



ARAŞTIRMA / RESEARCH

Parametrik olmayan yöntemlerin yanlış kullanımı ve doğurabileceği sonuçlar: simülasyon çalışması

Results of misuse of nonparametric methods: a simulation study

Didem Havva Ovla¹, Merve Türkegün¹, Ayça Asena Özdemir¹, Bennu Vedia Özcömert¹, Bahar Taşdelen¹

¹Mersin Üniversitesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin, Turkey

Cukurova Medical Journal 2017;42(2):241-248

Abstract

Purpose: Visual Analogue Scale has a narrow range which is used to evaluate the pain severity. In general, such type of data don't provide the parametric test hypothesis. But at the group comparisons the number of outlier could be excessive at one of the groups. Consequently, it is analyzed by using non-parametric methods. Therefore, in this study as comparing two groups which are narrow ranged rates and also have outliers (skew) it has been aimed that evaluating the performance of Student t test, Logistic transformed Student t test, Mann Whitney-U test and Mood's Median tests.

Material and Methods: The study was planned as a simulation study. Accordingly, it has been formed distributions which are at interval of 0-10, based on the mostly used Visual Analogue Scale rates which were used in evaluation of pain. Simulations were repeated 1000 times for 20 combinations which consisted of 4 different distributions and 5 different sample size.

Results: Type I error rates which obtained have varied as the sample size and the distribution type is changed and the test results haven't been changed though normalization method used in the type of data that is not concordance with normal distribution.

Conclusion: Misusing and overusing of non-parametric tests gives biased results in scientific studies. Although the general approach was using non-parametric tests in which datas were recorded at the interval of 0-10 cm, the simulation study showed that such datas could also be evaluated by using parametric tests.

Key words: Visual Analogue Scale, nonparametric statistics, pain

Öz

Amaç: Ağrı şiddetinin değerlendirilmesinde sıkça kullanılan Görsel Analog Skala dar aralığa sahiptir. Genellikle bu tip veriler parametrik test varsayımlarını sağlamaz. Grup karşılaştırmalarında ise grupların birinde aykırı değer sayısı fazla olabilir. Bu sebeplerle parametrik olmayan yöntemler kullanılarak analiz edilir. Buradan yola çıkarak, bu çalışmada dar aralıklı (range) ve aynı zamanda aykırı değerlere sahip (çarpık) iki grubun karşılaştırılmasında Student t testi, Logit dönüşümü yapılmış Student t testi, Mann Whitney U testi ve Mood's medyan testinin performansları değerlendirilmesi amaçlanmıştır.

Gereç ve Yöntem: Çalışma bir simülasyon çalışması olarak planlanmıştır. Bu doğrultuda, ağrı değerlendirmesinde en sık kullanılan Görsel Analog Skala değerleri temel alınarak, 0-10 aralığında dağılımlar türetilmiştir. Simülasyonlar dört farklı dağılım ve 5 farklı örnek genişliğinden oluşan 20 kombinasyon için 1000'er kez tekrarlanmıştır.

Bulgular: Elde edilen tip I hata oranları örnek genişliği ve dağılım tipi değişikçe farklılık göstermiştir. Ayrıca normal dağılıma uyumlu olmayan veri tipinde normalizasyon metodu kullanılsa bile test sonuçlarının değişmediği görülmüştür.

Sonuç: Parametrik olmayan testlerin uygun olmayan (misusing) ve gereksiz kullanımı (overusing) bilimsel çalışmalarda yanlış sonuçlar doğurmaktadır. Ölçek verilerinin 0-10 cm aralığında kaydedildiği çalışmalarda genel yaklaşım parametrik olmayan testlerin kullanılmasıyla beraber, yaptığımız simülasyon çalışması bu tip verilerin parametrik testlerle de değerlendirilebileceğini göstermektedir.

Anahtar kelimeler: Görsel Analog Skala, parametrik olmayan istatistikler, ağrı

Yazışma Adresi/Address for Correspondence: Dr. Didem Havva Ovla, Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin, Turkey. E-mail: didemovla@yahoo.com
Geliş tarihi/Received: 01.08.2016 Kabul tarihi/Accepted: 28.09.2016

GİRİŞ

Sıralı veri, mevcut istatistik analiz yöntemlerinin kullanımını için her zaman sorun olmuştur. Ne kategorik veri gibi değerlendirilebilir, ne de tam anlamıyla sürekli veridir. VAS (Visual Analogue Scale) ölçek verisi de Nöroloji başta olmak üzere Anesteziyoloji, Acil Tıp, Jinekoloji, psikiyatri, Fiziksel Tıp ve Rehabilitasyon gibi pek çok klinikte ağrı şiddetini değerlendirmek için sıkça kullanılan sıralı veriye tipik bir örnektir. VAS'ın dağılımının şekli kullanıldığı kliniğe göre değişir. Migren ağrısının şiddeti, acil baş ağrısının şiddeti ile aynı dağılıma sahip değildir. Dağılım şekli kullanılacak istatistiksel yöntemin seçimi için büyük önem arz etmektedir. Bu tip ölçekler genellikle likert tipindedir ve dar aralıklı olduğu için varyansların homojenliği varsayımı sağlanıyormuş gibi görünür. Oysaki dağılım şekli her bir grup için farklılık göstermektedir. Bu durumda bağımsız grupların karşılaştırılmasında parametrik olmayan yöntemler kullanılmaktadır.

Hastalardan ağrının (semptomun) şiddetini veya tedaviden memnuniyetlerini 0-10 santimetriklik (cm) bir ölçek üzerinde kendilerinin değerlendirmeleri istenir¹. Ancak ölçüm bir cetvel kullanılarak milimetre (mm) cinsinden 0-100 (mm) aralığında da değerlendirilebilir². Ağrı şiddetinin değerlendirilmesinde olması gerektiği gibi 0-100 (mm)'lik hassas ölçeğin kullanıldığı çalışmalar da mevcuttur. Ancak değerlendirmeler çoğunlukla 1 (cm)'lik aralıklara (0=Ağrı yok; 10=Dayanılmaz ağrı) yapılır. Bu durumda elde edilen veri sıralı veri olarak tanımlanabilir. Öyle ki; 9 her zaman 10'dan daha az ağrıyı tanımlarken 10 ve 9 arasındaki fark tam olarak açık değildir³.

Bununla birlikte ölçekler değerlendirilirken bir grupta diğerinden daha fazla aykırı değer bulunabilmektedir ve bu durum parametrik olmayan test sonuçlarının güvenilirliğini gölgelemektedir. Nitekim sonuçların grafik yöntem ile uyumsuzlukları da kaçınılmazdır. Doug Altman "Parametrik yöntemler kullanılmadan önce normal dağılım yaklaşımı her grubun kendi içinde yapılmasını, ham veriler bu koşulları sağlamazsa parametrik olmayan yöntemlerin kullanılmasını" önermiştir⁴. Parametrik yöntem yerine neden parametrik olmayan yöntem kullanıldığı açıkça belirtilmelidir. Oysa çok az çalışmada bu durum gerekçeleriyle verilir.

Bilindiği üzere, popülasyona ait olasılık dağılımı normal dağılıma uyumlu olduğu zaman, dağılımın normale yaklaştığını varsayan merkezi limit teoremini sağlayan en az örnek genişliğine sahip olduğunda, veri normal dağılıma uygun olmasa bile dönüşüm uygulanarak normale dönüştürülebildiğinde parametrik yöntemler kullanılabilir. Bunun yanı sıra, veri dönüşümle bile normal dağılıma yaklaşıyorsa, merkezi limit teoremini sağlayan en az örnek genişliğine bile sahip değilse, parametrik yöntemlerin uygulanmadığı bilinmeyen bir dağılımdansa, veri tipi sıralı veya nominal ise parametrik olmayan yöntemlerin kullanıldığı bilinir⁵. Parametrik testlerin varsayımları sağlandığı halde parametrik olmayan testler kullanılması durumunda testin gücünün düştüğü bilinir. Ancak VAS ölçeği (0-10 cm) gibi değişim aralığının küçük olduğu durumlarda örnek genişliğine bakılmaksızın, değerlendirme yapabilmek için geleneksel olarak parametrik olmayan yöntemlere başvurulur.

Bununla birlikte ölçekler değerlendirilirken bir grupta diğerinden daha fazla aykırı değer bulunabilmekte ve bu durum Mann-Whitney U test sonuçlarının güvenilirliğini gölgelemektedir. Üstelik verinin değişim aralığının dar olması da sonuçların güvenilirliğini olumsuz etkilemektedir. Nitekim sonuçların grafik yöntem ile uyumsuzlukları da kaçınılmazdır. Buradan yola çıkarak, bu çalışmada dar aralıklı (range) ve aynı zamanda aykırı değerlere sahip (çarpık) iki grubun karşılaştırılmasında Student t testi, Logit dönüşümü yapılmış Student t testi, Mann Whitney U testi ve Mood medyan testinin performansları değerlendirilmiştir.

GEREÇ VE YÖNTEM

Çalışma bir simülasyon çalışması olarak planlanmıştır. Pratikte çoğu kez, ağrı şiddeti 0 veya 10 olan hasta sayısı yok denecek kadar azdır. Bu doğrultuda, ağrı değerlendirmesinde en sık kullanılan VAS değerleri temel alınarak, 0-10 aralığında çarpık dağılımlar türetilmiştir.

Simülasyon adımları

İki grup karşıştırmalarında, ilgili hipotezleri test etmek için medyanları eşit olmak koşuluyla her grupta örnek genişliği n=5 olacak şekilde 0-10 aralığında;

Adım 1: İki de simetrik ($p=0.50$) binom dağılımına sahip veri üretilmiştir.

Adım 2: İki de pozitif çarpık ($p=0.25$) binom dağılımına sahip veri üretilmiştir.

Adım 3: İki de negatif çarpık ($p=0.75$) binom dağılımına sahip veri üretilmiştir.

Adım 4: Biri pozitif diğeri negatif çarpık dağılımlardan binom dağılımına sahip veri üretilmiştir.

Adım 5: Elde edilen verilerle grupların karşılaştırılması parametrik bağımsız iki grup t testi, parametrik olmayan Mann Whitney U testi ile Mood's medyan testi ve verilere logit dönüşüm yapıldıktan sonra parametrik bağımsız iki grup t testi ile değerlendirildi.

Adım 6: Karşılaştırmalar sonucu elde edilen Tip I hata oranları kaydedilmiştir. Bu oranlar $p \leq 0,05$ ve $p > 0,05$ olacak şekilde sayı ve yüzde olarak özetlendi.

Adım 7: Adım 1'e geri dönülerek işlemler $n=10$, $n=30$, $n=50$ ve $n=100$ için tekrarlanmıştır. Bu adımlar 20 kombinasyon için 1000'er kez tekrarlanmıştır.

Parametrik ve parametrik olmayan testler

Yapılan analiz sonuçlarında elde edilen Tip I hata oranları kaydedilmiştir. Tip I hata, gerçekte doğru olduğu halde elde edilen verilerle desteklenmeyen sonuçların oranıdır. Genel olarak parametrik test varsayımlarının başında örneklemin çekildiği popülasyonun normal dağılıma uyumlu olması gelir. Ayrıca karşılaştırılacak grupların varyanslarının homojen olması da önemli bir varsayımdır. Ancak varyansların homojenliği, normal dağılıma uyumlu olmayan gruplarda sağlansa bile tek başına yeterli bir varsayım değildir.

Parametrik test varsayımlarının sağlanmadığı durumlar için geliştirilmiş parametrik olmayan testler vardır. Parametrik olmayan testlerin kullanımı için örneklemin çekildiği popülasyonun dağılımıyla ilgili herhangi bir varsayım yoktur. Ancak $n > 30$ olduğu durumda popülasyonun dağılımı bilinmese bile, örneklem dağılımının normal dağılıma yakınsadığı bilinir ve parametrik testler kullanılabilir. Karşılaştırılmak istenen grup sayısı da burada önem kazanmaktadır. Parametrik test kullanımı için grup sayısı 2 olduğunda hem normallik hem de varyans homojenlik varsayımlarının aynı anda sağlanması beklenir. Oysa 2'den çok grup karşılaştırması

yapılmak istendiğinde bakılacak tek varsayım normalliğin sağlanmasıdır. Varyansların homojen olup olmaması, kullanılacak post-hoc seçimini etkilemektedir.

Hem parametrik hem de parametrik olmayan testler ilgili varsayımlar sağlandığı zaman güvenilir sonuçlar verir. Sonuçların güvenilirliği, karşılaştırılacak gruplardaki birim sayılarının birbirine eşit ya da yakın olmasıyla da artar. Çalışmamızda parametrik testlerden Student t testi, parametrik olmayan testlerden ise Mann Whitney U ve Mood's medyan testleri kullanılmıştır.

Student t testi

İlgilenilen bir değişken bakımından bağımsız 2 grup ortalamaları arasında fark olup olmadığı incelenir⁶.

Hipotez

H_0 : Grupların ortalamaları eşittir ($\mu_1 = \mu_2$)

H_1 : Grupların ortalamaları eşit değildir. ($\mu_1 \neq \mu_2$)

Logit dönüşüm

Simülasyon çalışması 0-10 aralığında VAS değerleri için yapıldığından en büyük VAS değeri 10 alınarak logit dönüşüm formülü,

$$y = \log\{(VAS)/(10 - VAS)\}$$

0-10 aralığında olan VAS ölçümlerini $-\infty$ ve $+\infty$ aralığına dönüştürür⁷.

Arcsin dönüşüm

VAS ölçümleri için alternatif bir dönüşüm de arcsin dönüşümüdür⁷.

$$\sin^{-1}(VAS/10)^{1/2}$$

formülü ile dönüşüm uygulanır.

Yapılan çalışmalarda arcsin dönüşümü yapılarak ve yapılmadan elde edilen sonuçların benzer olduğu görülmüştür⁸.

Mann Whitney U testi

Hipotez

H_0 : Grupların dağılımları aynıdır.

H_1 : Grupların dağılımları aynı değildir.

Aynı şekil ve eşit varyanslara sahip iki popülasyondan rasgele seçilmiş bağımsız örneklemelere veri tipi ordinal olduğu durumda

kullanılır. Örneklem dağılımlarını karşılaştırır⁹. Dağılımlar farklı olduğu durumda sonuçlar etkilenir.

Mood's Medyan testi

Hipotezler

H₀: Popülasyon medyanları eşittir. ($M_1=M_2$)

H₁: Popülasyon medyanları eşit değildir. ($M_1 \neq M_2$)

Yine aynı şekle ve eşit varyanslara sahip popülasyonlardan rasgele seçilmiş bağımsız örneklemeleri test eder. Ki-kare (χ^2) tabanlı bir istatistiktir.

Her bir örnekleme 5 ve daha fazla ölçüm olması gerekmektedir.

Örnek genişlikleri n_1 ve n_2 olan iki örnekleme birleştirilerek birleşik veri setinden birleşik (overall) medyan bulunur.

Bu değerden küçükler (-) ve büyüklerin (+) sayısı çıkartılıp hangilerinin hangi örnekleme ait olduğu tespit edilir. 2xk lık kontenjans tablosu elde edilir¹⁰.

Mann Whitney U testi pek çok farklı dağılımdan seçilmiş örneklem için Mood's medyan testinden daha güçlüdür. Ancak Mood's medyan testi uç değerlere karşı daha sağlamdır.

BULGULAR

Simetrik Dağılım

İlgili test sonuçları incelendiğinde Tip I hata oranları %3.5 ile %5.5 arasında değişmekte, genel olarak 3 testin de Tip I hata oranlarının benzer olduğu

görülmektedir (Tablo 1). Parametrik ve parametrik olmayan test sonuçları karşılaştırıldığında $n=5$ ve $n=10$ için Student t testinin Tip I hata oranlarının Mann Whitney U ve Mood's medyan testlerinden daha yüksek olduğu, $n=30$ olduğu durumda ise 3 testin de benzer sonuçlar verdiği görülmektedir

Örnek genişliğinin $n=50$ için en düşük Tip I hata oranı Mann Whitney U testinde, en yüksek ise Mood's medyan testinde görülmekteyken, $n=100$ olduğu durumda en düşük Tip I hata oranı Student t testinde, en yüksek ise Mann Whitney U testinde görülmektedir.

Negatif Çarpık Dağılım

Tüm testlerin sonuçları incelendiğinde anlamlılıklar %3.1 ile %6.5 arasında değişmektedir. Öncelikle parametrik ve parametrik olmayan test sonuçları karşılaştırıldığında Mann Whitney U test sonuçlarının $n=5$ haricinde Student t testi ile benzer olduğu görülmektedir. (Tablo 2).

Örnek genişliği $n=5$ durumunda ise değişim aralığının darlığı nedeniyle sapan değerlerin gözlenme olasılığının düşük olması ve Mann Whitney U testinin sıra değerlerine dayalı bir yöntem olması nedeniyle farklılık gözlenme olasılığı da düşmektedir. Ayrıca $n=5$ durumunda Mood's medyan testi için en yüksek anlamlılık yüzdesi elde edilmiştir. Bunun sebebi de Mood's medyan testinin Ki-Kare (χ^2) tabanlı bir test olması ve $n=5$ gözlem sayısının yeterli olmamasıdır. Örnek genişliği $n \geq 10$ olduğu durumlarda ise parametrik ve parametrik olmayan Tip I hata oranlarının birbirine çok yakın olduğu görülmektedir.

Tablo 1: Simetrik dağılım için tip I hata oranları

Simetrik Dağılım				Parametrik Test		Parametrik Olmayan Test			
				Student t Testi		Mann-Whitney U Testi		Mood's Medyan Testi	
		Ort±SS	Medyan	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05
n=5	Grup1	4.965±1.495	4.967	%5.5	%94.5	%3.5	%96.5	%5.3	%94.6
	Grup2	4.982±1.508	4.954						
n=10	Grup1	5.025±1.545	5.046	%5.4	%94.6	%4.5	%95.5	%4.1	%95.9
	Grup2	4.995±1.525	4.993						
n=30	Grup1	4.989±1.572	5.004	%5.0	%95.0	%4.7	%95.3	%5.1	%94.9
	Grup2	4.997±1.574	4.984						
n=50	Grup1	4.999±1.570	4.999	%4.0	%96.0	%3.8	%96.2	%5.4	%94.9
	Grup2	4.999±1.580	4.992						
n=100	Grup1	4.994±1.578	5.001	%4.1	%95.9	%4.9	%95.1	%4.4	%95.6
	Grup2	5.002±1.576	5.001						

(† : Ort±SS: Ortalama ± Standart Sapma)

Tablo 2: Negatif çarpık dağılım için tip I hata oranları

Negatif Çarpık Dağılım				Parametrik Test				Parametrik Olmayan Test			
				Student t Testi		Logistik Dönüşümlü t Testi		Mann-Whitney U Testi		Mood's Medyan Testi	
		Ort±SS	Medya n	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05
n=5	Grup1	7.502 ± 1.280	7.552	%4.8	%95.2	%4.9	%95.1	%3.1	%96.6	%6.5	%93.5
	Grup2	7.502 ± 1.320	7.572								
n=10	Grup1	7.482 ± 1.324	7.572	%4.4	%95.6	%5.3	%94.7	%4.1	%95.9	%4.4	%95.6
	Grup2	7.495 ± 1.331	7.567								
n=30	Grup1	7.505 ± 1.356	7.598	%3.5	%96.5	%4.4	%95.6	%3.5	%96.5	%4.9	%95.1
	Grup2	7.502 ± 1.365	7.619								
n=50	Grup1	7.494 ± 1.363	7.628	%4.8	%95.2	%5.3	%94.7	%4.7	%95.3	%5.8	%94.2
	Grup2	7.488 ± 1.367	7.617								
n=100	Grup1	7.495 ± 1.366	7.672	%4.3	%95.7	%3.8	%96.2	%4.5	%95.5	%5.5	%94.5
	Grup2	7.510 ± 1.363	7.724								

(† : Ort±SS: Ortalama ± Standart Sapma)

Parametrik test sonuçları kendi aralarında incelendiğinde ise dönüşümlü ve dönüşümsüz test sonuçlarının çok yakın oldukları görülmektedir.

Pozitif Çarpık Dağılım

Elde edilen sonuçlara göre anlamlılıklar %3.6 ile %6.1 arasında değişmektedir. En düşük ve en yüksek anlamlılıkların n=5 olduğu durumda ortaya çıktığı görülmektedir. Örnek genişliğinin n≤10 olduğu durumda Tip I hata oranı bakımından parametrik ve parametrik olmayan test sonuçlarının farklılaştığı görülmektedir. Parametrik test sonuçları kendi aralarında incelendiğinde ise dönüşümlü ve dönüşümsüz test sonuçlarının çok yakın oldukları görülmektedir (Tablo 3).

Mixed Dağılım

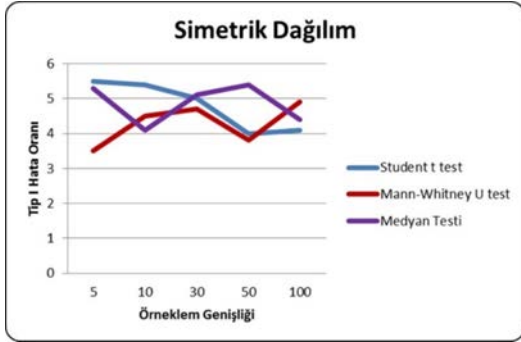
Medyanları eşit olan iki grubun dağılımlarından biri pozitif çarpık diğeri negatif çarpık olması

durumunda grupların karşılaştırmasına ait Tip I hata oranları Tablo 4'de verilmiştir.

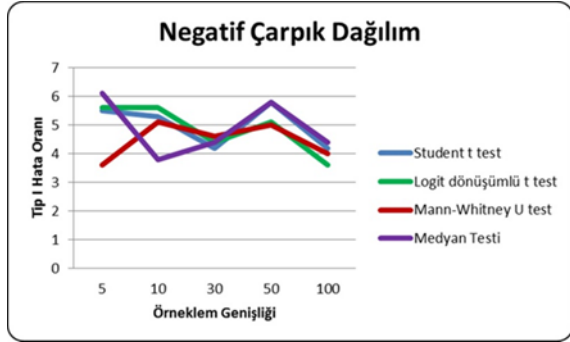
Öncelikle n=5 için parametrik ve parametrik olmayan test sonuçları incelendiğinde, dönüşüm uygulanmış test haricinde sonuçların benzer olduğu görülmektedir. Grupların dağılım şekli farklı olmakla beraber iç-içe geçmiş olmaları, değişim aralığının darlığı nedeniyle aralarında fark bulunma olasılığı çok düşüktür ($P_t=0.9$, $P_{MW}=0.2$, $P_M=0.5$).

Verilere dönüşüm uygulamak, ortalama etrafında yığılımı artırırken standart sapmayı da düşürmektedir. Bu nedenle de ortancaları benzer olan negatif ve pozitif çarpık iki dağılımın farklılaşmasına sebep olmaktadır ($p=0.555$).

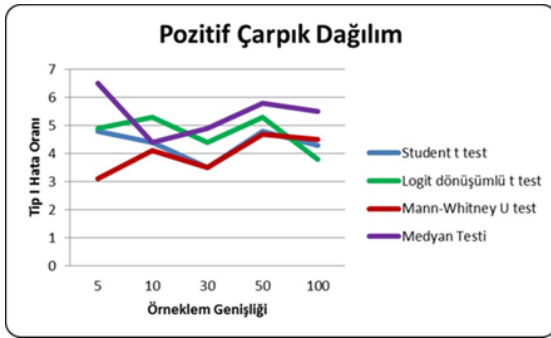
Örnek genişliği n≥10 için parametrik olmayan test sonuçları incelendiğinde ise örnek genişliği arttıkça aykırı değerlerin gözlenme olasılığı da arttığından, Tip I hata oranları da artarak, %78,3'e kadar çıkmaktadır. Bununla birlikte dönüşümlü ve dönüşümsüz test sonuçlarının benzer oldukları görülmektedir.



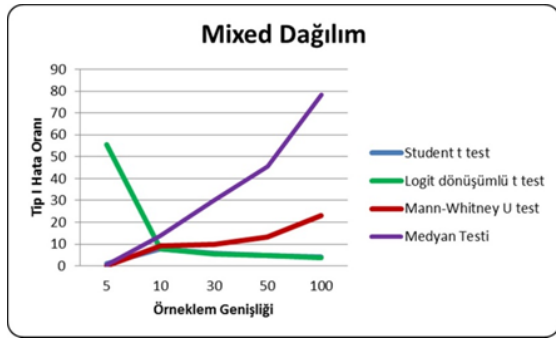
Şekil 1: Simetrik dağılım için sonuçların grafiksel gösterimi



Şekil 2: Negatif çarpık dağılım için sonuçların grafiksel gösterimi



Şekil 3: Pozitif çarpık dağılım için sonuçların grafiksel gösterimi



Şekil 4: Mixed dağılım için sonuçların grafiksel gösterimi

Tablo 3: Pozitif çarpık dağılım için tip I hata oranları

Pozitif Çarpık Dağılım				Parametrik Test				Parametrik Olmayan Test			
				Student t Testi		Logistik Dönüşümlü t Testi		Mann-Whitney U Testi		Mood's Medyan Testi	
	Grup	Ort±SS	Medyan	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05
n=5	Grup1	2.51±1.29	2.445	%5.5	%94.5	%5.6	%94.4	%3.6	%96.4	%6.1	%93.9
	Grup2	2.49±1.27	2.425								
n=10	Grup1	2.51±1.34	2.431	%5.3	%94.7	%5.6	%94.4	%5.1	%94.9	%3.8	%96.2
	Grup2	2.50±1.34	2.435								
n=30	Grup1	2.50±1.35	2.401	%4.2	%95.8	%4.4	%95.6	%4.6	%95.4	%4.4	%95.6
	Grup2	2.49±1.36	2.396								
n=50	Grup1	2.49±1.36	2.340	%5.8	%94.2	%5.1	%94.9	%5.0	%95.0	%5.8	%94.2
	Grup2	2.49±1.36	2.360								
n=100	Grup1	2.49±1.36	2.280	%4.2	%95.8	%3.6	%96.4	%4.0	%96.0	%4.4	%95.6
	Grup2	2.49±1.36	2,317								

(† : Ort±SS: Ortalama ± Standart Sapma)

Tablo 4: Mixed dağılım için tip I hata oranları

Mixed Dağılım				Parametrik Test				Parametrik Olmayan Test			
				Student t Testi		Logistik Dönüşümlü t Testi		Mann-Whitney U Testi		Mood's Medyan Testi	
		Ort±SS	Medyan	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05	p≤0.05	p>0.05
n=5	Grup1	4.485±0.646	4.666	%0.9	%99.1	%55.5	%44.5	%0.2	%99.8	%0.5	%99.5
	Grup2	4.506±0.643	4.307								
n=10	Grup1	4.519±0.629	4.733	%7.7	%92.3	%7.7	%92.3	%9.2	%90.8	%13.6	%86.4
	Grup2	4.489±0.636	4.277								
n=30	Grup1	4.503±0.661	4.847	%5.7	%94.3	%5.4	%94.6	%10.0	%90.0	%30.0	%70.0
	Grup2	4.502±0.664	4.147								
n=50	Grup1	4.499±0.667	4.919	%4.9	%95.1	%4.8	%95.2	%13.3	%86.7	%45.5	%54.5
	Grup2	4.499±0.664	4.102								
n=100	Grup1	4.506±0.666	4.977	%4.1	%95.9	%3.7	%96.3	%23.0	%77.0	%78.3	%21.7
	Grup2	4.499±0.668	4.028								

(† : Ort±SS: Ortalama ± Standart Sapma)

TARTIŞMA

Parametrik olmayan testlerin uygun olmayan (misusing) ve gereksiz kullanımı (overusing) bilimsel literatürde yanlış sonuçlar doğurmaktadır. VAS verilerinin 0-10 cm aralığında kaydedildiği çalışmalarda genel yaklaşım parametrik olmayan testlerin kullanılması olmakla beraber, yaptığımız simülasyon çalışması bu tip verilerin parametrik testlerle de değerlendirilebileceğini göstermektedir.

VAS aralıklı ve oransal ölçek özelliğinde olduğundan sayısal veri olarak da kabul edildiği çalışmalarla gösterilmiştir. Dexter ve Chestnus aynı veri setine hem parametrik (t test ve ANOVA) hem de parametrik olmayan (Mann Whitney U ve Kruskal Wallis) yöntemler uygulamışlar ve her iki yöntemin de benzer sonuçlar verdiğini görmüşlerdir¹¹. Örnek genişliği $n \geq 30$ olduğunda sonuçların benzerlik gösterdiği bizim çalışmamızda da gösterilmiştir.

Price ve arkadaşları VAS skorlarının ordinal özellikte olduğunu kabul eden ve bu yüzden parametrik yöntemlerin kullanılmayacağını belirten Kersten'e katılmadıklarını. VAS'ı oransal ölçek olarak kabul edip parametrik yöntemleri kullanılabileceğini belirtmişlerdir¹². Ayrıca, Bishop ve Herron parametrik testlerin, parametrik olmayan testlerden

daha güçlü olması sebebiyle VAS analizlerini parametrik testlerle değerlendirmeyi önermişlerdir¹³. Bu öneri, çalışmamızda elde edilen sonuçlarla da desteklenmektedir. Bishop ve Herron örnek genişliği küçük olduğunda 0-100 mm skalasının kullanımının daha doğru olduğunu ancak kesikli tipte (0-10 cm) skalanın da kullanılabileceğini belirtmişlerdir. Bizim çalışmamızda elde edilen sonuçlar da buna benzerlik göstermektedir. Bununla birlikte, örnek genişliği küçük olduğunda parametrik ve parametrik olmayan sonuçlar farklılık gösterdiği için 0-10 cm kullanımı yerine 0-100 mm kullanımının daha doğru olacağı söylenebilir.

Elde edilen sonuçların pratikte nasıl uygulandığını değerlendirmek amacıyla, VAS skorlarının sonuç değişken olarak kullanıldığı araştırmalardaki uygun olmayan istatistiksel yöntem kullanımını da incelediğimiz bu çalışmada, 100 VAS çalışmasından 26'sında normallik kontrolü yapıldığı tespit edilmiştir. Normallik kontrolü yapılan çalışmaların 7'sinde parametrik olmayan yöntemler kullanılmış ve bunların 2'sinde VAS'ın 100'lük ölçekte kullanıldığı görülmüştür. Normallik kontrolü yapılmayan 74 çalışmanın 21'inde parametrik olmayan yöntemler kullanılmış ve bunların 2'sinde VAS 100'lük ölçekte kullanılmıştır. Toplam 100 çalışmanın 28'inde parametrik olmayan yöntemlerin kullanıldığı tespit

edilmiştir. Buradan, VAS skorları için parametrik olmayan yöntemlerin oldukça yaygın ve gereksiz kullanıldığı sonucuna varılmaktadır.

KAYNAKLAR

1. Day S. Dictionary for Clinical Trials. 2nd ed. New York, Wiley, 2007.
2. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS) and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res. 2011;63:240-52.
3. Sullivan L. Nonparametric tests. Available from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/index.html. /Accessed on 1.8.2016)
4. Vickers AJ. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. BMC Med Res Methodol 2005;5:35.
5. Cardone R. Nonparametric: distribution-free. not assumption-free. Available from <http://www.isixsigma.com/tools-templates/hypothesis-testing/nonparametric-distribution-free-not-assumption-free/>. (Accessed on 1.8.2016)
6. Alpar R. Uygulamalı İstatistik ve Geçerlik-Güvenirlik. Ankara, Detay Yayıncılık, 2012.
7. Senn S. Cross Over Trials in Clinical Research. 2nd ed. London, Wiley, 2002.
8. Dexter F, Chestnut DH. Analysis of statistical tests to compare visual analog scale measurements among groups. Anesthesiology. 1995;82:896-902.
9. Kan İ. Biyoistatistik. Ankara, Nobel Yayın, 2006.
10. Mood's median test. Available from <http://www2.hawaii.edu/~taylor/z631/moods.pdf>. (Accessed on 1.8.2016)
11. Dexter F, Chestnut DH. Analysis of statistical tests to compare visual analog scale measurements among groups. Anesthesiology 1995;82:896-902.
12. Price DD, Staud R, Robinson ME. How should we use the visual analogue scale (VAS) in rehabilitation outcomes? II: Visual analogue scales as ratio scales: an alternative to the view of Kersten et al. J Rehabil Med. 2012;44:800-4.
13. Bishop PA, Herron RL. Use and misuse of the likert item responses and other ordinal measures. Int J Exerc Sci. 2015;8:297-302.