

İki Sonuçlu Nitel Veriler İçin Cohen Kappa Katsayısı ve Ayırt Edilebilirliğin Değerlendirilmesi: Bir Simülasyon Çalışması

Evaluation of Cohen Kappa Coefficient and Distinguishability for Binary Data: A Simulation Study

İsmet DOĞAN^a, Nurhan DOĞAN^a

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

ÖZET Amaç: Bu çalışmanın amacı, 2 sonuçlu nominal derecelendirme ölçekleri için Cohen kappa uyum katsayısı ile kategori ayırt edilebilirlik derecesini birlikte değerlendirmektir. **Gereç ve Yöntemler:** Çalışmada, Phyton-random kütüphanesi kullanılarak $10 \leq n \leq 1000$ aralığında yer alan 35 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 286, $n = 15$ için 815 ve $n \geq 20$ için 1000'er farklı veri seti çalışmada kullanılmıştır. **Bulgular:** Literatürde ayırt edilebilirlik derecesinin 0-1 aralığında değer aldığı ifade edilmektedir. Ancak 0 değeri içeren ya da ekstrem durumların söz konusu olduğu tablolarda ayırt edilebilirlik derecesinin tanım aralığının dışında, negatif değer aldığı belirlenmiştir. Bunun yanı sıra literatürde ayırt edilebilirlik derecesinin nasıl yorumlanacağı ile ilgili çok genel bilgiler dışında herhangi bir bilgi yer almamaktadır. Dolayısıyla Cohen kappa katsayısı için literatürde yer alan yorumlama seviyeleri esas alınarak, ayırt edilebilirlik derecesi için yorumlama seviyeleri belirlenmiştir. Cohen kappa katsayısı için simülasyon çalışmasından elde edilen toplam 34.097 sonucun %49,5'inde uyum yok, %22,1'inde önemsiz, %13,0'ında orta derecede, %8,3'ünde ekseriyetle, %5,2'sinde önemli derecede ve %2,0'ında neredeyse mükemmel uyum sonucu ile karşılaştırılmıştır. Ayırt edilebilirlik derecesi için ise %50,3'ünde ayırt edilemez, %12,4'ünde önemsiz, %11,8'inde orta derecede, %9,9'unda ekseriyetle, %7,6'sında önemli derecede ve %8,1'inde neredeyse mükemmel ayırt edilebilirlik sonucu ile karşılaştırılmıştır. Her bir örneklem büyüklüğü ayrı ayrı dikkate alındığında da elde edilen sonuçlar benzerlik göstermiştir. **Sonuç:** Kategoriler arasında ayırt edilebilirlik söz konusu değilse değerlendiriciler arasında kesinlikle uyum aranmamalıdır. Kategorilerin önemli ya da mükemmel derecede ayırt edilebilir olması da değerlendiriciler arasındaki uyumun her zaman önemli ya da mükemmel derecede yüksek olacağını göstermez.

ABSTRACT Objective: The aim of this study is to evaluate the Cohen kappa coefficient of agreement and the degree of category distinguishability for binary nominal rating scales. **Material and Methods:** In the study, data were derived for 35 different n values in the range of $10 \leq n \leq 1000$ using the Phyton-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 1000 different data sets for $n \geq 20$ were used in the study. **Results:** In the literature, it is stated that the degree of distinguishability takes a value in the range of 0-1. However, it was determined that the degree of distinguishability took a negative value outside the definition range in tables containing zero values or in extreme situations. In addition, there is no information in the literature other than very general information on how to interpret the degree of distinguishability. Therefore, interpretation levels for the degree of distinguishability were determined based on the interpretation levels in the literature for the Cohen kappa coefficient. Of the total 34,097 results obtained from the simulation study for the Cohen kappa coefficient; 49.5% were disagreement, 22.1% poor, 13.0% fair, 8.3% moderate, 5.2% substantial, and 2.0% perfect. For the degree of distinguishability; 50.3% were indistinguishable, 12.4% poor, 11.8% fair, 9.9% moderate, 7.6% substantial and 8.1% perfect. When each sample size was considered separately, the results obtained were similar. **Conclusion:** If there is no distinguishability between the categories, agreement should not be sought between the raters. The fact that the categories are important or perfectly distinguishable does not necessarily indicate that the agreement between the raters will be significant or perfectly high.

Anahtar kelimeler: Cohen kappa katsayısı; ayırt edilebilirlik; iki sonuçlu veri

Keywords: Cohen kappa coefficient; distinguishability; binary data

Correspondence: Nurhan DOĞAN

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

E-mail: nurhandogan@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

Received: 24 Feb 2022

Received in revised form: 26 May 2022

Accepted: 21 Jun 2022

Available online: 13 Sep 2022

2146-8877 / Copyright © 2022 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/>).

Değerlendiriciler arasındaki uyum, eğitim ve psikolojide olduğu kadar tıpta da iyi bilinen ve hâlen ilgi duyulan bir konudur. Değerlendiricilerin kullandıkları ölçülerin geçerli olduğundan emin olmaları gerekir ve değerlendiriciler arası güvenilirliğin zayıf olması güven eksikliğine yol açar.¹ Literatürde, 2 veya daha fazla değerlendirici tarafından verilen cevaplar arasındaki veya 2 veya daha fazla ölçüm cihazından elde edilen ölçümler arasındaki uyumu belirlemede yararlanılan birçok yöntem bulunmaktadır.² İki veya daha fazla ölçüm cihazı ya da 2 veya daha fazla değerlendiriciden elde edilen sonuçlar aynı olmak zorunda değildir. Bu durum ile özellikle derecelendirme ölçeğinin kullanıldığı durumlarda karşılaşmaktadır. Derecelendirmeler arasındaki farklılıklar, yalnızca değerlendiriciler tarafından yapılan sınıflandırma hatalarına değil, aynı zamanda nesnel olarak kesin bir tanımı olmayan kategorilere de atfedilebilir. Nominal ölçeğin kullanılması durumunda da farklı değerlendiriciler, kategorilerin ne anlama geldiği konusunda farklı algılara sahip olabilir ya da ortak bir algı olsa bile değerlendirme sonucunda verilen kararlar arasında farklılıklar oluşabilir. Belirli bir öznenin aynı değerlendirici tarafından tekrarlanan değerlendirmeleri dahi farklılık gösterebilir. Değerlendiriciler arası uyum analizinde ortaya çıkan sorunlar, aynı değerlendiricinin farklı değerlendirmeleri için yapılacak uyum analizi için de geçerlidir. Birçok tıbbi uygulamada değerlendirici uyumunun incelenmesi çok önemlidir. Bir semptomun varlığı veya bir hastalığın ciddiyeti ile ilgili değerlendirme sonuçları, özellikle ölçek sıralı olduğunda (örneğin bir semptomun varlığı olası değil, biraz olası, olası, çok olası ya da hastalığın şiddeti yok, hafif, orta, şiddetli gibi yanıtlara sahip kategorik bir ölçek kullanılarak ölçülebilir) çoğunlukla öznedir.³ Kategori ayırt edilebilirliğine ilişkin temel fikirler ilk kez Darroch ve McCloud tarafından ileri sürülmüştür. Onlara göre uygulamada kappa ve benzeri uyum katsayılarına olan ilgi, 2 değerlendiricinin birbiriyle olan uyumunu belirlemekten daha çok herhangi bir değerlendiricinin kategorileri birbirinden ne kadar iyi ayırt edebildiğini belirlemekten kaynaklanmaktadır. Birçok durumda öznel sınıflandırıldığı kategoriler, kesin nesnel tanımlara sahip değildir. Bu nedenle farklı değerlendiricilerin kategori tanımlarını farklı yorumlayabileceğini ve kategorilerin aynı değerlendirici tarafından bile birbirinden tamamen ayırt edilemeyeceğini kabul etmek gerekir.⁴ Darroch ve McCloud (1986) 2 değerlendirici tarafından rastgele seçilen bir konunun ortak sınıflandırma olasılıklarına yönelik modelleri aşağıda verilen özellikleri taşımaktadır;⁵

- Belirli bir konunun belirli bir değerlendirici tarafından sınıflandırılmasının rastgele olmasına izin verilir.
- Farklı değerlendiriciler aynı konu için farklı sınıflandırma olasılıklarına sahip olabilir.
- Sınıflandırma olasılıklarında konu etkileri ile değerlendirici etkileri arasında herhangi bir çarpımsal etkileşimin olduğu varsayılmaz.

Uyum çalışmalarında kategorilerin birbirinden ayırt edilebilir olup olmadığının belirlenmesi gerekmektedir. Kategoriler ayırt edilemez ise değerlendiricilerin algıları arasında bazı farklılıklar olabilir. Farklı değerlendiriciler kategorileri farklı anlayabilir veya aynı değerlendirici kategorileri doğru ayırt edemeyebilir. Bu 2 sorunun, değerlendiricilerin kendi alanlarında uzman olmamaları veya kategorileri ayırt etmenin zor olması nedeniyle oluşabileceği tartışılmaktadır.⁶ Uyum çalışmalarında, nesnel 2 değerlendirici tarafından (veya aynı değerlendirici tarafından 2 kez) bağımsız olarak değerlendirildiğinde, 2 kategorideki değerlendirmeler arasında bu 2 kategorinin bu değerlendiriciler için ayırt edilebilirliğini yansıtan bir ilişki ortaya çıkar. Kategorilerin ayırt edilebilirlik düzeyini hesaplamaya yönelik ölçü, ayırt edilebilirlik derecesi olarak adlandırılmaktadır.⁷ Kategorik ölçekler için değerlendiriciler arası uyumun; kategorilerin ayırt edilebilirliği ve ön yargı eksikliği olarak ifade edilen 2 bileşeni vardır. Bir değerlendiricinin diğerine göre ön yargısı, marjinal dağılımlar arasındaki tutarsızlıkları ifade eder. Marjinal dağılımlar daha eşit hâle geldikçe ön yargı azalır, ön yargı eksikliği her bir kategori için $p_{i+} = p_{+i}$ ($i = 1, 2$ olmak üzere i 'nci satır toplamı, i 'nci sütun toplamına eşit olacak) anlamına gelir. Buradaki "yanlılık" terimi, A ve B değerlendiricileri için marjinal heterojenliğe atıfta bulunur. Bunlar, derecelendirmeler arasındaki ilişkinin gücü ve marjinal dağılımlarının benzerliği ile ilgilidir. Nispeten büyük bir uyum olasılığı, hem güçlü bir ilişki hem de neredeyse marjinal homojenlik gerektirir. Darroch ve McCloud, uyumun kısmen her bir değerlendiricinin her bir kategori çiftini ne kadar iyi ayırt edebildiğine bağlı olduğunu belirtmiştir.³ Bu çalışmada, 2 sonuçlu nominal derecelendirme ölçeklerinde Cohen kappa uyum katsayısı ile kategori ayırt edilebilirliğinin birlikte değerlendirilmesi amaçlanmıştır. Kategori ayırt edilebilirliğinin güvenilirlik seviyesini nasıl etkilediği ve

düşük ayırt edilebilirliğin olası çözümlerinin neler olduğu incelenmiştir. Simülasyon verileri kullanılarak ayırt edilebilirlik derecesi ve kapa katsayıları hesaplanmış ve sonuçlar örneklem büyüklüğü üzerinden tartışılmıştır. Makalede, Helsinki Deklarasyonu Prensipleri dikkate alınmıştır.

GEREÇ VE YÖNTEMLER

Çalışmada, 2 (değerlendirici) \times 2 (kategoriler) tablolar için Cohen kapa ve ayırt edilebilirlik ölçülerini karşılaştırmak için bir simülasyon çalışması yapılmıştır. Çalışmada Phytion-random kütüphanesi kullanılarak $10 \leq n \leq 1000$ 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. $n = 10$ için 286, $n = 15$ için 815 ve $n \geq 20$ için 1.000'er farklı veri seti çalışmada kullanılmıştır. Veri üretimine ilişkin Python 3.9.10 programlama dili (Python Software Foundation, USA) kullanılarak yazılan programda yer alan simülasyon kurgusunun detayları aşağıdaki gibidir:

Adım 1. Veri üretimi 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 30 ise (0, 30) aralığında rastgele bir tam sayı seçilir.

Adım 4. Tüm kategoriler için 2 ve 3. adımlar tekrarlanır (k' inci) kategori için bir tam sayı seçildikten sonra seçilen sayıların toplamına bakılır. Eğer bu toplam, istenilen toplamdan fazla ise bu geçersiz bir veri olduğundan 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 1. adımdan itibaren program bir döngü içinde tekrar çalıştırılır. (Bu çalışmada, standardizasyonu sağlamak amacıyla her bir örneklem genişliği için 1.000 adet veri türetilmesi planlanmış ancak 4 kategori için toplamı 10 veya 15 olan ve tekrar etmeyen 1.000 tane veri bulmak imkânsız olduğundan $n = 10$ için 216 ve $n = 15$ için 815 adet veri türetilmiştir.)

Adım 7. Son olarak türetilen verilerden eşitlik 1 ve eşitlik 4 kullanılarak uyum ve ayırt edilebilirlik katsayıları hesaplanır.

Uyum ve ayırt edilebilirlik katsayılarının hesaplanmasında kullanılan formüller dikkate alındığında klasik 2×2 tablo örnekleri [Tablo 1a-1b](#)'de verildiği şekilde oluşturulmaktadır.

TABLO 1a: Uyum ve ayırt edilebilirlik derecesi hesaplamaları için 2×2 tablo örneği.

		Değerlendirici Y		
		+	-	Toplam
Değerlendirici X	+	a	b	$a + b = n_{1+}$
	-	c	d	$c + d = n_{2+}$
	Toplam	$a + c = n_{+1}$	$b + d = n_{+2}$	$n = a + b + c + d$

TABLO 1b: Uyum ve ayırt edilebilirlik derecesi hesaplamaları için 2×2 tablo örneği.

		Değerlendirici Y		
		+	-	Toplam
Değerlendirici X	+	$p_{11} = a/n$	$p_{12} = b/n$	p_{1+}
	-	$p_{21} = c/n$	$p_{22} = d/n$	p_{2+}
	Toplam	p_{+1}	p_{+2}	1

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. $a, b, c, ve d$ 'ye dayalı Cohen kappa katsayısının değeri eşitlik 1 kullanılarak elde edilmektedir.⁸

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (1)$$

$$P_e = p_{1+} * p_{+1} + p_{2+} * p_{+2}$$

$$P_0 = \frac{a + d}{n}$$

İki sonuçlu veriler için [Tablo 1b](#)'de yer alan olasılık değerleri kullanılarak bazı olasılık oranlarının analizi yoluyla kategori ayırt edilebilirliği kavramı ortaya konulmuştur. Darroch ve McCloud (1986), I kategorili bir sıralı derecelendirme ölçeğini ele alarak i 'nci satır j 'nci sütuna ait olasılık değerini p_{ij} ile göstermişler, p_{ij} değerleri ile ilgili olarak;

$$\Delta = \left\{ (p_{ij})_{cR^{I \times I}} : p_{ij} > 0, \sum_{i,j} p_{ij} = 1 \right\} \quad (2)$$

tanımlamasını yapmışlardır. Darroch ve McCloud'a (1986) göre i 'nci ve j 'nci kategoriler arasındaki ayırt edilebilir olma olasılığı eşitlik 3 kullanılarak hesaplanır.

$$\tau_{ij} = \frac{p_{ii}p_{jj}}{p_{ij}p_{ji}} \quad i, j = 1, 2, \dots, I, \text{ ve } i \neq j \text{ için}, \quad (3)$$

$\tau_{ij} = 1$ olması i 'nci ve j 'nci kategorilerin ayırt edilebilir olmadığını gösterir. Çapraz tablonun köşegen dışı öğelerinin küçük değerler alması ayırt edilebilirliğin yüksek olmasını sağlar. Bu nedenle ayırt edilebilirlik değerinin büyük olması, güçlü bir uyuma karşılık gelir. i 'nci ve j 'nci kategorilere ait ayırt edilebilirlik derecesi,

$$\delta = 1 - \tau_{ij}^{-1} \quad (4)$$

eşitliği ile verilmektedir. Dikkat edilmelidir ki $\tau_{ij} = 1$ olması yalnızca

$$p_{ii}p_{jj} - p_{ij}p_{ji} = 0. \quad (5)$$

olması durumunda mümkündür. Bu tanımdan da anlaşılacağı üzere ana köşegende yer alan değerlerin belli bir değerden küçük olması durumunda i 'nci ve j 'nci kategoriler ayırt edilemez. Eşitlik 5 aynı zamanda $I \times I$ tablolarda (i, i) , (i, j) , (j, i) , (j, j) şeklinde oluşturulabilecek tüm alt tabloların bağımsız olduklarını da göstermektedir.⁹ Eğer olumsuzluk tablosu 0 değeri içeriyorsa, bu durumda

$$\tau_{ij} = \frac{(p_{ii} + c) * (p_{jj} + c)}{(p_{ij} + c) * (p_{ji} + c)} \quad i, j = 1, 2, \dots, I, \text{ ve } i \neq j \text{ için}, \quad (6)$$

düzeltilmesi yapılmalıdır. Eşitlik 6'da yer alan c sabiti düzeltme faktörüdür. Sıfır değeri içeren çalışmalarda c sabiti için 0,2, 0,5 ya da 0'dan farklı en küçük değerin alınabileceği belirtilmektedir.¹⁰

BULGULAR

Cohen kappa katsayısı için literatürde yer alan yorumlama seviyeleri esas alınarak ayırt edilebilirlik derecesi için yorumlama seviyeleri belirlenmiş ve [Tablo 2](#)'de verilmiştir.

TABLO 2: Cohen kappa katsayısı ve ayırt edilebilirlik derecesine ait yorumlama seviyeleri.

Cohen kappa katsayısı		Ayırt edilebilirlik katsayısı	
Uyum yok	< 0	Ayırt edilemez	$\delta \leq 0$
Önemsiz uyum	0,00 – 0,20	Önemsiz ayırt edilebilirlik	$0,00 < \delta \leq 0,55$
Orta derecede uyum	0,21 – 0,40	Orta derecede ayırt edilebilirlik	$0,56 \leq \delta \leq 0,81$
Ekseriyetle uyum	0,41 – 0,60	Ekseriyetle ayırt edilebilirlik	$0,82 \leq \delta \leq 0,93$
Önemli derecede uyum	0,61 – 0,80	Önemli derecede ayırt edilebilirlik	$0,94 \leq \delta \leq 0,98$
Neredeyse mükemmel uyum	0,81 – 1,00	Neredeyse mükemmel ayırt edilebilirlik	$0,99 \leq \delta \leq 1,00$

Her bir örneklem büyüklüğü ayrı ayrı dikkate alındığında da elde edilen sonuçlar benzerlik göstermiştir. Simülasyon çalışmasından elde edilen toplam 34.101 sonuçtan 4 tanesinde ($n = 10$ için 2 tane, $n = 15$ için 2 tane) Cohen kappa katsayısının hesaplanmasında 0/0 belirsizliği ile karşı karşıya kaldığı için verilerin 34.097'si kullanılarak Cohen kappa katsayısı ve ayırt edilebilirlik derecesi yorumlama seviyeleri birbirleri ile karşılaştırılmış, elde edilen sonuçlar [Tablo 3a-3b](#)'de verilmiştir.

TABLO 3a: Cohen kappa ve ayırt edilebilirlik derecesinin yorumlama seviyeleri bakımından karşılaştırılması.

		Ayırt edilebilirlik derecesi						Toplam
		Ayırt edilemez	Önemsiz ayırt edilebilirlik	Orta derecede ayırt edilebilirlik	Ekseriyetle ayırt edilebilirlik	Önemli derecede ayırt edilebilirlik	Neredeyse mükemmel ayırt edilebilirlik	
Cohen kappa katsayısı	Uyum yok	16.863	2	4	6	2	0	16.877 %49,5
		99,92	0,012	0,024	0,036	0,012	0,0	
		98,36	0,047	0,099	0,18	0,077	0,0	
		49,46	0,006	0,012	0,018	0,006	0,0	
	Önemsiz uyum	282	4.213	1.848	393	220	577	7.533 %22,1
		3,74	55,93	24,53	5,22	2,92	7,66	
		1,64	99,93	45,9	11,7	8,47	20,95	
		0,83	12,36	5,42	1,15	0,65	1,69	
	Orta derecede uyum	0	1	2.174	1.517	280	454	4.426 %13,0
		0,0	0,023	49,12	34,27	6,33	10,26	
		0,0	0,024	54,0	45,16	10,78	16,49	
		0,0	0,0029	6,38	4,45	0,82	1,33	
	Ekseriyetle uyum	0	0	0	1.442	972	420	2.834 %8,3
		0,0	0,0	0,0	50,88	34,3	14,82	
		0,0	0,0	0,0	42,93	37,43	15,25	
		0,0	0,0	0,0	4,23	2,85	1,23	
	Önemli derecede uyum	0	0	0	1	1.122	637	1.760 %5,2
		0,0	0,0	0,0	0,057	63,75	36,19	
		0,0	0,0	0,0	0,029	43,2	23,13	
		0,0	0,0	0,0	0,0029	3,29	1,87	
	Neredeyse mükemmel uyum	0	0	0	0	1	666	667 %2,0
0,0		0,0	0,0	0,0	0,15	99,85		
0,0		0,0	0,0	0,0	0,04	24,18		
0,0		0,0	0,0	0,0	0,0029	1,95		
Toplam		17.145 %50,3	4.216 %12,4	4.026 %11,8	3.359 %9,9	2.597 %7,6	2.754 %8,1	34.097 %100

Not: Her bir hücrede yer alan ilk değer sayı, 2. değer satır içindeki yüzde, 3. değer sütun içindeki yüzde, 4. değer ise toplam içindeki yüzde değerini göstermektedir.

TABLO 3b: Yorumlama seviyelerine göre Cohen kappa ve ayırt edilebilirlik derecesine ait tanımlayıcı istatistikler.

		Ayırt edilebilirlik derecesi					
		Ayırt edilemez	Önemsiz ayırt edilebilirlik	Orta derecede ayırt edilebilirlik	Ekseriyetle ayırt edilebilirlik	Önemli derecede ayırt edilebilirlik	Neredeyse mükemmel ayırt edilebilirlik
Cohen kappa katsayısı	Uyum yok	16.863 -0,26±0,20 -57919,05±4212588,41	2 -0,23±0,01 0,20±0,04	4 -0,13±0,07 0,69±0,06	6 -0,14±0,05 0,88±0,04	2 -0,04±0,03 0,97±0,01	0 - -
	Önemsiz uyum	282 0,0009±0,004 -52,15±346,26	4.213 0,06±0,04 0,31±0,15	1.848 0,13±0,05 0,67±0,07	393 0,12±0,06 0,87±0,03	220 0,06±0,05 0,97±0,01	577 0,11±0,05 0,99±0,002
	Orta derecede uyum	0 - -	1 0,2007±0,00 0,5588±0,00	2174 0,27±0,04 0,73±0,06	1517 0,31±0,05 0,87±0,03	280 0,32±0,05 0,95±0,01	454 0,29±0,05 0,99±0,001
	Ekseriyetle uyum	0 - -	0 - -	0 - -	1442 0,47±0,05 0,90±0,02	972 0,52±0,05 0,96±0,01	420 0,50±0,05 0,99±0,002
	Önemli derecede uyum	0 - -	0 - -	0 - -	1 0,6003±0,00 0,9388±0,00	1122 0,67±0,05 0,97±0,01	637 0,71±0,06 0,99±0,003
	Neredeyse mükemmel uyum	0 - -	0 - -	0 - -	0 - -	1 0,8034±0,00 0,9886±0,00	666 0,88±0,06 0,99±0,002

Not: Her bir hücrede yer alan ilk değer frekans, ikinci değer Cohen Kappa katsayısı için aritmetik ortalama \pm standart sapma değerini, üçüncü değer ise ayırt edilebilirlik katsayısı için aritmetik ortalama \pm standart sapma değerini göstermektedir.

Gerek [Tablo 3a](#) gerekse [Tablo 3b](#)'den görüldüğü üzere uyum ve ayırt edilebilirlik birbiri ile ilişkili kavramlardır. Tablolarda verilen sonuçlara göre ayırt edilebilirlik arttıkça uyum artmaktadır. Ayırt edilemezliğin söz konusu olduğu durumlarda, uyumun olmadığı ya da önemsiz olduğu kesinlikle görülmektedir. Uyumun önemli ya da mükemmel olması ancak ayırt edilebilirliğin mükemmel olması durumunda söz konusu iken, ayırt edilebilirliğin mükemmel olması durumunda uyum önemsiz olabilir. Buna göre değerlendiricilerin kararları arasındaki uyum değerinin düşüklüğünün değerlendiricilerden kaynaklanabileceği, ancak buna tamamen değerlendiricilerin değil bir kısmına da kategorilerin ayırt edilebilir olmamasının sebep olabileceği söylenebilir. Her bir örneklem büyüklüğü için ayırt edilebilirlik derecesi ve kappa katsayıları hesaplanmış ve sonuçlar [Tablo 4](#)'te verilmiştir.

[Tablo 4](#)'te de görüldüğü gibi hemen tüm örneklem büyüklükleri için elde edilecek sonuçların %50'sinde hem ayırt edilebilirlik hem de uyum beklenmemelidir. Ayırt edilebilirlik ve uyum için "neredeyse mükemmel" yorumlama seviyesi karşılaştırıldığında ayırt edilebilirlik için daha yüksek bir yüzde söz konusu olmasına rağmen bu yorumlama seviyesinin ortaya çıkma yüzdesi hem uyum hem de ayırt edilebilirlik için örneklem büyüklüğü ile ters orantılı değer almaktadır.

TABLO 4: Örneklem büyüklüklerine göre uyum ve ayırt edilebilirlik yüzdeleri.

n	Uyum/ayırt edilebilirlik yorumlama seviyeleri					
	Yok	Önemsiz	Orta derecede	Ekseriyetle	Önemli derecede	Neredeyse mükemmel
10	41,9/52,5	30,6/6,0	10,6/7,7	7,4/6,0	6,3/4,9	3,2/22,9
15	45,5/51,1	25,6/8,8	11,8/7,9	8,4/6,4	4,9/5,7	3,9/20,1
20	46,7/50,9	24,8/9,6	11,8/9,0	8,4/6,9	5,9/5,4	2,3/18,2
25	47,0/49,4	23,8/10,3	12,7/9,7	8,0/7,8	5,8/7,0	2,7/15,8
30	50,4/52,8	23,2/10,5	12,3/10,4	8,2/8,0	3,7/5,2	2,2/13,1
35	49,4/50,9	21,7/11,3	12,7/9,6	9,2/9,3	4,7/6,6	2,3/12,3
40	48,9/51,8	23,8/12,6	11,8/10,2	8,5/10,5	5,3/7,0	1,7/7,9
45	48,2/49,6	22,5/12,5	14,3/12,6	7,3/7,4	5,5/6,1	2,2/11,8
50	50,0/50,8	21,0/10,4	12,4/11,5	9,6/9,9	4,8/6,3	2,2/11,1
55	49,4/49,8	22,7/12,3	11,9/12,0	6,8/7,9	7,1/8,2	2,1/9,8
60	49,4/50,2	22,4/12,1	12,4/12,5	8,1/8,6	5,6/7,7	2,1/8,9
65	49,8/50,4	21,3/11,9	15,4/13,5	8,1/10,0	3,6/5,4	1,8/8,8
70	51,0/51,8	21,4/12,4	12,5/11,8	8,9/9,7	4,6/7,2	1,6/7,1
75	48,7/49,6	22,4/13,4	13,2/11,4	7,3/9,9	5,9/8,3	2,5/7,4
80	51,1/51,3	21,5/12,3	11,9/11,8	8,8/9,6	5,0/7,4	1,7/7,6
85	48,6/48,7	23,2/13,4	13,2/12,5	9,0/9,7	4,5/8,1	1,5/7,6
90	48,6/48,9	21,8/13,3	13,5/12,2	8,7/9,2	5,0/7,9	2,4/8,5
95	48,5/47,4	21,5/12,4	13,9/12,6	9,6/11,0	5,0/8,2	1,5/8,4
100	53,0/53,2	20,7/11,1	11,2/12,1	7,1/9,1	6,3/7,5	1,7/7,0
125	50,7/50,7	21,9/12,4	12,2/13,0	7,1/9,3	5,8/7,9	2,3/6,7
150	50,0/50,0	20,8/13,1	13,8/11,9	10,2/10,4	3,3/8,6	1,9/6,0
175	52,2/52,3	21,1/12,8	12,3/12,4	7,3/9,9	5,2/7,2	1,9/5,4
200	48,5/48,7	22,0/14,1	13,0/11,0	8,5/11,2	5,9/7,8	2,1/7,2
225	51,3/51,4	20,0/12,6	13,2/11,5	9,5/11,0	4,4/8,1	1,6/5,4
250	52,0/52,0	19,6/11,4	13,4/12,9	8,3/10,8	4,7/8,1	2,0/4,8
300	51,3/51,3	20,9/11,5	11,9/13,4	8,9/11,8	5,3/7,4	1,7/4,6
350	51,4/51,4	19,2/11,1	14,3/12,4	8,5/11,7	5,3/9,4	1,3/4,0
400	49,7/49,7	20,6/12,4	15,6/14,0	8,6/10,7	4,6/9,0	0,9/4,2
450	47,7/47,7	23,6/15,6	12,2/11,9	8,4/10,7	5,5/8,0	2,6/6,1
500	49,9/49,9	23,5/16,0	13,7/11,0	6,4/10,1	5,0/7,9	1,5/5,1
600	47,3/47,3	21,8/14,2	14,4/13,8	8,7/10,8	5,6/8,5	2,2/5,4
700	49,8/49,8	21,4/13,8	13,8/11,1	7,8/11,3	5,6/9,3	1,6/4,7
800	47,8/47,9	23,3/13,2	12,6/13,6	8,4/10,3	6,2/10,4	1,7/4,6
900	50,5/50,5	21,4/14,1	13,6/11,6	8,2/12,0	4,7/8,2	1,6/3,6
1.000	50,0/50,0	23,0/12,6	12,9/13,1	8,1/12,5	4,8/8,4	1,2/3,4
Genel	49,3/50,3	22,3/12,2	12,9/11,7	8,3/9,8	5,2/7,6	2,0/8,4

Not: Her bir hücredeki ilk değer uyum, 2. değer ise ayırt edilebilirlik yüzdesini göstermektedir.

TARTIŞMA

Cohen kappa katsayısı 2 sonuçlu bir özelliğin 2 değerlendirici arasındaki uyum derecesinin değerlendirilmesinde yaygın olarak kullanılmaktadır. Ancak bazı yazarlar kappa ile ilgili olarak, gözlenen uyuma olasılığı yüksek olsa bile 0'a yakın bir değer alabileceği şeklinde bir soruna işaret etmişlerdir.¹¹ Cohen kappa katsayısı ile ilgili bir başka problem, p_e değerlerinin 0,5'i geçmemesi gerektiği gerçeğine rağmen marjinal olasılığa bağlı olarak 0'dan 1'e kadar çok geniş bir aralıkta değer aldığı, bunun da kappa için rastgele uyum olasılığı hesaplanırken uygulanan yanlış yöntemlerden kaynaklandığı şeklinde ifade edilmiştir.¹² Darroch ve McCloud (1986), kappa katsayısının bir dizi konuyu hangi değerlendirici çiftinin sınıflandırdığına büyük ölçüde bağlı olduğundan dolayı kappanın yerine ayırt edilebilirlik derecesinin

kullanılmasını tavsiye etmişlerdir.⁵ Çalışmadan elde edilen ve [Tablo 3a-3b](#)'de verilen bulgular dikkate alındığında ayırt edilebilirlik derecesinin düşük olduğu durumlarda değerlendiricilerin kararları arasında uyum beklenmemesi, ayırt edilebilirlik derecesinin yüksek olduğu durumlarda ise uyumun önemli derecede artmasının beklenebileceği söylenebilir. Buradan da anlaşılacağı üzere uyum, ayırt edilebilirlik derecesi ile göz ardı edilemeyecek bir birlikte hareket etme özelliğine sahiptir. Cohen kappa katsayısının kullanıldığı çalışmalarda elde edilen uyumun düşük/yüksek olması tek başına yorumlanmamalıdır. Ayırt edilebilirlik derecesi hesaplandıktan sonra yapılacak yorumlar ilgilenilen konu hakkında verilecek kararların isabetli olmasını sağlayacaktır. Örneklem büyüklüğünün ayırt edilebilirlik ve uyum değerleri üzerinde etkili olduğu her bir yorumlama seviyesi için söylenemez. Ancak elde edilen sonuçlara göre genel olarak tüm örneklem büyüklükleri için elde edilen sonuçların yaklaşık %50'sinde uyum ya da ayırt edilebilirlik beklenmemelidir. Ayırt edilebilirlik derecesi için "neredeyse mükemmel" yorumlama seviyesi ile karşılaşmak, uyum için aynı yorumlama seviyesi ile karşılaşmaya göre daha olası olmasına rağmen hem ayırt edilebilirlik hem de uyum için örneklem büyüklüğü arttıkça bu yorumlama seviyesi ile karşılaşmak güçleşmektedir.

SONUÇ

Ayırt edilebilirlik $-\infty \leq \delta \leq 1$ arasında değer alır. $\sum_{j=1}^c p_{ii} = 1$ koşulu ile $\delta = 1$ olması tüm kategori çiftlerinin tamamen ayırt edilebilir olduğunu, $\delta \leq 0$ olması ise tüm kategori çiftlerinin tamamen ayırt edilemez olduğunu gösterir. Kategori çiftlerinin ayırt edilememesi tüm i ve j değerleri için $p_{ij} = p_{i+}p_{+j}$ olması demektir. Kategoriler arasında ayırt edilebilirlik söz konusu değilse değerlendiriciler arasında kesinlikle uyum aranmamalıdır. Kategorilerin önemli ya da mükemmel derecede ayırt edilebilir olması da değerlendiriciler arasındaki uyumun her zaman önemli ya da mükemmel derecede yüksek olacağını göstermez. Dolayısıyla uyum çalışmalarında, uyum katsayılarının tek başına kullanılmasının yeterli olmayacağı, bu katsayılarla birlikte ayırt edilebilirlik derecesinin de kullanılmasının güvenilirlik bakımından gerekli olduğu 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

1. Sertdemir Y, Burgut HR, Alparslan ZN, Ünal İ, Günaştı S. Comparing the methods of measuring multi-rater agreement on an ordinal rating scale: a simulation study with an application to real data. J Appl Stat. 2013;40(7):1506-19. [\[Crossref\]](#)
2. Ato M, Lopez JJ, Benavente A. A simulation study of rater agreement measures with 2x2 contingency tables. Psicologica. 2011;32:385-402. [\[Link\]](#)
3. Agresti A. Modelling patterns of agreement and disagreement. Stat Methods Med Res. 1992;1(2):201-18. [\[Crossref\]](#) [\[PubMed\]](#)
4. Darroch JN, McCloud PI. Category distinguishability and observer agreement. Aust J Stat. 1986;28(3):371-88. [\[Crossref\]](#)
5. Shoukri MM. Coefficients of agreement for multiple raters and multiple categories. Measures of Interobserver Agreement. 1st ed. Boca Raton, Florida: Chapman & Hall/CRC, CRC Press LLC; 2004. p.42-3.
6. Yılmaz AE, Saraçbaşı T. Agreement and adjusted degree of distinguishability for square contingency tables. Hacettepe J Math Stat. 2019;48(2):592-604. [\[Crossref\]](#)

7. Valet F, Guinot C, Mary JY. Log-linear non-uniform association models for agreement between two ratings on an ordinal scale. *Stat Med*. 2007;26(3):647-62. [\[Crossref\]](#) [\[PubMed\]](#)
8. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46. [\[Crossref\]](#)
9. Carlini E, Rapallo F. Algebraic modelling of category distinguishability. In: Gibilisco P, Riccomagno E, Rogantin MP, Wynn HP, eds. *Part I Contingency Tables: Algebraic and Geometric Methods in Statistics*. 1st ed. Cambridge: Cambridge University Press; 2010. p.111-22. [\[Crossref\]](#)
10. Saraçbaşı T, Aktaş Altunay S. *Kategorik Veri Çözümlemesi*. 1. Baskı. Ankara: Hacettepe Üniversitesi Basımevi; 2016.
11. Ohya T. Statistical inference of Gwet's AC1 coefficient for multiple raters and binary outcomes. *Commun Stat Theory Methods*. 2021;50(15):3564-72. [\[Crossref\]](#)
12. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. 2013;13:61. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)