

Parametrik Olmayan Bağımsız İki Örneklem Testlerinde Örneklem Büyüklüğünün Beklenen Değer, Varyans ve Fraser Etkinliği Üzerine Etkileri

The Effects of Sample Size on Expected Value, Variance and Fraser Efficiency for Nonparametric Independent Two Sample Tests

İsmet DOĞAN,^a
Nurhan DOĞAN^a

^aBiyoistatistik ve Tıbbi Bilişim AD,
Afyon Kocatepe Üniversitesi
Tıp Fakültesi, Afyonkarahisar

Geliş Tarihi/Received: 22.01.2015
Kabul Tarihi/Accepted: 16.03.2015

Yazışma Adresi/Correspondence:
Nurhan DOĞAN
Afyon Kocatepe Üniversitesi
Tıp Fakültesi,
Biyoistatistik ve Tıbbi Bilişim AD,
Afyonkarahisar,
TÜRKİYE/TURKEY
nurhandogan@hotmail.com

ÖZET Amaç: İstatistikte temel problemlerden biri en etkin istatistiksel testi seçmektir. Asimtotik görelilik, aynı hipotezin test edilmesinde kullanılan iki farklı testin karşılaştırılmasına olanak sağlayan temel bir kavramdır. Testlere ait asimtotik etkinlik kavramı, tahminlerin asimtotik etkinliğinden daha karmaşıktır. Çalışmanın amacı dikkate alınan testlerde farklı örneklem büyüklüklerinin testlere ait beklenen değer ile varyans üzerindeki etkilerini incelemek ve Fraser Etkinlik değerini kullanarak farklı örneklem büyüklükleri için en etkin testi belirlemektir. **Gereç ve Yöntemler:** Testlerin karşılaştırılmasında güç değerinin hesaplanması çoğu zaman pratik olmadığından asimtotik görelilik değerlerinin kullanılması öne çıkmaktadır. Bu çalışmada, bağımsız iki grubun karşılaştırılmasında sıklıkla kullanılan ve beklenen değer ile varyans hesaplanmasında sadece örneklem büyüklüğünün etkili olduğu parametrik olmayan testler dikkate alınmıştır. Asimtotik görelilik, sıra sayıları testleri gibi çok sayıda sezgisel test içeren parametrik olmayan istatistikler için özellikle yarar sağlamaktadır. Bu çalışmada, $2 \leq n \leq 50$ örneklem büyüklüğü olarak belirlenmiştir. **Bulgular:** Gerek dengeli gerekse dengersiz durumlar için örneklem büyüklüğü arttıkça çalışmada dikkate alınan tüm testlerin beklenen değer ve varyans değerlerinin de arttığı tespit edilmiştir. Ayrıca, Fraser etkinliğine göre Mann-Whitney U testi, gerek dengeli gerekse dengersiz durumlarda hemen her örneklem büyüklüğü için bağımsız iki grubun karşılaştırılmasında kullanılan parametrik olmayan testler içerisinde en etkin test olarak belirlenmiştir. **Sonuç:** Fraser etkinliğine göre Mann-Whitney U testi, her örneklem büyüklüğü için bağımsız iki grubun karşılaştırılmasında kullanılan parametrik olmayan testler içerisinde en etkin testtir.

Anahtar Kelimeler: İstatistik, parametrik olmayan; örnek büyüklüğü; etkinlik

ABSTRACT Objective: Choosing the most efficient statistical test is one of the essential problems of statistics. Asymptotic relative efficiency is a notion which enables to implement in large samples the quantitative comparison of two different tests used for testing of the same statistical hypothesis. The notion of the asymptotic efficiency of tests is more complicated than that of asymptotic efficiency of estimates. This paper discusses the effect of sample size on expected values and variances of non-parametric tests for independent two samples and determines the most effective test for different sample sizes using Fraser efficiency value. **Material and Methods:** Since calculating the power value in comparison of the tests is not practical most of the time, using the asymptotic relative efficiency value is favorable. Asymptotic relative efficiency is an indispensable technique for comparing and ordering statistical test in large samples. It is especially useful in nonparametric statistics where there exist numerous heuristic tests such as the linear rank tests. In this study, the sample size is determined as $2 \leq n \leq 50$. **Results:** In both balanced and unbalanced cases, it is found that, as the sample size increases expected values and variances of all the tests discussed in this paper increase as well. Additionally, considering the Fraser efficiency, Mann-Whitney U test is found as the most efficient test among the non-parametric tests that are used in comparison of independent two samples regardless of their sizes. **Conclusion:** According to Fraser efficiency, Mann-Whitney U test is found as the most efficient test.

Key Words: Statistics, nonparametric; sample size; efficiency

doi: 10.5336/biostatic.2015-43724

Copyright © 2015 by Türkiye Klinikleri

Türkiye Klinikleri J Biostat 2015;7(2):103-7

Kullanılabilecek testler arasından en verimli olan istatistik testin istatistikçi tarafından seçilmesi, istatistiğin temel sorunlarından biridir. Klasik Neyman-Pearson teorisine göre en güçlü test en iyi olarak kabul edilmektedir. Ancak, bu tür güçlü testlerin teori ve pratikte ortaya çıkan sorunların çeşitliliğinden dolayı istatistiksel modellerin tamamını kapsamayan sınırlı sayıda mevcut olduğu bilinmektedir.¹

Bir istatistiksel analizin sonucunun uygunluğu, istatistiksel yöntemin seçiminin doğruluğuna bağlıdır. Dolayısıyla öncelikle parametrik ya da parametrik olmayan yöntemlerden hangisinin kullanılacağına karar verilmelidir. Çünkü istatistiksel analiz yöntemleri, bu yöntemlere ait varsayımların bozulması durumunda geçersiz olmaktadır. Bir araştırmadan elde edilen gözlem değerleri ile ilgili analiz yapılırken öncelikle gözlem değerlerinin çarpıklık, aykırı değer varlığı gibi sebeplerden dolayı normallikten uzaklaşıp uzaklaşmadığı incelenmelidir. Eğer gözlem değerleri normal dağılıma sahipse ve diğer varsayımlar da sağlanıyorsa parametrik testler parametrik olmayan testlere göre oldukça güçlüdür. Gözlem değerleri normal dağılmıyor ancak diğer kriterler sağlanıyorsa parametrik olmayan testlere ait sonuçlar daha güvenilir ve geçerli olmaktadır. Normallik varsayımı ile birlikte diğer varsayımların da sağlanmaması durumunda hem parametrik hem de parametrik olmayan testler birlikte uygulanmalı elde edilen sonuçlara göre karar verilmelidir. Bununla birlikte gözlem değerlerine karşı parametrik testlerin mi yoksa parametrik olmayan testlerin mi daha sağlam (robust) oldukları da belirlenmelidir.

Parametrik olmayan yöntemler arasında aynı amaca hizmet eden çok sayıda test bulunmaktadır. Parametrik ya da parametrik olmayan yöntemlere karar vermekten ziyade asıl zor olan aynı amaca hizmet eden testler arasında hangisinin kullanılması gerektiğine karar vermektir. Bu durum ile özellikle parametrik olmayan testlerde daha fazla karşılaşılmaktadır. Ne yazık ki aynı grupta yer alan benzer yöntemlerden hangisinin kullanılması gerektiğine karar vermek her zaman çok kolay değildir. Bu tür durumlarda sıklıkla ya testlerin gücü hesaplanmakta ya da etkinliklerine bakılmaktadır.

Gerek güç gerekse etkinlik kavramları ise doğrudan doğruya gözlem sayısı ile ilişkili kavramlardır.

Test yöntemlerinin birbirleri ile karşılaştırılmasında sıklıkla kullanılan etkinlik yöntemleri, Pitman Etkinliği (Asimtotik Göreli Etkinlik), Bahadur Etkinliği, Hodges-Lehmann Etkinliği ve Chernoff Etkinliği olarak ifade edilmektedir.¹ Literatürde çok fazla bahsedilmemesine rağmen etkinlik yöntemleri arasında hesaplanması en kolay olan ise Fraser Etkinliği'dir. A ve B gibi iki farklı test dikkate alındığında A testinin B testine göre Fraser Etkinliği, μ_t ve σ_t^2 sırasıyla ilgili teste ait ortalama ve varyansı göstermek üzere,

$$r(t) = \frac{\left[\frac{\partial}{\partial \xi} E\{t(X)\} \Big|_{\xi=0} \right]^2}{\text{var}\{t(X)\} \Big|_{\xi=0}} = \frac{\mu_t^2}{\sigma_t^2}$$

$$E(t_A, t_B) = \lim_{n \rightarrow \infty} \frac{r(t_A)}{r(t_B)}$$

şeklinde ifade edilmektedir.^{2,3} $0 < E(t_A, t_B) < 1$ olması A testinin B testine göre etkin olduğunu, $E(t_A, t_B) > 1$ olması ise B testinin A testine göre etkin olduğunu göstermektedir. Pitman, rastgelelik koşulu altında, bir testin başka bir teste göre asimptotik göreli etkinliğinin bu testlerin etkinliklerinin oranının limitine eşit olduğunu göstermiş, Gibbons ve Chakraborti tarafından da bu durum ispatlanmıştır.^{4,5}

Normallik varsayımının sağlanması durumunda parametrik testlerin hem güç hem de etkinlik bakımından parametrik olmayan testlere üstün olduğu bilinmektedir. Ancak aynı amaca hizmet eden parametrik olmayan testlerden hangisinin hangi durumda kullanılması gerektiği ile ilgili karşılaştırmalı bir çalışma bulunmamaktadır. Bu çalışmanın amacı, dengeli ve dengesiz durumlar için örneklem büyüklüğünün, çalışmada dikkate alınan testlerin beklenen değer ile varyansları üzerindeki etkilerini incelemek ve Fraser etkinlik değerini kullanarak bu testleri karşılaştırmaktır.

GEREÇ VE YÖNTEMLER

İlgilenilen hipoteze bağlı olarak farklı parametrik olmayan testler söz konusudur. Bu çalışmada, ba-

TABLO 1: Çalışmada dikkate alınan testlere ait beklenen değer ve varyans formülleri.

Test	Beklenen Değer	Varyans
Wilcoxon Sıra Sayıları Toplamı (Rank Sum) Testi	$\frac{n^*(n+1)}{2}$	$\frac{n_A n_B (n+1)}{12}$
Mann-Whitney U Testi	$\frac{n_A n_B}{2}$	$\frac{n_A n_B (n+1)}{12}$
Ansari-Bradley ve David-Barton Testi	$n_A + n_B = n$ çift ise,	
	$\frac{n^*(n+2)}{4}$	$\frac{n_A n_B (n^2-4)}{48(n-1)}$
	$n_A + n_B = n$ tek ise,	
	$\frac{n^*(n+1)^2}{4n}$	$\frac{n_A n_B (n+1)(n^2+3)}{48n^2}$
Mood Testi	$\frac{n^*(n^2-1)}{12}$	$\frac{n_A n_B (n+1)(n^2-4)}{180}$
Savage Testi	n^*	$\frac{n_A n_B}{n-1} \left[1 - \frac{1}{n} \left(\sum_{i=1}^n \frac{1}{i} \right) \right]$
Sukhatme Testi	$n_A n_B$	$\frac{n_A n_B (n+7)}{48}$
Wald-Wolfowitz Testi	$\frac{2n_A n_B}{n} + 1$	$\frac{2n_A n_B (2n_A n_B - n)}{n^2(n-1)}$

$n = n_A + n_B$, n_A : A grubundaki gözlem sayısı, n_B : B grubundaki gözlem sayısı.

n^* : Test istatistiğinin değerinin hesaplanmasında dikkate alınan gruba ait gözlem sayısı.

ğimsız iki grubun karşılaştırılmasında sıklıkla kullanılan ve beklenen değer ile varyans hesaplanmasında sadece örneklem büyüklüğünün etkili olduğu sekiz farklı parametrik olmayan test dikkate alınmıştır. Çalışmada dikkate alınan testlere ait beklenen değer ve varyans formülleri toplu olarak Tablo 1'de verilmiştir.^{6,7}

BULGULAR

Dikkate alınan testlerde farklı örneklem büyüklüklerinin testlere ait beklenen değer ve varyanslar üzerindeki etkileri incelenmiştir. Dengeli durumlar için elde edilen sonuçlar Tablo 2'de verilmiştir.

Dengesiz durumlar için elde edilen sonuçlar çok sayıda olmalarından dolayı burada verilmemiştir.

Fraser etkinlik değeri çalışmada dikkate alınan tüm testler için her bir örneklem büyüklüğüne göre ayrı ayrı hesaplanmıştır. Dengeli durumlar için

elde edilen Fraser etkinliği sonuçları Tablo 3'de, dengesiz durumlar için elde edilen Fraser etkinliği sonuçları ise Tablo 4'de verilmiştir (Tablo 3, Tablo 4).

SONUÇ

İstatistiksel testler birçok durumda, belirli bir örneklem büyüklüğü göz önüne alındığında hangi olasılıkla etkileri tespit ettiği veya kullanılan testten elde edilen sonuçların doğruluğuna ilişkin açıklayıcı bir değerlendirme yapılmaksızın kullanılmaktadır. İstatistikte aynı hipotezin test edilmesinde birkaç farklı testin kullanılabilirdiği durumlar ile sıklıkla karşılaşmaktadır. Bu tür durumlarda performansı daha iyi olan testin belirlenmesi ve kullanılması gerekir. Bunun en doğal yolu ise alternatif testlerin güçlerinin incelenmesidir. Ancak testlere ait gerçek güç değerlerinin belirlenmesi oldukça karmaşık hesaplamalar gerektirmektedir. Gerçek güç değerlerinin hesaplanmasındaki güçlük

TABLO 2: Dengeli durumlar için beklenen değer ve varyanslar.

	Wilcoxon		Mann Whitney		Mood		Savage		AB ve DB		Sukhatme		WW	
	$E(x)$	$V(x)$	$E(x)$	$V(x)$	$E(x)$	$V(x)$	$E(x)$	$V(x)$	$E(x)$	$V(x)$	$E(x)$	$V(x)$	$E(x)$	$V(x)$
$n_A = 2$ $n_B = 2$	5,0	1,7	2,0	1,7	2,5	1,3	2,0	0,8	3,0	0,3	4,0	0,9	3,0	0,7
$n_A = 3$ $n_B = 3$	10,5	5,3	4,5	5,3	8,8	11,2	3,0	1,3	6,0	1,2	9,0	2,4	4,0	1,2
$n_A = 4$ $n_B = 4$	18,0	12,0	8,0	12,0	21,0	48,0	4,0	1,7	10,0	2,9	16,0	5,0	5,0	1,7
$n_A = 5$ $n_B = 5$	27,5	22,9	12,5	22,9	41,3	146,7	5,0	2,1	15,0	5,6	25,0	8,9	6,0	2,2
$n_A = 6$ $n_B = 6$	39,0	39,0	18,0	39,0	71,5	364,0	6,0	2,6	21,0	9,5	36,0	14,3	7,0	2,7
$n_A = 7$ $n_B = 7$	52,5	61,3	24,5	61,3	113,8	784,0	7,0	3,1	28,0	15,1	49,0	21,4	8,0	3,2
$n_A = 8$ $n_B = 8$	68,0	90,7	32,0	90,7	170,0	1523,2	8,0	3,5	36,0	22,4	64,0	30,7	9,0	3,7
$n_A = 9$ $n_B = 9$	85,5	128,3	40,5	128,3	242,3	2736,0	9,0	4,0	45,0	31,8	81,0	42,2	10,0	4,2
$n_A = 10$ $n_B = 10$	105,0	175,0	50,0	175,0	332,5	4620,0	10,0	4,5	55,0	43,4	100,0	56,3	11,0	4,7
$n_A = 20$ $n_B = 20$	410,0	1366,7	200,0	1366,7	2665,0	145413,3	20,0	9,3	210,0	341,0	400,0	391,7	21,0	9,7
$n_A = 30$ $n_B = 30$	915,0	4575,0	450,0	4575,0	8997,5	1096780,0	30,0	14,2	465,0	1142,8	900,0	1256,3	31,0	14,7
$n_A = 40$ $n_B = 40$	1620,0	10800,0	800,0	10800,0	21330,0	4605120,0	40,0	19,2	820,0	2698,7	1600,0	2900,0	41,0	19,7
$n_A = 50$ $n_B = 50$	2525,0	21041,7	1250,0	21041,7	41662,5	14022166,7	50,0	24,1	1275,0	5258,8	2500,0	5572,9	51,0	24,7

AB: Ansari-Bradley; DB: David Barton; WW: Wald-Wolfowitz.

TABLO 3: Dengeli durumlar için fraser etkinliği bakımından yöntemlerin karşılaştırılması.

	Wilcoxon	Mood	Savage	Ansari-Bradley David-Barton	Sukhatme	Wald-Wolfowitz
Mann-Whitney	Mann-Whitney	Mann-Whitney	Mann-Whitney	Mann-Whitney	Mann-Whitney	Mann-Whitney
Wilcoxon		Mood	Savage	Wilcoxon	Sukhatme	Wald-Wolfowitz
Mood			$n \leq 5$ için Mood $n > 5$ için Savage	Mood	Mood	$n \leq 13$ için Mood $n > 13$ için Wald-Wolfowitz
Savage				Savage	Savage	Savage
Ansari-Bradley					Ansari-Bradley	
David-Barton					David-Barton	Wald-Wolfowitz
Sukhatme						Sukhatme

ve karmaşıklıktan dolayı testlerin karşılaştırılmasında güç çoğu zaman pratik olmamakta, görece etkinlik değerlerinin kullanılması öne çıkmaktadır.

Bu çalışmada, Fraser etkinliğine göre Mann-Whitney U testi, gerek dengeli gerekse dengesiz durumlarda hemen her örneklem büyüklüğü için

TABLO 4: Dengesiz durumlar için fraser etkinliği bakımından yöntemlerin karşılaştırılması.

	Ansari-Bradley					
	Wilcoxon	Mood	Savage	David-Barton	Sukhatme	Wald-Wolfowitz
Mann-Whitney	Mann-Whitney	Açıklama 1	Açıklama 2	Mann-Whitney	Mann-Whitney	Mann-Whitney
Wilcoxon		Mood	Savage	Açıklama 3	Açıklama 4	Açıklama 5
Mood			Savage	Mood	Mood	Açıklama 6
Savage				Savage	Açıklama 7	Açıklama 8
Ansari-Bradley						
David-Barton					Açıklama 9	Açıklama 10
Sukhatme						Wald-Wolfowitz

Açıklama 1: Gözlem sayısı büyük olan grup dikkate alındığında Mann-Whitney U testi, tersi durumda ise Mood testi daha etkindir.

Açıklama 2: Gözlem sayısı büyük olan grup dikkate alındığında Mann-Whitney U testi, tersi durumda ise Savage testi daha etkindir.

Açıklama 3: $n_A \leq 20$ veya $n_B \leq 20$ olması durumunda Wilcoxon testi, diğer durumlarda ise Ansari-Bradley ve David-Bartontestleri daha etkindir.

Açıklama 4: $n_A \geq 10$ veya $n_B \geq 10$ olması durumunda Wilcoxon testi, diğer durumlarda ise Sukhatme testi daha etkindir.

Açıklama 5: Gözlem sayısı büyük olan grup dikkate alındığında Wald-Wolfowitz testi, tersi durumda ise Wilcoxon testi daha etkindir.

Açıklama 6: Gruplardaki gözlem sayılarının 15'den büyük olduğu durumlarda Wald-Wolfowitz testi, gruplardaki gözlem sayılarının 15'den az olduğu durumlarda ise, yapılacak hesaplamalarda gözlem sayısı büyük olan grup dikkate alındığında Wald-Wolfowitz testi, tersi durumda ise Mood testi daha etkindir.

Açıklama 7: Gruplardaki gözlem sayılarının 10'dan büyük olduğu durumlarda Savage testi, gruplardaki gözlem sayılarının 10'dan az olduğu durumlarda ise, yapılacak hesaplamalarda gözlem sayısı büyük olan grup dikkate alındığında Sukhatme testi, tersi durumda ise Savage testi daha etkindir.

Açıklama 8: Gruplardaki gözlem sayılarının 10'dan büyük olduğu durumlarda Wald-Wolfowitz testi, gruplardaki gözlem sayılarının 10'dan az olduğu durumlarda ise, yapılacak hesaplamalarda gözlem sayısı büyük olan grup dikkate alındığında Wald-Wolfowitz testi, tersi durumda ise Savage testi daha etkindir.

Açıklama 9: Gruplardaki gözlem sayılarının 10'dan büyük olduğu durumlarda Ansari-Bradley ve David-Barton testleri, gruplardaki gözlem sayılarının 10'dan az olduğu durumlarda ise, yapılacak hesaplamalarda gözlem sayısı büyük olan grup dikkate alındığında Sukhatme testi, tersi durumda ise Ansari-Bradley ve David-Barton testleri daha etkindir.

Açıklama 10: Gruplardaki gözlem sayılarının 10'dan büyük olduğu durumlarda Wald-Wolfowitz testi, gruplardaki gözlem sayılarının 10'dan az olduğu durumlarda ise, yapılacak hesaplamalarda gözlem sayısı büyük olan grup dikkate alındığında Ansari-Bradley ve David-Barton testleri, tersi durumda ise Wald-Wolfowitz testi daha etkindir.

bağımsız iki grubun karşılaştırılmasında kullanılan parametrik olmayan testler içerisinde en etkin test olarak belirlenmiştir.

Ayrıca, gerek dengeli gerekse dengesiz du-

rumlar için örneklem büyüklüğü arttıkça çalışmada dikkate alınan tüm testlerin beklenen değer ve varyans değerlerinin de arttığı tespit edilmiştir.

KAYNAKLAR

1. Nikitin Y. Asymptotic efficiency of statistical tests and mathematical means for its computation. *Asymptotic Efficiency of Nonparametric Tests*, 1st ed. New York: Cambridge University Press; 1995. p. 1-36.
2. Fraser DAS. Large sample properties of tests. *Nonparametric Methods in Statistics*. 1st ed. New York: John Wiley & Sons Inc.; 1957. p. 266-93.
3. Sethuraman J, Rao JS. Pitman efficiencies of tests based on spacings, In: Puri ML, ed. *Nonparametric Techniques in Statistical Inference*, Cambridge: Cambridge University Press; 1970. p. 405-15.
4. Clarke KA. A simple distribution-free test for nonnested hypotheses. *Political Analysis* 2007;15(3):347-63.
5. Gibbons JD, Chakraborti S. Asymptotic relative efficiency. *Nonparametric Statistical Inference*. 4th ed. New York: Marcel Dekker Inc.; 2003. p. 494-503.
6. Doğan İ, Doğan N. [Nonparametric statistical methods with step by step solution]. *Bağımsız İki Örneklem Testleri*. 1. Baskı. Ankara: Detay Yayıncılık; 2014. p. 9-66.
7. Conover WJ. Some methods based on ranks, two independent samples. *Practical Nonparametric Statistics*, 2nd ed. New York: John Wiley & Sons Inc.; 1980. p. 215-229.