

İkili Veriler İçin Uyum Katsayılarının Değerlendirilmesi: Bir Benzetim Çalışması

Evaluation of Agreement Coefficients of Binary Data: A Simulation Study

İsmet DOĞAN^a, Nurhan DOĞAN^a, Taylan DOĞAN^b

^aAfyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Afyonkarahisar, TÜRKİYE

^bKPN B.V. Mobile Telecommunications Company, Amsterdam, HOLLANDA

ÖZET Amaç: Bu çalışmanın amacı, türetilmiş veri setleri kullanarak, farklı n, a, b, c ve d değerleri için belirlenen 23 farklı ikili uyum katsayısını tanıtmak, özelliklerini ortaya koyarak değerlendirmektir. **Gereç ve Yöntemler:** Bu çalışmada, ikili veriler için ileri sürülen uyum katsayıları ele alınmıştır. Çalışmada, Python-random kütüphanesi kullanılarak, $10 \leq n \leq 1.000$ aralığında yer alan 35 farklı n değeri için veri türetilmiştir. Verilerin türetilmesinde, önce a, b, c ve d ile gösterilen gözelerden hangisine değer atacağı, sonra da ilgili gözeye atanacak değer belirlenmiştir. $n = 10$ için 286, $n = 15$ için 815 ve $n \geq 20$ için 1.000'er farklı veri seti çalışmada kullanılmıştır. Katsayıların alacağı değerlerin tahmininde pozitif ve negatif eşleşme frekansları ile örneklem büyüklüğünün etkisi belirlenmiştir. Ayrıca hiyerarşik kümeleme analizi ile tüm simülasyon sonuçları dikkate alınarak, katsayıların benzerlikleri ortaya konmuştur. **Bulgular:** İkili veriler için tüm uyum katsayılarının değer aralığının 0 (uyum yok)-1 (tam uyum) olması beklenmesine rağmen tüm katsayılar için bu aralık geçerli değildir. Dikkate alınan 23 farklı katsayı içerisinde 6 tanesi bu aralıkta değer almaktadır. Hiyerarşik kümeleme analizine göre uyum katsayılarının çoğu birbirine benzememekte ve hiçbir n sayısından etkilenmemektedir. **Sonuç:** Genel olarak hemen tüm katsayılara ait değerler, örnekler daha benzer hâle geldikçe sabit bir minimumdan sabit bir maksimuma doğru artmaktadır. Ancak Andrés Marzo, Bennett, Brennan Prediger, Byrt ve ark., Gwet, Janson Vegelius, Osgood Holsti, Potter Levine-Donnerstein ve uyum yüzdesi katsayıları, tüm n değerleri için uyum ile doğrusal olarak sorunsuz bir şekilde artmaktadır. Değer aralığının 0-1 olması ve uyum artışı ile paralellik göstermesinden dolayı Osgood Holsti ve uyum yüzdesi katsayıları, tüm katsayılar içerisinde öne çıkmaktadır.

ABSTRACT Objective: The aim of this study is to introduce 23 different binary agreement coefficients determined for different n, a, b, c and d values by using derived data sets and to evaluate them by revealing their properties. **Material and Methods:** In this study, the agreement coefficients put forward for binary data are considered. In the study, data were derived for 35 different n values in the range of $10 \leq n \leq 1,000$ using the Python-random library. In the derivation of the data, firstly, which cell shown with a, b, c and d will be assigned value, then the value to be assigned to the relevant cell was determined. 286 for $n = 10$, 815 for $n = 15$ and 1,000 different data sets for $n \geq 20$ were used in the study. The effect of positive and negative matching frequencies and sample size on the estimation of the values of the coefficients was determined. In addition, the similarities of the coefficients were revealed by considering all the simulation results with hierarchical clustering analysis. **Results:** Although it is expected that the value range of all agreement coefficients for binary data is 0 (no agreement) to 1 (exact agreement), this range is not valid for all coefficients. Of the 23 different coefficients considered, 6 of them take values in this range. According to the hierarchical cluster analysis, most of the agreement coefficients are not similar to each other and none of them are affected by the number n . **Conclusion:** In general, the values for almost all coefficients increase from a constant minimum to a constant maximum as the samples become more similar. However, the coefficients of Andrés Marzo, Bennett, Brennan Prediger, Byrt et al, Gwet, Janson Vegelius, Osgood Holsti, Potter Levine-Donnerstein, and percent agreement increase linearly and smoothly with agreement for all n values. Osgood Holsti and percent agreement coefficients stand out among all coefficients, since the value range is 0-1 and is in parallel with the increase in agreement.

Anahtar kelimeler: Uyum katsayısı; hiyerarşik kümeleme; ikili veri

Keywords: Agreement coefficient; hierarchical clustering; binary data

Correspondence: Nurhan DOĞAN

Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Afyonkarahisar, TÜRKİYE/TURKEY

E-mail: nurhandogan@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

Received: 22 Jun 2021 **Received in revised form:** 13 Sep 2021 **Accepted:** 14 Sep 2021 **Available online:** 29 Sep 2021

2146-8877 / Copyright © 2021 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/>).

Uyum, 2 farklı değerlendirici (karar verici, cihaz vb.) arasındaki tutarlılığı ifade etmektedir ve sonuçlar arasındaki güvenilirliği değerlendirmek için yaygın olarak kullanılan bir kavramdır. Uyum katsayısı, 2 farklı değerlendiricinin farklı zaman, mekân ve durumlarda ilgilenilen olaya aynı tepkiyi verme derecesini göstermektedir. İki veya daha fazla değerlendirici arasındaki uyumun belirlenmesi, istatistikçiler, klinisyenler, epidemiyologlar, psikologlar ve diğer birçok bilim insanı için öncelikli konulardan biridir. Çünkü 2 karar vericinin kararları ya da 2 farklı cihazdan elde edilen ölçüm değerlerinin uyumlu olması, geçerlilik çalışmaları, randomize-kontrollü çalışmalar ve sağlık hizmetlerinin verimli bir şekilde sunulmasını sağlamak için esastır.¹ Uyum ölçüleri, istatistikte özel bir rol oynar. İyi tanımlanmış bir uyum ölçüsü, bir değerlendiriciden herhangi bir konuya ilişkin elde edilen sonucun, diğer değerlendiricilerin rapor ettikleriyle ne kadar uyduğuna veya bir değerlendiricinin, bir konu ile ilgili 2 farklı zamana ait değerlendirme sonuçlarının ne kadar uyduğuna tanımlar. Değerlendiricilere ait sonuçlar için uyum değerinin belirlenmesi, elde edilen sonuçların farklı zaman veya farklı yerlerde ne kadar tekrarlanabilir olabileceğini gösterir. Dolayısıyla uyum ölçüsü, bir sonucun genelleştirilebilirliğinin göstergesi olarak kullanılabilir.² İlişki ve uyum kavramları genellikle yanlış anlaşılmaktadır ve doğru kullanılmamaktadır. İlişki ve uyum birbirinden farklı 2 kavramdır. Uyum varlığı, ilişkinin varlığını ima eder, ancak tersi doğru olmayabilir.³ İlişki ve uyum kavramları, değerlendiricilerden elde edilen sonuçlar arasındaki ilişkinin gücünü belirtmek için kullanılır, ancak kavramsal olarak farklıdır ve bu nedenle farklı istatistiklerin kullanılmasını gerektirir. İlişki, genellikle farklı yapıların değerlendirilmesinden elde edilen 2 sonuçtaki değişikliklerin ilişkisine odaklanır. Uyum da sonuçlar arasındaki ilişkileri değerlendirir, ancak adından da anlaşılacağı gibi vurgu, 2 veya daha değerlendiriciden elde edilen sonuçlardaki uyum derecesi üzerindedir.⁴ Hemen hemen tüm araştırmalarda elde edilen sonuçlardan hem ilişki hem de uyum ölçütleri elde edilebilmektedir. İlişki, bir sonucun kategorisini diğer sonucun kategorisinden tahmin edebilmektir. Uyum ise 2 sonucun aynı özdeş kategoride yer almasıdır. Bu tanımlardan da anlaşılacağı üzere 2 sonuç arasında bir uyum varsa ilişki zaten var demektir, ancak sonuçlar arasında güçlü bir uyum olmamasına rağmen güçlü bir ilişki olabilir.⁵ Tutarlılık, uygun bir ölçü seçimine dayandığından birçok araştırmacı, 100 yılı aşkın bir süredir en anlamlı uyum katsayısını bulmak için ayrıntılı çaba sarf etmiştir. Çeşitli alanlarda ikili veriler için çok sayıda uyum katsayısı önerilmiş olmasına rağmen bunların birlikte ele alınarak değerlendirildiği sadece birkaç karşılaştırmalı çalışma bulunmaktadır.⁶⁻⁸ Uyum katsayıları, tutarlılık değerlendirmesini teşvik eden çok çeşitli alanlarda kullanılmaktadır. Örneklerin ikili durumunu evet/hayır, doğru/yanlış veya var/yok olarak ifade eden ikili özellikler, çok çeşitli verileri temsil etmek için yaygın olarak kullanılmıştır. İkili özellik matrisleri tarafından temsil edilen verileri değerlendirmek için uygun bir uyum katsayısının seçimi gereklidir, çünkü farklı uyum katsayıları çelişkili sonuçlar verebilmektedir. İki farklı cihazdan elde edilen sonuçları ya da 2 farklı karar vericiye ait kararları birbiriyle örtüşmeyen birçok kategoriden birine sınıflandırmada uyum miktarını belirlemek, biyoistatistik ve epidemiyoloji için giderek daha önemli hâle gelen bir sorundur. Literatürde istatistik, biyoistatistik, psikoloji, psikiyatri, eğitim ve sosyoloji alanlarındaki araştırmacılar tarafından önerilmiş çok sayıda uyum ölçütü bulunmaktadır.⁹ Genel olarak 2 nesnenin bir dizi öznelilik bakımından evet/hayır, doğru/yanlış veya var/yok şeklinde karşılaştırılması söz konusu ise veya bir nesne kümesi, 2 sonuçlu 2 farklı değişken için çapraz sınıflandırılırsa, 2×2 bir tablo elde edilir. Uyum katsayılarının hesaplanmasında kullanılan formüller dikkate alındığında klasik 2×2 tablo örnekleri, [Tablo 1](#) ve [Tablo 2](#)'de verildiği şekilde oluşturulmaktadır.

TABLO 1: Uyum katsayısı hesaplamaları için 2×2 tablo örneği.

		<i>Değerlendirici Y</i>		
		+	–	<i>Toplam</i>
<i>Değerlendirici X</i>	+	<i>a</i>	<i>b</i>	$a + b = n_{1+}$
	–	<i>c</i>	<i>d</i>	$c + d = n_{2+}$
	<i>Toplam</i>	$a + c = n_{+1}$	$b + d = n_{+2}$	$n = a + b + c + d$

TABLO 2: Uyum katsayısı hesaplamaları için 2×2 tablo örneği.

		<i>Değerlendirici Y</i>		
		+	–	<i>Toplam</i>
<i>Değerlendirici X</i>	+	$p_{11} = a/n$	$p_{12} = b/n$	$p_{1.}$
	–	$p_{21} = c/n$	$p_{22} = d/n$	$p_{2.}$
	<i>Toplam</i>	$p_{.1}$	$p_{.2}$	1

Çalışmalarda, genellikle 2 veya daha fazla değerlendiriciden elde edilen sonuçlar arasındaki uyumun, tesadüfi uyumdan veya önceden belirlenmiş herhangi bir uyum seviyesinden önemli ölçüde farklı olup olmadığını belirlemek istenmektedir.¹⁰ Uyum katsayılarının hesaplanması; a, b, c ve d frekanslarına bağlıdır. a, b, c ve d ile ifade edilen 4 frekans, X ve Y değerlendiricilerinin kararlarının ortak dağılımını karakterize eder. a ve d frekansları sırasıyla pozitif ve negatif eşleşmeler olarak adlandırılır, b ve c ise uyumsuzlukları gösterir. 2×2 tablolar için uyum katsayıları, 2 nesnenin birbiri ile ne ölçüde tutarlı olduklarını ölçen fonksiyonlardır. Uyum katsayıları bağımsız değişken olarak a, b, c ve d frekanslarını dikkate alan ve değişkenler arasındaki tutarlılık arttıkça daha yüksek sayısal değerlere döndüren işlevlerdir. Mevcut onlarca uyum katsayısı içerisinde hangi katsayının kullanılması gerektiği konusunda çok fazla kafa karışıklığı vardır. Uyum katsayıları, bazı istatistiksel parametrelerin tahmin edicileri değil, esas olarak tanımlayıcı katsayılar oldukları için özel katsayı türleridir. Çoğu uyum katsayısı için güvenilir güven aralıkları vermek zordur ve olası hatalar, yalnızca bir tür randomizasyon işlemi ile tahmin edilebilir. İkili veriler için tüm uyum katsayılarının aralığının 0 (uyum yok)-1,0 (tam uyum) olması beklenir. Ancak bu beklenti, tüm katsayılar için geçerli değildir. Tüm uygulama alanları dikkate alındığında uyum ölçülerinin taşınması beklenen 2 özelliği vardır. İlk olarak uyum ölçüsü, örnek büyüklüğünden bağımsız olmalıdır. İkinci olarak 2 topluluk örneği daha tutarlı hâle geldikçe, ölçü sabit bir minimumdan sabit bir maksimuma sorunsuz bir şekilde artmalıdır. İkili veriler için bugüne kadar literatürde yer alan her biri kendi matematiksel özelliklerine sahip olan ve farklı bilimsel alanlarda kullanılan 23 farklı uyum katsayısı belirlenmiştir. Çalışmada, ikili veriler için belirlenen 23 uyum katsayısı tanımlanmış ve birbirleriyle karşılaştırılmıştır. Bu çalışmanın amacı, türetilmiş veri setleri kullanarak, farklı n, a, b, c ve d değerleri için belirlenen 23 farklı ikili uyum katsayısını tanıtmak, özelliklerini ortaya koyarak değerlendirmektir.

GEREÇ VE YÖNTEMLER

Bu çalışmada, özellikle ikili veriler için kullanılması önerilen uyum katsayıları ele alınmıştır, çünkü ikili veriler, çeşitli verileri temsil etmek için yaygın olarak kullanılmaktadır. İkili veriler için geliştirilmiş uyum katsayıları evet/hayır, doğru/yanlış veya var/yok biçiminde veriler mevcut olduğunda kullanılır ve bu nedenle nominal ölçek için uygundur. Çalışmada, Python-random kütüphanesi kullanılarak, $10 \leq n \leq 1.000$ aralığında yer alan 35 farklı n değeri (10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1.000) 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 286, $n = 15$ için 815 ve $n \geq 20$ için 1.000'er farklı veri seti çalışmada kullanılmıştır. $a, b, c,$ ve d 'ye dayalı 23 farklı uyum katsayısına ait formüller [Tablo 3](#)'te verilmiştir.

Ayrıca çalışmada dikkate alınan uyum katsayılarının hangilerinin birbirine benzediğini belirlemek amacıyla hiyerarşik kümeleme yöntemi kullanılarak dendrogramlar çizilmiştir. Kümelerin belirlenmesinde yöntem olarak Centroid bağlantı yöntemi, uzaklık ölçüsü olarak ise karesel Öklid uzaklığı kullanılmıştır. Dendrogramların elde edilmesinde SPSS 20.0 (SPSS Inc., SPSS 20.0 for Windows, 2011) paket programından yararlanılmıştır.

Makalede, Helsinki Deklarasyonu Prensipleri dikkate alınmıştır.

TABLO 3: Uyum katsayıları.

No	Adı	Formül	Yıl	Değer aralığı
1	Andrés Marzotta ¹¹	$\delta = \frac{k(a+d) - n}{n}$	2004	$\langle -1, +1 \rangle$
2	Bangdiwala ¹²	$B = \frac{a^2 + d^2}{n_1 + n_{+1} + n_2 + n_{+2}}$	1985	$\langle 0, +1 \rangle$
3	Benini ¹³	$\beta = 1 - \frac{[(p_{12} + p_{21}) - p_{12} - p_{21}]}{[n_{+1}n_{2+} + n_1n_{+2} - p_{12} - p_{21}]}$	1901	$(-\infty, +1 \rangle$
4	Bennett ¹⁴	$S = \frac{k}{k-1} \left(\sum_{i=1}^k p_{ii} - \frac{1}{k} \right)$	1954	$\langle -1, +1 \rangle$
5	Brennan Prediger ¹⁵	$k_n = \frac{\sum_{i=1}^k p_{ii} - \frac{1}{k}}{1 - \frac{1}{k}}$	1981	$\langle -1, +1 \rangle$
6	Byrt ve ark. ¹⁶	$PABAK = \frac{(2t/n) - 0,5}{1 - 0,5}$ $t = \frac{a+d}{2}$	1993	$\langle -1, +1 \rangle$
7	Cohen ¹⁷	$\kappa = \frac{P_a - P_c}{1 - P_c}$ $P_c = p_{1.} * p_{.1} + p_{2.} * p_{.2}$ $P_a = \frac{a+d}{n}$	1960	$\langle -1, +1 \rangle$
8	Gini I ^{18,19}	$G_1 = \frac{\sum p_{ii} - \sum p_i p_i}{1 - \sum p_i p_i - \frac{1}{2} \sum p_i - p_i }$	1914-1915	$(-\infty, +1 \rangle$
9	Gini II ^{18,19}	$G_2 = \frac{\sum p_{ii} - \sum p_i p_i}{\sqrt{(1 - \sum p_i^2)(1 - \sum p_i^2)}}$	1914-1915	$\langle -1, +1 \rangle$
10	Gini III ^{18,19}	$G_3 = \frac{\sum p_{ii} - \sum p_i p_i}{1 - \frac{1}{2} \sum p_i^2 - \frac{1}{2} \sum p_i^2}$	1914-1915	$\langle -1, +1 \rangle$
11	Goodman Kruskal ²⁰	$\lambda = \frac{\sum_{i=1}^k p_{ii} - \frac{1}{2} \max(p_j + p_j)}{1 - \frac{1}{2} \max(p_j + p_j)}$	1954	$\langle -1, +1 \rangle$
12	Gwet ²¹	$AC_1 = \frac{P_a - P_c}{1 - P_c}$ $P_c = 2\Pi_1(1 - \Pi_1)$ $\Pi_1 = \frac{p_{1.} + p_{.1}}{2}$ $P_a = \frac{a+d}{n}$	2008	$\langle -1, +1 \rangle$

13	Holley Guilford ²²	$\kappa_2 = \frac{P_a - 0,5}{1 - 0,5}$ $P_a = \frac{a + d}{n}$	1964	< -1, +1 >
14	Janson Vegeius ²³	$C = \frac{2(a + d) - n}{n}$	1979	< -1, +1 >
15	Krippendorff ²⁴	$\alpha_k = \frac{P'_a - P_c}{1 - P_c}$ $P'_a = (1 - \varepsilon_n)P_a + \varepsilon_n$ $\varepsilon_n = \frac{1}{2n}$ $P_c = \Pi_1^2 + (1 - \Pi_1)^2$ $\Pi_1 = \frac{p_{1.} + p_{.1}}{2}$ $P_a = \frac{a + d}{n}$	1970	< -1, +1 >
16	Maxwell ²⁵	$RE = \frac{2(p_{11}p_{22} - p_{12}p_{21})}{p_{1.}p_{.2} + p_{1.}p_{.2}}$	1977	< -1, +1 >
17	Osgood Holst ^{26,27}	$CR = \frac{2(a + d)}{2n}$	1959, 1969	< 0, +1 >
18	Perreault Leigh ²⁸	$I_r = \sqrt{\left(\frac{a + d}{n} - \frac{1}{k}\right)\left(\frac{k}{k - 1}\right)}$	1989	< 0 + 1 >
19	Potter Levine- Donnerstein ²⁹	$Rdf - P_i = \frac{\sum_{i=1}^k p_{ii} - \left(\frac{1}{k^{n-1}}\right)}{1 - \left(\frac{1}{k^{n-1}}\right)}$	1999	< -1, +1 >
20	Rogot Goldberg ³⁰	$A_1 = \frac{1}{4}\left(\frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{c + d} + \frac{d}{b + d}\right)$	1966	< 0, +1 >
21	Rogot Goldberg II ³⁰	$A_2 = \frac{1}{2}\left(\frac{2a}{2a + b + c} + \frac{2d}{2d + b + c}\right)$	1966	< 0, +1 >
22	Scott ³¹	$\kappa_{scott} = \frac{P_a - P_c}{1 - P_c}$ $P_c = \Pi_1^2 + (1 - \Pi_1)^2$ $\Pi_1 = \frac{p_{1.} + p_{.1}}{2}$ $P_a = \frac{a + d}{n}$	1955	< -1, +1 >
23	Uyum yüzdesi ³²	$a_0 = \frac{a + d}{n}$	Pre 1901	< 0, +1 >

k: Kategori sayısı.

BULGULAR

Bu çalışmada, dikkate alınan uyum katsayılarının çoğu iyi bilinmekte ve yaygın olarak kullanılmaktadır. Herhangi bir model kullanmadan kapsamlı ve varsayımsal olarak rastgele elde edilen uyum matrislerine dayalı 23 farklı uyum katsayısına ait değerler elde edilmiş ve karşılaştırılmıştır. n , a , b , c ve d 'nin farklı değerleri için elde edilen sonuçlar [Tablo 4](#)'te verilmiştir.

Otuz beş farklı durum için elde edilen uyum katsayılarına ait değerler dikkate alınarak dendrogramlar çizilmiştir. Her bir örnek büyüklüğü için ayrı ayrı elde edilen dendrogramlar incelenerek, birbirine benzeyen uyum katsayıları belirlenmiştir. Otuz beş farklı n değeri için elde edilen dendrogramlar, sayfa sayısı fazlalığından dolayı ayrı ayrı çalışmada sunulmamıştır. Ancak dendrogramlardan elde edilen sonuçlar, [Tablo 5](#)'te verilmiştir.

TABLO 4: Özel durumlar için uyum değerleri.

	$a = n$	$b = n$	$c = n$	$d = n$	$a + d = n$	$b + c = n$	$a = b = c = d$	$a + d$
Andrés Marzo	1	-1	-1	1	1	-1	0	C_{a+d}
Bangdiwala	1	NC	NC	1	1	0	0,25	C_{a+d}
Benini	NC	NC	NC	NC	1	-1	0	C_{a+d}
Bennett	1	-1	-1	1	1	-1	0	C_{a+d}
Brennan Prediger	1	-1	-1	1	1	-1	0	C_{a+d}
Byrt ve ark.	1	-1	-1	1	1	-1	0	C_{a+d}
Cohen	NC	0	0	NC	1	C	0	C_{a+d}
Gini I	NC	NC	NC	NC	1	C	0	C_{a+d}
Gini II	NC	NC	NC	NC	1	-1	0	C_{a+d}
Gini III	NC	NC	NC	NC	1	-1	0	C_{a+d}
Goodman Kruskal	1	-1	-1	NC	1	-1	0	C_{a+d}
Gwet	1	-1	-1	1	1	-1	0	C_{a+d}
Holley Guilford	1	-1	-1	1	1	-1	0	C_{a+d}
Janson Vegelius	1	-1	-1	1	1	-1	0	C_{a+d}
Krippendorff	NC	C_n	C_n	NC	1	C_n	C_n	C_{a+d}
Maxwell	NC	NC	NC	NC	1	-1	0	C_{a+d}
Osgood Holsti	1	0	0	1	1	0	0,5	C_{a+d}
Perreault Leigh	1	NC	NC	1	1	NC	0	C_{a+d}
Potter Levine	1	-1	-1	1	1	-1	0	C_{a+d}
Rogot Goldberg I	NC	NC	NC	NC	1	0	0,5	C_{a+d}
Rogot Goldberg II	NC	0	0	NC	1	0	0,5	C_{a+d}
Scott	NC	-1	-1	NC	1	-1	0	C_{a+d}
Uyum yüzdesi	1	0	0	1	1	0	0,5	C_{a+d}

NC: Hesaplanamaz; C: n , b ve c değerlerine bağlı olarak değeri büyük/küçülür; C_n : n sayısına göre farklı değerler alır, n sayısı büyüdükçe değeri büyür; C_{a+d} : $(a + d)$ değeri arttıkça değeri büyür.

TABLO 5: Hiyerarşik kümeleme analizi sonuçları.

Küme no	Yöntemler
1	Andrés Marzo, Bennett, Brennan Prediger, Byrt ve ark., Holley Guilford, Janson Vegelius, Potter Levine-Donnerstein
2	Krippendorff, Scott, Cohen, Gini II, Gini III, Maxwell
3	Osgood Holsti, uyum yüzdesi
4	Rogot Goldberg I, Rogot Goldberg II
5	Benini, Gini I
6	Gwet
7	Bangdiwala
8	Perreault Leigh
9	Goodman Kruskal

TARTIŞMA

Değerlendiricilerden elde edilen sonuçlar arasındaki uyumu belirlemek için kullanılacak çok sayıda yöntem söz konusudur ve araştırmacıların, en uygun olanı seçmeye büyük özen göstermeleri gerekir. En uygun uyum ölçüsünün hangisi olduğu konusunda bir yargıya ulaşmak için katı ve hızlı bir kural veya nesnel bir kriter bulunmamaktadır. Popping tarafından yapılan çalışmada, uyum katsayılarının taşınması gereken 10 farklı özellik dikkate alınarak uyum katsayıları birbirleri ile karşılaştırılmış, Cohen kappa katsayısının bu özelliklerin tamamını sağladığı ifade edilmiştir.⁶ Zhao ve ark. tarafından yapılan çalışmada, uyum katsayılarından elde edilen tahminler, liberal ve tutucu hiyerarşiler bakımından karşılaştırılmış, uyum yüzdesi, Osgood Holsti ve Rogot Goldberg I katsayılarının en liberal, Goodman Kruskal katsayısının ise en tutucu olduğu sonucuna ulaşmışlardır.³² Feng, uyum katsayılarını duyarlılık ve özgüllük bakımından karşılaştırmış, Maxwell RE katsayısının diğer katsayılardan daha üstün olduğunu belirlemiştir.⁷

SONUÇ

Bu çalışmada, her bir uyum katsayısı ayrı ayrı ele alınmış ve özellikleri ortaya konulmuştur. İkili veriler için uyum katsayılarının $0 \leq \text{uyum katsayısı} \leq 1$ aralığında değer alması beklenmesine rağmen tüm katsayılar için bu aralık geçerli değildir. Dikkate alınan 23 farklı katsayı içerisinde yalnızca 6 tanesi bu aralıkta değer almaktadır. Hiyerarşik kümeleme analizine göre uyum katsayıları birbirine benzememektedir. Özellikle Gwet, Bangdiwala, Perreault Leigh ve Goodman Kruskal katsayıları, tüm n değerleri için diğer katsayılardan farklılık göstermektedir. Uyum katsayılarının hiçbiri n sayısından etkilenmemektedir. Dolayısıyla uyum katsayılarının tamamının örnek büyüklüğünden bağımsız oldukları belirlenmiştir. Genel olarak hemen hemen tüm katsayılar a ait değerler, örnekler daha benzer hâle geldikçe sabit bir minimumdan sabit bir maksimuma doğru artmaktadır. $a + d$ değeri arttıkça Bangdiwala, Benini, Cohen, Gini I, Gini II, Gini III, Maxwell, Perreault Leigh ve Rogot Goldberg I katsayılarının değerleri doğrusal olmayan artış göstermekte, Andrés Marzo, Bennett, Brennan Prediger, Byrt ve ark., Gwet, Janson Vegelius, Osgood Holsti, Potter Levine-Donnerstein ve Uyum Yüzdesi katsayıları ise, tüm n değerleri için uyum ile doğrusal olarak artmaktadır. Goodman Kruskal, Krippendorff, Rogot Goldberg II ve Scott katsayıları ise $n \leq 50$ için doğrusal, $n > 50$ için ise doğrusal olmayan bir artış göstermektedir. Değer aralığının $0 - 1$ olması ve uyum artışı ile doğrusal olarak artmasından dolayı Osgood Holsti ve uyum yüzdesi katsayıları, tüm katsayılar içerisinde öne çıkmaktadır.

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 etkileyecek 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

Fikir/Kavram: İsmet Doğan, Nurhan Doğan; **Tasarım:** İsmet Doğan, Nurhan Doğan; **Denetleme/Danışmanlık:** İsmet Doğan, Nurhan Doğan, Taylan Doğan; **Veri Toplama ve/veya İşleme:** İsmet Doğan, Nurhan Doğan, Taylan Doğan; **Analiz ve/veya Yorum:** İsmet Doğan, Nurhan Doğan, Taylan Doğan; **Kaynak Taraması:** İsmet Doğan, Nurhan Doğan; **Makalenin Yazımı:** İsmet Doğan, Nurhan Doğan, Taylan Doğan; **Eleştirel İnceleme:** İsmet Doğan, Nurhan Doğan, Taylan Doğan; **Kaynaklar ve Fon Sağlama:** İsmet Doğan, Nurhan Doğan; **Malzemeler:** İsmet Doğan, Nurhan Doğan.

KAYNAKLAR

1. Costa-Santos C, Bernardes J, Ayres-de-Campos D, Costa A, Amorim-Costa C. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. *J Clin Epidemiol*. 2011;64(3):264-9. Erratum in: *J Clin Epidemiol*. 2011;64(6):703. Costa, Célia [corrected to Amorim-Costa, Célia]. Erratum in: *J Clin Epidemiol*. 2011;64(9):1049. [[Crossref](#)] [[PubMed](#)]
2. Bloch DA, Kraemer HC. 2x2 kappa coefficients: measures of agreement or association. *Biometrics*. 1989;45(1):269-87. [[Crossref](#)] [[PubMed](#)]
3. Teles J. Concordance coefficients to measure the agreement among several sets of ranks. *J Appl Stat*. 2012;39(8):1749-64. [[Crossref](#)]
4. Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry*. 2016;28(2):115-20. [[PubMed](#)] [[PMC](#)]
5. Adejumo AO, Heumann C, Toutenburg H. A review of agreement measure as a subset of association measure between raters. *Collaborative Research Center* 386. 2004:Discussion Paper 385. [[Link](#)]
6. Popping R. On agreement indices for nominal data. In: Saris WE, Gallhofer IN, eds. *Sociometric Research: Vol.1: Data Collection and Scaling*. 1st ed. London: The Macmillan Press Ltd.; 1988. p.90-105. [[Link](#)]
7. Feng GC. Factors affecting intercoder reliability: a Monte Carlo experiment. *Qual Quant*. 2013;47:2959-82. [[Crossref](#)]
8. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol*. 2016;16:93. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
9. Aickin M. Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa. *Biometrics*. 1990;46(2):293-302. [[Crossref](#)] [[PubMed](#)]
10. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychol Bull*. 1971;76(5):365-77. [[Crossref](#)]
11. Andrés AM, Marzo PF. Delta: a new measure of agreement between two raters. *Br J Math Stat Psychol*. 2004;57(Pt 1):1-19. [[Crossref](#)] [[PubMed](#)]
12. Bangdiwala S. A graphical test for observer agreement. In: *Proceedings of the 45th International Statistical Institute Meeting*. August 1985; Amsterdam; 1985. [[Link](#)]
13. Benini R. *Principii di demographia. Manuali Barbèra di scienze giuridiche, sociali e politiche*. 1st ed. Serie teorica; 29. Firenze: G. Barbèra; 1901. [[Link](#)]
14. Bennett EM, Alpert R, Goldstein AC. Communications through limited-response questioning. *Publ Opin Q*. 1954;18:303-8. [[Crossref](#)]
15. Brennan RL, Prediger DJ. Coefficient kappa: some uses, misuses, and alternatives. *Educ Psychol Meas*. 1981;41(3):687-99. [[Crossref](#)]
16. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-9. [[Crossref](#)]
17. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46. [[Crossref](#)]
18. Gini C. Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Series 8*. 1914-1915; 74: 583-610. [[Link](#)]
19. Gini C. Nuovi contributi alla teoria delle relazioni statistiche. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Series 8*. 1914-1915; 74: 1903-42. [[Link](#)]
20. Goodman LA, Kruskal WH. Measures of association for cross classifications. *J Am Stat Assoc*. 1954;49(268):732-64. [[Crossref](#)]
21. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61(Pt 1):29-48. [[Crossref](#)] [[PubMed](#)]
22. Holley JW, Guilford JP. A note on the G index of agreement. *Educ Psychol Meas*. 1964;24(4):749-53. [[Crossref](#)]
23. Janson S, Vegelius J. On generalizations of the G index and the phi coefficient to nominal scales. *Multivariate Behav Res*. 1979;14(2):255-69. [[Crossref](#)] [[PubMed](#)]
24. Krippendorff K. Bivariate agreement coefficients for reliability data. In: Borgatta ER, Bohmstedt GW, eds. *Sociological Methodology*. 1st ed. San Francisco: Jossey-Bass; 1970. p.139-50. [[Crossref](#)]
25. Maxwell AE. Coefficients of agreement between observers and their interpretation. *Br J Psychiatry*. 1977;130:79-83. [[Crossref](#)] [[PubMed](#)]
26. Osgood CE. The representational model and relevant research. In: de Sola Pool I, ed. *Trends in Content Analysis*. 1st ed. Urbana: University of Illinois Press; 1959. p.33-8.
27. Holsti OR. *Content Analysis for the Social Sciences and Humanities*. 1st ed. Reading, Mass: Addison-Wesley; 1969.
28. Perreault WD, Leigh LE. Reliability of nominal data based on qualitative judgments. *J Mark Res*. 1989;26(2):135-48. [[Crossref](#)]
29. Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. *J Appl Commun Res*. 1999;27(3):258-84. [[Crossref](#)]
30. Rogot E, Goldberg ID. A proposed index for measuring agreement in test-retest studies. *J Chronic Dis*. 1966;19(9):991-1006. [[Crossref](#)] [[PubMed](#)]
31. Scott WA. Reliability of content analysis: The case of nominal scale coding. *Publ Opin Q*. 1955;19(3):321-5. [[Crossref](#)]
32. Zhao X, Liu JS, Deng K. Assumptions behind Intercoder Reliability Indices. In: Salmon CT, ed. *Communication Yearbook 36*. 1st ed. New York: Routledge; 2013. p.419-80. [[Crossref](#)]