğin kullanımı, sunumu ve sık yapılan hatalar. *Sağlık Bilimlerinde Süreli Yayıncılık–2005. 3. Ulusal Sempozyumu’nda Sunulmuş Bildiri*. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir.
- Kutsal, A., Muluk, F. Z. (1978). *Uygulamalı temel istatistik*. Ankara: H.Ü. Fen Fakültesi Yay. Ders Kitapları Dizisi: 8. [http://www.ulakbim.gov.tr/cabim/vt/uv/tip/sempozyum3/013\\_ergun.pdf](http://www.ulakbim.gov.tr/cabim/vt/uv/tip/sempozyum3/013_ergun.pdf)
- Lang, T. (2004). Twenty statistical errors even YOU can find in biomedical research articles. *Croatian Medical Journal*, 45(4), 361–370. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.cmj.hr/2004/45/4/15311405.pdf>
- Lang, T. ve Lantz, C. J. (2003). 20 Statistical errors even YOU can find. *Science Editor*, 26(6), 191. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.councilscienceeditors.org/members/securedDocuments/v26n6p191.pdf>
- Oğuz, N. Y. (1999). Bilimsel yayın etiği. *Klinik Psikiyatri*, 2, 153–159. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. [www.klinikpsikiyatri.org/index.php?name=ozet&yid=34](http://www.klinikpsikiyatri.org/index.php?name=ozet&yid=34)
- Özdamar, K. (2004). *Paket programlar ile istatistiksel veri analizi (Çok Değişkenli Analizler)* (Beşinci Baskı). Eskişehir: Kaan Kitabevi.

- Saraçbaşı, T. ve Kutsal, A. (1987). *Betimsel istatistik*. Ankara: H.Ü. Fen Fakültesi Yay. Ders Kitapları Dizisi: 17.
- Sargut, A. S. (2006). Buluşsal arayış ya da sav: bilimsel makale niteliği üzerine düşünceler. *Sosyal Bilimlerde Süreli Yayıncılık, I. Ulusal Kurultayı'nda Sunulan Bildiri*. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.ulakbim.gov.tr/cabim/vt/uvvt/sbvt/kurultay1.pdf>
- Sayın, S. (2008baskıda). Bilimsel arařtırmalarda yapılan bazı istatistiksel ve yöntembilimsel hatalar-II: güvenilirlik kestirimlerine yönelik hatalar. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 9(15) sayısında yayımlanacak.
- Spiegel, M. R., Stephens, L. J. (1999). *Teori ve problemlerle istatistik* (Üçüncü Baskı). (Çeviri Ed: A. Esin ve S. Çelebiođlu). Ankara: Nobel Yayın Dağıtım.
- Strasak, A. M. ve diđerleri. (2007). Statistical errors in medical research –a review of common pitfalls. *Swiss Medical Weekly*, 137, 44–49. İnternette (son olarak) 01 Aralık 2007 tarihinde elde edilmiştir. <http://www.smw.ch/docs/pdf200x/2007/03/smw-11587.PDF>



### *Summary*

#### **SOME OF THE STATISTICAL AND METHODOLOGICAL MISTAKES MADE IN SCIENTIFIC RESEARCHES – II: INTERPRETATION AND REPRESENTATION MISTAKES ON GRAPHICS AND TABLES**

**Saime SAYIN\***

#### **Aim**

Every science field or discipline tries to produce valid and reliable information. But each valid and reliable information may not be scientific. The knowledge on criteria, methodology or method (how the data has been obtained) differentiates scientific data from other types of data (c/f, Güvenç, 1994). Every attempt reducing the value and trustability of a research is called as scientific misconduct. A researcher may produce non-reliable information unintentionally if the knowledge on research planning, selection of appropriate method, implementation of method and knowledge on analysis and interpretation of results doesn't exist. Research Training can be provided to researches of such so-called cases of sloppy research, and they can be taught on research discipline which in turn can help to correct that type of scientific misconducts and such researchers can be guided to the benefit of science (Kansu, 1994). But first, there is a need of detecting such cases and defining the shortcomings. Every scientist should assume responsibility on inspecting the scientific studies, follow up publications on their fields of interest, conduct detailed investigations in order to detect non-scientific assertion or mistakes in studies and should declare such cases on scientific publications (Oğuz 1999). Hence, it's important to evaluate studies on education with a critical view in order to eliminate shortcomings and mistakes. But, mistakes made specific to statistical and methodological approach are especially faced in a few researches made in the area of medical science. For example, Erkuş (2004), in his study, had investigated some articles in recent medical journals from statistical and methodological view and draw some conclusions. Yet, Erkuş (2004), in his study briefly pointed out to some mistakes but didn't go into details and didn't

---

Address for correspondence: \*Yrd. Doç. Dr., Zonguldak Karaelmas Üniversitesi, Ereğli Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Psikolojik Hizmetler Anabilim Dalı, Kdz. Ereğli-Zonguldak, saimesay@gmail.com

specify how they can be corrected. Karaođlu (2005) tried to indicate to which statistical operations attentions should be paid attention in clinical research and underlined what mistakes can be faced otherwise. Demirbađ (2007), on the other hand, investigated the result of 222 sequential medical articles which had been sent to 4 medical journals and tried especially to define the subjects in which mistakes are made. Demirbađ (2007) has underlined only the mistakes which have been made. In the area of education, researches were made by Bademci (2005a, 2005b, 2006a, 2006b and 2007). Bademci (2005a, 2005b, 2006a, 2006b and 2007) has clearly defined some mistake made related to measurement and research methodology and explained solutions and corrections for those mistakes in details. Erdođan and Kabaca (2006) selected randomly 100 theses out of master and doctorate theses submitted in the area of secondary science and mathematics education in Marmara University Institute of Educational Sciences between the years of 1996 and 2006 and investigated the mistakes made related to validity and reliability of data collection tools, samples, usage of descriptive statistics, determination of normal distribution, selection of parametric tests, usage of multivariate statistics, usage of non-parametric tests and structural mistakes (such as mistakes related to structuring the tables, etc.)

The focus of this study is to detect and correct statistical mistakes made on researches concerning field of education. As a consequence, it's expected to prevent researches producing non-reliable information as a result of their studies leading to an increase in the scientific quality of researches carried out in the field of education.

### **Method**

The focus of this study is also to detect and correct statistical mistakes with in the researches concerning field of education. For that purpose, the scope has been limited to articles concerning sub-field-science education- published between 1999 and 2007 in national journals with referees, and reachable on internet. The *methodology and findings* sections of the articles in the area of science education have been examined and nine examples from articles, for each year in the given period, were selected out of those articles having statistical mistakes especially on graphics, tables and evaluation. Samples were taken out of these selected articles, detected mistakes have been provided together with critics and corrections on these mistakes.

### **Findings**

The result of the investigation carried on nine sample articles subject to this study are as follows;

1. Graphical representations aren't appropriate with the aim and not evaluated properly,
2. Since the meaning of correlation coefficient isn't known, expressions like difference or indifference are used instead of correlation while interpreting the correlation coefficient.
3. Concepts of correlation and difference are mistakenly used instead of each other,
  - a) The same test (t test, dependent or independent sample t test) is being used for two different purposes ( in control of correlation and difference),
  - b) Two concepts (correlation and difference) were mistakenly used instead of each other in evaluation of Turkey test performed during and after Analysis of Variance,
4. Degrees of freedom concerning utilised test isn't provided correctly,
  - a) Even though the test conducted has been dependent to samples t test , degrees of freedom of independent two samples t test is provided and used
  - b) Degrees of freedom of the 2\*2 Chi square table was provided as 4,
5. Basic statistical and methodological concepts were used in a wrong way ( as in the example of; *...null hypothesis and t test has been utilised in order to test significance and ...t tests defining the correlation has been calculated*)
6. Deep concern for the quality of decision made on accepting/rejecting hypothesis based on the result of statistical test or analysis (such as reporting a significant difference of 0.05 level for a probability of  $p=0.06$ ),
7. Mistakes made on representation of significant level under the tables (like using  $p<.05$  representations even though  $*p>.05$  and irrespective of whether differences between means are significant or not).

In addition, findings has been detected which shows that researches may experience difficulties in using computers for statistical data analysis, especially when entering the data.

### Discussion and Result

Based on the statistical mistakes detected on examined studies, one can concluded that if other researches utilize the results of these previously published articles in national journals, new articles may contain similar mistakes. Detection and sharing of such mistakes with researches working in the field of education can help to reduce the threat to certain extend as well as helping to prevent repetition of such mistakes made unintentionally by some researchers and help to eliminate their shortcomings. Hence, these researches can turn to became beneficial for the sake of the science as stated by Kansu (1994).