

Küçük Örneklem Büyüklüğüne Sahip 2x2 Çapraz Tablolar İçin Ki-Kare Yöntemlerinin Karşılaştırılması: Bir Simülasyon Çalışması

Comparison of Chi-Square Methods for 2x2 Crosstabs with Small Sample Size: A Simulation Study

İsmet DOĞAN^a, Nurhan DOĞAN^a

^aAfyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, Afyonkarahisar, Türkiye

ÖZET Amaç: Bu çalışmanın amacı, 2x2 çapraz tablolarda kullanılan klasik ve süreklilik düzeltilmeli ki-kare testlerini tanıtmak ve bunları karşılaştırmaktır. **Gereç ve Yöntemler:** Çalışmada serbestlik derecesi 1 (bir) olan ki-kare testleri dikkate alınmıştır. Çünkü bu testler, verilerin kesikli olmasından ciddi şekilde etkilenmektedir. Python-random kütüphanesi kullanılarak $10 \leq n \leq 25$ aralığında yer alan 4 farklı n değeri için veri türetilmiştir. Verilerin türetilmesinde önce a, b, c ve d ile gösterilen gözelerden hangisine değer atanacağı sonra da ilgili göze atanacak değer belirlenmiştir. $n=10$ için 246, $n=15$ için 756, $n=20$ için 958, $n=25$ için 963 farklı veri seti çalışmada kullanılmıştır. Yöntemlerin karşılaştırılmasında hem her bir yöntemin hipotetik H_0 hipotezini farklı örneklem büyüklüğü ve önemlilik düzeyleri için reddetme yüzdeleri hem de yöntemlerin birbirlerine göre hipotetik H_0 hipotezini *ret/ret*, *ret/kabul*, *kabul/ret* ve *kabul/kabul* oranları kullanılmıştır. **Bulgular:** Çalışmada dikkate alınan yöntemlere ait sonuçlar, yöntemlerden birinin önerilen tüm yöntemler içerisinde en iyi yöntem olarak seçilmesini sağlamamaktadır. Farklı yöntemler farklı örneklem büyüklüklerinde ve önemlilik düzeylerinde öne çıkmaktadır. Bu durum, bir araştırmadan elde edilecek sonucunun doğru şekilde yorumlanamayacağı anlamına gelir. Tüm yöntemlerin örneklem büyüklüğü ve önemlilik düzeyinden etkilendiği, örneklem büyüklüğünün artması ve önemlilik düzeyinin de 0,01'den 0,10'a doğru değişmesi durumunda H_0 hipotezinin reddedilme oranlarının da arttığı belirlenmiştir. **Sonuç:** Ki-kare testinin büyük örneklem için uygun olduğu, beklenen değerlerden en az birinin 5'ten küçük olması durumunda gerek klasik gerekse süreklilik düzeltilmeli ki-kare yöntemlerinin kullanılmaması gerektiği sonucuna ulaşılmıştır.

ABSTRACT Objective: The aim of this study is to introduce and compare the classical and continuity corrected chi-square tests used in 2x2 contingency tables. **Material and Methods:** Chi-square tests with 1 (one) degree of freedom were taken into consideration in the study. Because these tests are seriously affected by the discontinuity of the data. Using the Python-random library, data was derived for 4 different values in the range of $10 \leq n \leq 25$. In deriving the data, first the value to be assigned to the cells indicated by a , b , c and d was determined, and then the value to be assigned to the relevant cell was determined. 246 different data sets for $n=10$, 756 for $n=15$, 958 for $n=20$, and 963 for $n=25$ were used in the study. In the comparison of the methods, both the rejection percentages of each method for the hypothetical H_0 hypothesis for different sample sizes and significance levels, and the *rejection/rejection*, *rejection/acceptance*, *acceptance/rejection* and *acceptance/acceptance* rates of the hypothetical H_0 hypothesis of the methods relative to each other were used. **Results:** The results of the methods considered in the study do not enable one of the methods to be chosen as the best method among all the proposed methods. Different methods stand out at different sample sizes and significance levels. This means that the result of a study cannot be interpreted correctly. It has been determined that all methods are affected by the sample size and significance level, and if the sample size increases and the significance level changes from 0.01 to 0.10, the rejection rates of the H_0 hypothesis also increase. **Conclusion:** It was concluded that the chi-square test is suitable for large samples, and if at least one of the expected values is less than 5, both the classical and continuity corrected chi-square methods should not be used.

Anahtar kelimeler: Hipotez testi; ki-kare testi; 2x2 çapraz tablo; süreklilik düzeltmesi

Keywords: Hypothesis test; chi-square test; 2x2 contingency table; continuity correction

KAYNAK GÖSTERMEK İÇİN:

Doğan I, Doğan N. Küçük örneklem büyüklüğüne sahip 2x2 çapraz tablolar için ki-kare yöntemlerinin karşılaştırılması: Bir simülasyon çalışması. Türkiye Klinikleri J Foren Sci Leg Med. 2024;16(2):71-83.

Correspondence: Nurhan DOĞAN

Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, Afyonkarahisar, Türkiye

E-mail: nurhandogan@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

Received: 18 Dec 2023

Received in revised form: 13 Mar 2024

Accepted: 13 Mar 2024

Available online: 14 Jun 2024

2146-8877 / Copyright © 2024 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Ki-kare (χ^2) testi, biyoloji, tıp (genellikle klinik araştırmalarda) ve sosyal bilimler alanındaki araştırmacılar tarafından yaygın olarak kullanılan parametrik olmayan bir testtir. Ki-kare yalnızca gözlenen farklılıkların önemi hakkında bilgi vermekle kalmaz, aynı zamanda bulunan farklılıkların tam olarak hangi kategori ya da kategorilerden kaynaklandığı konusunda da ayrıntılı bilgi sağlar. Dolayısıyla, bu istatistiğin sağlayabileceği bilgi miktarı ve ayrıntısı, onu en kullanışlı araçlardan biri hâline getirir.¹ Nominal ölçekte bir veri örneğinin elde edilmesi ve bu verinin geldiği popülasyonun özel bir teorik dağılıma uyup uymadığı çıkarımının yapılması ya da bir değişkenin çeşitli kategorilerindeki ortaya çıkma sıklıklarının, ikinci değişkendeki sıklıklardan bağımsız olduğu hipotezinin test edilmesi sıklıkla istenmektedir. Her iki durum için de en yaygın prosedür ki-kare testinin kullanılmasıdır. Bağımsızlık testi olarak kullanıldığı durumlarda çapraz tablo adı verilen tabloların oluşturulması gerekir. Olabilecek en küçük çapraz tablo 2 satır ve 2 sütundan oluşan tablodur. Bu tablolar genel olarak 2x2 çapraz tablo veya olumsuzluk tablosu ya da 4 hücreli tablo olarak ifade edilmektedir.² 2x2 tablo örneği [Tablo 1](#)'de verilmiştir.

TABLO 1: 2x2 çapraz tablo.

Satır değişkeni	Sütun değişkeni		Satır toplamı
	Durum 1	Durum 2	
Durum 1	<i>a</i>	<i>c</i>	<i>m</i>
Durum 2	<i>b</i>	<i>d</i>	<i>n</i>
Sütun toplamı	<i>r</i>	<i>s</i>	<i>N</i>

Birçok istatistikçi, çapraz tablolarda beklenen değerlerin (e_{ij}) her biri en az 5 olduğunda ki-kare testinin kullanılmasını tavsiye etmiştir. Fisher, tarafından χ^2 testinin yalnızca $\min(e_{ij}) > 5$ olduğunda kullanılması gerektiği belirtilmiş, ancak sonraki araştırmacılar tarafından bu durum yavaş yavaş gevşetilmiştir.³ Cochran $\min(e_{ij}) > 1$ 'i önermiş, Lewontin ve Felsenstein bazı $e_{ij} = 0,5$ 'e eşit olmasını tatmin edici bulmuş, Yarnold ise $\min(e_{ij}) > 5s/rc$ 'den (r satır sayısını, c sütun sayısını, s ise beklenen değeri 5'ten az olan hücre sayısını göstermektedir) daha büyük olması gerektiğini öne sürmüştür.⁴⁻⁷ Sıklıkla kullanılan ve yararlı bir istatistik olmasına rağmen beklenen değerlerin küçük olduğu çapraz tablolara uygulandığında Pearson ki-kare test istatistikleri, H_0 hipotezi altındaki dağılım için tatmin edici bir ki-kare yaklaşımının mevcut olduğu durumlarda bile asimptotik olarak tutarsız olabilir.⁸ Ki-kare istatistiğinden yararlanmak 2x2 çapraz tablosunda yer alan ve Binom dağılımına sahip gözlenen değerlerin kesikli (ayrık, discrete) olasılıklarının, sürekli ki-kare dağılımı ile yaklaşık olarak tahmin edilebileceğini varsaymayı gerektirir. Bu varsayım tamamen doğru değildir ve bazı hatalara neden olur. Bir başka ifade ile kesikli frekans verilerinden hesaplanan gerçek ki-kare istatistikleri yalnızca kesikli değerler alır ve bu da düzensiz kademeli dağılımlar üretir. Teorik ki-kare dağılımları ise düzgün ve sürekli. Sonuç olarak deneye dayalı ki-kare fonksiyonu ile olasılıkların değerlendirildiği sürekli teorik dağılım arasında önemli farklılıklar mevcut olabilir. Süreklilik düzeltmeleri bu tutarsızlık için yapılan düzeltmelerdir. Düzeltme, hesaplanan ki-karenin büyüklüğünü dolayısıyla önemli bulunma olasılığını azaltır.⁹ Yaklaşımındaki hatayı azaltmak için birçok süreklilik düzeltmesi veya ki-kare testinin varyantları önerilmiştir. 2x2 çapraz tablolarda ki-kare bağımsızlık testinin kullanımına ilişkin uzun ve kapsamlı literatüre rağmen süreklilik düzeltmesinin uygunluğuna ilişkin tartışmalar hâlâ devam etmektedir.¹⁰ Yates, 2x2 çapraz tablolarda kullanılmak üzere Pearson ki-kare istatistiğinin ortaya çıkardığı hatayı azaltmak için gözlenen her değer ile beklenen değer arasındaki farktan 0,5 değerini çıkararak ayarlayan bir süreklilik düzeltmesi önermiştir. Bu düzeltme, elde edilen ki-kare değerini azaltmakta p -değerini ise artırmaktadır.¹¹ 2x2 tablolarla ilgili olarak, bazı istatistiksel metinler Yates'in düzeltmesinin aşırı tutucu olma eğiliminde olması nedeniyle kullanılmaması tavsiyesinde bulunurken, diğerleri daha da vurgulu bir şekilde "düzeltmenin her zaman kullanılması gerektiğini" savunmaktadır.¹⁰ Schouten ve ark., tarafından Yates tarafından önerilen ve koşullu bir test olarak ki-kare testine uygulanan düzeltmenin H_0 hipotezine göre tutucu olduğu ifade edilmektedir. Schouten ve ark. na göre koşulsuz olarak bakıldığında, $p_1 = p_2$ değerine

bağlı olan testin gerçek boyutu, önem seviyesinden çok daha küçüktür ve bu da güçte önemli bir azalmaya yol açmaktadır.¹² Yates'in kesikli bir dağılıma sürekli bir dağılım ile yaklaşıldığında süreklilik için bir düzeltme kullanılması gerektiğini belirtmesinin üzerinden neredeyse yüzyıl geçmesine rağmen Yates düzeltmesinin genellikle kesikli bir rastgele değişkenin birikimli dağılım fonksiyonuna yaklaşımı iyileştirdiği kabul edilse de istatistiksel bir test gerçekleştirirken bu süreklilik düzeltmesinin kullanılması hâlâ büyük bir tartışma konusudur.¹³ Yates'in süreklilik düzeltmesi, satır ve sütun toplamları değişmeyecek şekilde farklı hücre değerlerine sahip 2x2 tablolarındaki olasılıkları yaklaşık olarak hesaplamak için geliştirilmiştir. Bu tür tablolar, Fisher'in kesin ki-kare testinin esas aldığı hipergeometrik dağılım için de uygundur. Dolayısıyla, hem Yates düzeltmeli ki-kare testinin hem de Fisher kesin ki-kare testinin, satır ve sütun toplamlarının değiştiği durumlarda kullanılması uygun değildir.⁹ 2x2 tablolar üzerinde ki-kare testiyle elde edilen yaklaşımlar, süreklilik düzeltmesinin yardımıyla önemli ölçüde iyileştirilmiştir. İki yönlü bir test yapılırken geleneksel Yates düzeltmesi yetersizdir.¹⁴ Yates düzeltmesinin etkisi, küçük veri örnekleri için istatistiksel önemliliğin eksik tahmin edilmesini önlemektir. Ne yazık ki Yates düzeltmesi, p değerinin aşırı düzeltilmesine dolayısıyla H_0 hipotezinin reddedilmesi gerektiğinde kabul edilmesine (Tip II hata) neden olabilir. Fisher kesin (exact) ki-kare testi aslında ki-kare testinin uygun olmadığı durumlarda, yani örneklem büyüklüğünün küçük olduğu ve 2x2'lik bir çapraz tablonun herhangi bir hücresindeki beklenen değer 5 'in altında olduğu durumlarda kullanılan altın test olarak kabul edilir.¹⁵ Matchima ve ark. tarafından, beklenen frekansın küçük olması durumunda Tip-I hatanın kontrol edilememesi ki-kare testinin en büyük sınırlaması olarak ifade edilmiş ve bu zafiyetin Tip-I hata ile p değerinin benzer değerlere sahip olmasını sağlayacak bir test ile giderilebileceği belirtilmiştir.¹⁶

Bu çalışmada, serbestlik derecesi 1 (bir) olan ki-kare testleri dikkate alınmıştır. Çünkü bu testler, verilerin kesikli olmasından ciddi şekilde etkilenmektedir. Çalışmanın amacı, 2x2 çapraz tablolarında kullanılan klasik ve süreklilik düzeltmeli ki-kare testlerini tanıtmak ve bunları karşılaştırmaktır. Çalışma Helsinki Deklarasyonu prensipleri ilkelerine uygun olarak yürütülmüştür.

GEREÇ VE YÖNTEMLER

Python-random kütüphanesi kullanılarak 4 farklı örneklem büyüklüğü ($n = 10, n = 15, n = 20$ ve $n = 25$) için veri türetilmiştir. Veri türetimine ilişkin Python 3.9.10 (Python Software Foundation, ABD) programlama dili kullanılarak yazılan programda yer alan simülasyon kurgusunun detayları aşağıdaki gibidir:

Adım 1. Veri türetimi için gerekli değerler (toplam kaç adet veri türetilmek istendiği, kategorilere atanacak değerlerin toplamı, kategori sayısı) programa girilir.

Adım 2. Python random kütüphanesindeki choice fonksiyonu kullanılarak rastgele bir şekilde kategori seçimi yapılır.

Adım 3. Choice fonksiyonu tekrar kullanılarak seçilen kategori için 0 ile istenen toplam arasında rastgele bir tam sayı seçimi yapılır. Örneğin istenen toplam 15 ise $[0, 15]$ aralığında rastgele bir tam sayı seçilir.

Adım 4. Tüm kategoriler için ikinci ve üçüncü adımlar tekrarlanır. k 'nci kategori için bir tam sayı seçildikten sonra seçilen sayıların toplamına bakılır. Eğer bu toplam, istenilen toplamdan fazla ise bu geçersiz bir veri olacağından işleme baştan başlanır. Son kategoriye atanacak tam sayı ise istenilen toplamdan mevcut toplamın çıkarılmasıyla bulunur.

Adım 5. Türetilen verinin daha önce türetilip türetilmediğine bakılır. Aynı veri daha önce türetilmişse veri setine eklenmez.

Adım 6. İstenilen toplam veri sayısına ulaşana kadar birinci adımdan itibaren program bir döngü içinde tekrar çalıştırılır (Bu çalışmada standardizasyonu sağlamak amacıyla her bir örneklem genişliği için 1.000 adet veri türetilmesi planlanmıştır. Ancak, gerek ki-kare değerlerinin hesaplanamaması gerekse toplamı 10, 15, 20 veya 25 olan ve tekrar etmeyen 1000 tane veri bulmak imkânsız olduğundan $n=10$ için 246, $n=15$ için 756, $n=20$ için 958 ve $n=25$ için 963 adet veri hesaplamalarda kullanılmıştır).

Adım 7. Son olarak türetilen verilerden çalışmada dikkate alınan yöntemler için ki-kare değerleri hesaplanır.

Çalışmada dikkate alınan yöntemlere ilişkin hesaplama formülleri kronolojik sıra ile [Tablo 2](#)'de verilmiştir.

TABLO 2: Kronoloji ve formüller.

Kronoloji	Formül
Pearson, ¹⁷	$\frac{(ad - bc)^2 * N}{mnr s}$
Yates, ¹¹	$\frac{\left(ad - bc - \frac{N}{2}\right)^2 * N}{mnr s}$
Fisher, ¹⁸	$\frac{m! n! r! s!}{a! b! c! d! N!}$
Boschloo, ¹⁹	$\left(\left a - \frac{mr}{N}\right - \frac{1}{2}\right)^2 * \frac{N^3}{mnr s}$
Garside, ²⁰	$\frac{N(ad - bc - \lambda N)^2}{mnr s}$ λ 'nın değeri r , s ve α değerlerine göre farklılık göstermektedir.
Pirie-Hamdan, ²¹	$\frac{N\left(ad - bc - \frac{1}{4}N\right)^2}{mnr s}$
Williams, ²²	$\frac{\sum_i \sum_j 2n_{ij} \log\left(\frac{n_{ij}n_{..}}{n_{i.}n_{.j}}\right)}{1 + \frac{(N^2 - rs)(N^2 - mn)}{6mnr s N}}$ n_{ij} : i . satır j . sütunda yer alan gözlenen değerler (a, b, c, d), $n_{..}$: tüm gözlenen değerlerin toplamı (N), $n_{i.}$: i . satırda yer alan gözlenen değerlerin toplamı ($m \cdot n$), $n_{.j}$: j . sütunda yer alan gözlenen değerlerin toplamı ($r \cdot s$),
Schouten ve ark., ¹²	$\frac{N\left\{ ad - bc - \frac{1}{2} \min(m, n)\right\}^2}{mnr s}$
Upton, ²³	$\frac{(N - 1)(ad - bc)^2}{mnr s}$
Double Dichotomy ²³	$\frac{N\left(ad - bc - \frac{1}{2}\right)^2}{mnr s}$
Overall-Rhoades, ²⁴	$\frac{(N + 2)[(a + 1)(d + 1) - bc - C]^2}{(r + 1)(s + 1)(m + 1)(n + 1)}$ $C = \frac{N}{2}(0,75 + 5 \alpha) - (r - s)(m - n)$
Serra, ¹⁵	$\frac{16}{N^3}(ad - bc)^2$

[Tablo 2](#)'de yer almasına rağmen Garside ve Mack tarafından önerilen yönteme ilişkin hesaplamalar, formülde yer alan λ 'nın değeri r , s ve α değerlerine göre farklılık gösterdiği için gerçekleştirilememiştir. Overall ve Rhoades tarafından önerilen yöntem için C değerinin hesaplanmasında $\alpha = 0,05$ olarak alınmıştır.^{20,24}

BULGULAR

Çalışmada dikkate alınan yöntemlerin örneklem büyüklüğü ve önemlilik değerlerine göre yapılan analizler sonucunda elde edilen sonuçlar [Tablo 3](#)'te verilmiştir.

[Tablo 3](#)'e göre tüm yöntemlerin örneklem büyüklüğü ve önemlilik düzeyinden etkilendiği, örneklem büyüklüğünün artması ve önem düzeyinin de 0,01'den 0,10'a doğru değişmesi durumunda H_0 hipotezinin reddedilme oranlarının da arttığı belirlenmiştir.

TABLO 3: Hipotetik H_0 hipotezini reddetme yüzdeleri.

Yöntem	n	p			Yöntem	n	p		
		0,01	0,05	0,10			0,01	0,05	0,10
Pearson	10	10,6	25,2	35,8	Yates	10	0,0	5,7	12,2
	15	16,9	31,2	42,3		15	6,9	16,4	24,9
	20	23,6	35,5	46,0		20	11,0	22,5	28,7
	25	26,2	40,8	49,3		25	17,0	28,7	35,7
Fisher	10	4,1	12,2	21,1	Boschloo	10	0,0	5,7	12,2
	15	10,6	22,8	33,3		15	6,9	16,4	24,9
	20	15,7	28,6	37,5		20	11,0	22,5	28,7
	25	20,8	35,0	45,7		25	17,0	28,7	35,7
Pirie-Hamdan	10	5,7	17,1	22,0	Double-Dichotomy	10	4,5	11,8	19,5
	15	11,1	22,2	30,7		15	8,2	14,3	20,4
	20	15,9	28,5	35,5		20	10,5	18,1	22,9
	25	20,7	34,9	41,2		25	13,3	20,7	26,5
Williams	10	0,0	0,8	4,1	Schouten	10	7,3	18,7	26,8
	15	0,0	4,8	9,0		15	14,3	27,5	35,4
	20	0,5	6,2	13,3		20	19,1	31,7	39,7
	25	2,3	10,3	17,4		25	23,6	37,3	44,5
Upton	10	7,3	25,2	30,9	Overall-Rhoades	10	11,0	24,8	37,8
	15	15,3	31,2	40,2		15	19,3	33,9	43,4
	20	22,0	35,1	44,6		20	22,8	38,3	45,1
	25	25,2	39,8	48,4		25	28,1	43,0	50,7
Serra	10	4,1	13,8	22,8					
	15	9,0	20,1	30,7					
	20	12,6	24,3	30,3					
	25	17,0	27,9	36,7					

Yöntemlerin birbirlerine göre hipotetik H_0 hipotezini ret/kabul oranları $p = 0,01$, $p = 0,05$ ve $p = 0,10$ için sırasıyla [Tablo 4](#), [Tablo 5](#), [Tablo 6](#)'da verilmiştir. [Tablo 4](#), [Tablo 5](#), [Tablo 6](#)'da yer alan her bir hücredeki birinci (ret/ret) ve dördüncü (kabul/kabul) değerler satır ve sütunda yer alan yöntemlerin benzer yönde,

ikinci (ret/kabul) ve üçüncü (kabul/ret) değerler ise ters yönde verdiği kararların oranlarını göstermektedir. Dolayısıyla, birinci ve dördüncü değerlerin büyük ikinci ve üçüncü değerlerin ise küçük olması yöntemlerin birbirine benzerliğini belirtmektedir. Tüm yöntemler için sonuçlar elde edilmiş olmasına rağmen Pearson, Yates ve Fisher yöntemleri referans standartlar olarak kullanıldığından diğer yöntemlerin bu üç yöntemle benzerliği yorumlanmış, diğer yöntemlere ilişkin yorumlar okuyucuya bırakılmıştır.

TABLO 4: $p = 0,01$ için hipotetik H_0 hipotezi ret-kabul yüzdelерinin karşılaştırılması.

	Yates	Fisher	Boschloo	Pirie-Hamdan	Double Dichotomy	Williams	Schoten	Upton	Overall-Rhoades	Serra
Pearson	11,0	15,1	11,0	15,4	10,1	0,9	18,3	20,1	17,1	12,4
	10,6	6,6	10,6	6,3	11,5	20,7	3,3	1,5	4,6	9,2
	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	5,6	0,0
	78,4	78,4	78,4	78,4	78,2	78,4	78,4	78,4	72,8	78,4
Yates		10,9	11,0	11,0	5,6	0,9	11,0	11,0	10,7	9,5
		0,0	0,0	0,0	5,3	10,1	0,0	0,0	0,2	1,5
		4,1	0,0	4,4	4,7	0,0	7,4	9,1	11,9	2,9
		84,9	89,0	84,6	84,3	89,0	81,7	79,9	77,1	86,1
Fisher			10,9	14,3	7,6	0,9	14,7	15,1	13,1	11,9
			4,1	0,8	7,4	14,1	0,4	0,0	2,0	3,1
			0,0	1,1	2,7	0,0	3,7	5,1	9,6	0,5
			84,9	83,9	82,2	84,9	81,3	79,9	75,4	84,4
Boschloo				11,0	5,6	0,9	11,0	11,0	10,7	9,5
				0,0	5,3	10,1	0,0	0,0	0,2	1,5
				4,4	4,7	0,0	7,4	9,1	11,9	2,9
				84,6	84,3	89,0	81,7	79,9	77,1	86,1
Pirie-Hamdan					7,9	0,9	15,4	15,4	13,5	11,6
					7,4	14,4	0,0	0,0	1,8	3,8
					2,4	0,0	3,0	4,8	9,1	0,8
					82,2	84,6	81,7	79,9	75,5	83,8
Double Dichotomy						0,4	9,4	10,0	10,1	6,5
						9,9	1,0	0,4	0,3	3,8
						0,5	9,0	10,2	12,6	5,9
						89,2	80,7	79,5	77,1	83,8
Williams							0,9	0,9	0,9	0,9
							0,0	0,0	0,0	0,0
							17,4	19,2	21,7	11,5
							81,7	79,9	77,4	87,6
Schouten								18,3	15,4	11,8
								0,0	2,9	6,5
								1,8	7,3	0,6
								79,9	74,4	81,1
Upton									16,3	12,4
									3,8	7,8
									6,3	0,1
									73,6	79,8
Overall-Rhoades										10,7
										11,9
										1,7
										75,7

Her bir hücrede yer alan ilk değer: Ret/Ret"; ikinci değer: Ret/Kabul"; üçüncü değer: Kabul/Ret"; dördüncü değer: Kabul/Kabul" yüzdelерini göstermektedir. "satırdaki, "ise sütundaki yöntemi ifade etmektedir.

TABLO 5: $p = 0,05$ için hipotetik H_0 hipotezi ret-kabul yüzdelerinin karşılaştırılması.

	Yates	Fisher	Boschloo	Pirie-Hamdan	Double Dichotomy	Williams	Schoten	Upton	Overall-Rhoades	Serra
Pearson	21,4	27,8	21,4	28,0	16,6	6,7	31,4	34,8	29,6	23,5
	13,9	7,5	13,9	7,3	18,7	28,6	3,9	0,5	5,6	11,7
	0,2	0,0	0,2	0,0	0,9	0,0	0,0	0,0	7,9	0,0
	64,6	64,7	64,6	64,7	63,9	64,7	64,7	64,7	56,8	64,7
Yates		21,3	21,6	21,4	11,1	6,7	21,4	21,4	21,2	19,2
		0,2	0,0	0,2	10,4	14,8	0,2	0,2	0,3	2,4
		6,5	0,0	6,6	6,3	0,0	10,0	13,4	16,4	4,4
		71,9	78,4	71,8	72,2	78,4	68,5	65,0	62,1	74,1
Fisher			21,3	26,5	13,8	6,7	27,4	27,8	24,7	23,3
			6,5	1,3	14,0	21,1	0,4	0,0	3,1	4,5
			0,2	1,5	3,6	0,0	4,0	7,0	12,8	0,2
			71,9	70,6	68,6	72,2	68,2	65,2	59,4	71,9
Boschloo				21,4	11,1	6,7	21,4	21,4	21,2	19,2
				0,2	10,4	14,8	0,2	0,2	0,3	2,4
				6,6	6,3	0,0	10,0	13,4	16,4	4,4
				71,8	72,2	78,4	68,5	65,0	62,1	74,1
Pirie-Hamdan					14,6	6,7	28,0	28,0	25,4	22,5
					13,4	21,3	0,0	0,0	2,6	5,5
					2,8	0,0	3,4	6,8	12,1	1,0
					69,1	72,0	68,6	65,2	59,8	71,0
Double Dichotomy						3,6	15,8	16,4	16,6	12,0
						13,9	1,6	1,0	0,8	5,4
						3,1	15,5	18,4	20,9	11,6
						79,4	67,1	64,2	61,6	71,0
Williams							6,7	6,7	6,7	6,7
							0,0	0,0	0,0	0,0
							24,7	28,1	30,9	16,8
							68,6	65,2	62,4	76,5
Schouten								31,4	27,5	23,1
								0,0	3,9	8,3
								3,4	10,1	0,4
								65,2	58,6	68,2
Upton									29,3	23,5
									5,5	11,3
									8,2	0,0
									57,0	65,2
Overall-Rhoades										21,3
										16,2
										2,2
										60,2

Her bir hücrede yer alan ilk değer: Ret/Ret"; ikinci değer: Ret/Kabul"; üçüncü değer: Kabul/Ret"; dördüncü değer: Kabul/Kabul" yüzdelerini göstermektedir. "sattırdaki, "ise sütündeki yöntemi ifade etmektedir.

TABLO 6: $p = 0,10$ için hipotetik H_0 hipotezi ret-kabul yüzdelerinin karşılaştırılması.

	Yates	Fisher	Boschloo	Pirie-Hamdan	Double Dichotomy	Williams	Schoten	Upton	Overall-Rhoades	Serra
Pearson	28,3	37,7	28,3	35,0	21,6	12,8	39,1	43,6	37,9	31,9
	17,0	7,6	17,0	10,3	23,7	32,5	6,2	1,7	7,4	13,4
	0,3	0,0	0,3	0,0	1,5	0,0	0,0	0,0	8,0	0,0
	54,4	54,7	54,4	54,7	53,2	54,7	54,7	54,7	46,7	54,7
Yates		28,3	28,6	28,3	15,1	12,8	28,3	28,3	27,7	26,2
		0,3	0,0	0,3	13,5	15,9	0,3	0,3	0,9	2,4
		9,4	0,0	6,7	8,0	0,0	10,8	15,2	18,1	5,6
		62,0	71,4	64,7	63,4	71,4	60,6	56,1	53,2	65,7
Fisher			28,3	34,1	19,3	12,8	35,7	37,4	33,1	31,6
			9,4	3,6	18,4	25,0	2,0	0,3	4,6	6,2
			0,3	0,9	3,8	0,0	3,4	6,1	12,8	0,3
			62,0	61,4	58,5	62,3	58,9	56,1	49,5	62,0
Boschloo				28,3	15,1	12,8	28,3	28,3	27,7	26,2
				0,3	13,5	15,9	0,3	0,3	0,9	2,4
				6,7	8,0	0,0	10,8	15,2	18,1	5,6
				64,7	63,4	71,4	60,6	56,1	53,2	65,7
Pirie-Hamdan					18,6	12,8	35,0	35,0	32,0	29,8
					16,4	22,2	0,0	0,0	3,0	5,2
					4,5	0,0	4,1	8,6	13,9	2,1
					60,5	65,0	60,9	56,4	51,1	62,9
Double Dichotomy						6,7	20,0	21,3	21,9	16,4
						16,4	3,1	1,8	1,2	6,7
						6,0	19,1	22,2	23,9	15,5
						70,9	57,8	54,6	52,9	61,4
Williams							12,8	12,8	12,8	12,8
							0,0	0,0	0,0	0,0
							26,3	30,8	33,1	19,1
							60,9	56,4	54,1	68,1
Schouten								39,1	34,7	30,7
								0,0	4,4	8,4
								4,4	11,2	1,1
								56,4	49,7	59,8
Upton									36,9	31,9
									6,6	11,7
									9,0	0,0
									47,5	56,4
Overall-Rhoades										28,4
										17,4
										3,4
										50,7

Her bir hücrede yer alan ilk değer: Ret/Ret"; ikinci değer: Ret/Kabul"; üçüncü değer: Kabul/Ret"; dördüncü değer: Kabul/Kabul" yüzdelerini göstermektedir. "satırdaki, "ise sütündeki yöntemi ifade etmektedir.

Tablo 4'e göre Ret/Ret oranı benzerliği bakımından:

1. Williams, Double Dichotomy, Yates, Boschloo, Serra, Fisher, Pirie-Hamdan, Overall-Rhoades, Schouten ve Upton yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Williams, Double Dichotomy, Serra, Overall-Rhoades, Fisher, Boschloo, Pirie-Hamdan, Schouten ve Upton yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Williams, Double Dichotomy, Boschloo, Serra, Overall-Rhoades, Pirie-Hamdan, Schouten ve Upton yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Ret/Kabul oranı benzerliği bakımından:

1. Upton, Schouten, Overall-Rhoades, Pirie-Hamdan, Fisher, Serra, Yates, Boschloo, Double-Dichotomy ve Williams yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Yates yöntemi, Fisher, Boschloo, Pirie-Hamdan, Schouten ve Upton yöntemlerine tıpatıp benzemektedir. Overall-Rhoades, Serra, Double Dichotomy ve Williams yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Fisher yöntemi Upton yöntemine tıpatıp benzemektedir. Schouten, Pirie-Hamdan, Overall-Rhoades, Serra, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Kabul/Ret oranı benzerliği bakımından:

1. Pearson yöntemi, Yates, Fisher, Boschloo, Pirie-Hamdan, Williams, Schouten, Upton ve Serra yöntemlerine tıpatıp benzemektedir. Double Dichotomy ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.
2. Yates yöntemi, Boschloo ve Williams yöntemlerine tıpatıp benzemektedir. Serra, Pirie-Hamdan, Double Dichotomy, Fisher, Schouten, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.
3. Fisher yöntemi, Williams yöntemine tıpatıp benzemektedir. Boschloo, Serra, Pirie-Hamdan, Schouten, Double Dichotomy, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.

Kabul/Kabul oranı benzerliği bakımından:

1. Yates, Fisher, Boschloo, Pirie-Hamdan, Williams, Schouten, Upton, Serra, Double Dichotomy ve Overall-Rhoades yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Boschloo, Williams, Serra, Fisher, Pirie-Hamdan, Double Dichotomy, Schouten, Upton ve Overall-Rhoades yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Boschloo, Williams, Serra, Pirie-Hamdan, Double Dichotomy, Schouten, Upton ve Overall-Rhoades yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Tablo 5'e göre Ret/Ret oranı benzerliği bakımından:

1. Upton, Schouten, Overall-Rhoades, Pirie-Hamdan, Fisher, Serra, Yates, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Boschloo, Pirie-Hamdan, Schouten, Upton, Fisher, Overall-Rhoades, Serra, Double Dichotomy ve Williams yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Upton, Schouten, Pirie-Hamdan, Overall-Rhoades, Serra, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Ret/Kabul oranı benzerliđi bakımından:

1. Upton, Schouten, Overall-Rhoades, Pirie-Hamdan, Fisher, Serra, Yates, Boschloo, Double-Dichotomy ve Williams yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Yates yöntemi, Boschloo yöntemine tıpatıp benzemektedir. Fisher, Pirie-Hamdan, Schouten, Upton, Overall-Rhoades, Serra, Double Dichotomy ve Williams yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Fisher yöntemi Upton yöntemine tıpatıp benzemektedir. Schouten, Pirie-Hamdan, Overall-Rhoades, Serra, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Kabul/Ret oranı benzerliđi bakımından:

1. Pearson yöntemi, Fisher, Pirie-Hamdan, Williams, Schouten, Upton ve Serra yöntemlerine tıpatıp benzemektedir. Yates, Boschloo, Double Dichotomy ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.
2. Yates yöntemi, Boschloo ve Williams yöntemlerine tıpatıp benzemektedir. Serra, Double Dichotomy, Fisher, Pirie-Hamdan, Schouten, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.
3. Fisher yöntemi, Williams yöntemine tıpatıp benzemektedir. Boschloo, Serra, Pirie-Hamdan, Double Dichotomy, Schouten, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.

Kabul/Kabul oranı benzerliđi bakımından:

1. Fisher, Pirie-Hamdan, Williams, Schouten, Upton, Serra, Yates, Boschloo, Double Dichotomy ve Overall-Rhoades yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Boschloo, Williams, Serra, Double Dichotomy, Fisher, Pirie-Hamdan, Schouten, Upton ve Overall-Rhoades yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Williams, Boschloo, Serra, Pirie-Hamdan, Double Dichotomy, Schouten, Upton ve Overall-Rhoades yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

[Tablo 6](#)'ya göre Ret/Ret oranı benzerliđi bakımından:

1. Upton, Schouten, Overall-Rhoades, Fisher, Pirie-Hamdan, Serra, Yates, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Boschloo, Fisher, Pirie-Hamdan, Schouten, Upton, Overall-Rhoades, Serra, Double Dichotomy ve Williams yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Upton, Schouten, Pirie-Hamdan, Overall-Rhoades, Serra, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Ret/Kabul oranı benzerliđi bakımından:

1. Upton, Schouten, Overall-Rhoades, Fisher, Pirie-Hamdan, Serra, Yates, Boschloo, Double-Dichotomy ve Williams yöntemleri sırasıyla Pearson yöntemine en çok benzeyen yöntemlerdir.
2. Yates yöntemi Boschloo yöntemi ile birbirine tıpatıp benzemektedir. Fisher, Pirie-Hamdan, Schouten, Upton, Overall-Rhoades, Double Dichotomy ve Williams yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Upton, Schouten, Pirie-Hamdan, Overall-Rhoades, Serra, Boschloo, Double Dichotomy ve Williams yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

Kabul/Ret oranı benzerliđi bakımından:

1. Pearson yöntemi, Fisher, Pirie-Hamdan, Williams, Schouten, Upton ve Serra yöntemlerine tıpatıp benzemektedir. Yates, Boschloo, Double Dichotomy ve Overall-Rhoades yöntemleri ise sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Yates yöntemi, Boschloo ve Williams yöntemlerine tıpatıp benzemektedir. Serra, Pirie-Hamdan, Double Dichotomy, Fisher, Schouten, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.
3. Fisher yöntemi, Williams yöntemine tıpatıp benzemektedir. Boschloo, Serra, Pirie-Hamdan, Schouten, Double Dichotomy, Upton ve Overall-Rhoades yöntemleri ise sırasıyla benzeyen yöntemlerdir.

Kabul/Kabul oranı benzerliđi bakımından:

1. Fisher, Pirie-Hamdan, Williams, Schouten, Upton, Serra, Yates, Boschloo, Double Dichotomy ve Overall-Rhoades yöntemleri sırasıyla Pearson yöntemine benzeyen yöntemlerdir.
2. Boschloo, Williams, Serra, Pirie-Hamdan, Double Dichotomy, Fisher, Schouten, Upton ve Overall-Rhoades yöntemleri sırasıyla Yates yöntemine benzeyen yöntemlerdir.
3. Williams, Boschloo, Serra, Pirie-Hamdan, Schouten, Double Dichotomy, Upton ve Overall-Rhoades yöntemleri sırasıyla Fisher yöntemine benzeyen yöntemlerdir.

TARTIŞMA

İki tedavi grubunda pozitif ve negatif tepkilerin göreceli sıklıkları arasındaki farkların büyük olması ve verilerin maliyetli olması beklendiğinde, araştırmacılar gerçek tedavi etkilerinin varlığını belgelemek amacıyla minimum örneklem büyüklüklerini kullanma eğilimindedir. 2x2 çapraz tablolarda Pearson ki-kare testinin özellikle küçük örneklemelerde çok fazla Tip 1 hata üretme eğiliminde olmasından dolayı süreklilik düzeltmeleri önerilmiştir. Süreklilik düzeltmesi, gözlenen değer ile beklenen değer arasındaki farkı azaltma düşüncesini esas almaktadır. Ki-kare testi için düzeltme değeri, çapraz tablonun yapısına bağlıdır. Uygun düzeltme değerinin mutlaka 0,5'lik Yates düzeltme değerine eşit olması gerekmez. Yates tarafından önerilen süreklilik düzeltmesi, p değerinin aşırı düzeltilmesine neden olduğundan kullanımı önerilmemektedir.¹⁶ Fisher kesin ki-kare testi ve Yates düzeltmeli ki-kare testi pratik olarak aynıdır. Her ikisi de son derece tutucudur. Plackett ve Grizzle'nin her ikisi de $m = n$ olduğu durumlarda Yates düzeltmesinin uygun olmadığı sonucuna ulaşmışlardır. Tek yönlü bir test olarak kullanıldığında Yates tarafından önerilen testin herhangi bir sorun yaratmadığı ancak iki yönlü bir test olarak sorunlu bir test olduğu ifade edilmiştir.^{25,26} Pirie ve Hamdan tarafından $ad - bc$ 'nin farklı değerler almasının ancak $h = \frac{1}{2}N$ değeri değıştikçe mümkün olacağı, dolayısıyla Yates tarafından önerilen düzeltmenin de düzeltilmesi gerektiđi belirtilmiştir.²¹ Liddell tarafından ise Pirie ve Hamdan tarafından önerilen testin $m \neq n$ olduğu durumlarda bile tutucu bir test olduğu öne sürülmüştür.²⁷ Reddetme bölgesinin büyüklüğü bakımından yöntemler karşılaştırılmış, küçük m ve n değerleri için birçok test aynı reddetme bölgesine sahip olmasına rağmen büyük m ve n değerleri için her testin farklı reddetme bölgelerine sahip olduğu belirtilmiş, genel olarak, Fisher kesin ki-kare ve Yates düzeltmeli ki-kare testlerine ait ret bölgelerinin Pearson ki-kare testine göre daha az 2x2 çapraz tablo içerdiği ifade edilmiştir.^{23,28} Çalışmadan elde edilen sonuçlar, Pearson ki-kare testinin küçük örneklemelerde H_0 hipotezini reddetme, Yates düzeltmeli ki-kare testinin kabul etme eğilimine sahip olduğunu, Fisher kesin ki-kare testinin ise bu iki test arasında kaldığını göstermektedir. Tip 1 hatanın büyüklüğünün belirlenen p önem seviyesine bağlı olduğu bilinmektedir. p önem seviyesinde bir test yapıldığında, dolaylı olarak Tip 1 hatanın büyüklüğünün gerçekten p olduğu varsayılır, ancak kesikli dağılımlarda bunun tam olarak böyle olması pek olası değildir. [Tablo 4](#), [Tablo 5](#), [Tablo 6](#)'dan çalışmada dikkate alınan tüm yöntemlerin, p 'nin artan değerleri için liberal olmaya

eğilimli oldukları görülmektedir. Bir testin sonucunun genel olarak doğruluğu p önem seviyesine bağlı olduğundan rakip testler hakkında kesin ifadelerde bulunmak mümkün değildir. Ayrıca, iki farklı testi güçlerini dikkate alarak karşılaştırabilmek için bu testlerin aynı Tip 1 hataya sahip olması gerekir. Ancak bu çalışmada dikkate alınan testler için Tip 1 hata, bilinmeyen p 'nin bir fonksiyonudur ve yalnızca özdeş testler, p 'nin tüm değerleri için aynı Tip 1 hataya sahip olabilir. Bu nedenle doğrudan güç karşılaştırmaları mümkün değildir. Dolayısıyla çalışmada dikkate alınan yöntemler hakkında gerek genel doğruluk bakımından gerekse güçleri bakımından karşılaştırma yapılmamıştır.

SONUÇ

Çalışmada dikkate alınan yöntemlere ait sonuçlar, yöntemlerden birinin önerilen tüm yöntemler içerisinde en iyi yöntem olarak seçilmesini sağlamamaktadır. Farklı yöntemler farklı örneklem büyüklüklerinde ve önemlilik düzeylerinde öne çıkmaktadır. Bu durum, bir araştırmadan elde edilecek sonucunun doğru şekilde yorumlanamayacağı anlamına gelir. Dolayısıyla ki-kare testinin büyük örneklem için uygun olduğu, beklenen değerlerden en az birinin 5'ten küçük olması durumunda gerek klasik gerekse süreklilik düzeltmeli ki-kare yöntemlerinin kullanılmaması gerektiği sonucuna ulaşılmıştır.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Çıkar Çatışması

Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.

Yazar Katkıları

Bu çalışma hazırlanırken tüm yazarlar eşit katkı sağlamıştır.

KAYNAKLAR

- McHugh ML. The chi-square test of independence. *Biochem Med (Zagreb)*. 2013;23(2):143-9. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
- Zar HJ. *Biostatistical Analysis*. 3rd ed. New Jersey: Prentice Hall. Inc.; 1974.
- Fisher RA. *Statistical Methods for Research Workers*. 1st ed. Edinburgh: Oliver and Boyd; 1925.
- Cochran WG. Some methods of strengthening the common χ^2 tests. *Biometrics*. 1954;10(4):417-51. [\[Crossref\]](#)
- Lewontin RC, Felsenstein J. The robustness of homogeneity tests in 2xN tables. *Biometrics*. 1965;21(1):19-33. [\[Crossref\]](#)
- Yarnold JK. The minimum expectation in χ^2 goodness-of-fit tests and the accuracy of approximations for the null distribution. *J Amer Statist Assoc*. 1970;65(330):864-86. [\[Crossref\]](#)
- Lawal HB, Upton GJG. On the use of χ^2 as a test of independence in contingency tables with small cell expectations. *Austral J Statist*. 1984;26(1):75-85. [\[Crossref\]](#)
- Haberman SJ. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *JASA*. 1988;83(402):555-60. [\[Crossref\]](#)
- Haviland MG. Yates's correction for continuity and the analysis of 2 x 2 contingency tables. *Stat Med*. 1990;9(4):363-7; discussion 369-83. [\[Crossref\]](#) [\[PubMed\]](#)
- Kroll NEA. Testing independence in 2x2 contingency tables. *J Educ Stat*. 1989;14(1):47-79. [\[Crossref\]](#)
- Yates F. Contingency tables involving small numbers and the χ^2 test. *J R Stat Soc*. 1934;1(2):217-35. [\[Crossref\]](#)
- Schouten HJA, Molenaar IW, Van Strik R, Boomsma A. Comparing two independent binomial proportions by a modified chi square test. *Biom J*. 1980;22(3):241-8. [\[Crossref\]](#)
- Haber M. The continuity correction and statistical testing. *Int Stat Rev*. 1982;50(2):135-44. [\[Crossref\]](#)
- Haber M. A comparison of some continuity corrections for the chi-squared test on 2x2 tables. *JASA*. 1980;75(371):510-5. [\[Crossref\]](#)
- Serra N. A significant minimization of Pearson's χ^2 statistics in 2x2 contingency tables: preliminary results for small samples. *EBPH*. 2018;15(3):e12949-1. [\[Crossref\]](#)

16. Matchima K, Vongprasert J, Chutiman N. The development of a correction method for ensuring a continuity value of the chi-square test with a small expected cell frequency. *J Sci Technol.* 2018;26(1):98-105. [\[Link\]](#)
17. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos Mag.* 1900;50(302):157-75. [\[Link\]](#)
18. Fisher RA. *The Design of Experiments.* 1st ed. Edinburgh: Oliver and Boyd; 1935.
19. Boschloo RD. Raised conditional level of significance for the 2x2 table when testing for the equality of two probabilities. *Stat Neerl.* 1970;21(1):1-35. [\[Link\]](#)
20. Garside GR, Mack C. Actual type 1 error probabilities for various tests in the homogeneity case of the 2x2 contingency table. *Am Stat.* 1976;30(1):18-21. [\[Crossref\]](#)
21. Pirie WR, Hamdan MA. Some revised continuity corrections for discrete distributions. *Biometrics.* 1972;28(3):693-701. [\[Crossref\]](#)
22. Williams DA. Improved likelihood ratio tests for complete contingency tables. *Biometrika.* 1976;63(1):33-7. [\[Crossref\]](#)
23. Upton GJG. A comparison of alternative tests for the 2x2 comparative trial. *J R Stat Soc Ser A Stat Soc.* 1982;145(1):86-105. [\[Crossref\]](#)
24. Overall JE, Rhoades HM, Starbuck RR. Small-sample tests for homogeneity of response probabilities in 2 X 2 contingency tables. *Psychol Bull.* 1987;102(2):307-14. [\[Crossref\]](#) [\[PubMed\]](#)
25. Plackett RL. The continuity correction in 2x2 tables. *Biometrika.* 1964;51:327-37. [\[Crossref\]](#)
26. Grizzle JE. Continuity correction in the χ^2 test for 2x2 tables. *Amer Statist.* 1967;21(4):28-32. [\[Crossref\]](#)
27. Liddell D. Practical tests of 2x2 contingency tables. *The Statistician.* 1976;25(4):295-304. [\[Crossref\]](#)
28. Rhoades HM, Overall JE. A sample size correction for Pearson chi-square in 2x2 contingency tables. *Psychol Bull.* 1982;91(2):418-23. [\[Crossref\]](#)